Modeling Heavy Metals in Soil Using Spatial Regression Analysis

by

Steeve Deschênes
BSc, University of Victoria, 2010

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in the Department of Geography

Supervisory Committee

Modeling Heavy Metals in Soil Using Spatial Regression Analysis

by

Steeve Deschênes
BSc, University of Victoria, 2010

Supervisory Committee

Dr. Eleanor Setton, Department of Geography
Co-Supervisor

Dr. Peter Keller, Department of Geography
Co-Supervisor

Dr. Julie Zhou, Department of Mathematics and Statistics
Outside Member

Abstract

Supervisory Committee
Dr. Eleanor Setton, Department of Geography
Co-Supervisor
Dr. Peter Keller, Department of Geography
Co-Supervisor
Dr. Julie Zhou, Department of Mathematics and Statistics
Outside Member

High levels of toxic heavy metals in the environment are a major concern and our knowledge about their adverse impacts and distribution patterns is improving. To mitigate human exposure for large regions, understanding the spatial distribution of metals in soil is key. Several types of models are available to predict the concentration levels, but they are often complex and data-intensive.

The objective of this research is to explore the application of a simple method that produces geographically referenced predictions of surface soil concentrations of heavy metals. The approach uses publicly-available Canadian soil sample data, Geographic Information Science, statistical correlation and regression analyses.

Geographically Weighted Regression (GWR) was used to investigate the spatial variability of the relationship between surface and the subsurface soil metal concentrations. Correlation analysis (Pearson's) between the log of concentration levels of the two layers shows relationships of 0.51 for arsenic (As), and 0.23 for lead (Pb). Although the correlation results showed levels in the two layers are related, GWR analysis illustrates that the degree of this relation varies geographically. This study suggests that factors (natural and anthropogenic) other than the subsurface concentration itself are contributing to the concentration surface levels for all of the studied metals in this dataset.

Based on the above findings, two linear regression models were developed to predict As and Pb levels in surface soil. Independent variables in the models were developed using geographic data on factors hypothesized to influence surface levels, an

approach that has been extensively used for modelling air pollution and known as Land Use Regression (LUR).

For the LUR analysis, the results show that industrial activities account for more than 70% of the variation of Pb concentrations in surface soil. Interestingly, the LUR model for As suggests that the bedrock geology and the total length of road at a location are the main factors. Both variables account for more than 40% of the variations of the As levels in surface soil in BC. The LUR results suggest that regional scale modeling of As and Pb surface soil concentrations can provide information about their spatial patterns that may be useful for understanding potential human exposure and the conduct of environmental epidemiological studies.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank my friends and my lab peers for their constant support: Maéva, Annie, Adam, Norma, Karla, Caty, Aisley, Basil, Cloe, and Rose - thank you very much, you are awesome. A special thanks goes to my doppelgänger Kristi, without which I would not have been able to make it in “Henglish”. Thank you for showing me the fun and the beauties of English editing. The UVic Statistical Consulting Center went a long way to help me understanding the statistical pieces of my project. Mary and Linghong, you deserve a special thanks for your patience. To my supervisors Dr. Setton, Dr. Keller, and Dr. Zhou; thank you for your patience and repeated explanations over the past few years. I am very grateful to have received your precious advices and support.

Finally and more importantly, I have to thank Camile, my daughter. Your happiness and all the fun we had every weekend was more than needed. Thank you to understand every time I told that I was too busy to play with you. I have neglected you so much and took too much time away from you. To my mother, despite the distance you always been there listening and supporting me. Thank you all for being in my life!!

Chapter 1 Introduction

Heavy metals are known to be present in the environment and to be toxic (ATSDR, 2013). Their adverse effects on plants and animals, including humans, raise concerns about elevated levels of exposure. Therefore, understanding where metal levels may be elevated can aid in targeting actions to reduce human exposure, and improve the quality of the environment by providing useful information to support remediation and emission regulations. On one hand, soil sampling is widely used to provide information about the concentration levels, but is very costly to undertake at large regional scale and the results are typically only valid at specific locations. On the other hand, models that attempt to predict concentration levels may be able to provide estimates for exposure assessment, while reducing cost and time of surveying.

This thesis is divided into four chapters and three appendices. This first chapter serves as an introduction, providing a general overview of metals in the environment, soil characteristics, and approaches predicting metal concentration in soils. The chapter also describes the methods and the data used for the analyses.

The second chapter documents the use of Geographically Weighted Regression (GWR) to analyze the spatial relationship between surface and subsurface concentrations of heavy metals in two Canadian regions. The objective is to evaluate the spatial variability of the relationship between the different metal concentrations between two soil layers (surface and B horizon). This study proposes that if the two layers are to a certain degree related, the subsurface would be an important variable to include in the modelling of surface concentrations.

The third chapter is presented as a research paper. It builds on the findings of the GWR analysis presented in Chapter 2. The study investigates how well Land Use Regression (LUR) can predict concentration levels of Arsenic (As) and Lead (Pb) in soil using variables representing specific natural and anthropogenic sources. For this study, Geographical Information System (GIS) is used to develop a series of predictive geographical variables around metal sample locations. The geographical factors hypothesized to influence the concentration of metals in soil - transport networks, land use, mines and tailing sites, major industrial emission locations (Murray et al., 2004; Adriano, 2001) - are derived using circular buffers of different radii (Jerrett et al., 2005; Su et al., 2009). Other geographic variables used are site-specific information (i.e., bedrock and surficial geology, elevation, slope, and precipitation). Linear regression is applied to develop a deterministic equation, which is then used to calculate predicted metal concentration levels for a finely spaced grid of points for the study region.

The fourth chapter presents the conclusion of the research. The first part summarizes the findings for both GWR and LUR results. The chapter goes on to discuss the limitations and the challenges of modeling heavy metal concentrations in soil. Finally, a future research opportunity for improving the models concludes the chapter.

Appendix A includes detailed information about the sources and statistical descriptions of the soil sample locations, and contains maps and graphs of the metal concentration levels. Appendix B contains a full description of the concentration levels data, the regression results, and the residual analysis of the GWR. Appendix C covers the results of the LUR models for As and Pb, including residuals and bootstrap analyses.

The remainder of this introductory chapter is organized into four main sections. The first section presents the research questions in detail for developing both GWR and LUR models. The second section provides a general description of metals in soil, including impacts on health, exposure pathways, physical and chemical processes in soil, and emission sources. The third section reviews techniques and modeling approaches to predict the concentration of metals in soil and to relate their spatial distribution to natural and anthropogenic sources. This section also highlights the potential of regression analysis to predict the concentration levels of metals in soil and describes some limitations of this technique. Finally, section four provides a comprehensive description of the concentration data, including the method used to derive the independent variables for the LUR models.

1.1 Research objectives

This thesis explores the capacity of geographically-based regression analyses to identify variables influencing the spatial distribution of metals, and uses these relevant variables to predict the concentration levels in Canadian surface soils. In this research, two linear regression techniques - Geographically Weighted Regression (GWR) and Land Use Regression (LUR) - are employed to analyze publicly-available, government-sourced data. GWR is used to determine the relationship between the surface and the subsurface (B-horizon) concentration levels. For this analysis, the data includes surface and subsurface concentration levels for six metals: Arsenic (As), Cadmium (Cd), Cobalt (Co), Chromium (Cr), Nickel (Ni), and Lead (Pb). In addition, LUR, which is an approach widely used in modeling air pollution for exposure assessment, is used to model the contributions of potential emission sources on the spatial distribution of As and Pb in surface soil using variables derived from geographic data. The goals of this thesis are

three-fold: 1) to explore the relationship of the subsurface levels with the surface levels, 2) to develop maps of predicted soil surface concentrations at a regional level based on models using geographic variables, which could be useful to determine the potential exposure levels of Canadians to As and Pb, and 3) to assess the strengths and weakness of this approach.

1.2 Heavy metals in the environment

Heavy metals are natural elements in which toxicity, sources, mobility, chemistry, and exposure pathways are often very different from each other, and at this point, not fully understood. Before a predictive model can be developed, it is important to understand the metal characteristics and environmental conditions that affect their mobility, bioavailability, and concentration levels at the surface soil. This section provides a review of soil processes and mobility, adverse health effects, exposure pathways, and emission sources associated with six metals: As, Cd, Co, Cr, Ni, and Pb; and describes the metal characteristics in more details.

1.2.1 –Soil chemistry and mobility factors

Soil is a highly variable and dynamic environment (Brady and Weil, 2002). Soil formation processes create horizontal layers called “horizons”. Soil horizons are the different horizontal layers that have specific chemical and physical properties (Brady and Weil, 2002). The horizons run more or less parallel to the surface layer. The depth or thickness of each horizon is determined by the soil formation processes, so depending on the soil type, certain horizons may or may not be present, determining their position in the soil classification (Brady and Weil, 2002). Figure 1 shows an example of a soil profile showing different horizons. Its characteristics vary greatly in both time and space, and thus influence the types and magnitudes of chemical reactions among the different soil components (Brady and Weil, 2002).

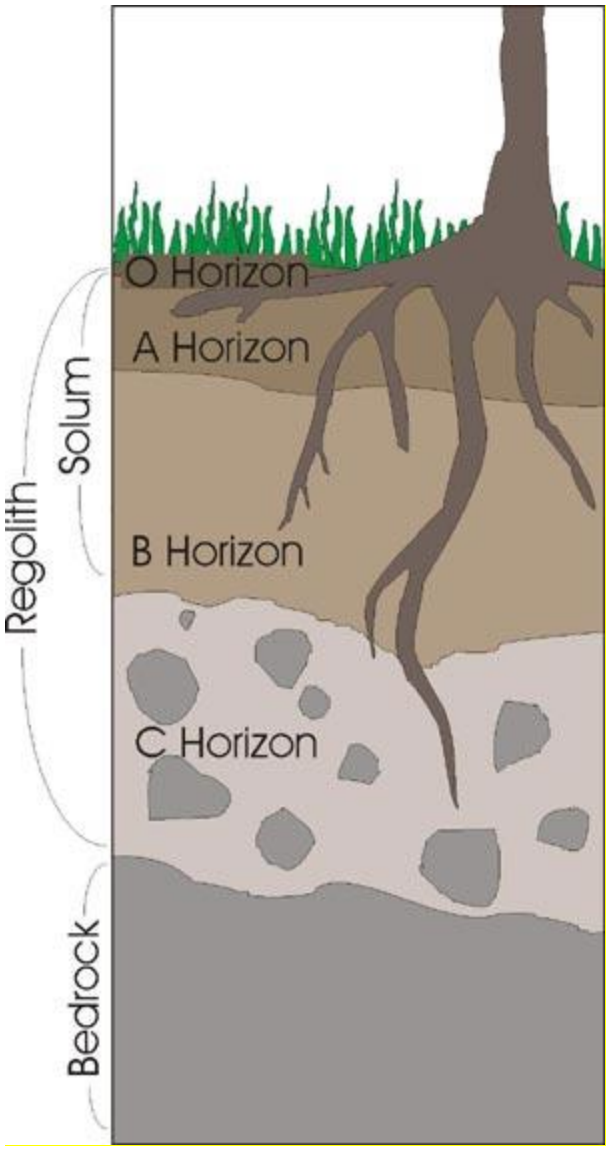
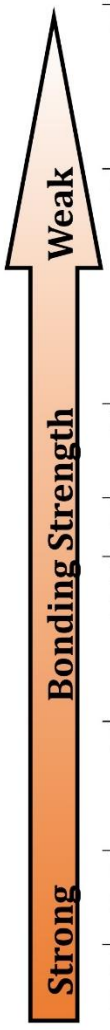


Figure 1 : Soil horizons – (Source : Leslie Dampier http://www.landfood.ubc.ca/soil200/classification/soil_horizon.htm)

There are eight major soil components – primary and secondary minerals, humic and fluvic acids, biomass, precipitates, colloids, and solution (Table 1). Each component is characterized by the different possible chemical bonds of the metals with the other soil constituents. The strength of bonding determines the mobility and the bioavailability of the

metal in soil (Table 1). Variation of magnitudes and types of chemical reactions affects the metal's form, which changes its mobility in soil.

Table 1 : Soil components (adapted from Adriano, 2001)



Component	Description
Soil solution	Aqueous phase of the soil constituents; refers as soil moisture or soil water. It contains dissolved organic and inorganic substances, gases from the atmosphere, and plant and microbial activities. Sites of intense geochemical reactions. Metal reaction occurs in the aqueous phase.
Colloids	Mobile solid phase, which includes the very fine particulates (0.001 to 0.1 μm). Occur as discrete layer of silicate, oxides, or humus or a combination of them. Their surface can be coated with humic material or oxides, which block their reactive sites and impede cation exchange. Identified as effective transporter, colloids can leach in the vadose zone and to the groundwater.
Precipitates	Includes carbonates, phosphates, and sulfides. Their quantities are controlled by pH and redox potential.
Biomass	Includes all living organisms and their remains.
Humic and Fluvic acid	Formed by the microbial degradation of organic matter. Humic acid (dark brown, not very soluble at normal pH) and fluvic acid (light brown-yellow, soluble at all pH condition) are very chemically reactive in soil.
Humus	Soil Organic Matter (SOM), which is a heterogeneous mixture of products resulting from microbial and chemical transformations of organic residues
Secondary minerals	Clay size elements. High surface area and are very chemically reactive compared to primary minerals
Primary minerals	Least reactive constituent of soil and mostly found as sand fraction. It includes silicates minerals, feldspar, olivines, proxenes, and micas

Figure 1 shows the key processes controlling the partitioning of heavy metals between the different phases in soil (Adriano, 2001). The variability of soil properties and the chemical reactions are the main parameters that control the metals' mobility in soil (Fritsh et al., 2010). Similar patterns are recognizable among certain metals but each element is chemically distinct. Thus, their mobility can be different under the same soil conditions. For example, several

studies demonstrate the strong affinity of Cd and Pb with organic matter (Adriano, 2001). In surface soil, Pb mobility is restrained under normal pH levels, while Cd can be uptaken by plants (Adriano, 2001). Meanwhile, As leaches more easily and eventually accumulates in the lower soil layers (Adriano, 2001).

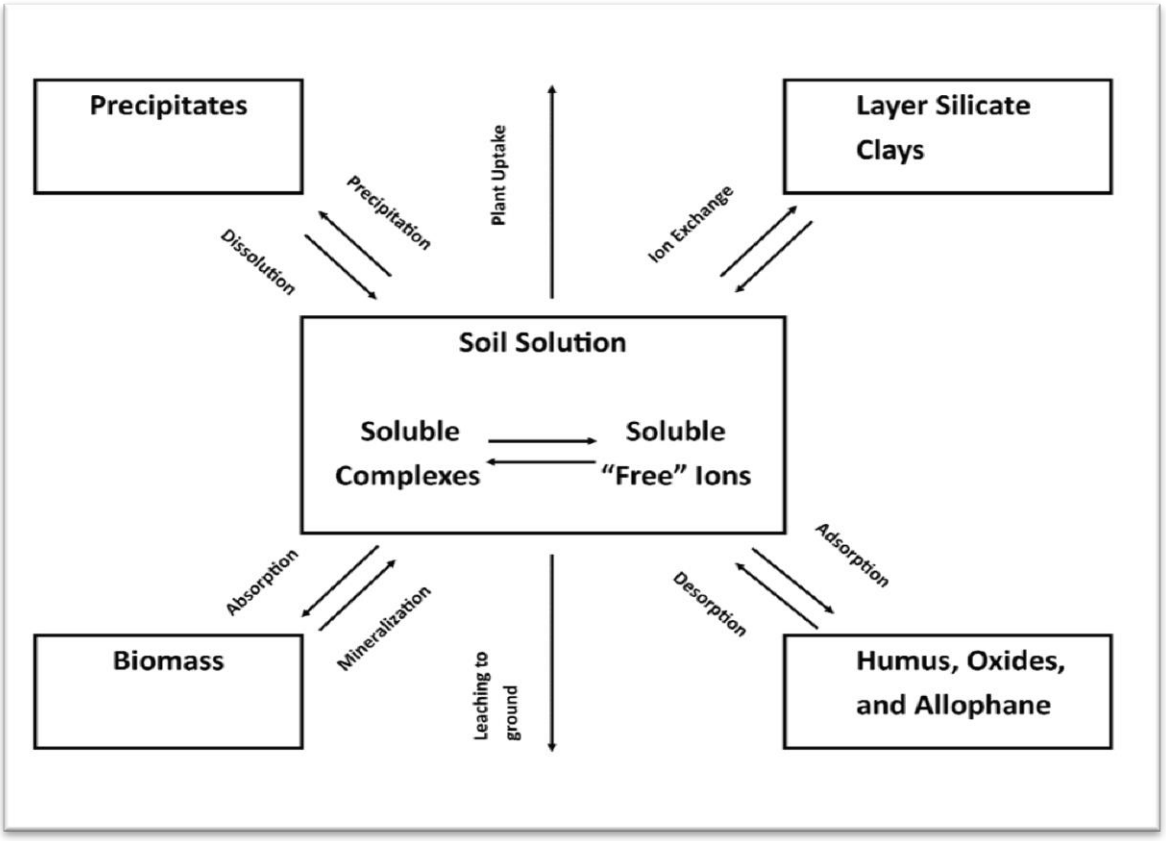


Figure 2 : Interactive key processes affecting the partitioning of trace metal (from Adriano, 2001. p.35)

Figure 2 shows the three components controlling the spatial variability of metal concentrations in soil: 1) background level variations, 2) emission and deposition characteristics, and 3) the retention, mobility, and cycling process at each location (Fritsch et al., 2010). Each component has high spatial variability. The background levels are not constant over space, the emission depositions are not evenly distributed, and the soil conditions

controlling the mobility of metals also vary both in space and time (Adriano, 2001; Fritsch et al., 2010).

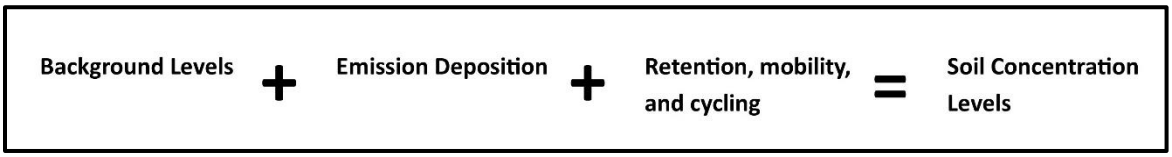


Figure 3 : Factors controlling the spatial distribution of metals in soil

The identification of parameters, natural or anthropogenic, that control the spatial distribution of metal in soil is required to develop potential exposure models (Fritsch et al., 2010). In this thesis, the goal is not to explain the movement of metals at a micro scale, but to explore any existing relationships with emission factors and spatial patterns of metal concentration levels at the soil surface over large regions. At the regional level, spatial variations of metal concentrations are potentially related to several different factors acting simultaneously at various magnitudes. In addition, chemical differences among the metals would likely alter the distribution of metal concentrations in soil. Tables 2 – 7 provide more details for the different metals and their mobility in soil, including natural and anthropogenic sources.

Table 2 : Arsenic description (from Adriano (2001), except where cited).

Description	Soil content	Biogeochemical processes (mobility)	Sources
Steel-grey, brittle, crystalline metalloid **Oxidation states:** III, 0, III, and IV	World soil As content is averaged at 10 ppm, but near ore deposit or contaminated sites, soil levels have reached 400 to 900 ppm Canadian uncontaminated soil and sediment range from 4 to 150 mg/kg (Wang and Mulligan, 2006) Average of 6.3 ppm total As was reported for agricultural soils in Ontario	In coarse-texture soils, As can move downward in the soil profile with leaching water - High clay content reduces As mobility Leaching of As in soil is continuous, most of the As at the surface layer leached in the 20-40 cm depth Highly related to the adsorption/desorption reactions (Zhang and Selim, 2006) The slow downward movement of As has been observed in contaminated sites and it is controlled by the soil conditions (Zhang and Selim, 2006) Larger concentrations due to surface contamination remain in the upper layers and the levels decrease with increasing depth, suggesting a slow downward movement (Zhang and Selim, 2006).	**Natural:** Weathering and erosion of rocks containing As - arsenopyrite, realgar, and orpiment (Wang and Mulligan, 2005) **Anthropogenic:** Combustion of coal Pesticide production and application Wood preservative production and application Mining and smelting operations and tailings Fossil fuel processing and combustion. Disposal and incineration of waste

Table 3 : Cadmium description (from Adriano (2001), except where cited).

Description	Soil content	Biogeochemical processes (mobility)	Sources
Soft, ductile, silver-white lustrous and electropositive element	Concentration level is largely influenced by the parent material Canadian soil concentrations range from 0.001 to 0.1 ppm and between 0.01 to 0.7 ppm in glacial tills Northern Canadian surface soil was averaged at 1.8 ppm Soil near ore deposits can reach up to 40 ppm In areas affected by smelting operation surface soils range from 0.2 to 350 ppm	More labile in soil and is more bioavailable (Van der Perk, 2006) In contaminated areas, elevated levels were found up to 2 m depth, but the majority remains in the top 20 cm (Sterckeman et al., 2000) Remains in the upper layer because of affinity with the solid phase of soils (Sterckeman et al., 2000) The mobility is facilitated at low pH (Sterckeman et al., 2000) Higher concentration at the surface produces deeper leaching, and higher quantity in solution (Sterckeman et al., 2000) More a vertical movement than horizontal Food is the most common source of exposure for humans (Robards and Worsfold, 1991). Bio magnification through the food chain makes it dangerous	**Natural:** Weathering of Cd-containing parent material **Anthropogenic:** By-product of Zn, Pb, and Cu refining Metal plating, batteries, and for paint, printer ink and plastic pigments Phosphate fertilizers and sewage sludge Burning of fossil fuels - power plants, furnaces, stoves, cars, etc. Incineration of municipal waste materials

Table 4 : Cobalt description (from Adriano (2001), except where cited).

Description	Soil content	Biogeochemical processes (mobility)	Sources
Silvery white Chemically similar to Ni Rank 30th in abundance of elements in the earth's crust Large deposits in Canada	World soils range from 2 to 40 ppm with levels up to 1000 ppm	Cobalt accumulates in Fe and Mn oxides mostly located in the B-horizon	**Natural:** Cobalt is associated with As, Ni, Pb, Cu, and Fe ores Soils developed from ultrabasic rocks are usually enriched in Co, and also in Ni and Cr. **Anthropogenic**: Production of high grade steels, alloys, super alloys, and magnetic alloys Catalyst in the petroleum industry

Table 5 : Chromium description (from Adriano (2001), except where cited).

Description	Soil content	Biogeochemical processes (mobility)	Sources
Silvery, lustrous, malleable metal that takes a high polish Dissolves readily in non-oxidizing mineral acid, but not in cold Aqua Regia Ranks 21st in abundance Distribution levels in the soil profile are inconsistent, but most of the time, the concentration is higher in the B and C horizons Oxidation states: 0, III, and VI, where III is the most stable and VI is the most toxic	Concentration was averaged at 14 ppm in agriculture soils in Ontario Some agricultural soils in Ontario, Canada exhibit a uniform distribution for the A, B, and C horizons. Smallest concentration in organic rich layer (0-6 cm depth) Total Cr concentration decrease with depth in contaminated urban areas (Imperato et al., 2003)	Concentration levels are determined by the parent material Associated with the clay fraction Controlled by Eh, pH, oxidation state, and CEC. The pH is the most important driver, where acidic conditions enhance adsorption	**Natural:** Ultramafic igneous rocks and serpentine **Anthropogenic:** Mining, smelting, and metal works Paper products industries Petrochemical, inorganic chemicals, fertilizer Textile mill products, leather tanning and finishing Power coal plants

Table 6 : Nickel description (from Adriano (2001), except where cited).

Description	Soil content	Biogeochemical processes (mobility)	Sources
Silvery white hard, malleable, ductile, ferromagnetic metal, which is relatively resistant to corrosion and water insoluble Canada is the largest producer of Ni in the world. The major deposits are found in Sudbury, Ontario and Thompson, Manitoba 23rd most common element in the earth's crust Nickel is closely related to Co in both its chemical and biochemical properties	In unpolluted Canadian soils, Ni concentration is around 20 ppm with a range of 5 to 50 ppm. However, high levels of Ni were found around the smelter in Sudbury, Ontario, with levels up to 3000 to 5000 ppm The concentration heterogeneity in the parent material is the main factor that controls the variability in soil concentration No distinctive pattern of distribution in unpolluted soil profiles In Podzol soils: 1-Ni concentration increases with depth from the A to B horizons; 2-uniform distribution throughout the profile; 3-decrease concentration with increasing depth, and; 4-increasing concentration followed by a decrease	Residual fraction (50%), bound with Fe-Mn oxides (20%), in carbonate fraction (~30%), and in exchangeable and organic fraction (`1%) Concentration was more elevated in the humus layer than the parent till in a highly contaminated zone around metal processes industry complex Industrial emissions were correlated with the deposition rate of Ni in the humus layer (Räisäsen et al., 1997) Under anaerobic conditions, sulfides control the solubility The acidification of surface soil might be the explanation of mobilized Ni that leaches down the soil profile	**Natural:** Weathering of igneous rocks **Anthropogenic:** Electroplating, alloy production and fabrication, batteries, electronic components, and stainless steel production Mining and smelting of Ni-bearing ores Oil and coal combustion Sewage sludge application Ni deposition might reach up to 150km around point source (Reimann and Garrett, 2005)

Table 7 : Lead description (from Adriano (2001), except where cited).

Description	Soil content	Biogeochemical processes (mobility)	Sources
Bluish-grey, bright luster, soft, highly malleable, ductile, and poor conductor of electricity	Average content in the earth's crust is 13 to 16 ppm. Coal and shale have higher content than the bedrock Background level in Canada is estimated to an average of 20 ppm, with values ranging from 5 to 50 ppm Agriculture soils of Ontario averaged 46 ppm, with a range of 1.5 to 888 ppm (data from 1976) Soils from fruit orchards had the highest average content of 123 ppm (range 4.4 to 888 ppm) from the use of lead arsenate The mean residence time has been estimated to range from 150 to 500 years	Low solubility in hydroxide, carbonate, and phosphate forms (Van der Perk, 2006) Strongly adsorbed on mineral and organic surfaces Concentration decreases steeply with depth (0 to 30 cm) in contaminated urban area (Imperato et al., 2003) Extractability is decreased by high pH, high phosphate content, organic matter (OM), or clay contents Pedogenic processes, climatic conditions, topography, and microbial activities influence the distribution in the soil profile The mineral layer can serve as the final sink for Pb because of the presence of OM and Fe oxides	**Natural:** Present in moderate quantities in more than 200 minerals, which includes igneous and sedimentary rocks **Anthropogenic:** Fuel combustion - Pb-gasoline, oil, and coal Batteries Production of chemical additives - fuel, paint pigments, pigments for glazing ceramics Steel and non-ferrous metal production - pipes, sheets, solders, ammunitions Incineration of material containing Pb

1.2.2 –Health effects and bioavailability

Heavy metals are non-biodegradable elements which have a long environmental persistence and biological half-life. These two factors are responsible for their bioaccumulation and resultant adverse effects in organisms (Robards and Worfold, 1991), as summarized in Table 8. Bio-accumulation and food chain transfer (bio-magnification) increase the concentration levels within an organism (Robards and Worfold, 1991). Bio-magnification greatly affects top food chain consumers, which are more likely to have greater concentrations in their tissues than in their surrounding media (Dokmeci et al., 2009). The toxicity of each metal depends on its potency level and its concentration in the organism's tissues (ATSDR, 2013).

Table 8 : Health effects of metals (from Adriano (2001), except where cited).

Metals	Toxic species	Health effects
Arsenic	Both III and IV	Causes death at high doses, chronic exposure causes cancer.
Cadmium	All forms	Highly toxic to plants and animals. Chronic exposure results in kidney damage and bone deformation (Itai-Itai disease) (Van der Perk, 2006).
Chromium	VI	Although Cr (III) is essential, Cr (VI) is a potent carcinogen. It can damage kidneys, gastrointestinal tracts and circulatory systems.
Cobalt	All forms	Essential in trace amounts. Limited toxicity to plants and animals.
Nickel	All forms	Essential in trace amounts. Toxic at high exposures. Increases the risk of lung and nasal cancer by inhalation.
Lead	All forms	Toxic to all living organisms. Affects the neuronal system and kidneys (Van der Perk, 2006)

Quantifying metal bio-availability is not only challenging due to the limitations of the analytical techniques, but also because complex chemical interactions of heavy metals occur

under various conditions (Tack and Verloo, 1995). The heterogeneity and dynamism of soil conditions constantly affect the metals' speciation and therefore, their bioavailability (Clemente et al., 2008). Moreover, the bioavailability of each metal not only depends on its concentration in soil, but also on the concentration of various elements on which they sorbed on (Tack and Verloo, 1995). In other words, metals can remain strongly attached to sorbents, which makes them inassimilable by an organism's metabolism even if the metal is present in sufficient quantity. Nevertheless, while it is recognized that the total content of metal in soil is in general a poor indicator of toxicity, it can highlight potential risks of exposure to metals.

1.2.3 –Exposure pathways

From soil, there are three main exposure pathways: 1) inhalation, 2) dermal contact, and 3) ingestion (either direct or through the food chain) (Adriano, 2001). Ingestion is the most important pathway for the general population, while inhalation and dermal contact are more often of occupational concern (Adriano, 2001). However, inhalation of re-suspended contaminated soil, especially urban soil, has been identified as a significant source of exposure to metal (Laidlaw and Filippelli, 2008; Mielke and Reagan, 1998).

Briefly, inhalation is controlled by the particle size (generally <10µm), which can reach the lung alveoli and transfer to the circulatory system. Absorption via the gastrointestinal tracts occurs by way of diffusion, which follows a concentration gradient between the intestine and the blood system (Adriano, 2001). Metals reach the circulatory system through the lungs, skin, or gastrointestinal tracts before accumulating in different tissues (brain, fat, bones), while some are excreted through defecation, urination, respiration, and secretion (Adriano, 2001). The rate

of absorption within an organism varies greatly among metals, individual diet, and physiological conditions (Adriano, 2001).

1.2.4 –Natural sources

In the Earth's crust, metal concentrations vary both within and between different rock types (Reimann and Garrett, 2005). Weathering and erosion processes release metals in the surrounding environment. Thus, the levels in soil, groundwater, surface water, and stream sediments in the surrounding environment can be naturally increased (Reimann and Garrett, 2005; Adriano, 2001), and elevated levels can be reached in ore-rich areas.

The concentration of an element in soil less all potential contamination inputs from human activities is what defines the geochemical background (Reimann and Garrett, 2005). The difficulties in determining the background value of an element two-fold: 1) the retention of metal inputs in soil from anthropogenic sources are difficult to calculate with precision; and 2) contaminants are not evenly distributed over space. Estimated background values often assume a homogenous regional geochemistry, whereas a range of values would typically better define the geochemical background rather than an absolute value (Reimann and Garrett, 2005). In addition, long distance atmospheric deposition studies have proven that remote areas are also affected by human's metal emissions, which further complicates the determination of natural background values (Reimann and Garrett, 2005; Steinnes and Friedland, 2006).

1.2.5 –Anthropogenic sources

Human activities can disturb natural cycles and distributions of metals in soil, as well as in air and water (Adriano, 2001). These disturbances create an unbalanced input/output ratio in the metal cycle, where inputs are more important and result in an accumulation in the media

(Adriano, 2001). Human sources mainly relate to agriculture, transportation, and industrial activities (Fritsh et al., 2010; Imperato et al., 2003). These sources can emit significant quantities of metals into the atmosphere and increase the surrounding metal concentrations in soil via contaminated depositions. Moreover, direct dumping in water and soil also contribute to an increase metal concentration levels in the surrounding environment. The accumulation of the metals in soil from human sources typically decreases with distance from the emission point (Räisäsen et al., 1997).The spatial distribution of metals in soil around an emission source is influenced by emission intensity, wind direction and speed (for air emissions), frequency and quantity of precipitation, and interception by the surrounding landscape (Fritsh et al., 2010).

Fortunately, improvement of waste air, water purification, waste recycling and implementation of stricter environmental regulations have reduced the direct emission of metals. Nonetheless, heavy metals are not biodegradable and elevated concentrations can remain hazardous for living organisms over long periods (Van der Perk, 2006).

1.3 Approaches for modelling soil concentrations

Field sampling is key to establishing the levels of metals in soil and for providing information about their spatial variability. The data can be projected on a map or linked with other variables (land use, parent material, or blood concentration levels) for further analysis (Ajmone-Marsan and Biasioli, 2010; Murray et al., 2004). Moreover, basic statistical tests are often performed between concentration levels and predictor variables or health outcomes to highlight any potential relationships. For example, Murray et al. (2004) studied the influence of pH, grain size, and human impact (urbanization) on the variability of metal concentration throughout the soil profile over a watershed. Simple descriptive statistics (average, ratio, and

outlier) are most often used to compare the sample results with each other (Murray et al., 2004; Preciado et al., 2007; Krcmova et al., 2009). Although sampling offers accurate information, the approach only provides information about the sample location itself. The elevated cost and time-consuming nature of sampling when the study area of interest is large justifies the use of models to predict the spatial distribution of metal in soil.

Models can be valuable tools for exposure assessment and to support risk management decisions (Van de Perk, 2006). Although models are only a representation of the reality, they can also be useful for identifying dominant processes as well as providing useful insight about the spatial distribution of metals for human exposure assessment. The following section describes possible approaches for modelling levels of heavy metals in soil.

1.3.1 –Mass balance models

This approach is based on the fugacity concept (Mackay, 1979). Mathematical algorithms simulate chemical processes (vapour pressure, decay) among a few or several compartments (air, surface and groundwater, soil, sediment, and diverse biota). The fugacity models give estimates about the concentrations of contaminants for each of the compartments at a given scale, based on input data about emission rates for all sources and flows between compartments. Adding more compartments increases the complexity of mass balance models rapidly (Cahill and Mackay, 2003).

While fugacity models have been used for organic contaminants, the applicability of this method for metals has not yet provided compelling results, due to their chemical specificity (low fugacity level) (Jeon et al., 2008). For example, in a study of the San Francisco Bay area, a fugacity model identified sources and concentration level patterns of Mercury (Hg), but had

major discrepancies between estimated and measured concentration levels in every compartment (MacLeod et al., 2005).

1.3.2 –Dispersion models

Dispersion models integrate meteorological variables with emission rates to predict deposition patterns around point sources (Williams and Ogston, 2002). These models have successfully improved our understanding about deposition patterns to soil from atmospheric emission sources (Gerritse, 1996; Islam et al., 2001). In soil studies, dispersion models have been used to calculate the percolation rate of contaminants in the soil column from a point source over relatively small distances – for example, predicting leachate coming from landfills (Islam et al., 2001).

Mainly based on Darcy's infiltration law, these models calculate the rate and direction of infiltration within the soil column; however, the heterogeneity of soil's characteristics (grain size, moisture level) greatly reduces the prediction quality of these models (Islam et al., 2001). Moreover, dispersion models become very complex when emission rates vary and more than one point-source is considered. For many epidemiological studies, their outputs may not be ideal because these models cannot consider all sources (point or diffuse) or multiple controlling factors. However, Schmitt *et al.* (1979) suggest that dispersion models can improve sampling strategies around point emission sources.

1.3.3 –Geostatistical models

Geostatistical methods, mainly inverse distance weighted (IDW) and Kriging, use sample data with geographic coordinates to interpolate metal levels between sample locations, thus providing a continuous surface or predicted concentrations for the sampling region (Imrie et al.,

2008; Li and Heap, 2008). More recently, Lambert et al. (2011) used IDW to predict As, Pb, and polycyclic aromatic hydrocarbon (PAH) levels in surface soil in Nova Scotia, Canada. In the case of IDW, the method uses only sample location information and employs a basic assumption of the metal concentration variation between sample points.

Kriging has several advantages over simple IDW interpolation. In addition to a prediction surface, Kriging produces an uncertainty surface that maps the locations where the model does not predict well (Goovaerts, 1999). Kriging models can also incorporate explanatory variables. For example, Lado et al. (2008) used regression-Kriging to predict soil surface concentration of 8 different metals for 26 countries. Their model satisfactory predicted the concentration of As, Ni, and Pb (45-52% of total variance) in the topsoil. In another case, logistic regression and Kriging methods were uses to predict the probability of exceeding standard levels of As and Pb (Lin et al., 2011), although concentration maps were not produced because no predicted values were above the pollution control standards of the Taiwan Environmental Protection Agency.

In general, the accuracy of both IDW and Kriging models is directly linked with the density of the sampled sites. In fact, Kriging models need a relative high density of data points to produce a low uncertainty level. Although Kriging models provide useful information about the spatial distribution of metals, their complexity can dramatically increase with larger surfaces or by considering more controlling variables. To remain relatively simple, Kriging models rarely integrate many variables, usually only one or two potential drivers (Goovaerts, 1999); these models are therefore best used for relatively small regions.

1.3.4 –Statistical models

Statistical methods, including analysis of variance (Murray et al., 2004), multivariate analysis (Facchinelli et al., 2001), principal component analysis (Davis et al., 2009), and cluster analysis (Zhang, 2006), have been widely applied to understand the relationships between soil concentration and influencing factors. For example, Murray et al. (2010) classified sample locations based on different land uses, soil characteristics, and urbanization intensity. The authors showed positive relationships between urbanization intensity and soil concentration levels. None of these analyses, however, have focussed on incorporating maps of surface concentration levels, which limits the application of the results for human exposure.

1.4 Methods and data

For this thesis, two adaptations of linear regression are applied for spatial analysis: 1) GWR – the local regression approach that subsamples the point locations and describes their spatial relationships; and 2) LUR – the geographic approach that derives geographic variables based on features around sample points using GIS in order to produce high resolution maps of predicted concentration levels.

Linear regression investigates the relationship between a dependent and one or more independent variables. A regression model is described as follows:

$$y = \alpha + \beta x + \epsilon, \qquad [1]$$

where y is the dependent variable, α is the intercept, β is the regression coefficient, x is the independent variable and ϵ is the random error term (Montgomery et al, 2006). The validity of the equation [1] requires four specific conditions: 1-a linear relationship between the dependent and the independent variables; 2-a constant variance; 3-a normal distribution; and

4-independence of the error term. Moreover, for spatial analysis, the residuals should not display any spatial autocorrelation (Zhang et al, 2009). Spatial autocorrelation represents a problem in regression analysis as it violates the independence assumption (Legendre, 1993).

1.4.1 -Land Use Regression

Regression analysis using geographically-based variables is increasingly being applied in environmental sciences. In air pollution, for example, LUR predicted the spatial distribution of contaminants with comparable results to other more complex dispersion models (Jerrett et al., 2005). More recently, in soil modeling, LUR predicted the spatial variation of Pb concentration in soil in soil within a city (Wu et al., 2010) and at a regional scale (Deschênes et al., 2012). Interestingly, these studies have demonstrated the possibility of using relevant geographic features related to Pb emission sources and of predicting the concentration of the metal.

1.4.2 -Geographically Weighted Regression

GWR is a local form of linear regression (Fotheringham, Brunsdon, & Charlton, 2002). The method recognizes the existence of spatial variation in the relationship among the variables (also called spatial non-stationarity) and provides a mean to calculate this variability. For each point location, GWR applies a kernel to subset the dataset and calculates a regression equation using a decay function (Fotheringham, Brunsdon, & Charlton, 2002). The number of points per equation is determined by the bandwidth of the kernel, which itself is calculated by using Least Squares Cross Validation or Akaike Information Criterion corrected (AICc) (Gao and Li, 2011; Leung, Mei and Zhang, 2000).

The GWR equation is defined as follows:

$$y_i = \beta_0(u_i, v_i) + \Sigma_k \beta_k(u_i, v_i)\, x_{ik} + \epsilon_i, \quad [2]$$

here, (u_i, v_i) represents the spatial location of the i^{th} case. Each location has a unique set of sample points obtained with the defined adaptive kernel. y_i is the value of the dependent variable for the i^{th} case. x_{ik} is the independent variable k for the i^{th} case, β_0 is the regression intercept at (u_i, v_i) location, β_k is the regression coefficient at (u_i, v_i) location, and ϵ_i is the random error of the i^{th} case (Fotheringham, Brunsdon, & Charlton, 2002). This method allows the parameters, coefficients, and intercept to vary in space, while showing spatial patterns.

GWR can provide a better understanding of local factors acting on the dependent variable (Tu and Xia, 2008). In opposition, global regression or Ordinary Least Squared (OLS) regression is considered an average that includes all data points, which prevents any representation of spatial variability. With GWR, a distance weighting function calibrates the number of points for each model and adjusts the kernel sizes. Therefore, if the bandwidth is too small, not enough data points would be included, which would drag the fitted value to the actual value. Too large, the kernel would include almost all the points and hide any local spatial patterns (Fotheringham, Brunsdon, & Charlton, 2002).

A limitation of the kernel subsampling is the edge effect in which data samples located at the outskirt of the study area would contain fewer data points, so a fixed window might not be appropriate for sparse datasets. To correct this problem, GWR uses variations of “adaptive Gaussian spatial kernel” according to the density of the data points (Fotheringham, Brunsdon, & Charlton, 2002). Thus, areas with sparser points would have relatively similar amounts of

data points in the kernel compared to more clustered areas. As a consequence, the standard errors are reduced at each sample point (Fotheringham, Brunsdon, & Charlton, 2002).

GWR has been used to reveal spatial patterns in different environmental studies. For example, Tu (2011) used GWR to model the spatial variability of water quality based on various land uses in an urbanized watershed. In another study, Zhang et al. (2009) identified Pb outliers in soil using aluminum (Al) concentrations as the dependent variable.

1.4.3 –Soil Concentration

Concentration data used for this thesis includes soil surveys available online from the Ontario Geological Survey and the Ministry of Environment of British Columbia, which were downloaded in 2010. (Appendix A contains all the details about the sources, analytical and digestion methods, number of samples, and statistical descriptions for all the metals used in the GWR and LUR analyses.)

In the case of the samples from the Ministry of Environment of British Columbia, each sample site represents four soil samples that were collected over a surface area of 80 m^2. The total surface area was divided into four quadrants and one sample taken in each quadrant. The approximate center of the four quadrants was used as the sample site location.

To account for a random sampling assumption in linear regression analysis, all transects and clusters were removed from the data set. Blanks, zeros, or values below the detection limit were excluded. Multiple samples taken at the same location were averaged. Samples analysed using the *Aqua Regia* method were retained, as it is the most widely used method for dissolving the soil samples.

After that, data using Inductively Coupled Plasma (ICP) methods (ICP-MS and ICP-OES) were pooled as they provide comparable results for this type of study, both in accuracy and precision (Baffi, Bettinelli, Beone and Spezia, 2002). Moreover, Paya-Peres et al. (1993) analyzed Cr, Ni, Cd, and Pb in soil by these different ICP methods and did not find any significant discrepancies. For GWR, Instrumental Neutron Activation Analysis (INAA) method was used for As because not enough samples analyzed by ICP methods remain after removing blanks and values below the detection limit for both the surface and the subsurface samples.

Table 9 summarizes the concentration levels of As, Cd, Co, Cr, Ni, and Pb for both surface and subsurface. Figures 3 and 4 show the spatial distribution of Arsenic in both surface and subsurface soil levels.

Table 9 : Metal concentration levels for both surface and subsurface for GWR analysis (ppm)

Metals	Layers	Minimum	1st quartile	Median	Mean	3rd quartile	Max
Arsenic	Surface	1.10	2.60	3.40	3.90	4.50	79.00
	Subsurface	1.00	1.50	1.90	3.84	2.80	520.00
Cadmium	Surface	0.01	0.21	0.40	0.53	0.60	3.85
	Subsurface	0.02	0.08	0.15	0.22	0.30	0.87
Cobalt	Surface	1.00	2.00	3.00	4.82	5.00	249.00
	Subsurface	1.07	6.00	7.00	7.94	8.82	48.70
Chromium	Surface	1.00	7.00	9.00	20.23	15.00	412.00
	Subsurface	2.78	28.00	33.00	39.96	41.00	758.00
Nickel	Surface	2.00	9.00	12.00	15.93	16.00	256.00
	Subsurface	3.33	16.00	20.00	23.40	26.00	330.00
Lead	Surface	0.2	34.75	54.00	58.09	75.00	621.20
	Subsurface	0.80	7.00	9.00	9.88	11.00	76.00

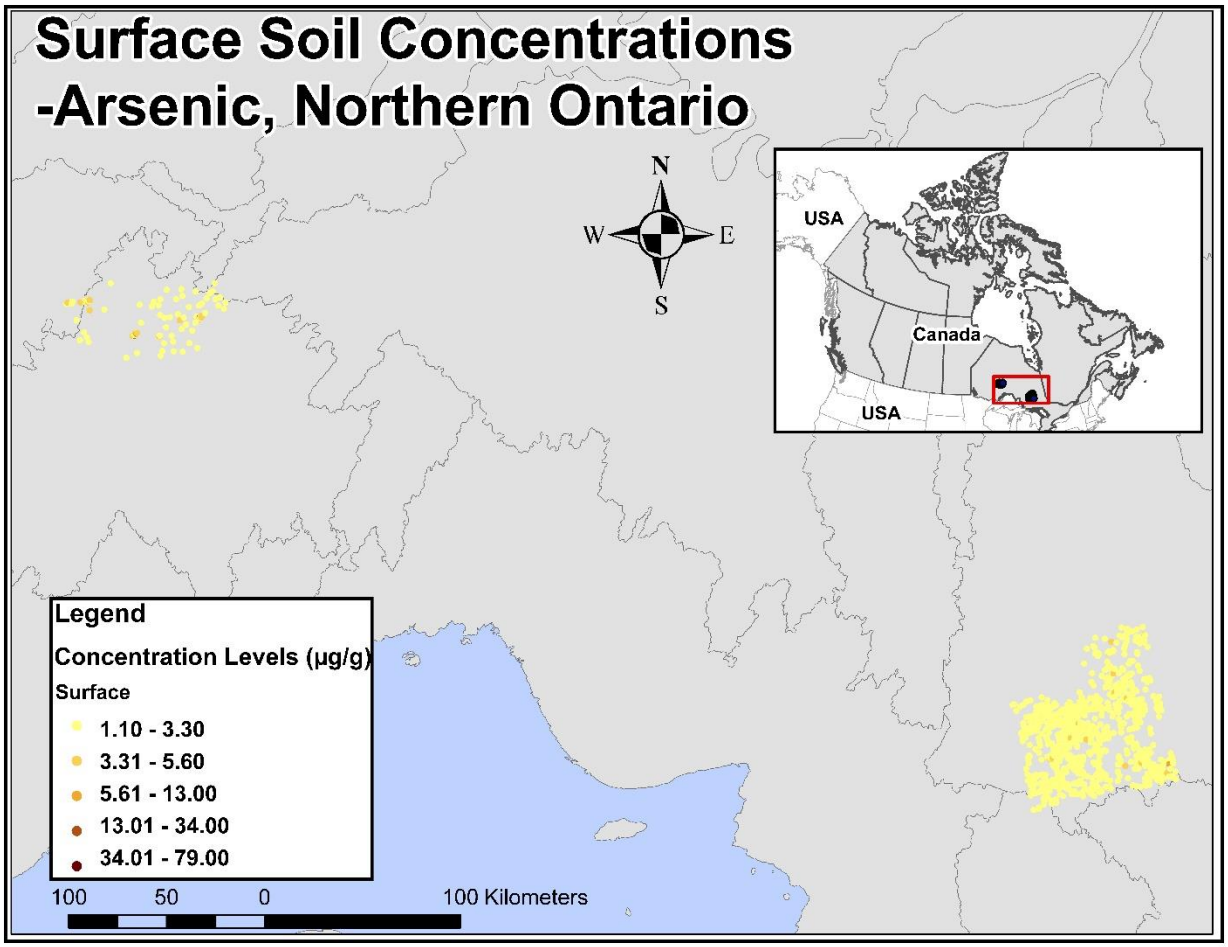


Figure 4: Surface concentration levels for the Arsenic GWR analysis - Ontario

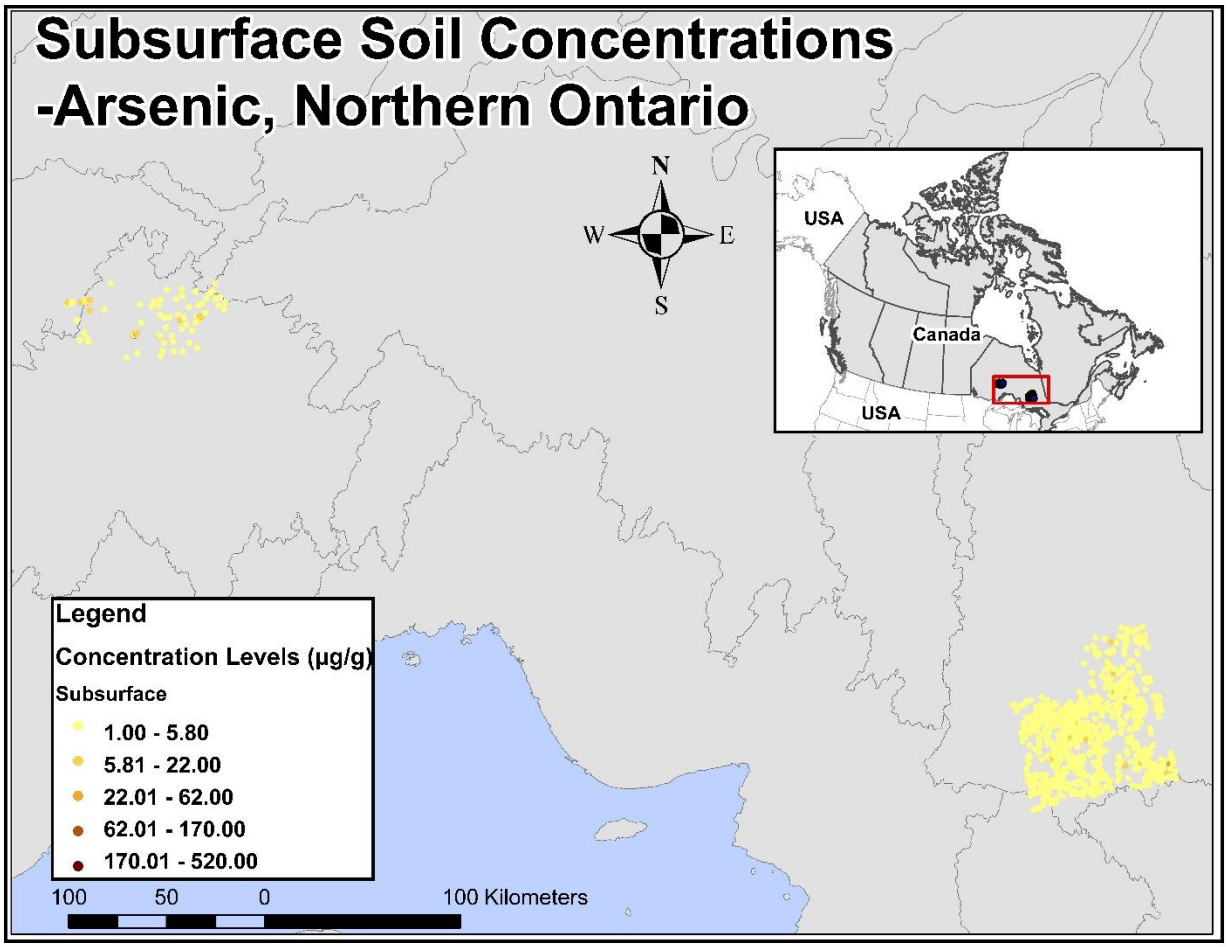


Figure 5: Subsurface concentration levels for the Arsenic GWR analysis - Ontario

Table 10 provides summaries of the final concentration data used for the LUR analysis. Figures 5 and 6 show the concentration levels used in the LUR analysis for As and Pb, respectively.

Table 10 : Concentration levels description for Arsenic and Lead for LUR analysis (µg/g)

Metals	Minimum	1st quartile	Median	Mean	3rd quartile	Max
Arsenic	0.25	3.05	5.50	6.85	9.05	25.70
Lead	1.63	6.79	9.90	45.66	32.12	621.20

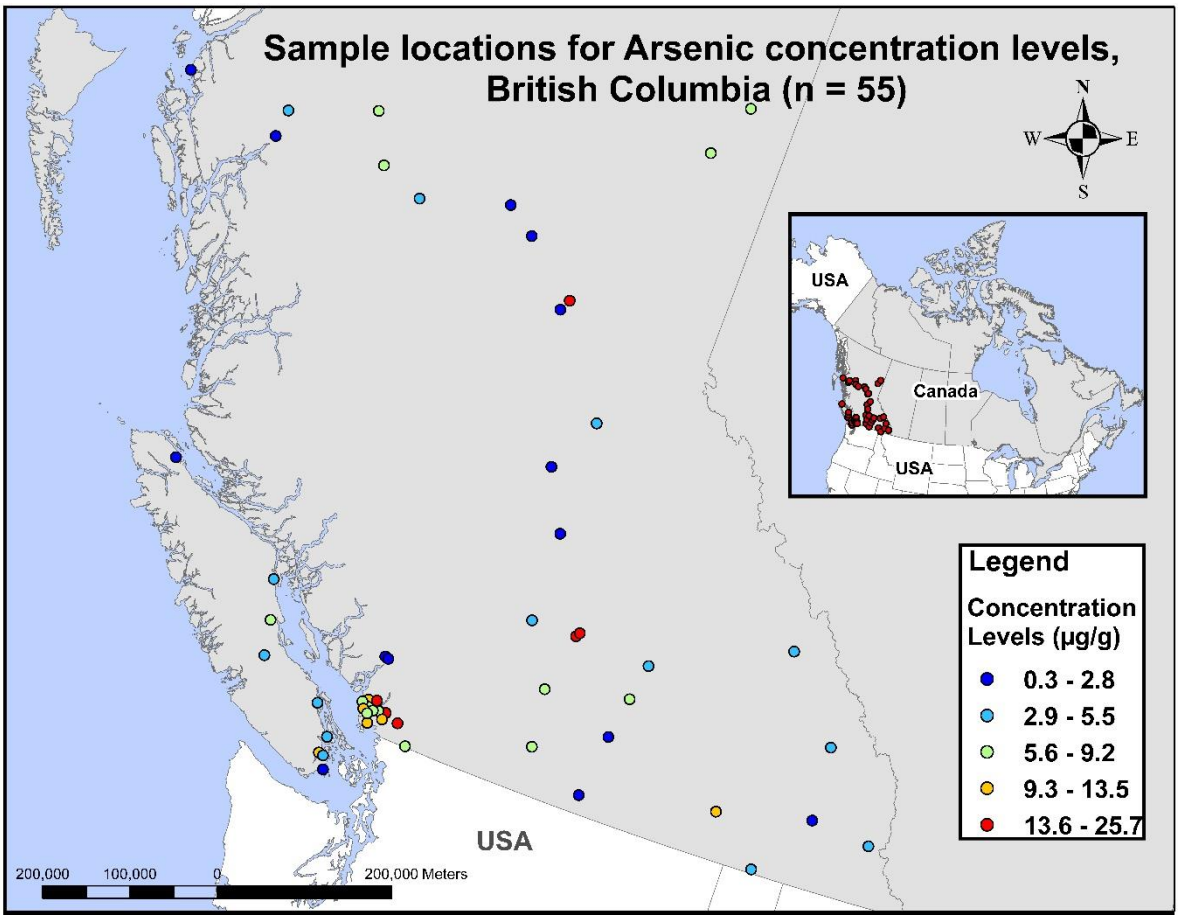


Figure 6 : Concentration level locations for the LUR Arsenic analysis – British Columbia

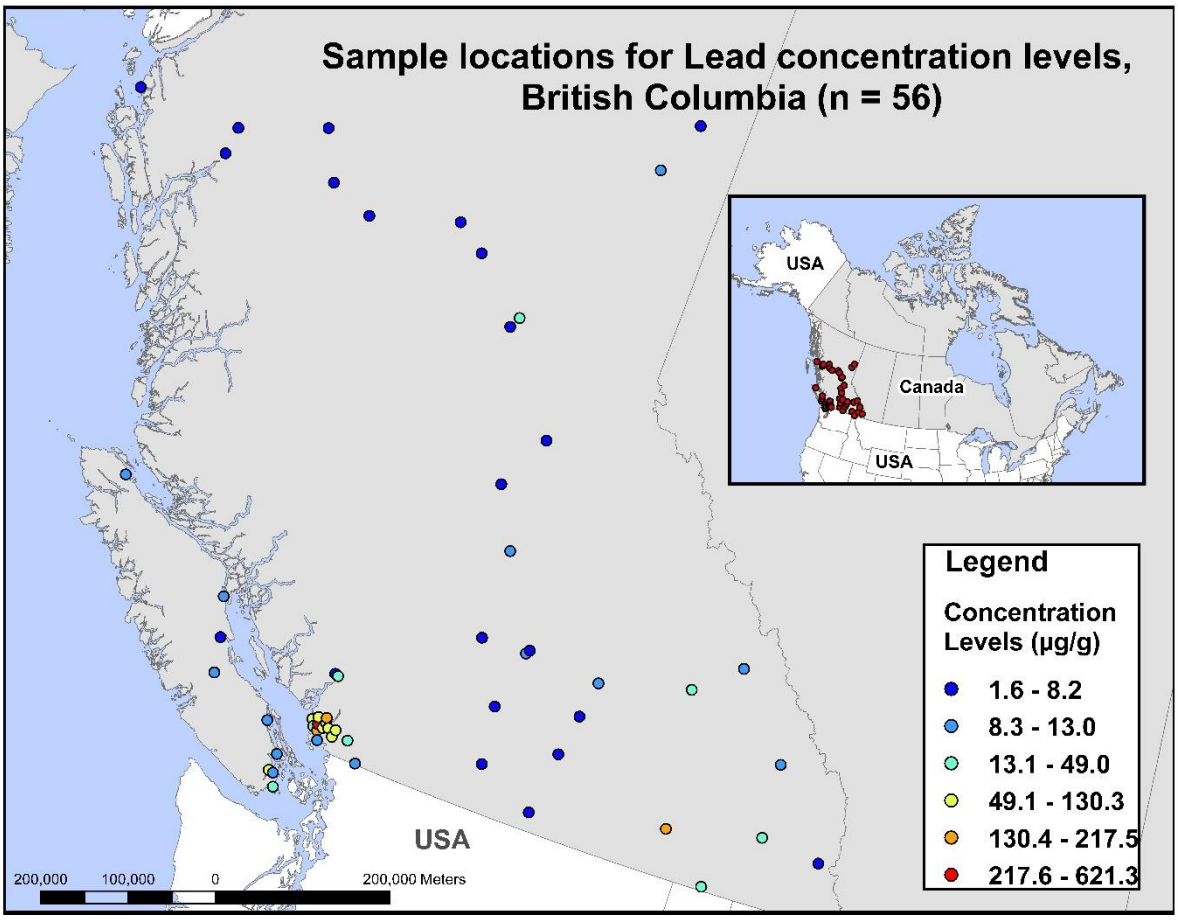


Figure 7 : Concentration level locations for the LUR Lead analysis – British Columbia

1.4.4 –Independent variables for LUR

A comprehensive literature review identified a number of important factors that influence the distribution of As and Pb in soil. Independent variables for the LUR models were developed using available geographic datasets. These factors are described below and include transportation networks, land use, industrial and extraction activities, precipitation, surficial and bedrock geology, elevation and slope, and population density.

1.4.4.1 –Transportation networks

Lead has been phased out of use in gasoline for many years in Canada. However, past deposition remains at the soil surface and is still a potential exposure source of Pb

(Clemente et al., 2008; Sterckeman et al., 2000). Leaded gasoline played a major role in soil contamination along roadsides and urban areas (Mielke et al., 2010; Wu et al., 2010). Therefore, urban areas are more likely to have higher concentrations of Pb in soil than rural areas that are lacking an industrial point source (Adriano, 2001). Although metal contamination along roads decreases rapidly with distance from the roadway (Pagotto et al., 2000), a high density of roads along with heavy traffic can have a significant impact on metal levels in soil (Imperato et al., 2003; Wu et al., 2010).

The settling gradient of vehicle emission ranges from 10% within 0.10 km, 45% within 20 km, 10% between 20 km and 200 km, and up to 35% for further distances (Adriano, 2001). Despite the reduction of Pb emissions, elevated levels are still present in urban areas (Imperato et al., 2003). Furthermore, other metals (Cd, Copper, Cr, Ni, and Zinc) are found in elevated concentrations proximate to roads (Hjortenkrans et al., 2008; Preciado et al., 2005).

1.4.4.2 –Land use

The transformation of the landscape by human activities influences metal concentrations in soil. In fact, urban development, either for industrial, commercial, or residential uses, affects the natural cycle of metals in soil (Tu, 2011). For example, Murray et al. (2004) demonstrated that land uses have different effects on metal soil concentrations for Cd, Cr, Pb, and Ni. It was also demonstrated that spatial patterns of metal levels in topsoil vary with different land uses (agricultural, urban, forested, etc.) (Fritsch et al., 2010).

1.4.4.3 –Industrial and extraction activities

The atmospheric emissions from smelters and the windblown dust from tailing ponds and waste rock piles are important sources of soil contamination (Garcia-Sanchez & Alvarez-Ayuso, 2003 & Nannoni et al., 2011). Moreover, Taylor et al. (2010) showed that mining-related activities have a significant influence within a 2 km radius on soil, for both Cd and Pb concentrations. Garcia-Sanchez & Alvarez-Ayuso (2003) estimate that for As, accumulation usually occurs within 500 meters of the emission source. The concentration levels mostly depend on the distance from the smelter, wind direction and strength, and the organic carbon content of the soil (Fritsh et al., 2010).

Furthermore, the emission of acidification elements (SO_x, NO_x, and CO_2) increases the solubility of metal in soil, and thus influences their mobility in surface soil. Variables include the location of mines and smelters (past and present) and emissions data from the NPRI. This data includes the quantity of metal emitted into the air, water, and soil, as well as the quantity of metals in the tailing ponds and waste rock piles as a result of mining activity.

1.4.4.4 –Precipitation

Precipitation plays an important role in the intensity of wet deposition of contaminants (Novak et al, 2010). Suspended metals are carried down through the atmosphere by water droplets and reach the surface soil. More precipitation around one emission source implies that more contaminants are deposited within a close proximity of that source. In the case of As, precipitation can play an important role in the leaching of the metalloid from the surface. In fact, Arsenite [As (III)] is very soluble and easily leaches

into the groundwater (Brady & Weil, 2002). The data used in this thesis include the average annual precipitation records for 2005 and 2006. Annual averages, including winter and summer averages, were included to control for the type of precipitation.

1.4.4.5 –Surficial and bedrock geology

The weathering and erosion of bedrock releases metals into the environment. Generally, elevated concentrations of metals are likely to be located near an ore-rich deposit. The intensity of weathering and erosion processes controls the spatial distribution of metals in soil (Adriano, 2001).The nature of the parent material from which it was developed has an influence on the resulting soil (Boyle et al., 1998). For this study, it is assumed that Canadian post-glaciated soils (~10,000 years) might still be affected by lower geological layers.

1.4.4.6 –Elevation and slope

According to Adriano (2001), high elevations have a positive influence on metal levels because the precipitation patterns they create trap long range emissions. Steep slopes can also reflect the roughness of a landscape. Elevation variability can have significant influence on the rate of metal deposition. In fact, variability of elevation can increase the surface roughness of a landscape, thus influencing the rate of deposition of pollutants from air to soil and water (Fritsh et al., 2010; Steinnes and Friedland, 2006).

1.4.4.7 –Population density

Higher concentration levels of metals are more common in urban areas (Mielke et al., 2010 and Murray et al., 2001). An increasing gradient of metal concentrations from

rural to urban is widely reported: in the United States (Yesilonis et al., 2008), in China (Gong et al., 2009), and in Italy (Biasioli et al., 2005). The density of urbanization is an important factor that influences the levels of metal in soil. We assumed that higher population density would characterize more elevated concentration of metals in soil, especially for Pb. In the present research, the Canadian census (2006) represents the distribution of the population, which reflects the level of urbanization in a study area.

Table 11 shows a description and the sources of the independent variables for the LUR analysis. The next section describes the pertinence of each independent variable to the metal concentration levels.

Table 11 : Independent variables used in the LUR analysis

Name	Description	Extraction method	Sources (date)
Transportation network (Rails, highway, major and local roads)	Total length of roads (m)	Buffers 500m, 1, 2, 5, 10, 25km	DMTI Spatial (2007)
Land use (Residential, commercial, and industrial)	Total area (m^2)	Buffers 500m, 1, 2, 5, 10, 25km	DMTI Spatial (2007)
National Pollution Release Inventory (NPRI)	Sum to total emission of each metal to air, water, and soil (T)	Buffers 500m, 1, 2, 5, 10, 25km	Environment Canada (from 1994 to 2010)
Mines (past and active)	Presence of mining activity	Buffers 500m, 1, 2, 5, 10, 25km	Ministry of Energy and Mines and responsible for Housing of British Columbia (downloaded in 2012)
Tailing ponds and Waste rock piles	Total amount of stored metal in	Buffers 500m, 1, 2, 5, 10, 25km	Environment Canada (2010)

	tailing and waste rock pills		
Population density	Total population	Buffers 500m, 1, 2, 5, 10, 25km	Statistics Canada (2006)
Total precipitation	Annual, summer, and winter average precipitation (mm)	Closest station	Environment Canada (2006)
Surficial geology	Surficial geology at sample site	Value at point location	Natural Resources Canada (1993)
Bedrock geology	Bedrock geology at sample site	Value at point location	Natural Resources Canada (1993)
Elevation	Elevation (m)	Value at point location	Natural Resources Canada (2002)
Slope	Slope (%)	Value at point location	Natural Resources Canada (2002)

Chapter 2 Modelling Soil Surface and Subsurface Metals Concentration Levels using Geographically Weighted Regression

2.1 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 the soil's parent materials and soil formation processes, but their distribution patterns can also be greatly altered by human activities (Murray et al., 2004). Anthropogenic emissions, from both diffuse and point sources, affect the levels of metal concentrations in soil and make 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 potentially-elevated levels of metals in surface soils can help target efforts to reduce exposure risk for populations by targeting remediation and prevention efforts.

Although heavy metals may have some similarities to each other, they can all behave differently under the same conditions. Thus, the patterns of the concentration levels of the metals are not expected to be the same. Moreover, soil properties are known to vary, both vertically and horizontally, which also influences the metals' mobility. Investigating the concentration variability between surface and subsurface soils can help to determine the degree of relationship between the two layers. For soil exposure assessment, a lack of relationship could be an indicator of anthropogenic influences, 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. The analyses are based on soil surveys where surface and subsurface samples were collected at the same location.

2.2 Methods

2.2.1 –Data

Data were obtained 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). More details about the sources, analytical and digestion methods, concentration levels descriptions, and year of collection of the soil samples are available in Appendix A. The data, which cover several regions of Canada, were available online and obtained between March and September 2010 at (http://www.gov.bc.ca/env/ and http://www.ontario.ca/en/your_government/009883). It is important to note here that these datasets have different spatial distributions. Maps showing the concentration levels at each sample site are available in Appendix B. The Ontario datasets are from geological surveys where one large cluster is present and the rest are distributed throughout the province, whereas the British Columbia sample sites are more evenly spread across the province. These differences are due to the different purposes of the surveys: the samples

collected in Ontario were for mineral explorations, whereas the samples collected in BC are dispersed throughout the province. The samples for BC were collected to determine the background levels of metals for remediation standard purposes.

First, the files were combined together in a matrix. The matrix was sorted by sampling depth (surface and B-horizon), by type of digestion, and by analytical methods. To account for the random sampling assumption in the linear regression analysis, all transects and clusters (50 to 100 samples within a 1 km^2 area) were removed from the data set. Blanks, zeros, or values below the detection limit were excluded. Multiple samples taken at the same location and 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, zeroes, 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 relationship. Before the analysis, the concentration data

were log- and square-transformed to normalize the residuals distribution. Appendix B includes statistic descriptions, concentration maps, scatterplots, and boxplots of the concentration level samples of all the metals used in the GWR analyses.

2.2.2 –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 (Fotheringham, Brunsdon, & Charlton, 2002). The number of points included per equation was determined by the bandwidth of the kernel, which itself was determined by Akaike Information Criterion corrected (AICc).

GWR allows the parameters to vary in space, while showing spatial patterns and providing a better understanding about local phenomena which are 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, as it maximizes the number of sample points to include in the regression as described by Fotheringham, Brunsdon, and Charlton (2002). Finally, the results were analyzed to assess the validity of the regression residuals for both their normality and their spatial autocorrelation.

2.3 Results

The correlation (Pearson's r) between the subsurface and surface concentration levels (log) was 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. The correlations suggest that subsurface levels could be important predictors of surface levels in linear regression models. However, our GWR results showed non-normal residuals for all metals except As, which indicates that linear regression is not effective for predicting surface concentrations of any of these metals (other than As) using only subsurface concentrations. Therefore, only the results of the As model are presented here in greater detail (Table 12). The results for As, Cd, Cr, Co, Ni, and Pb are presented in Appendix B, and include concentration maps, regression results, and residual analysis for all metals in this dataset.

For As, the R-squared at each sample location, based on subsample size of 142 locations (neighbours), 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). Regression and residual analyses, including spatial autocorrelation results for all the metals are available in Appendix B.

Table 12: GWR analysis for Arsenic (log)

Arsenic (n = 877)	Results
Neighbours	142
R^2 adjusted	0.34
Intercept range	0.544 to 1.611
Coefficient range	-0.249 to 0.7555

R^2 range	0.00 to 0.63
Moran's I p-value	0.929112

Figure 7 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 there; however, in the western region, the R-squared values are typically above 0.5.

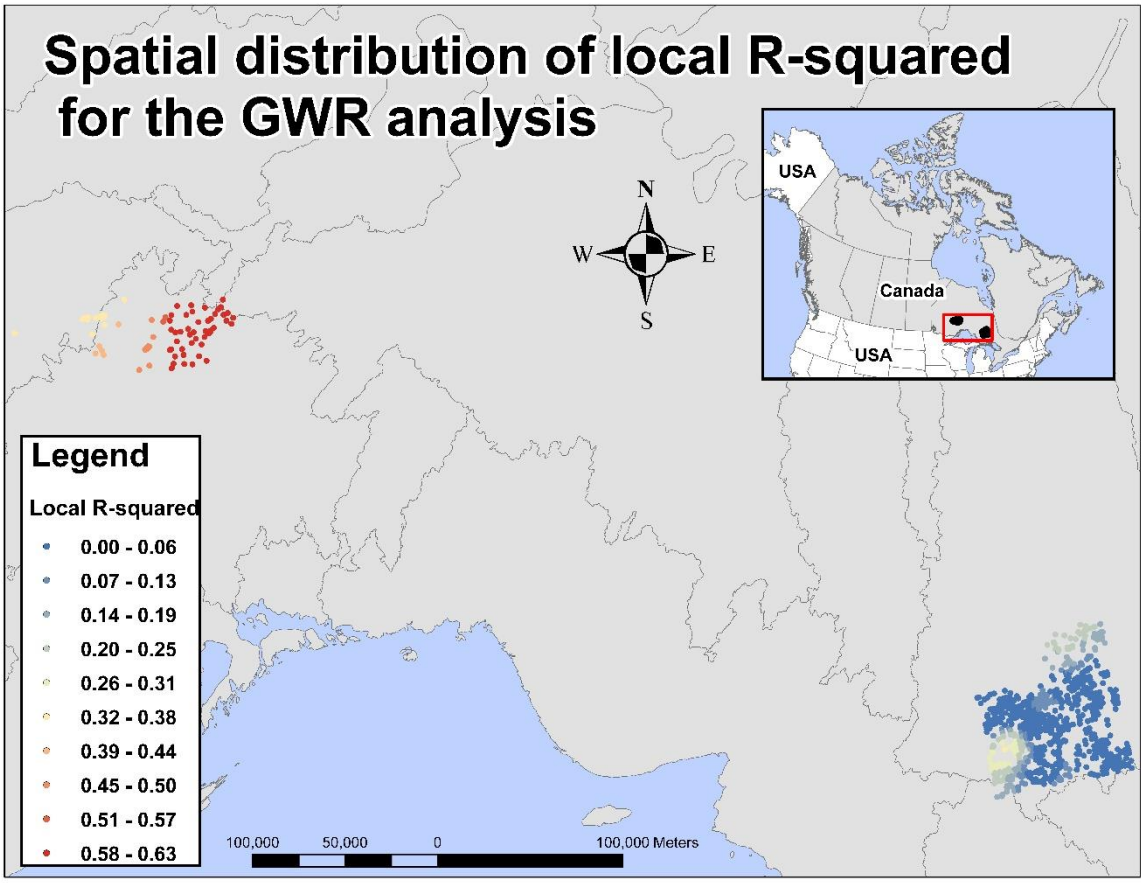


Figure 8: Spatial distribution of the local R^2 from the GWR analysis

Figure 8 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 clusters of negative values by using sample locations with coefficients equal to zero with the majority of points having a positive coefficient. The pattern suggests that for most of the sample sites, the surface level is higher than the subsurface level.

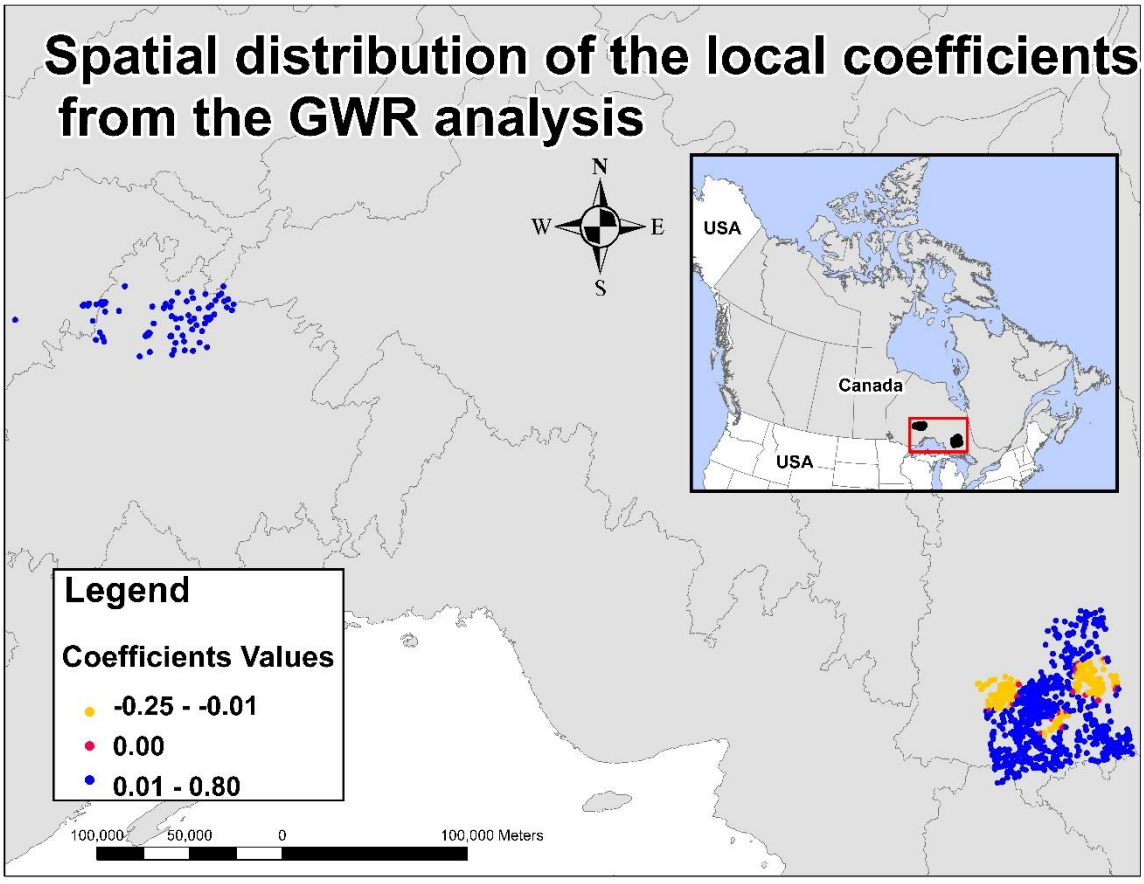


Figure 9: Spatial distribution of the regression coefficients from the GWR analysis

2.4 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 heavy metal surface concentrations at a regional level.

Generally, soil properties and conditions vary both horizontally and vertically in soil (Murray et al., 2004), often over very small areas, and this can influence the relationship between subsurface and surface levels for all studied metals. Our analysis showed that the concentration levels between the two layers are correlated for all the metals, but the level of correlation varies (0.23 to 0.52).

The variation in correlations was not entirely unexpected. In fact, the specific chemical properties of each metal are different under the same soil conditions, which also vary in time and space (Adriano, 2001; Svendsen et al., 2007). Therefore, the relationship between the surface and the subsurface levels should also vary for each metal.

For example, Pb and Cd depositions are mainly from anthropogenic sources and are the most likely to remain bound with the organic matter at the surface. This leads to an accumulation of these metals over long periods of time (Steinnes and Friedland, 2006). Also, the relationship between surface and subsurface may vary extensively over space (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 vertical mobility of the metalloid is better understood, so the relationship between the two soil layers was expected to be more evident. As reported by Adriano (2001), As tends to be more uniformly distributed throughout the soil column. Its presence at the surface level is known to be controlled by both the underlying rocks as well as the atmospheric deposition from nearby emission sources (Murray et al., 2004). Generally, As concentration tends to be lower in the upper layer above the B-horizon due to both its volatilization and its downward leaching movement (Wang and Mulligan, 2006; Zhang and Selim, 2006).

The GWR analysis of As suggests that the relationship between the surface and the subsurface concentration levels varies geographically, and several factors act simultaneously at different scales. However, data on factors that may be involved in the distribution of As in soil (pH, organic content, and soil texture) was not available for this study. That is because this information is difficult to interpolate with accuracy for large geographic regions, as these factors are more site specific. Moreover, the chemical characteristics would be only available at the sample locations and any interpolation between the points will have a definite degree of uncertainty.

Moreover, human activities are well-known to affect metal distribution in soil. In fact, anthropogenic emissions of metal and other human disturbances increase the concentration levels in soil. Human activities may also change the chemical and physical

condition of the soil. Important sources of local contamination include the proximity of mines, industrial emissions, and urban areas (Adriano, 2001). Therefore, spatial data on location and emission intensity as well as from natural sources (geology) may help to improve the prediction of soil surface concentration of As and the other metals.

Furthermore, the INAA method can be a better indicator than other partial digestion analyses, as it represents the total content of As in the soil sample. The total content could be an important factor as the speciation of As influences its mobility in soil (Adriano, 2001). By contrast, all the other metals were analysed by ICP, which represents a only fraction of the total content of the metal in the soil sample.

Finally, any comparison between the surface and the B-horizon might not be adequate to compare the mobility of metals. 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, these C-horizon values were not widely available.

There is also some ongoing debate about whether subsurface levels can 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 at the subsurface. The authors explain that the chemical compositions of the different layers are not chemically similar and therefore, they cannot be compared.

2.5 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 results show that the B-horizon is, to various degrees, related to the surface concentration levels, and may even contribute to the surface concentration levels. Nonetheless, only As showed a statistically robust relationship between the two soil layers using GWR analysis.

The results for As suggest that the metalloid is more mobile through the soil column and its concentration levels are more evenly distributed. For Co, Cr, and Ni, many studies have reported contradictory results regarding the vertical distribution of the metals' concentrations (Adriano, 2001). In the case of Cd and Pb, the capacity of the long-range atmospheric deposition to bind with the organic matter at the soil's surface was suspected to obscure any potential relationships. Therefore, no specific patterns were expected for these metals, except for As. This suggests that other factors than the subsurface itself control the spatial distribution of the metals other than As.

The heterogeneity of soil conditions, chemical and physical, plays a key role in the mobility of all metals. In addition of anthropogenic sources, the dynamic condition of the soil alters the strength of chemical bonds, reducing or increasing the metals' mobility (Adriano, 2001). Unfortunately, the soil parameters (pH, organic content) were only available for a few surveys (MRD136, Background Soil Quality Study) and could not be used for this analysis.

Finally, the concentration data include a compilation of different surveys, between which sample collection methods and laboratory analysis standards may differ. Moreover, there is a gap of ten years between the first survey (MRD015) and the last one (MRD136). These factors might have had significant impacts on the results.

Chapter 3 Modelling Arsenic and Lead Surface Soil Concentrations using Land Use Regression

This chapter contains some duplication of previously reported data and methods because it was written in a format for journal publication.

3.1 Abstract

Elevated concentrations of metal-contaminated soil can be a serious threat to the quality of ecosystems and human health. Exposure assessments often require models that predict concentration levels to estimate the exposure risks for a given population. In environmental exposure assessments, Land Use Regression (LUR) models are increasingly used to predict the concentration of contaminants in outdoor air. We explore the use of LUR as an alternative to more complex models to predict the concentration of metals in surface soil. Here, we used 55 soil samples of As and 56 of Pb collected between 1994 and 1996 across British Columbia (BC), Canada by the Ministry of Environment. Predictor variables were derived either with different buffer sizes or identity values at each sample site using Geographic Information System. For the Pb model (R^2 =0.78), the predictor variables are the total surface area of industrial land use (m^2) within 5 km , the emissions of Pb (t) within 10 and 25 km, and the presence of closed mines within 50 km. For As (R^2 = 0.44), the resulting linear regression model includes the total length of roads (m) within 25 km, and bedrock geology. The LUR model for Pb offers a strong explanation for the variation of the metal in surface soil, which can be partially explained by the strong influence of human emissions. For As, both natural and human sources seem to dictate the surface soil levels.

3.2 Introduction

Both acute and chronic exposure to metals in soil can cause adverse effects on human population and ecosystem health. People living proximate to contaminated sites or emission sources can develop serious health problems, including cancer and neurological disorders (ATSDR, 2013). Unfortunately, an estimation of exposure is difficult to establish because of the complexity of the different pathways and reliable models. An increasing number of studies demonstrate a significant relationship between metal concentrations in soil and health problems (Aelion et al., 2008; Mielke 1998; Madrid et al., 2008; Carrizales et al., 2006).

For example, Aelion et al. (2008) demonstrated the potential relationship between elevated levels of metals in soil and an increased prevalence of mental retardations among young children. The authors estimated the surface concentration levels with Inverse Distance Weighted method to predict the exposure levels and conducted a Principal Component Analysis to identify the emission sources (Aelion et al., 2008). Models that interpolate metal concentration in surface soils are often data and computer intensive, while the spatial resolution of the predicted surfaces are directly related to the size of the study area and the density of sample points. While all models have their uncertainty and complexity levels, the simplification of the model-building process can ease exposure assessments for human health.

First used in air pollution modelling, Land Use Regression (LUR) has since been employed to predict ambient concentrations using simple geographical variables, with results comparable to other more complex models (Jerrett et al., 2005). Once developed, LUR models can be used to predict concentration levels of a pollutant wherever the relevant geographical variables are available, and thereby derive a modelled surface concentration to aid in exposure

assessment for health-related studies. Recently, Wu et al. (2010) applied a LUR approach to predict the concentration of Pb in soil in Los Angeles. The authors used bivariate linear regression to develop a Pb surface concentration based on the relative age of the buildings and traffic variables (Wu et al., 2010). More recently, LUR model predicted the Pb levels in surface soil using traffic, land use, and emission sources variable at a regional scale (Deschênes et al., 2012). These results suggest that LUR may be an effective tool to provide potential exposure estimates of various pollutants in soil for epidemiological studies.

The goal of this study is to apply LUR to predict the concentrations of As and Pb in surface soil across BC. The model variables include publicly available soil samples and GIS-derived metal emission sources and other controlling factors. The objective is to develop a simple and reliable method to predict the concentration levels of As and Pb in surface soil across large areas that might be suitable for an environmental epidemiological study.

3.3 Methods

3.3.1 - Data

Surface soil samples (As = 55 and Pb = 56) were collected in BC between 1994 and 1996 by the Ministry of Environment (Figures 9 and 10). The samples were analysed by ICP-OES, after an *Aqua Regia* digestion. For more information about the collection, preparation, and analysis of the samples, refer to the Ministry website (http://www.env.gov.bc.ca/). Multiple samples taken at the same location were averaged and values below the detection limits were excluded. Table 13 shows the concentration value range, mean and median for both As and Pb.

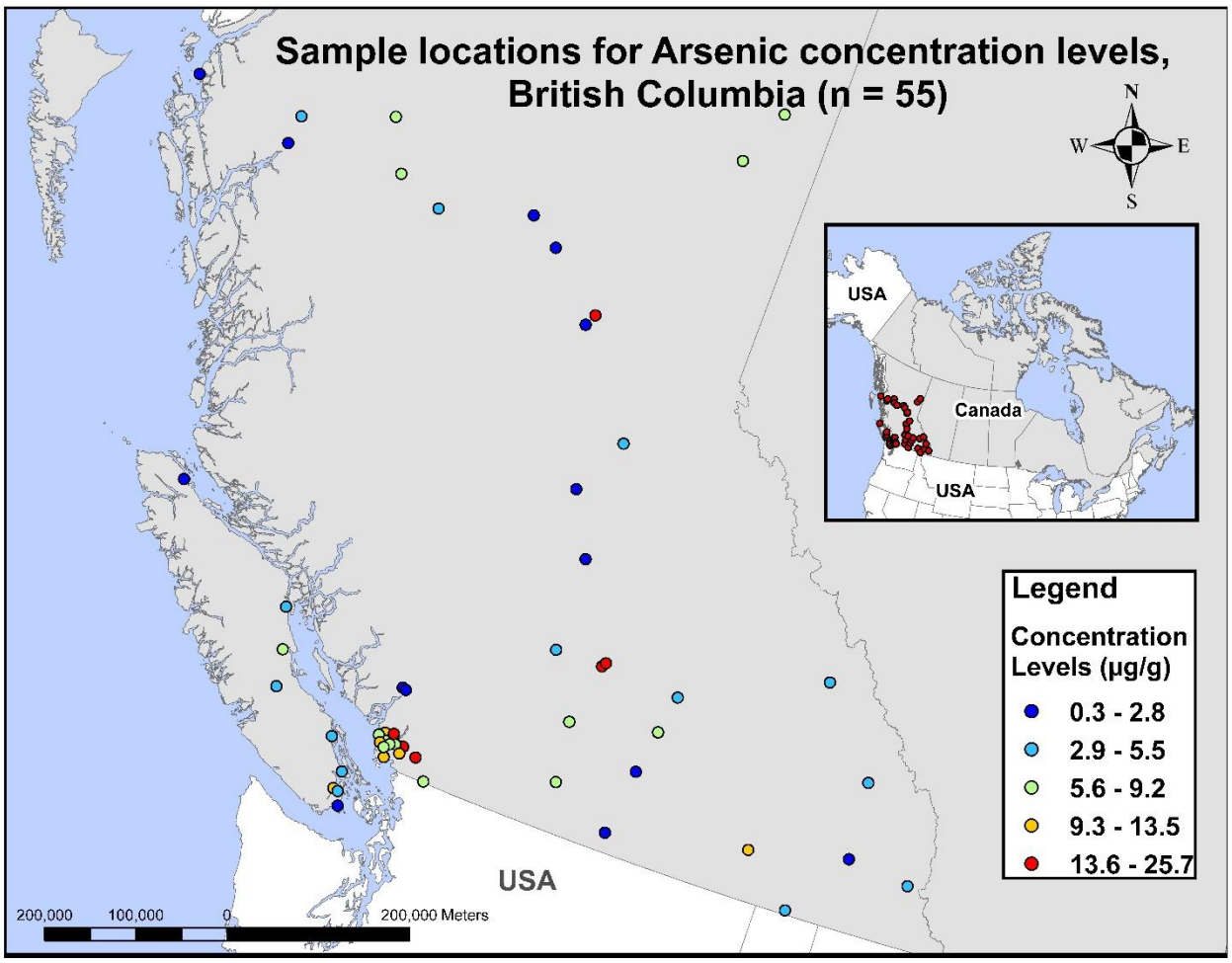


Figure 10 : Concentration level locations for the LUR Arsenic analysis – British Columbia

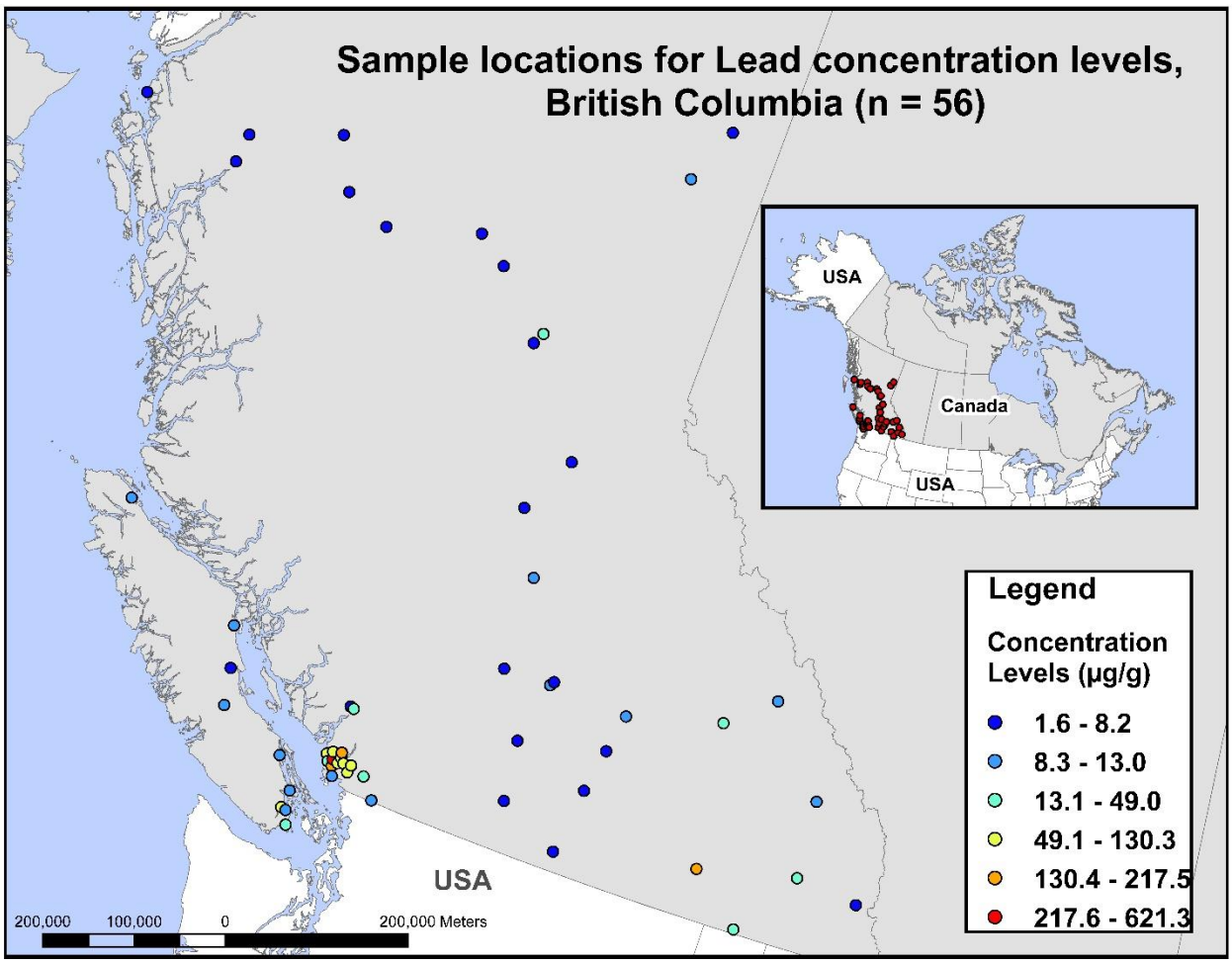


Figure 11 : Concentration level locations for the LUR Lead analysis – British Columbia

Table 13 : Concentration levels description for Arsenic and Lead for LUR analysis (µg/g)

Metals	Minimum	1st quartile	Median	Mean	3rd quartile	Max
Arsenic	0.25	3.05	5.50	6.85	9.05	25.70
Lead	1.63	6.79	9.90	45.66	32.12	621.20

At each sample location, the predictor variables (Table 14) were derived using a range of circular buffer sizes (100m to 50km). The buffer sizes were chosen to represent a proximity gradient of deposition around emission sources as developed by Su et al. (2009). Other site specific variables were derived using the sample location value of the geographic variables (elevation, bedrock geology, etc.). The variables are known sources of metals, natural and

anthropogenic, or surrogates. For example, the Canadian Census, land use types, and road networks can all represent the degree of urbanization, which is an important factor that contributes to higher levels of metals in soil (Murray et al., 2004; Adriano, 2001).

In other cases, the locations of mines, direct emissions, tailing ponds, and waste rock piles are the primary sources of metals in the environment (Adriano, 2001). The intensity of emissions from these sources is directly related to the surrounding soil concentration levels (Steinnes & Friedland, 2006; Adriano, 2001). Other natural factors were also included to represent the roughness of the landscape. In fact, the terrain relief dictates the degree of deposition; slope and elevation are used as roughness indicators (Adriano, 2001; Fritsch et al., 2010). All these features have a potential impact on the level of metals in surface soil.

Table 14 : Independent variables used in the LUR analysis

Name	Description	Extraction method	Sources (date)
Transportation network (Rails, highway, major and local roads)	Total length of roads (m)	Buffers 500m, 1, 2, 5, 10, 25km	DMTI Spatial (2007)
Land use (Residential, commercial, and industrial)	Total area (m^2)	Buffers 500m, 1, 2, 5, 10, 25km	DMTI Spatial (2007)
National Pollution Release Inventory (NPRI)	Sum to total emission of each metal to air, water, and soil (T)	Buffers 500m, 1, 2, 5, 10, 25km	Environment Canada (from 1994 to 2010)
Mines (past and active)	Presence of mining activity	Buffers 500m, 1, 2, 5, 10, 25km	Ministry of Energy and Mines and responsible for Housing of British Columbia (downloaded in 2012)
Tailing and Waste rocks	Total amount of	Buffers 500m,	Environment Canada

	stored metal in tailing and waste rock pills	1, 2, 5, 10, 25km	(2010)
Population density	Total population	Buffers 500m, 1, 2, 5, 10, 25km	Statistics Canada (2006)
Total precipitation	Annual, summer, and winter average precipitation (mm)	Closest station	Environment Canada (2006)
Surficial geology	Surficial geology at sample site	Value at point location	Natural Resources Canada (1993)
Bedrock geology	Bedrock geology at sample site	Value at point location	Natural Resources Canada (1993)
Elevation	Elevation (m)	Value at point location	Natural Resources Canada (2002)
Slope	Slope (%)	Value at point location	Natural Resources Canada (2002)

3.3.2 - Modeling approach

Concentration levels were transformed using the natural logarithm because they were positively skewed. The boxplots showing the distribution of the concentration levels are available in Appendix C (Figures 184, 185, 203, and 204). Bivariate analysis identified the most correlated buffer size per category of explanatory variables with the concentration levels (log). The most correlated variables within a category were included in the regression analysis for the first variable selection as per Su et al. (2009). The correlation analysis between the independent and the dependent variables are available in Appendix C (Figures 186 -196 and 205 – 218). The other site specific variables (precipitation, elevation, slope, surficial and bedrock geology) were also added to the first model.

Next, backward and stepwise selections were performed to identify the most significant variables. Then, a variable that was collinear (variance inflator factor > 10), not significant, or

had a relationship contrary to that expected was eliminated. The two previous steps were repeated until a final model including only significant variables was produced (Henderson et al., 2007). The residuals were analysed to insure that the models met all the regression assumptions, including normally distributed residuals without spatial autocorrelation.

The sensitivity of the models was examined by bootstrap analysis to assess how the model parameters (coefficients and R^2) varied when the LUR models were sub-sampled. Random selections of the soil concentration samples, with replacements, were repeated 10,000 times and the respective parameter results for 95% confidence interval were recorded. Statistical analyses were conducted using the open-source software R 2.14.1 (R Core Team, 2013). For the bootstrap analysis, the residuals normality and the spatial autocorrelation for both the As and Pb models were not calculated.

3.4 Results

The final LUR models explain the variation of the metals in the surface soil of As and Pb at 44% and 78%, respectively. Tables 15 and 16 summarize the model parameters. The significant predictors (the variance explained by the variable is in brackets) for As include the length of local roads (log) within 25km (28%), and the bedrock geology (9%). The parameters for Pb include industrial land use within 5km (10%), the industrial emissions of Pb within 25 km (log) (4%) and 10 km (7%), and the presence of closed mines within 50 km (12%).

Table 15 : Regression analysis results for Arsenic

As Model	Value	SE	t	p
Intercept	-5.47857	1.36517	-4.013	0.0002
Length of Roads (m) within 25 km (log)	0.49839	0.09166	5.437	2.39e-6
Mafic intrusive rocks – dorite[1]	-1.36899	0.65123	-2.102	0.04143

Table 16 : Regression analysis results for Lead

Pb Model	Value	SE	t	p
Intercept	1.82	1.35e-1	13.464	< 2e-16
Industrial land use within 5 km	1.49e-7	2.94e-8	5.069	5.62e-6
Industrial Emissions within 25 km (log)	9.76e-2	2.96e-2	3.303	0.00175
Industrial Emissions within 10 km	5.26e-5	1.21e-5	4.344	6.67e-5
Presence of Closed Mines within 50 km	1.199e-2	2.18e-3	5.501	1.23e-6

Table 17 shows the bootstrap 95% confidence intervals for the R^2 and the individual coefficient variables. The R^2 ranges from 0.3 to 0.9 with a mean of 0.63 for As and from 0.35 to 0.95 with a mean of 0.80 for Pb. For As, bedrock geology is sensitive to resampling because the coefficients of many categorical variables ranged from negative to positive values and were not always significant, except for mafic intrusive rocks – diorite and gabbro. For the Pb model, all the coefficients remained positive suggesting an adequate consistency in the resampling process.

[1] Other categorical variable values are not shown here, only the significant.

Table 17 : Bootstrap results for Arsenic and Lead

As Model	Model Value	95% CI[a]
R^2	0.44	0.31 to 0.89
Intercept	-5.479	-8.23 to -2.72
Length of Roads (m) within 25 km (log)	0.498	0.31 to 0.68
Mafic intrusive rocks – dorite[2]	-1.369	-2.68 to -0.06
Pb Model		
R^2	0.78	0.35 to 0.95
Intercept	1.82	1.59 to 2.14
Industrial land use (m^2) within 5 km	1.49e-7	9.04e-8 to 2.12e-7
Industrial Emissions within 25 km (log)	9.76e-2	3.50e-2 to 1.51e-1
Industrial Emissions within 10 km	5.26e-5	2.56e-5 to 7.81e-5
Presence of Closed Mines within 50 km	1.199e-2	4.69e-3 to 1.58e-2

Figures 11 and 12 show the predicted concentrations at the soil surface for As and Pb for southwest BC. For the Pb model, predicted values (27 points) were truncated to the highest concentration level because they were above the maximum level of the original sample data. This manipulation is necessary because values outside the concentration range might not be linear. Detailed results for both the As and Pb models are available in Appendix C.

[2] Only significant categorical variable of the model

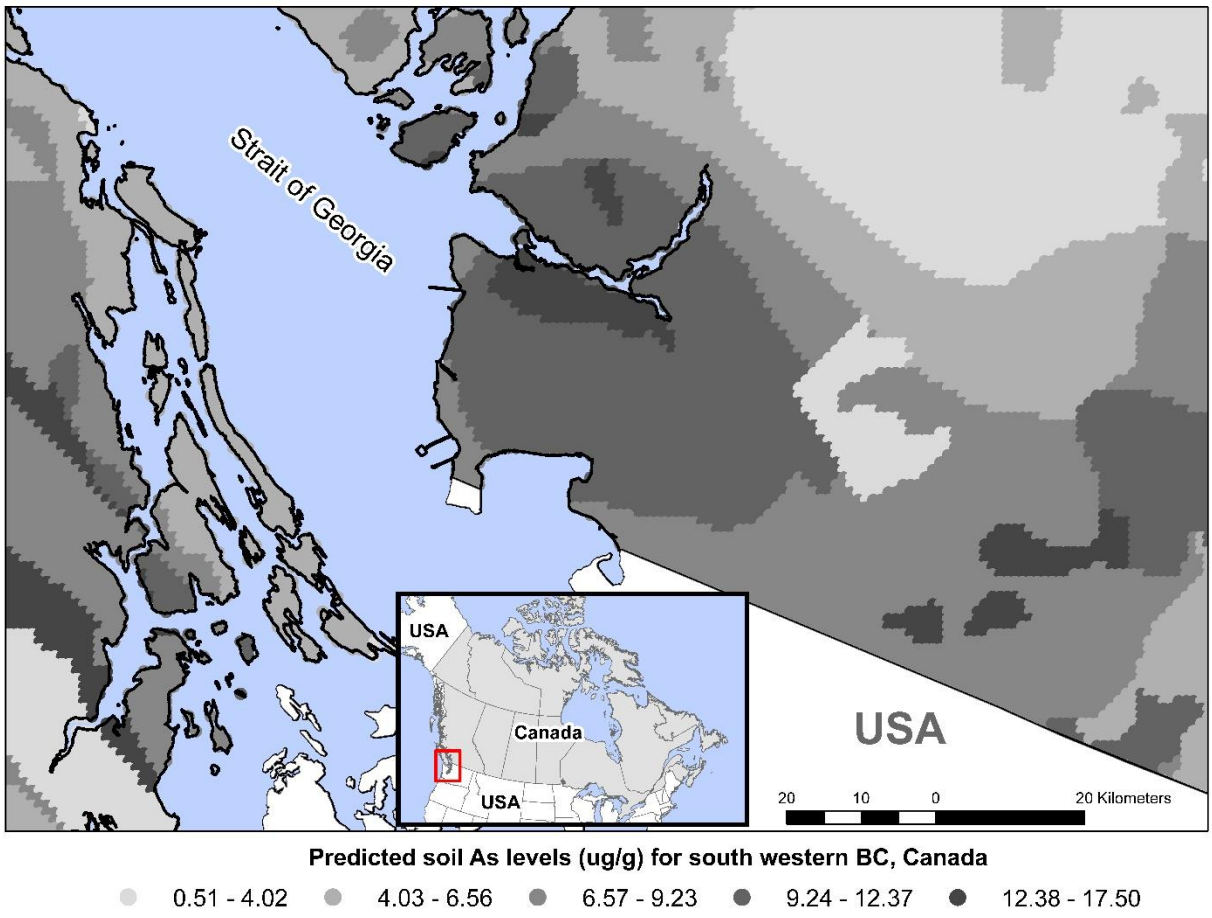


Figure 12 : Map of predicted soil Arsenic levels (µg/g) for Southern British Columbia

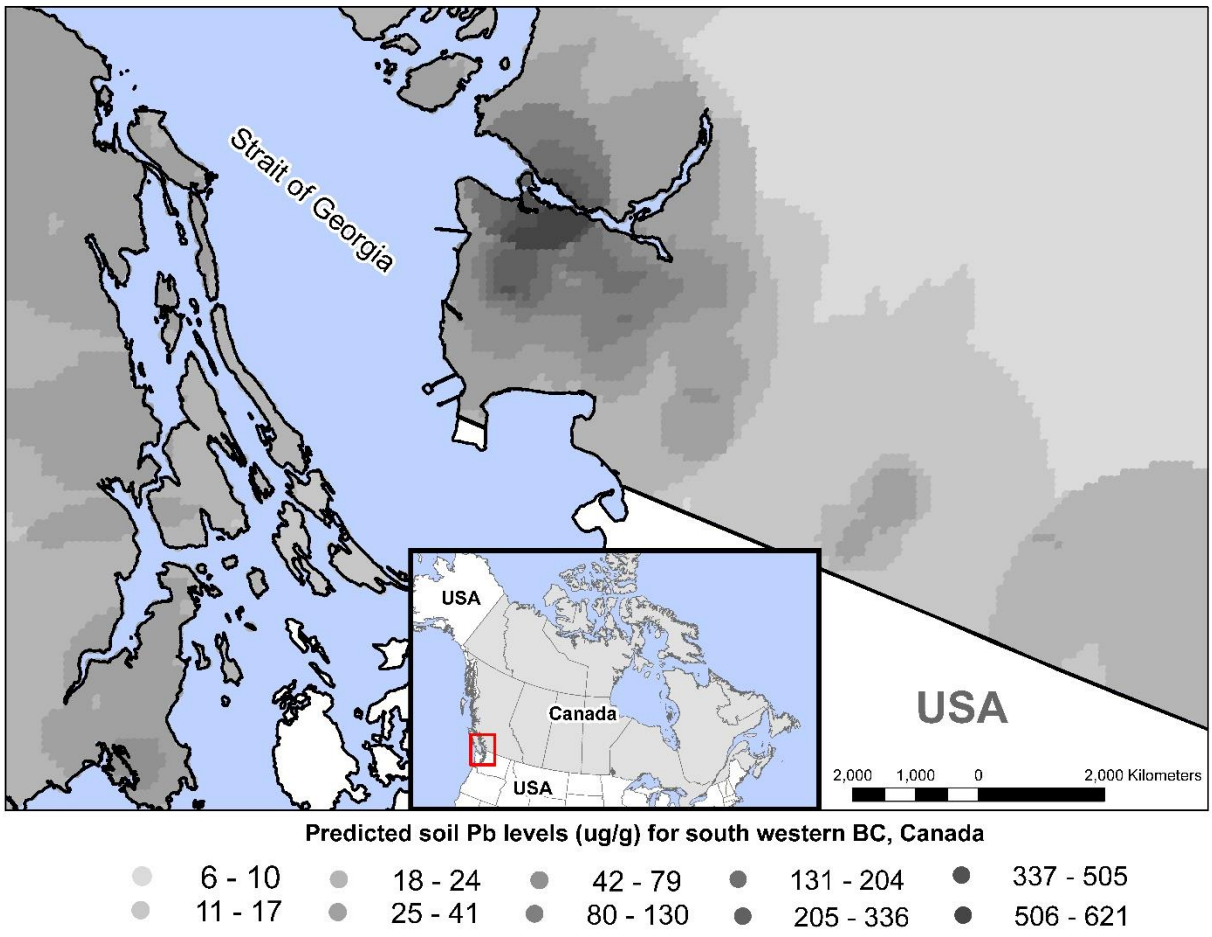


Figure 13 : Map of predicted soil Lead levels (µg/g) for Southern British Columbia

3.5 Discussion

This study shows the potential of using LUR to predict the concentration of As and Pb in surface soil at regional levels. Overall model performance (As R^2 = 0.44 and Pb R^2 = 0.78) is similar to that obtained for air pollution models (Jerrett et al., 2005; Su et al., 2009).

The bedrock variable includes Canadian bedrock formation at or near the surface landmass (Wheeler et al., 1997). For the As model, bedrock geology reflects the natural source of the metalloid in soil, which agrees with other studies (Garcia-Sanchez and Alvarez-Ayuso, 2003; Grosz et al., 2004); however, the bedrock variable is qualitative and assumes continuous and homogenous As concentration in the bedrock unit, which may not always be the case

(Garcia-Sanchez and Alvarez-Ayuso, 2003). Although the accuracy of the geological map was improved (Wheeler et al., 1997), the bedrock type boundaries might not be exactly representative.

In fact, few sample sites were not exactly overlapping the geological map and thus, were given the value of the closest polygon. Furthermore, the bedrock geology variable represents different categories of rocks, which can be rough descriptors of As content levels. Rock type categories would have been a better indicator of As variation between geological features than bedrock categories. The road variable is considered a surrogate for human activities related to the utilization of As-containing pesticide and wood preservatives, combustion of fossil fuels (mainly coal plants), and mining activities that are known to increase the concentration levels of As in the surrounding soils (Wang and Mulligan, 2006; Garcia-Sanchez and Alvarez-Ayuso, 2003). However, the relative higher mobility of As in soil may explain the lesser performance of the model compaired to Pb, as past emissions may have leached down beneath the surface.

For Pb, the predominant predictors are industrial land use surface area and Pb emission point sources (NPRI). This finding agrees with other studies that demonstrated a strong influence of industrialization on the Pb surface soil concentrations (Murray et al., 2004; Salonen and Korkka-Niemi, 2007; Wu et al., 2010).

Wu *et al.* (2010) used a spatial approach similar to the current study, but at local scale. The authors suggest that Pb concentration levels in surface soil are strongly related to proximity to roads and land use type. Interestingly, the current Pb model does not include any transportation related variables, which is unexpected as other studies reported the legacy of Pb-gasoline emissions from vehicles (Adriano, 2001; Deschênes et al, 2012; Wu et al., 2010).

Unlike more complex dispersion models, the LUR approach does not typically consider factors controlling the deposition rate and direction from sources. Many studies reported that topography roughness, along with wind speed and direction, influence atmospheric deposition – and thus the concentration of As or Pb in surface soil (Adriano, 2001; Fritsch et al., 2010). The use of circular buffers to define geographic variables can result in distribution patterns in the predicted surface that may not reflect actual distribution. This effect is clearly visible in the Pb concentration surface map (Figure 12) where circular concentration levels patterns are present.

Other researchers have used regression-Kriging to interpolate metal concentration in surface soil over a large area (Lado et al., 2008). This method could not be applied with the present dataset because of the low spatial density of the sample locations; 56 samples over 949110 km^2 giving one sample per 16,948 km^2. In contrast, in a study performed by Lado et al. (2008), the study area was 4,217,241 km^2 for a total of 1588 sample sites, giving one sample per 2,655 km^2.

In this study, there is a discrepancy between the date of the data collection of the soil samples (1994 - 1996) and certain predictor variables (see Table 14). The predictor variables were used as indicators of human activities. In this study, we assume that the changes in the transportation network, population density, and land use type were minimal. For the NPRI, emission levels are used to indicate the location of the source. The emission data are reported from 1994, but we assume that the emission sources were active prior the beginning of the monitoring program and could be used as emission intensity indicators. Finally, the predictive maps exclusively represent the situation at the time of the data collection (1994 - 1996). Any

remediation or contamination that took place after the data collection is not accounted for in the resulting map of predicted concentrations.

In conclusion, while limitations exist, LUR soil models for As and Pb may be useful as screening tools for remediation strategy and, more important, for human exposure assessments. The simplicity of LUR and the increasing availability of accurate digital data can provide cheap, easy, and reliable exposure models for large areas.

Chapter 4 Summary and conclusion

Although all metals occur naturally in the environment, anthropogenic activities can dramatically increase their concentration levels, especially around emission sources. Elevated exposure to heavy metals poses a serious threat to human health and the health of other species. Exposure models can help to determine potential exposure levels and are important tools for risk assessments to these populations.

Soil models can provide valuable information about exposure levels. In fact, soil is an important pathway, as it plays the role of a metal sink as well as a direct source of metal exposure (Mielke & Reagan, 1998). The distribution patterns of soil metal concentrations is key to providing accurate evaluations of exposure risk. Simplification of exposure models, while maintaining high levels of accuracy, can reduce the cost and time of risk assessments.

This thesis presents the results of spatial regression analysis on soil metal concentration levels from publicly-available data. First, GWR was used to investigate the spatial variation of the relationships between the surface and subsurface concentration levels of As, Cd, Co, Cr, Ni, and Pb, in order to inform subsequent modelling. Second, LUR was used to develop two predictive surface soil models for As and Pb across BC, using a series of anthropogenic and natural variables. To our knowledge, this is the first ever study to explore the feasibility of predicting surface soil concentration levels for epidemiological research with publicly-available data at a regional scale.

The concluding chapter to follow provides a summary of the research, discusses the connection between Chapters 2 and 3, addresses the limitations of the study, and outlines opportunities for future research.

4.1 Summary of the findings

4.1.1 –Geographically Weighted Regression

In Chapter 2, GWR was used to assess the spatial variability of the relationships between the surface and the subsurface concentration levels of As, Cd, Co, Cr, Ni, and Pb. Despite a significant degree of correlation between the metal levels of the two soil layers, the spatial variation of Cd, Cr, Co, Ni, and Pb concentrations calculated with GWR, was not associated. Only As had a valid regression model, while exhibiting a low spatial relationship level between the two soil layers.

The GWR was a rapid test designed to highlight the potential influence of the subsurface to the surface concentration levels. The main goal of the research was to estimate the human influence on the soil surface concentration levels. If the subsurface alone could explain the spatial variation of the metal levels, it would have been irrelevant to develop LUR models.

4.1.2 –Land Use Regression

In Chapter 3, we explored the capacity of LUR to predict the soil surface concentration levels of As and Pb using soil samples and emission factors. The emission factors were chosen because of their impacts on the soil metal concentration levels. Two soil models developed with LUR predicted the concentration of As and Pb across BC. The As model contains bedrock geology and the total length of roads as independent variables, while the Pb model includes industrial-activity-related independent variables (the number of closed mines, any industrial

land use, and any industrial Pb emissions). From the models, predicted concentration maps were developed as potential risk assessment tools for southwest BC.

4.2 Research limitations and opportunities

Models are only one representation of reality, so they all have their limitations. The next section notes the limitations of the GWR and LUR models and discusses challenges encountered during the research. The main issues were data quality and availability, soil and metal chemical complexity, temporal discrepancies between the variables, metal deposition patterns, and soil sample spatial distribution.

4.2.1 –Data quality

Combining different sources of data is not always possible because of the many differences between the various surveys. Surveys often differ from each other by their sample location patterns and their soil analytical and digestion methods.

This study would be improved with the use of more comprehensive analyses of the soil samples. This would include the physical and chemical properties of soil (texture, pH, moisture and organic content, and soil type).

Moreover, the independent variables also had discrepancies. For example, a transportation network is considered a simple linear feature, without any variation of emission intensity along the line segments. Certain portions of the road would have a higher impact, as the density of the traffic and the width of the road would also change. These factors are not reflected in the LUR model.

Likewise, the land use variables are a broad category. The different emission factors within a category are not considered. For example, industrial land use includes various activities

that are not necessarily metal-emitting; however, this problem is partially solved by adding the NPRI emission factors to the analysis.

Although the NPRI database helped to determine the exact locations of emission sources, the dispersion patterns due to wind speed and direction were not included in the analysis. For this research, developing a simple model was the primary goal. Moreover, complex dispersion models would only be useful around a local emission source and would not be applicable at the regional scale.

The precipitation variable was an annual average of the 2005 and 2006 data collection. Although the yearly amount of precipitation differs for each station, the models aimed to include climate variations between the regions where the soil samples were collected. The models assume that 2005 and 2006 are representative of annual patterns. In addition, many rain-gauge stations were located far from the soil sample locations. There is likely to be some discrepancy between the actual precipitation value at the sample location and the closest rain-gauge station.

In the case of the surficial and bedrock geology, the variables used were polygon vectors. There was a lack of accuracy between certain sample locations and the polygons. Certain samples were not overlying any polygons and therefore were attributed to the closest polygon value. To improve the relevance of these geology variables, a suggestion is to record the actual type of geology *in-situ* while collecting the soil sample.

4.2.2 –Soil and metals dynamic complexity

In the case of GWR, the heterogeneity of the soil characteristics combined with the dynamics of soil conditions create a complex system that might not be explainable by

exclusively comparing the surface and subsurface soil layers. Factors such as human emissions and disturbances, plant uptake, and soil processes all alter the distribution patterns of metal concentrations in soil. Metal concentrations in soil are not expected to be in a state of equilibrium, as the physical and chemical conditions of soil are constantly being modified. Thus, the vertical and horizontal patterns are also consistently variable.

Moreover, the digestion and analysis methods of the soil samples from the surveys varied. For example, As samples were analysed by INAA, while the other metals' soil samples were dissolved by *Aqua Regia* and analysed by different ICP methods. These discrepancies decrease the certainty of the adequacy of these GWR analyses.

4.2.3 –Temporal discrepancy

This analysis would greatly improve with more recent soil samples. A central challenge of this study was the temporal discrepancies between the date of collection of the soil samples and the independent variables. For example, the NPRI emission data were from 1994 to 2010 and the collection dates were from 1994 to 1996. Although the Pb model predicted well, it was a major assumption to simply accept the latest NPRI emission factors.

4.2.4 –Concentration level patterns

For the LUR models, the circular buffers do not necessarily represent the spatial zone of influence of the emission sources and do not consider the landscape morphology, which also affects the atmospheric depositions. Integration of more representative buffer shapes, which would include atmospheric deposition patterns and the morphologic characteristics of the landscape at each sample location, might improve the LUR models.

4.2.5 –Sample spatial distribution

Publicly-available soil surveys are usually not available for urban areas. Most of the surveys obtained for this study were for mineral extraction purposes, thus the most samples were taken in remote areas. In the case of As, Garcia-Sanchez & Alvarez-Ayuso (2003) suggest that the accumulation of As around a mineis higher within 500 meters of it. Moreover, soil samples used to develop the LUR models were not collected near emission sources. In our study, most soil samples were collected in remote areas that were too far away from point emission sources. As a result, the analysis would not integrate As mining activities in the surrounding areas.

Another problem with these surveys is the clustered patterns of the sample locations over distinct and relatively small, remote areas. In these areas, there is a low population density and little man-made infrastructure. This reduces the spatial variation of the independent variables and reduces the effectiveness of the analysis. The spatial distribution of these soil surveys might not be adequate for this type of study. Combining different surveys together created patches of clusters over large regions. The patchiness of the surveys produces clusters of points that greatly contribute to high levels of spatial autocorrelation.

Furthermore, the training documents provided by Arc Map 10 (ESRI. Redlands, CA) suggest that at least several hundred points are required to obtain valid results with the GWR tool. Although all surveys combined satisfy this criteria (except for Cd), most of the surveys taken individually did not have enough soil samples to produce meaningful results. Only the survey MRD015 from the Ministry of Ontario had more than 700 points.

4.3 Conclusion

This research analyzed the spatial distribution of publicly-available soil samples with two types of regression methods at regional scale. GWR can easily integrate other factors controlling the mobility of metals in soil and improve our understanding about their spatial distribution. Meanwhile, LUR models are a potential tool to monitor human exposure levels for purposes of risk assessment by cheaply predicting metal concentrations in surface soil.

Future efforts to develop metal soil models would require up-to-date soil samples and model validation. LUR is a relatively inexpensive and simple tool for exposure risk assessments and can potentially be developed at a national scale. Heavy metals exposure poses a serious risk to Canadian health. Chronic and acute exposures of heavy metals are dangerous; their emission levels should be monitored and reduced.

Bibliography

ATSDR (Agency for Toxic Substances and Disease Registry). U.S. Department of Health and Human Services. http://www.atsdr.cdc.gov/ (accessed February 2013).

Adriano, D. C. (2001). *Trace Elements in Terrestrial Environments: Biogeochemistry, Bioavailability, and Risks of Metals*. (2nd edition ed.) Springer.

Aelion, C. M., Davis, H. T., McDermott, S., & Lawson, A. B. (2008). Metal concentrations in rural topsoil in South Carolina: Potential for human health impact. *Science of The Total Environment, 402,* 149-156.

Ajmone-Marsan, F. & Biasioli, M. (2010). Trace Elements in Soils of Urban Areas. *Water Air and Soil Pollution, 213,* 121-143.

Baffi, C., Bettinelli, M., Beone, G. M., & Spezia, S. (2002). Comparison of different analytical procedures in the determination of trace elements in lichens. *Chemosphere, 48,* 299-306.

Boyle, D. R., Turner, R. J. W., & Hall, G. E. M. (1998). Anomalous arsenic concentrations in groundwaters of an island community, Bowen Island, British Columbia. *Environmental Geochemistry and Health, 20,* 199-212.

Brady, N. & Weil, R. (2002). *The Nature and Properties of Soils*. (13 ed.) Pearson Education Inc.

Cahill, T. M. & Mackay, D. (2003). Complexity in multimedia mass balance models: When are simple models adequate and when are more complex models necessary? *Environmental Toxicology and Chemistry, 22,* 1404-1412.

Carrillo-Gonzalez, R., Simunek, J., Sauve, S., & Adriano, D. (2006). *Mechanisms and pathways of trace element mobility in soils*. (vols. 91) SAN DIEGO: ELSEVIER ACADEMIC PRESS INC.

Clemente, R., Dickinson, N. M., & Lepp, N. W. (2008). Mobility of metals and metalloids in a multi-element contaminated soil 20 years after cessation of the pollution source activity. *Environmental Pollution, 155,* 254-261.

Davis, H. T., Aelion, C. M., McDermott, S., & Lawson, A. B. (2009). Identifying natural and anthropogenic sources of metals in urban and rural soils using GIS-based data, PCA, and spatial interpolation. *Environmental Pollution, 157,* 2378-2385.

Deschênes, S., Setton, E., Hystad, P., Demers, P., & Keller, P. (2012). Proceedings of the ICHMET-15 Conference Heavy Metals in the Environment. Gdansk: University of Technology, Poland.

Dokmeci, A. H., Ongen, A., & Dagdeviren, S. (2009). Environmental Toxicity of Cadmium and Health Effect. *Journal of Environmental Protection and Ecology, 10,* 84-93.

Facchinelli, A., Sacchi, E., & Mallen, L. (2001). Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. *Environmental Pollution, 114,* 313-324.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). Geographically Weighted Regression: the analysis of spatially varying relationships. John Wiley & Sons Ltd.

Fritsch, C., Giraudoux, P., Coeurdassier, M., Douay, F., Raoul, F., Pruyot, C. et al. (2010). Spatial distribution of metals in smelter-impacted soils of woody habitats: Influence of landscape and soil properties, and risk for wildlife. *Chemosphere, 81,* 141-155.

Gao, J. B. & Li, S. C. (2011). Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression. *Applied Geography, 31,* 292-302.

Garcia-Sanchez, A. & Alvarez-Ayuso, E. (2003). Arsenic in soils and waters and its relation to geology and mining activities (Salamanca Province, Spain). *Journal of Geochemical Exploration, 80,* 69-79.

Gerritse, R. G. (1996). Dispersion of cadmium in columns of saturated sandy soils. *Journal of Environmental Quality, 25,* 1344-1349.

Goovaerts, P. (1999). Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma, 89,* 1-45.

Grosz, A. E., Grossman, J. N., Garrett, R., Friske, P., Smith, D. B., Darnley, A. G. et al. (2004). A preliminary geochemical map for arsenic in surficial materials of Canada and the United States. *Applied Geochemistry, 19,* 257-260.

Henderson, S. B., Beckerman, B., Jerrett, M., & Brauer, M. (2007). Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environmental Science & Technology, 41,* 2422-2428.

Hjortenkrans, D. S. T., Bergback, B. G., & Haggerud, A. V. (2008). Transversal immission patterns and leachability of heavy metals in road side soils. *Journal of Environmental Monitoring, 10,* 739-746.

Imperato, M., Adamo, P., Naimo, D., Arienzo, M., Stanzione, D., & Violante, P. (2003). Spatial distribution of heavy metals in urban soils of Naples city (Italy). *Environmental Pollution, 124,* 247-256.

Imrie, C. E., Korre, A., Munoz-Melendez, G., Thornton, I., & Durucan, S. (2008). Application of factorial kriging analysis to the FOREGS European topsoil geochemistry database. *Science of The Total Environment, 393,* 96-110.

Islam, J., Singhal, N., & O'Sullivan, M. (2001). Modeling biogeochemical processes in leachate-contaminated soils: A review. *Transport in Porous Media, 43,* 407-440.

Jeon, W., Shin, C. B., Kim, J. H., Kwak, B. K., Yi, J., Lee, J. L. et al. (2008). Analysis of the distribution of lead concentration under steady state conditions in urban multimedia environment. *Korean Journal of Chemical Engineering, 25,* 1401-1406.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T. et al. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology, 15,* 185-204.

Krcmova, K., Robertson, D., Cveckova, V., & Rapant, S. (2009). Road-deposited sediment, soil and precipitation (RDS) in Bratislava, Slovakia: compositional and spatial assessment of contamination. *Journal of Soils and Sediments, 9,* 304-316.

Lado, L. R., Hengl, T., & Reuter, H. I. (2008). Heavy metals in European soils: A geostatistical analysis of the FOREGS Geochemical database. *Geoderma, 148,* 189-199.

Laidlaw, M. A. S. & Filippelli, G. M. (2008). Resuspension of urban soils as a persistent source of lead poisoning in children: A review and new directions. *Applied Geochemistry, 23,* 2021-2039.

Lambert, T. W., Boehmer, J., Feltham, J., Guyn, L., & Shahid, R. (2011). Spatial Mapping of Lead, Arsenic, Iron, and Polycyclic Aromatic Hydrocarbon Soil Contamination in Sydney, Nova Scotia: Community Impact From the Coke Ovens and Steel Plant. *Archives of Environmental & Occupational Health, 66,* 128-145.

Legendre, P. (1993). Spatial Autocorrelation - Trouble Or New Paradigm. *Ecology, 74,* 1659-1673.

Leung, Y., Mei, C. L., & Zhang, W. X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A, 32,* 9-32.

Mackay, D. (1979). Finding Fugacity Feasible. *Environmental Science & Technology, 13,* 1218-1223.

Madrid, L., Diaz-Barrientos, E., Ruiz-Cortes, E., Reinoso, R., Biasioli, M., Davidson, C. M. et al. (2006). Variability in concentrations of potentially toxic elements in urban parks from six European cities. *Journal of Environmental Monitoring, 8,* 1158-1165.

MacLeod, M., McKone, T. E., & Mackay, D. (2005). Mass Balance for Mercury in the San Francisco Bay Area. *Environmental Science & Technology, 39,* 6721-6729.

Matschullat, J., Ottenstein, R., & Reimann, C. (2000). Geochemical background - can we calculate it? *Environmental Geology, 39,* 990-1000.

Mielke, H. W., Gonzales, C. R., Cahn, E., Brumfield, J., Powell, E. T., & Mielke, P. W. (2010). Soil arsenic surveys of New Orleans: localized hazards in children's play areas. *Environmental Geochemistry and Health, 32,* 431-440.

Mielke, H. W., Laidlaw, M. A. S., & Gonzales, C. (2010). Lead (Pb) legacy from vehicle traffic in eight California urbanized areas: Continuing influence of lead dust on children's health. *Science of The Total Environment, 408,* 3965-3975.

Mielke, H. W. & Reagan, P. L. (1998). Soil is an important pathway of human lead exposure. *Environmental Health Perspectives, 106,* 217-229.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to Linear Regression Analysis*. (4th edition ed.) John Wiley & Sons, Inc.

Murray, K. S., Rogers, D. T., & Kaufman, M. M. (2004). Heavy metals in an urban watershed in southeastern Michigan. *Journal of Environmental Quality, 33,* 163-172.

Nannoni, F., Protano, G., & Riccobono, F. (2011). Fractionation and geochemical mobility of heavy elements in soils of a mining area in northern Kosovo. *Geoderma, 161,* 63-73.

Novak, M., Erbanova, L., Fottova, D., Voldrichova, P., Prechova, E., Blaha, V. et al. (2010). Increasing arsenic concentrations in runoff from 12 small forested catchments (Czech Republic, Central Europe): Patterns and controls. *Science of The Total Environment, 408,* 3614-3622.

Pagotto, C., my, N., Legret, M., & Le, C. P. (2001). Heavy Metal Pollution of Road Dust and Roadside Soil near a Major Rural Highway. Environmental Technology 22[3], 307-319.

Paya-Perez, A., Sala, J., & Mousty, F. (1993). Comparison of ICP-AES and ICP-MS for the Analysis of Trace Elements in Soil Extracts. *International Journal of Environmental Analytical Chemistry, 51,* 223-230.

Preciado, H. & Li, L. (2006). Evaluation of Metal Loadings and Bioavailability in Air, Water and Soil Along Two Highways of British Columbia, Canada. *Water Air Soil Pollut, 172,* 81-108.

Preciado, H., Li, L., & Weis, D. (2007). Investigation of Past and Present Multi-metal Input along Two Highways of British Columbia, Canada, Using Lead Isotopic Signatures. *Water Air Soil Pollut, 184,* 127-139.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/.

Räisäsen, M. L., Kashulina, G., & Bogatyrev, I. (1997). Mobility and retention of heavy metals, arsenic and sulphur in podzols at eight locations in northern Finland and Norway and the western half of the Russian Kola Peninsula. *Journal of Geochemical Exploration, 59,* 175-195.

Reimann, C. & Garrett, R. G. (2005). Geochemical background□Çöconcept and reality. *Science of The Total Environment, 350,* 12-27.

Robards, K. & Worsfold, P. (1991). Cadmium - Toxicology and Analysis - A Review. *Analyst, 116,* 549-568.

Salonen, V. P. & Korkka-Niemi, K. (2007). Influence of parent sediments on the concentration of heavy metals in urban and suburban soils in Turku, Finland. *Applied Geochemistry, 22,* 906-918.

Schmitt, N., Philion, J. J., Larsen, A. A., Harnadek, M., & Lynch, A. J. (1979). Surface Soil As A Potential Source of Lead-Exposure for Young-Children. *Canadian Medical Association Journal, 121,* 1474-1478.

Steinnes, E. & Friedland, A. J. (2006). Metal contamination of natural surface soils from long-range atmospheric transport: Existing and missing knowledge. *Environmental Reviews, 14,* 169-186.

Sterckeman, T., Douay, F., Proix, N., & Fourrier, H. (2000). Vertical distribution of Cd, Pb and Zn in soils near smelters in the North of France. *Environmental Pollution, 107,* 377-389.

Su, J. G., Jerrett, M., Beckerman, B., Wilhelm, M., Ghosh, J. K., & Ritz, B. (2009). Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy. *Environmental Research, 109,* 657-670.

Svendsen, M. L., Steinnes, E., & Blom, H. A. (2007). 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, 16,* 585-603.

Tack, F. M. & Verloo, M. G. (1995). Chemical speciation and fractionation in soil and sediment heacy metal analysis:A review. International Journal of Environmental Analytical Chemistry 59, 225-238.

Taylor, M. P., Mackay, A. K., Hudson-Edwards, K. A., & Holz, E. (2010). Soil Cd, Cu, Pb and Zn contaminants around Mount Isa city, Queensland, Australia: Potential sources and risks to human health. *Applied Geochemistry, 25,* 841-855.

Tu, J. (2011). Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. *Applied Geography, 31,* 376-392.

Tu, J. & Xia, Z. G. (2008). Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. *Science of The Total Environment, 407,* 358-378.

Van der Perk, M. (2006). *Soil and Water Contamination: From molecular to catchment scale*. Taylor & Francis Group.

Wang, S. & Mulligan, C. N. (2006). Occurrence of arsenic contamination in Canada: Sources, behavior and distribution. *Science of The Total Environment, 366,* 701-721.

Wheeler, J.O., Hoffman, P.F., Card, K.D., Davidson, A., Sanford, B.V., Okulitch, A.V., and Roest, W.R. (comp.)1997: Geological Map of Canada, Geological Survey of Canada, Map D1860A.

Williams, F. L. R. & Ogston, S. A. (2002). Identifying populations at risk from environmental contamination from point sources. *Occupational and Environmental Medicine, 59,* 2-8.

Wu, J., Edwards, R., He, X. Q., Liu, Z., & Kleinman, M. (2010). Spatial analysis of bioavailable soil lead concentrations in Los Angeles, California. *Environmental Research, 110,* 309-317.

Yesilonis, I. D., Pouyat, R. V., & Neerchal, N. K. (2008). Spatial distribution of metals in soils in Baltimore, Maryland: Role of native parent material, proximity to major roads, housing age and screening guidelines. *Environmental Pollution, 156,* 723-731.

Zhang, C. S. (2006). Using multivariate analyses and GIS to identify pollutants and their spatial patterns in urban soils in Galway, Ireland. *Environmental Pollution, 142,* 501-511.

Zhang, L. J., Ma, Z. H., & Guo, L. (2009). An Evaluation of Spatial Autocorrelation and Heterogeneity in the Residuals of Six Regression Models. *Forest Science, 55,* 533-548.

Appendix A – Data description

List of Tables and Figures

This section contains information on the sources of the soil samples, the files' names, analytical and digestion methods, and years of collection for both the LUR and GWR analyses.

Table 18 shows the websites where the data were obtained and the file names. Table 19 describes the concentration levels analytical methods, year of collection, and the number of sample used for the LUR analysis. Table 20 presents the basic descriptive statistics of the data for As and Pb concentration levels.

Table 18 : Data Sources – LUR and GWR

Agency	Website	Files Names
Ministry of Environment of British Columbia	http://www.gov.bc.ca/env/	Background Soil Quality Database
Ministry of Northern Development, Mines and Forestry of Ontario	http://www.mndmf.gov.on.ca/	MRD015, MRD021, MRD061, and MRD136

Table 19 : Data Description - LUR

Metal	Number of samples	Analytical method	Digestion method	Year of collection
Arsenic	55	ICP-OES	*Aqua Regia*	1994 to 1996
Lead	56	ICP-OES	*Aqua Regia*	1994 to 1996

Table 20 : Concentration Levels Description for Arsenic and Lead (µg/g) - LUR

Metal	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
Arsenic	0.25	3.05	5.50	6.85	9.05	25.70
Lead	1.63	6.79	9.90	45.66	32.12	621.20

Figure 15 shows the soil sample locations for all the metal datasets, including LUR and GWR analysis from the Ministry of Northern Development, Mines and Forestry of Ontario and the Ministry of Environment of British Columbia.

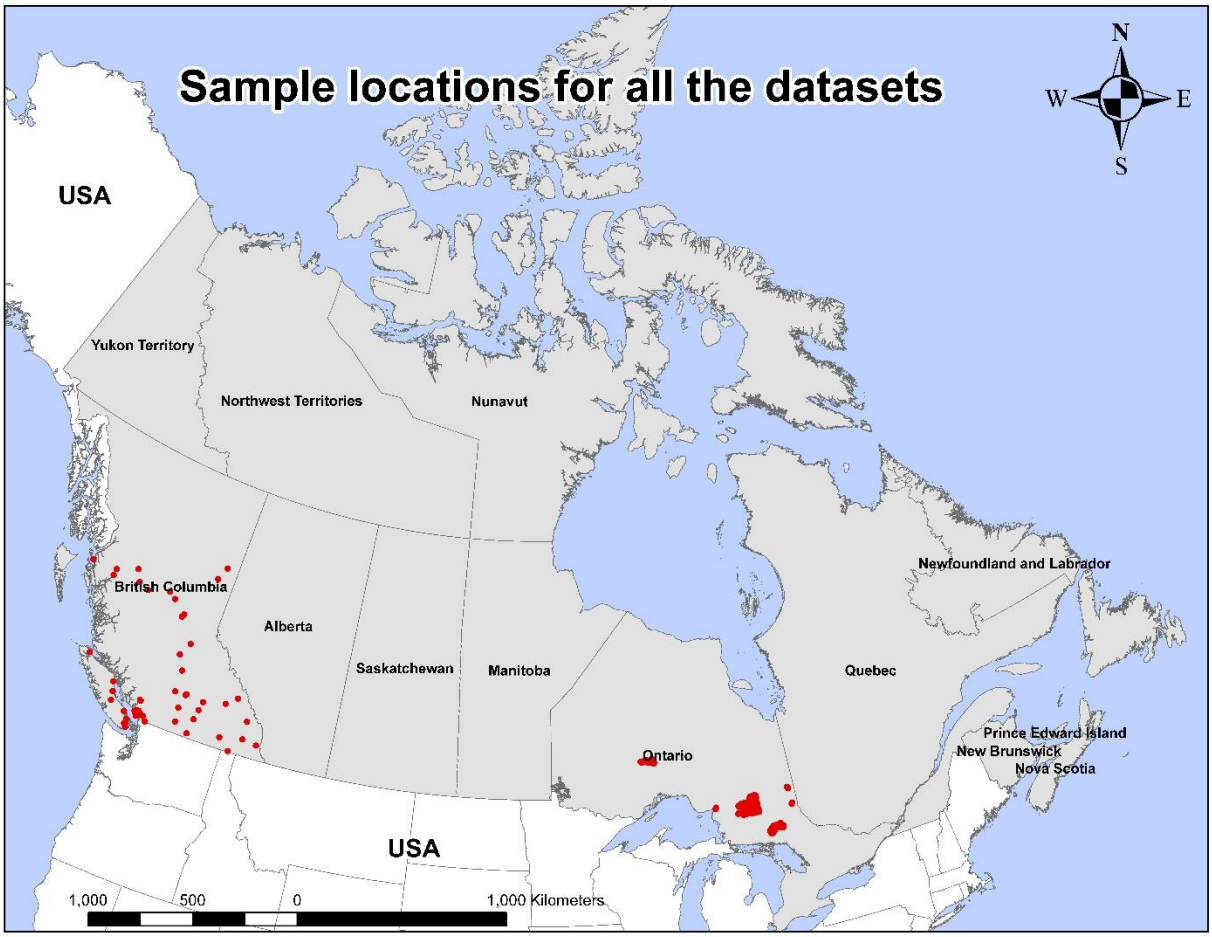


Figure 14 : Soil sample locations from all metal datasets, including LUR and GWR analysis

Table 21 presents in detail the datasets used for the GWR analysis, including the number of samples, the file names, the analytical methods and the year of collection.

Table 21 : Soil metal samples description - GWR analysis

Metals (n=number of samples)	Agency	Name of the file	Number of samples per files	Analytical method	Year of collection
Arsenic (n=877)	Ontario Geological Survey	MRD015	724	INAA	1993
		MRD021	153	INAA	1996
Cadmium (n=81)	Ministry of Environment of British Columbia	Background Soil Quality Study	54	ICP-OES	1996
	Ontario Geological Survey	MRD015	9	ICP-OES	1993
		MRD021	2	ICP	1996
		MRD136	16	ICP-MS	2003
Cobalt (n=839)	Ministry of Environment of British Columbia	Background Soil Quality Study	55	ICP-OES	1996
	Ontario Geological Survey	MRD015	720	ICP-OES	1993
		MRD021	22	ICP	1996
		MRD136	42	ICP-MS	2003
Chromium (n=899)	Ministry of Environment of British Columbia	Background Soil Quality Study	56	ICP-OES	1996
	Ontario Geological Survey	MRD015	779	ICP-OES	1993
		MRD021	22	ICP	1996
		MRD136	42	ICP-MS	2003
Nickel (n=860)	Ministry of Environment of British Columbia	Background Soil Quality Study	55	ICP-OES	1996

	Ontario Geological Survey	MRD015	741	ICP-OES	1993
		MRD021	22	ICP	1996
		MRD136	42	ICP-MS	2003
Lead (n=860)	Ministry of Environment of British Columbia	Background Soil Quality Study	56	ICP-OES	1996
		MRD015	741	ICP-OES	1993
		MRD021	22	ICP	1996
		MRD136	41	ICP-MS	2003

Appendix B – GWR results

This Appendix contains surface and subsurface concentration levels, including boxplots, scatterplots, and maps. For each metal, regression analysis and model adequacy results are presented for normal, logarithm- and squared-transformed concentration levels. Appendix B is divided in 6 sections by metal type (As, Cd, Cr, Co, Ni, and Pb). Each section shows Box Plots and maps of the concentration levels, correlation analysis graphs, results for the ordinary least squares results, the GWR results, including residuals analysis (Histograms, QQ Plots, and variance of errors). Maps and histograms showing the distribution of the coefficients and R-squared can be found in Figures 41 to 44.

In summary, the histograms, the QQ plots, and the variance of errors show that the data for Cd, Cr, Co, Ni, and Pb are not suitable for regression analysis. In all case, except for Arsenic, the residuals of the GWR analysis were not normally distributed as shown in Figures 54 to 71 for Cadmium, 81 to 98 for Chromium, 108 to 125 for Cobalt, 135 to 152 for Nickel, and 162 to 179 for Lead. Also included in Figures 41 to 44, the distribution of the coefficients and R-squared for Arsenic as described in Chapter 2.

List of Tables and Figures

Arsenic

Concentration levels and maps

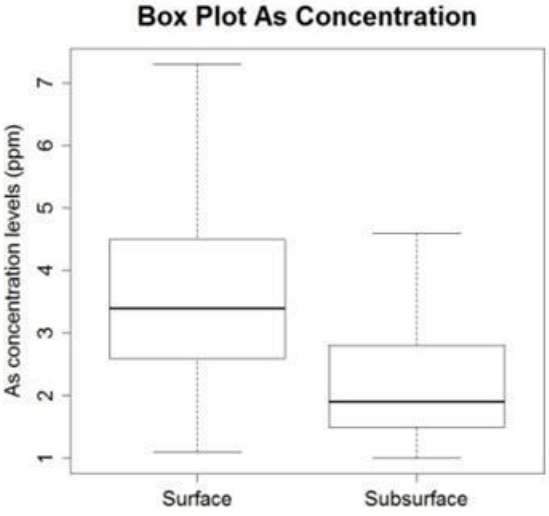


Figure 15: Box Plot of Arsenic Concentration without outliers

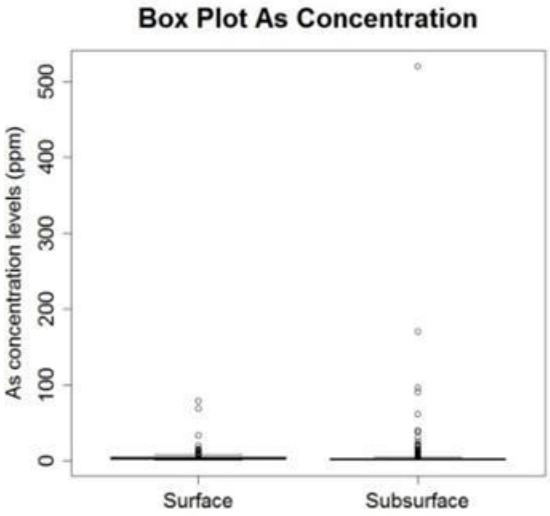


Figure 16: Box Plot of Arsenic Concentration with outliers

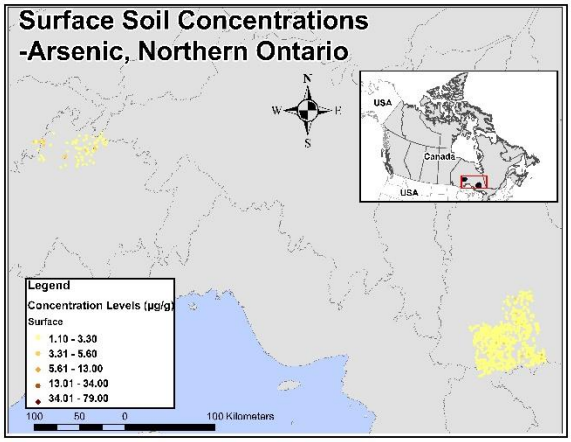


Figure 17: Surface concentration levels for Arsenic

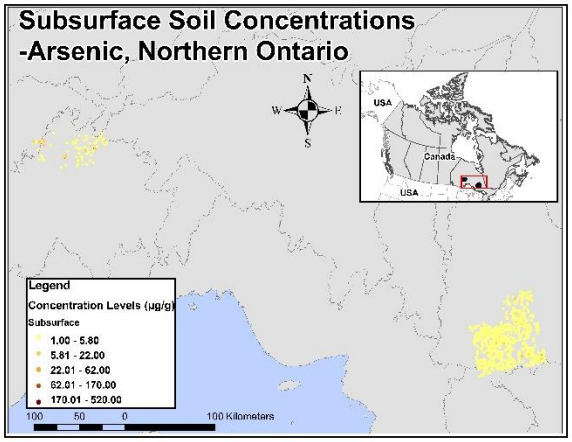


Figure 18: Subsurface concentration levels for Arsenic

Correlation analysis

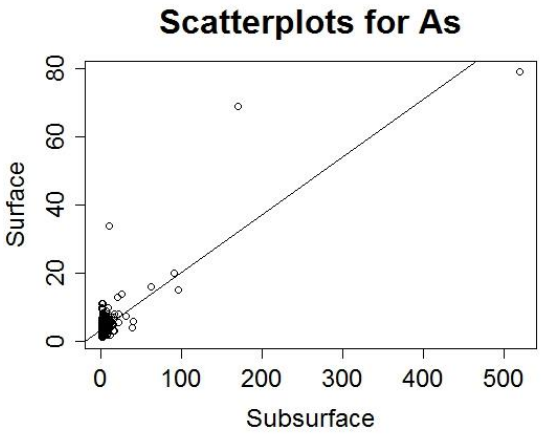


Figure 19 : Correlation between Surface and Subsurface Concentration for Arsenic

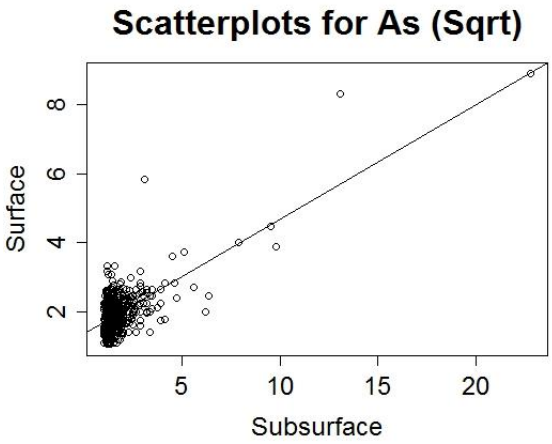


Figure 21 : Correlation between Surface and Subsurface Concentration for Arsenic (Sqrt)

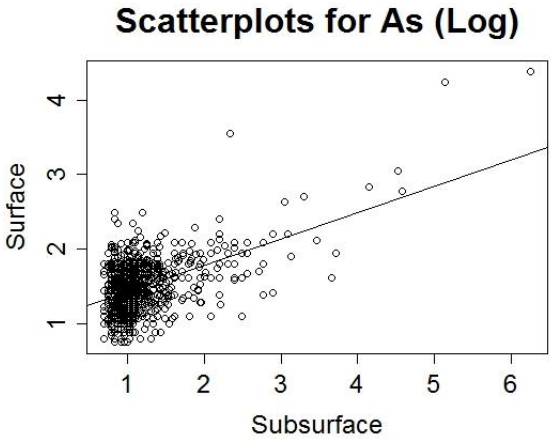


Figure 20 : Correlation between Surface and Subsurface Concentration for Arsenic (Log)

Table 22 : Correlation analysis between surface and subsurface Arsenic concentration levels

Data	Correlation levels
Normal	0.83
Log-transformed	0.51
Squared-transformed	0.68

Ordinary Least Squares results

Table 23: OLS Analysis for Normal, Logarithmic, and Squared Transform

Arsenic (n = 877)	Normal	Log	Squared
R^2 adjusted	0.68*	0.26*	0.46*
AICc	3889	444	876
Coefficient	0.169123	0.354419	0.331981
Standard Errors	0.003899	0.020176	0.01221
t-statistic	43.37*	17.566759*	27.19*
F-statistics (p-value)	1881*	308*	739*

*** indicates significance (<0.05)**

Table 24: Residual Analysis

Arsenic (n = 877)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.5872 (<2.2e-16)	0.9868 (4.13e-07)	0.9374 (<2.2e-16)
Moran's I	0.004376	0.085373	0.065989
Expected Index	-0.001142	-0.001142	-0.001142
Variance	0.000104	0.000121	0.00012
z-score	0.540317	7.880475	6.139715
p-value	0.588978	0.000000	0.00000

Geographically Weighted Regression results

Table 25: GWR Analysis for Normal, Log, and Squared Transform

Arsenic (n = 877)	Normal	Log	Squared
Neighbors	140	142	142
Residual Squares	3584	71.96	118.16
Effective Number	35.84	38.38	37.42
Sigma	2.064	0.293	0.375
AICc	3781	358	791
R^2 adjusted	0.72	0.34	0.52
Coefficient range	-0.3709 – 2.489	-0.249 – 0.7555	-0.2814 – 1.163
R^2 range	0.00 – 0.82	0.00 – 0.63	0.00 – 0.81

Table 26: Residuals Analysis for GWR

Arsenic (n = 877)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.6515 (<2.2e-16)	0.9934 (0.0006557)	0.9607 (1.353e-14)
Moran's I	-0.062887	-0.002118	-0.014435
Expected Index	-0.001142	-0.001142	-0.001142
Variance	0.000103	0.000121	0.00012
z-score	-6.077494	-0.088962	-1.212903
p-value	0.000000	0.929112	0.225167

Histograms - residuals analysis

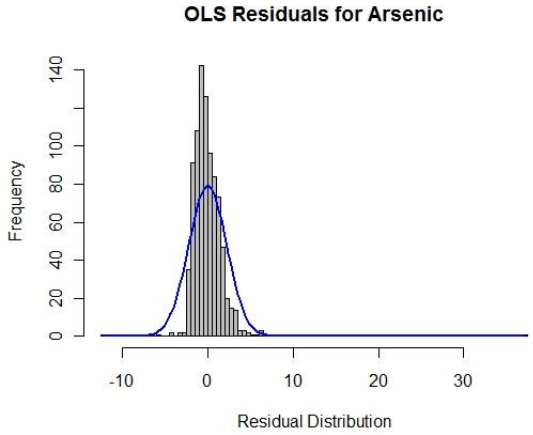


Figure 22 : Histogram for Arsenic Residuals – OLS

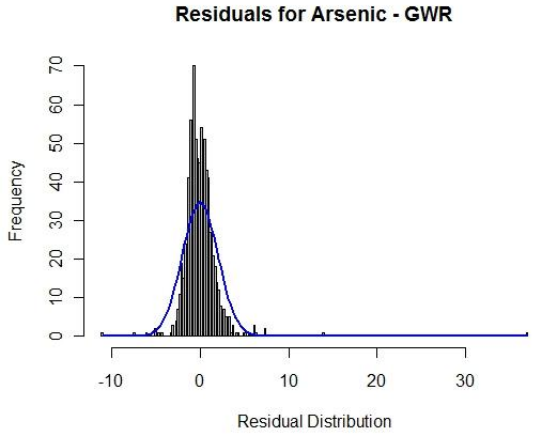


Figure 23 : Histogram for Arsenic Residuals – GWR

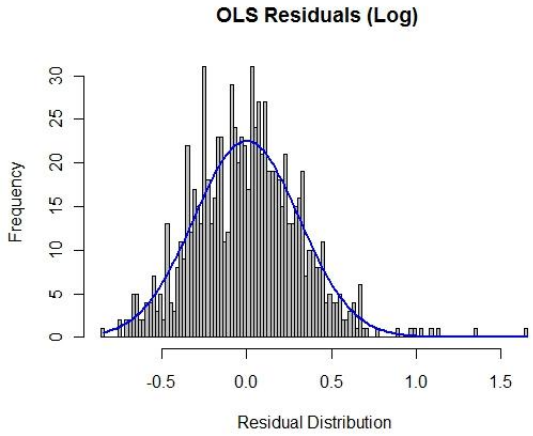


Figure 24 : Histogram for Arsenic Residuals - OLS (Log)

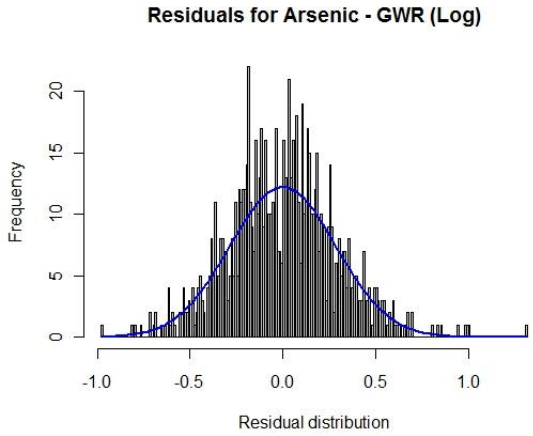


Figure 25 : Histogram for Arsenic Residuals - GWR (Log)

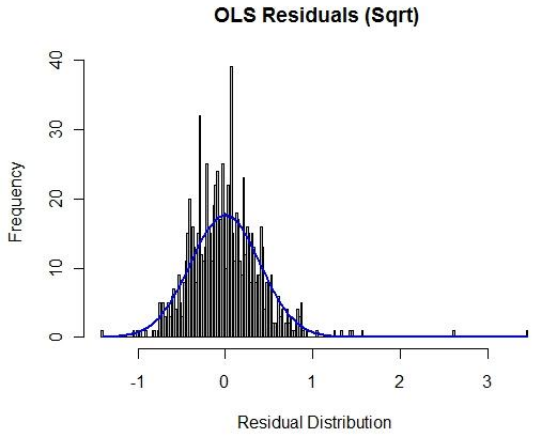


Figure 26 : Histogram for Arsenic Residuals - OLS (Sqrt)

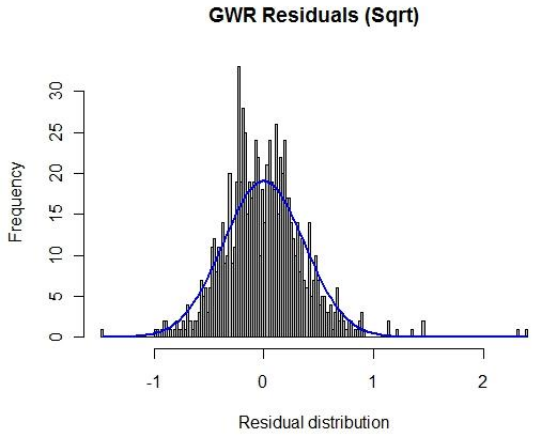


Figure 27 : Histogram for Arsenic Residuals - GWR (Sqrt)

QQ plots - residuals analysis

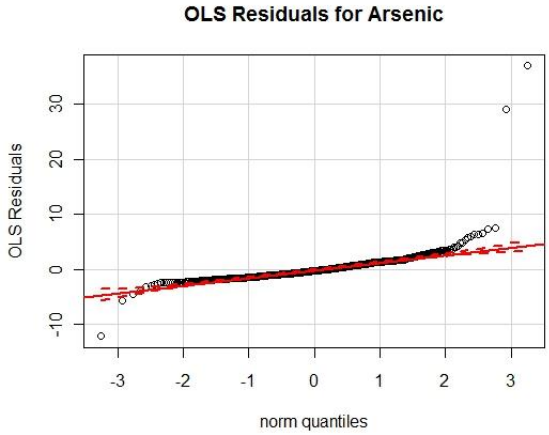


Figure 28 : QQ plot for Arsenic residuals – OLS

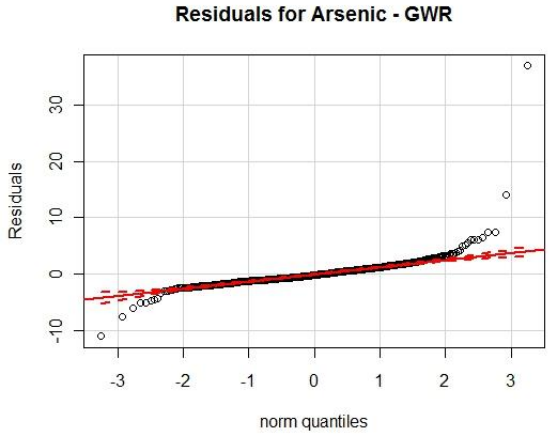


Figure 29 : QQ plot for Arsenic residuals - GWR

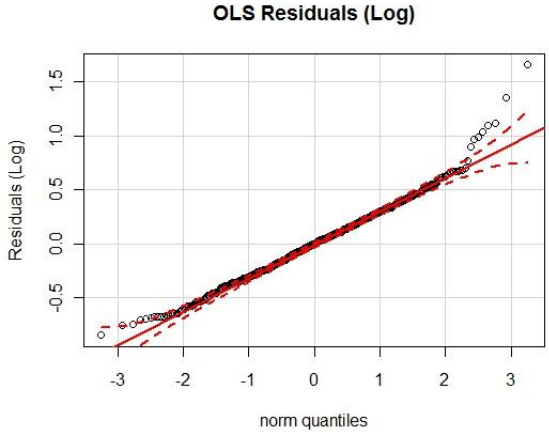


Figure 30 : QQ plot for Arsenic residuals - OLS (Log)

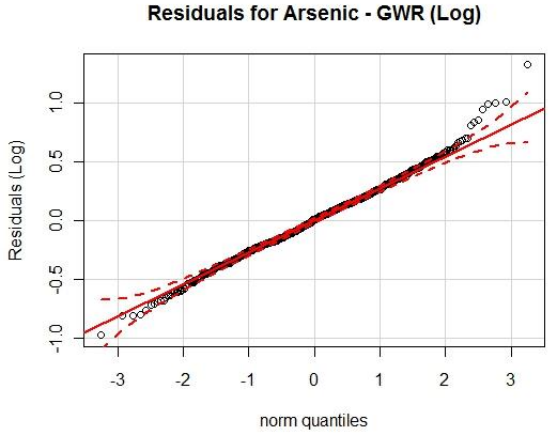


Figure 31 : QQ plot for Arsenic residuals - GWR (Log)

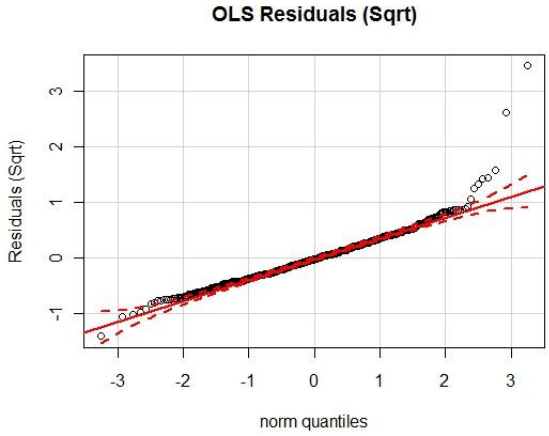


Figure 32 : QQ plot for Arsenic residuals - OLS (Sqrt)

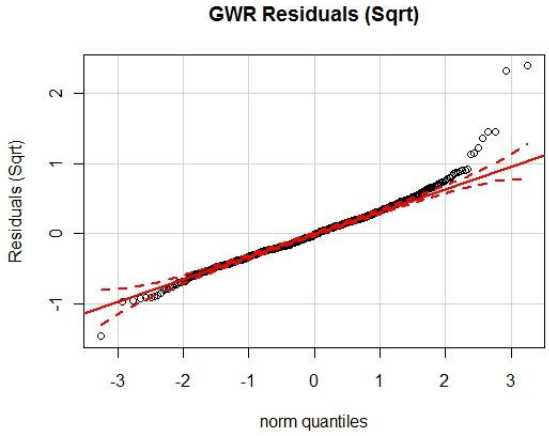


Figure 33 : QQ plot for Arsenic residuals - GWR (Sqrt)

Variance of errors

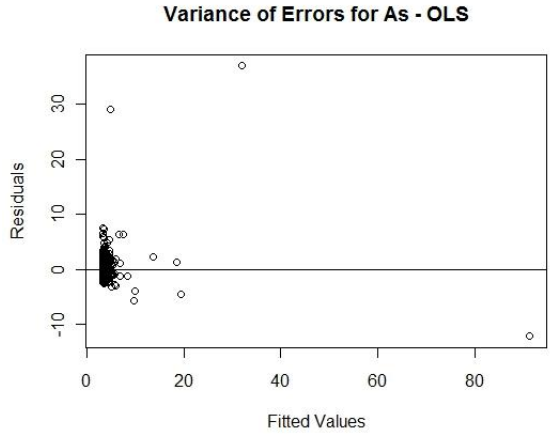


Figure 34 : Variance of Errors for Arsenic – OLS

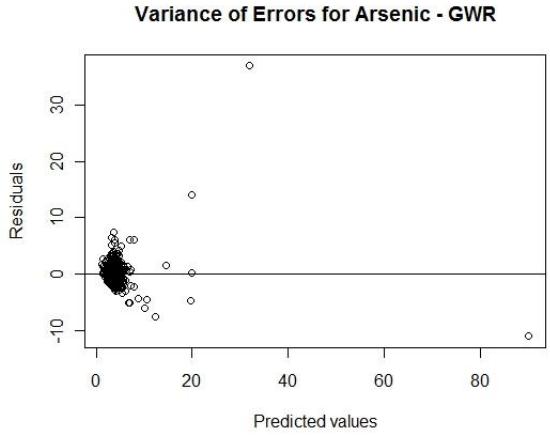


Figure 35 : Variance of Errors for Arsenic – GWR

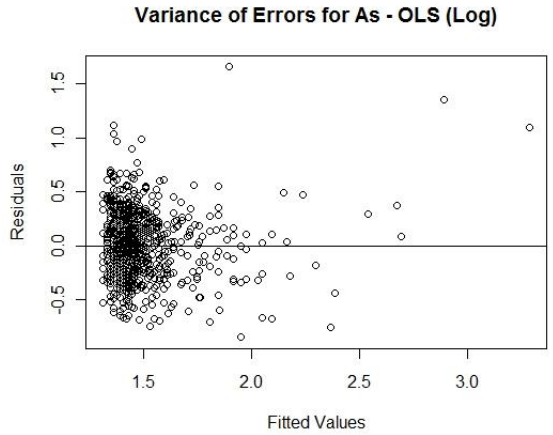


Figure 36 : Variance of Errors for Arsenic – OLS (Log)

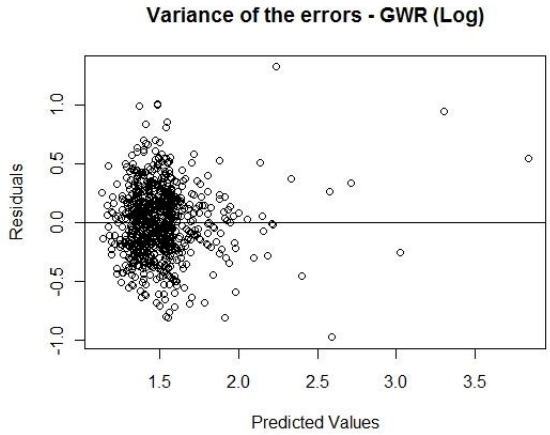


Figure 37 : Variance of Errors for Arsenic – GWR (Log)

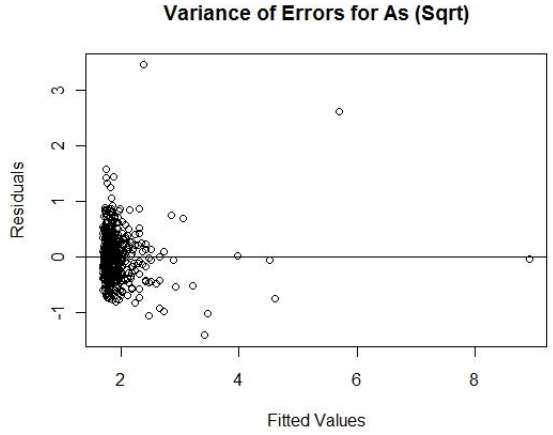


Figure 38 : Variance of Errors for Arsenic – OLS (Sqrt)

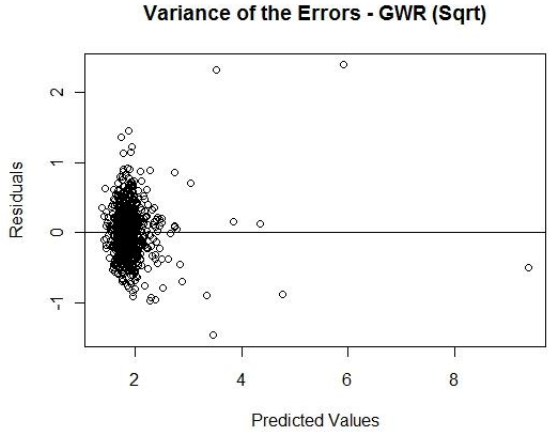


Figure 39 : Variance of Errors for Arsenic – GWR (Sqrt)

Distribution of the coefficient

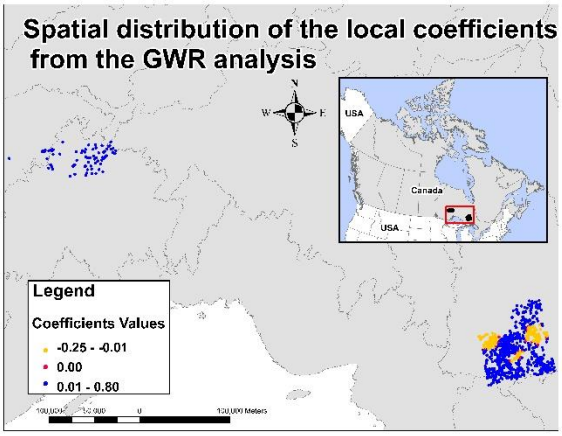


Figure 40 : Spatial distribution of the coefficients - GWR (Log)

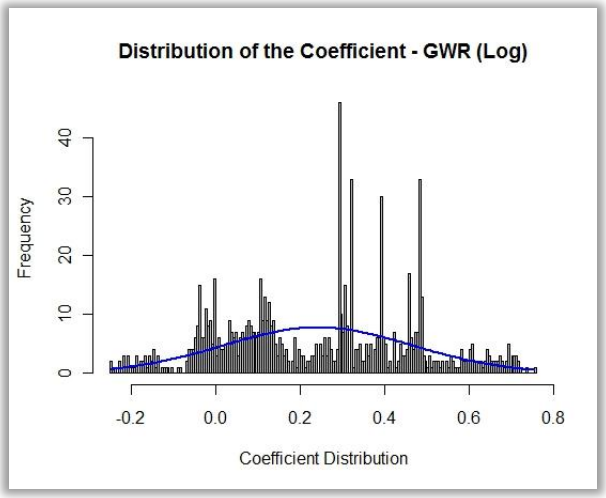


Figure 41 : Distribution of the coefficient for Arsenic - GWR (Log)

R-squared Distribution

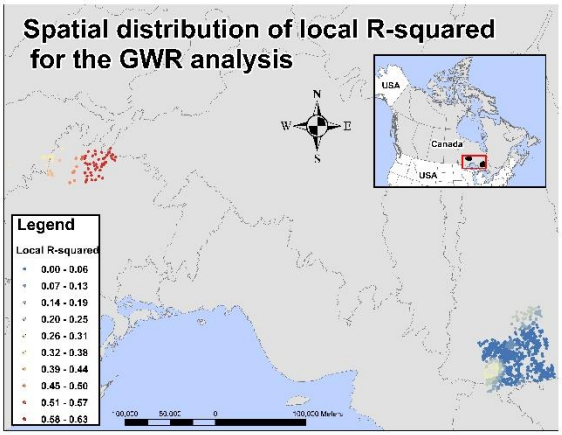


Figure 42 : Spatial distribution of the R-squared for Arsenic - GWR (Log)

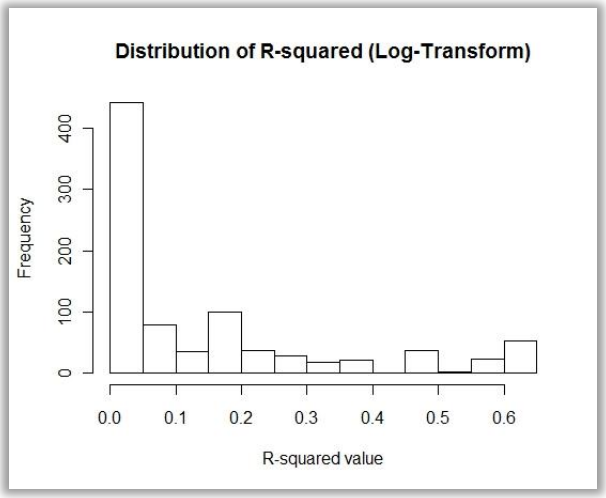


Figure 43 : Distribution of the R-squared for Arsenic - GWR (Log)

Cadmium

Concentration levels and maps

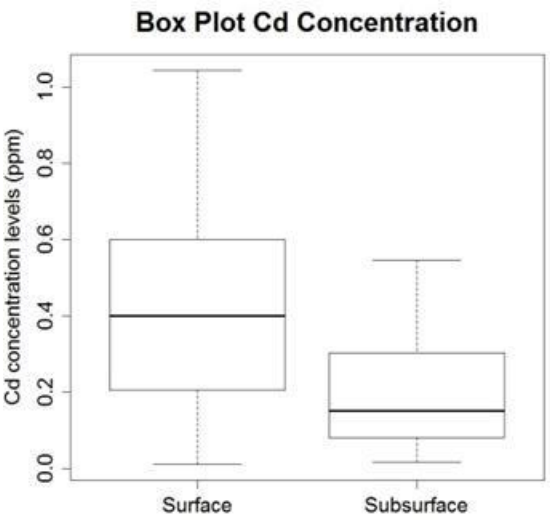


Figure 44: Box plot of Cd concentration without outliers

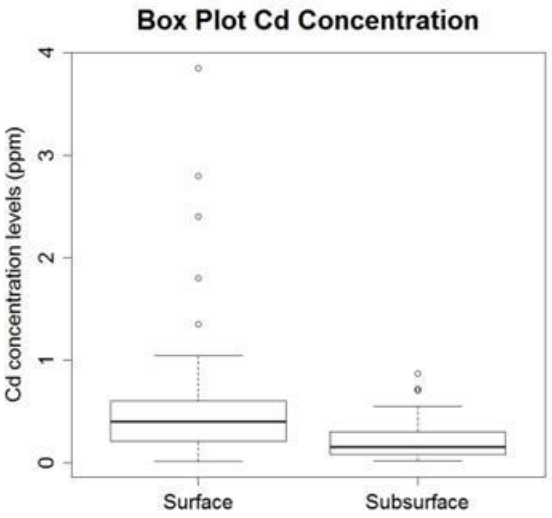


Figure 45: Box plot of Cd concentration with outliers

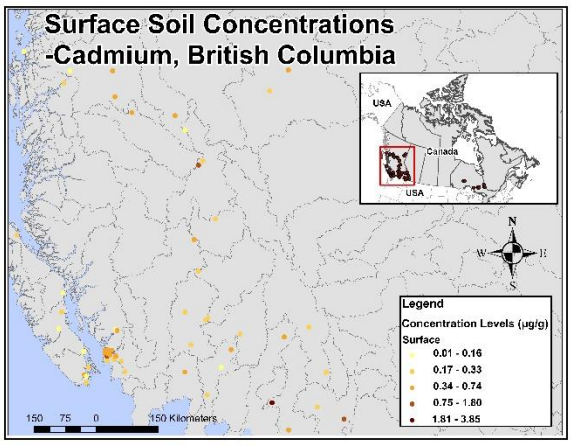


Figure 46: Surface concentration levels for Cadmium (British Columbia)

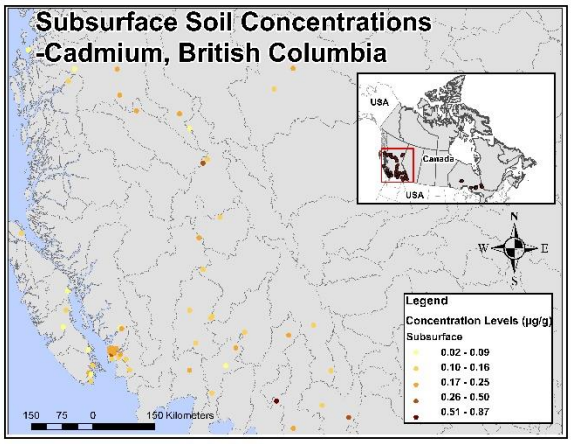


Figure 47: Subsurface concentration levels for Cadmium (British Columbia)

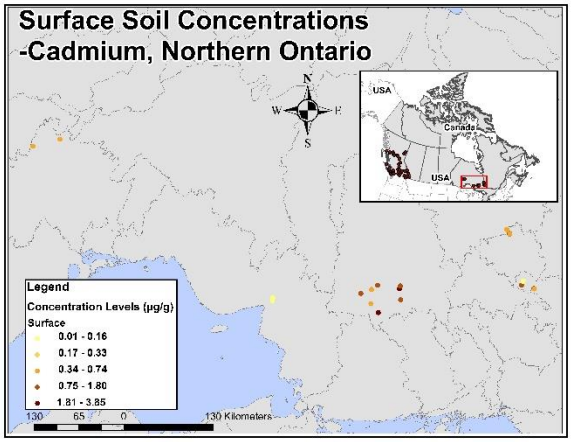


Figure 48 : Surface Concentration Levels for Cadmium (Ontario)

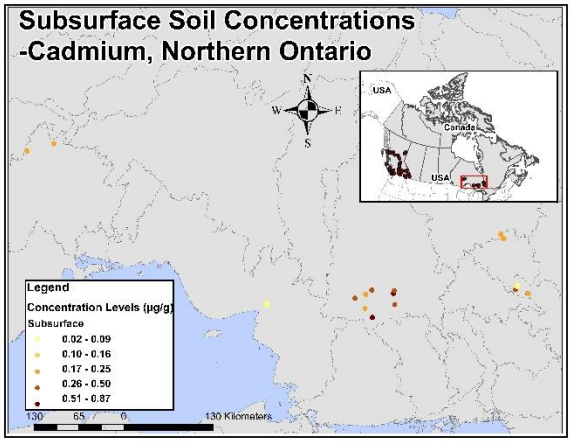


Figure 49 : Subsurface Concentration Levels for Cadmium (Ontario)

Correlation analysis

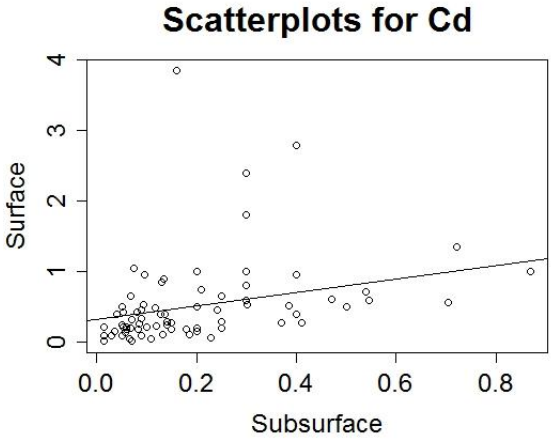


Figure 50 : Correlation between Surface and Subsurface Concentration for Cd

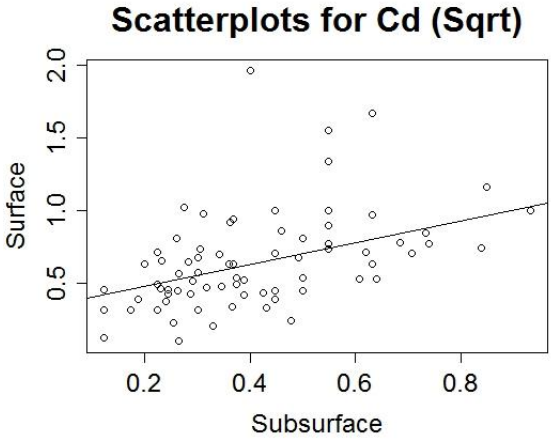


Figure 52 : Correlation between Surface and Subsurface Concentration for Cd (Sqrt)

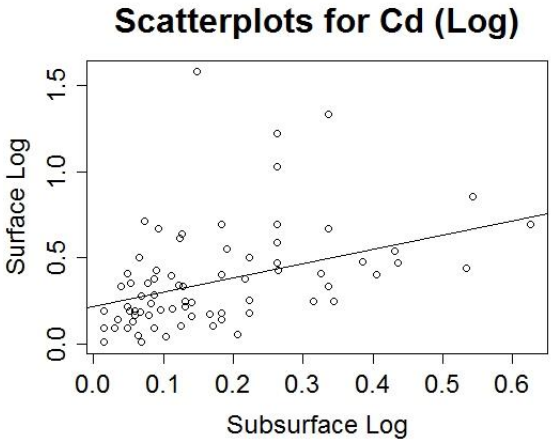


Figure 51 : Correlation between Surface and Subsurface Concentration for Cd (Log)

Table 27 : Correlation analysis between surface and subsurface Cadmium concentration levels

Data	Correlation levels
Normal	0.29
Log-transformed	0.40
Squared-transformed	0.43

Ordinary Least Squares results

Table 28: OLS Analysis for Normal, Log, and Squared Transform

Cadmium (n = 81)	Normal	Log	Squared
R^2 adjusted	0.07*	0.15*	0.18*
AICc	144	18	33
Coefficient	0.956251*	0.818273*	0.765835*
Standard Errors	0.357477	0.213054	0.174449
t-statistic	2.675001	3.84069	4.275369
F-statistics (p-value)	7.156*	14*	18.279*

*** indicates significance (<0.05)**

Table 29: Residual Analysis for OLS

Cadmium (n = 81)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.5973 (1.436e-13)	0.7928 (2.565e-09)	0.8526 (1.725e-07)
Moran's I	0.055886	0.102503	0.128122
Expected Index	-0.012500	-0.0125	-0.0125
Variance	0.029239	0.033501	0.033903
z-score	0.399930	0.628316	0.763723
p-value	0.689208	0.529797	0.445032

Geographically Weighted Regression results

Table 30: GWR Analysis for Normal, Log, and Squared Transform

Cadmium (n = 81)	Normal	Log	Squared
Neighbors	81	81	81
Residual Squares	24.77	5.08	6.26
Effective Number	4.158	4.157	4.146
Sigma	0.5677	0.257	0.285
AICc	144	15	32
R^2 adjusted	0.10	0.19	0.20
Coefficient range	0651 – 0.874	0.6064 – 0.7659	0.5571 – 0.7788
R^2 range	0.04 – 0.06	0.10 – 0.13	0.13 – 0.16

Table 31: Residuals Analysis for GWR

Cadmium (n = 81)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.6156 (3.052e-13)	0.8192 (1.470e-08)	0.8631 (4.027e-07)
Moran's I	0.124383	0.146798	0.156055
Expected Index	-0.012500	-0.0125	-0.0125
Variance	0.028090	0.032858	0.03341
z-score	0.816726	0.878798	0.92215
p-value	0.414085	0.379511	0.35645

Histograms - residuals analysis

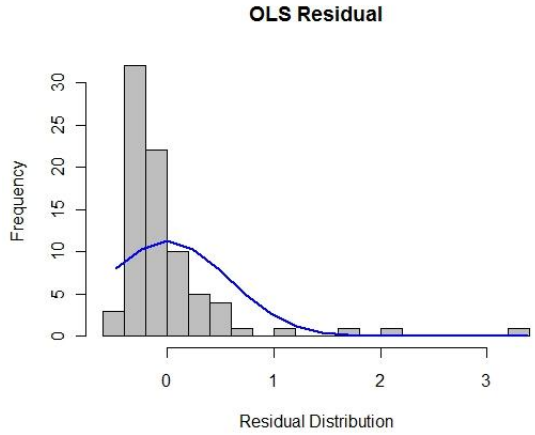


Figure 53 : Histogram for Cadmium Residuals – OLS

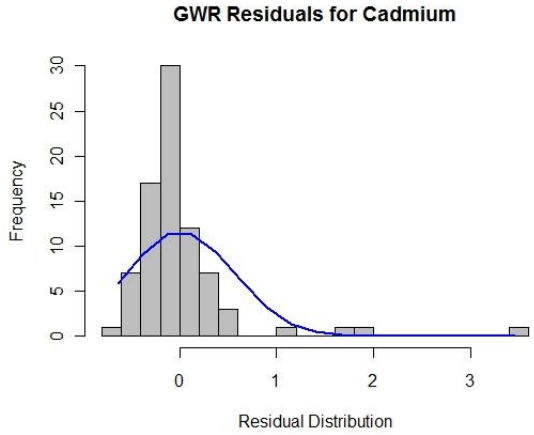


Figure 54 : Histogram for Cadmium Residuals – GWR

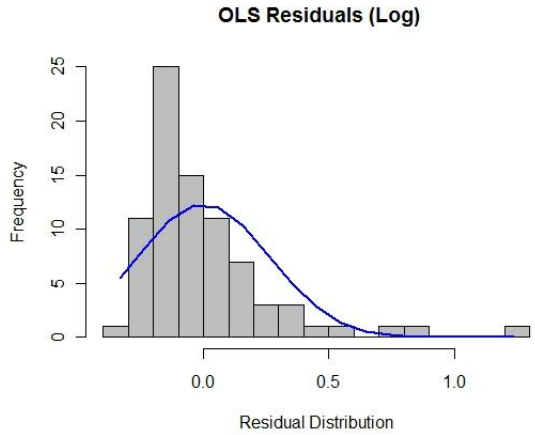


Figure 55 : Histogram for Cadmium Residuals – OLS (Log)

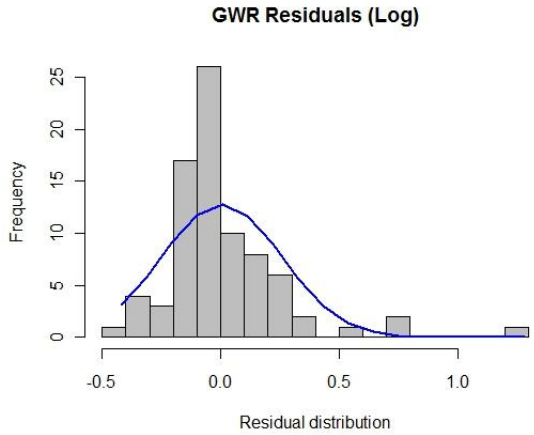


Figure 56 : Histogram for Cadmium Residuals – GWR (Log)

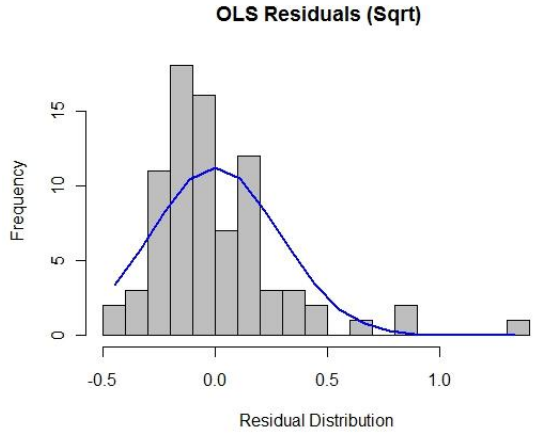


Figure 57 : Histogram for Cadmium Residuals – OLS (Sqrt)

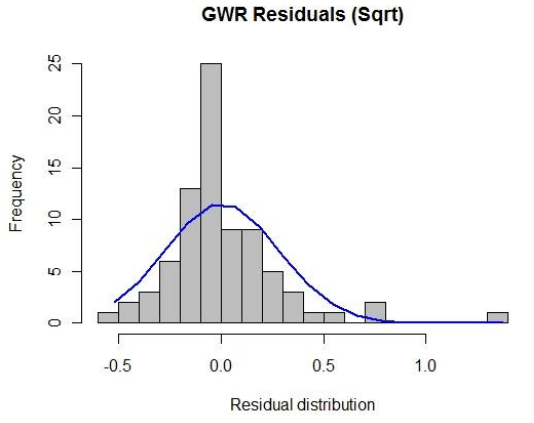


Figure 58 : Histogram for Cadmium Residuals – GWR (Sqrt)

QQ plots - residuals analysis

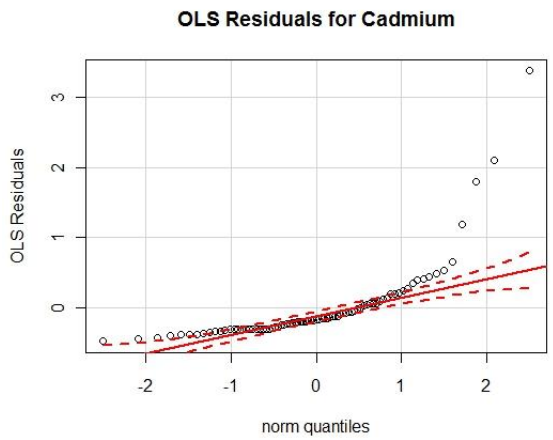


Figure 59 : QQ plot for Cadmium residuals – OLS

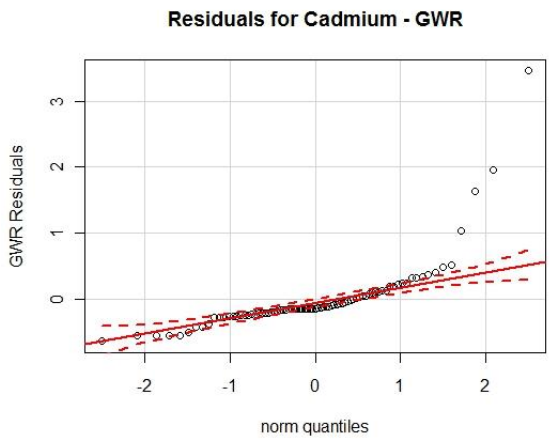


Figure 60 : QQ plot for Cadmium residuals – GWR

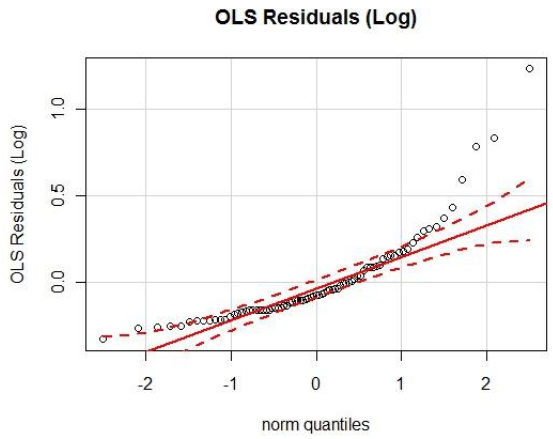


Figure 61 : QQ plot for Cadmium residuals – OLS (Log)

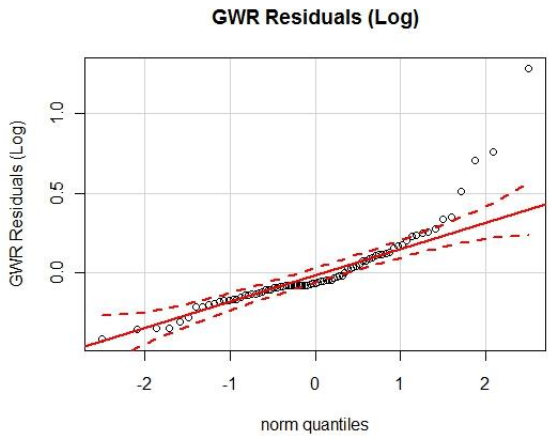


Figure 62 : QQ plot for Cadmium residuals – GWR (Log)

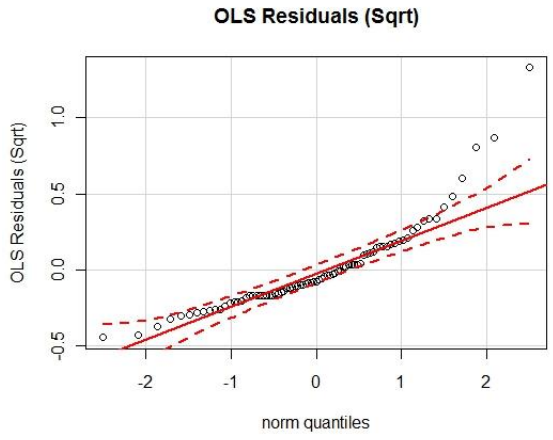


Figure 63 : QQ plot for Cadmium residuals – OLS (Sqrt)

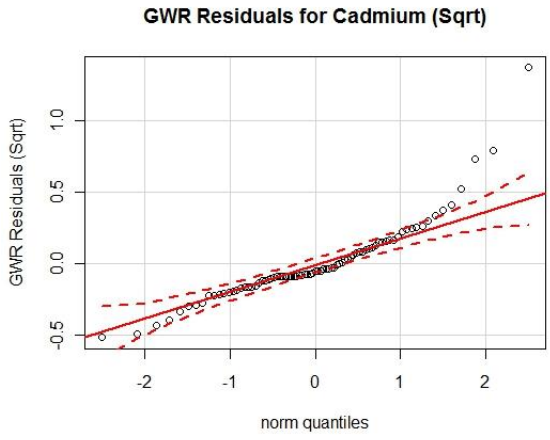


Figure 64 : QQ plot for Cadmium residuals – GWR (Sqrt)

Variance of errors

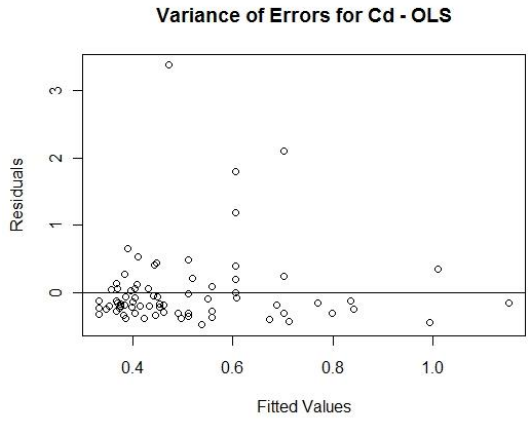


Figure 65 : Variance of Errors for Cadmium – OLS

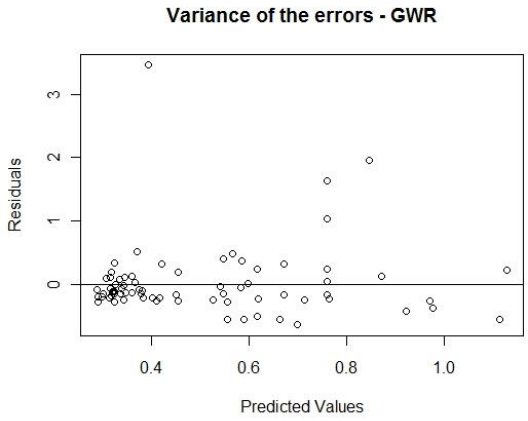


Figure 66 : Variance of Errors for Cadmium – GWR

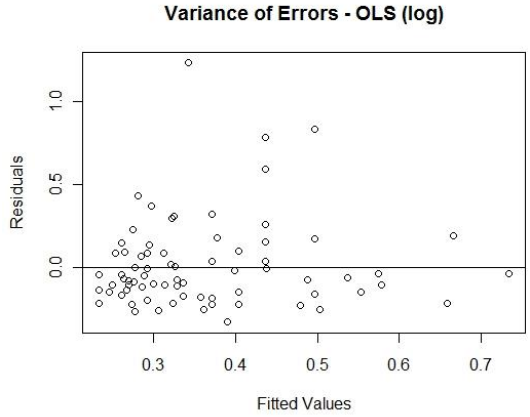


Figure 67 : Variance of Errors for Cadmium - OLS (Log)

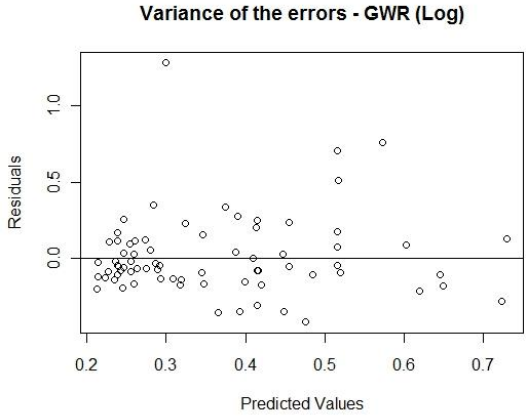


Figure 68 : Variance of Errors for Cadmium - GWR (Log)

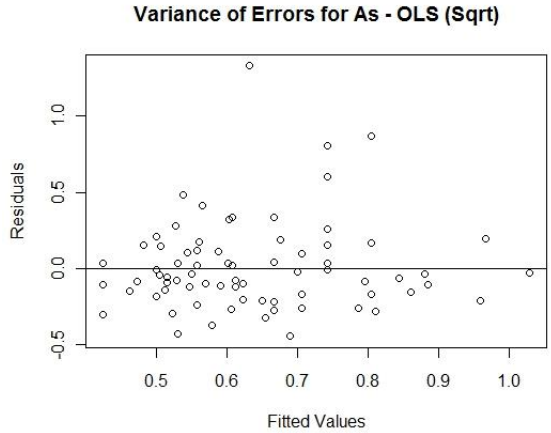


Figure 69 : Variance of Errors for Cadmium - OLS (Sqrt)

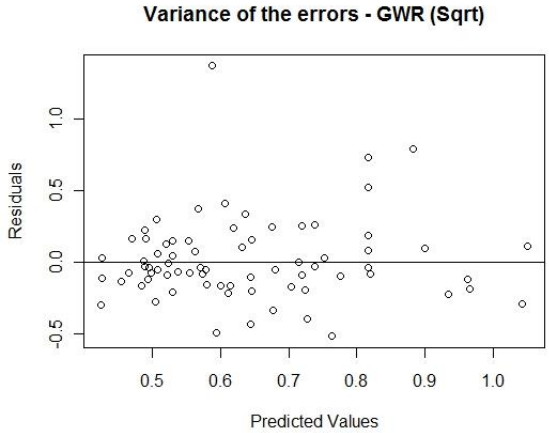


Figure 70 : Variance of Errors for Cadmium - GWR (Sqrt)

Chromium

Concentration levels and maps

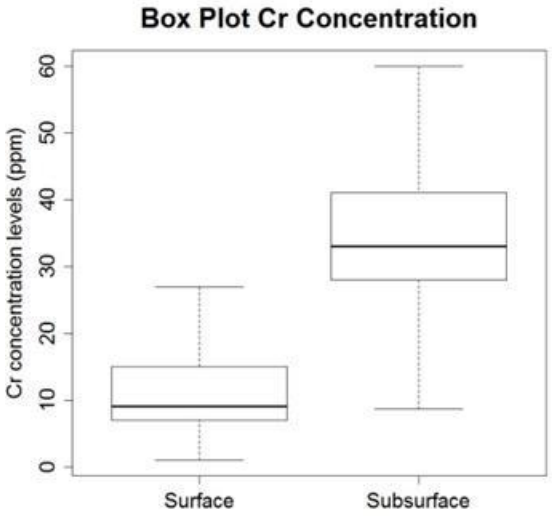


Figure 71: Box Plot of Cr Concentration without outliers

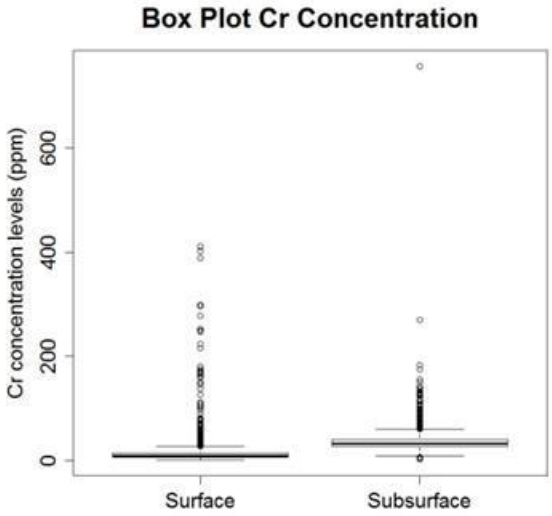


Figure 72: Box Plot of Cr Concentration with outliers

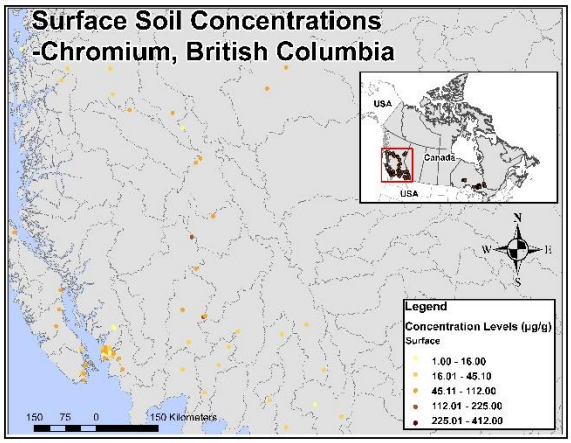


Figure 73 : Surface Concentration Levels for Chromium (British Columbia)

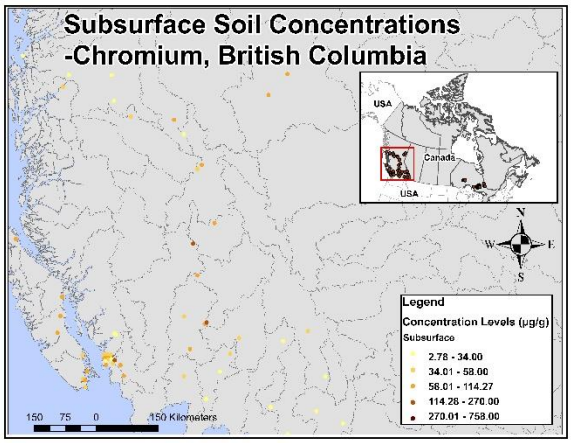


Figure 74 : Subsurface Concentration Levels for Chromium (British Columbia)

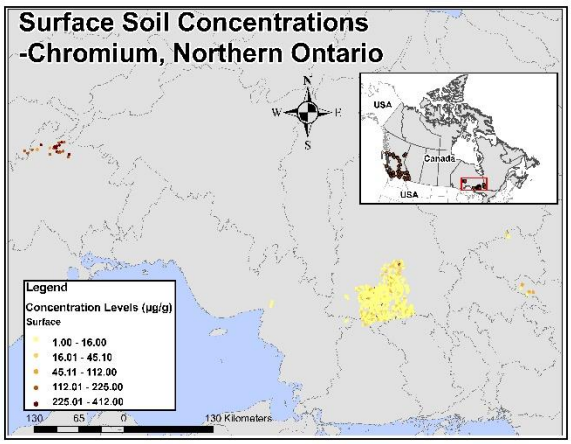


Figure 75 : Surface Concentration Levels for Chromium (Ontario)

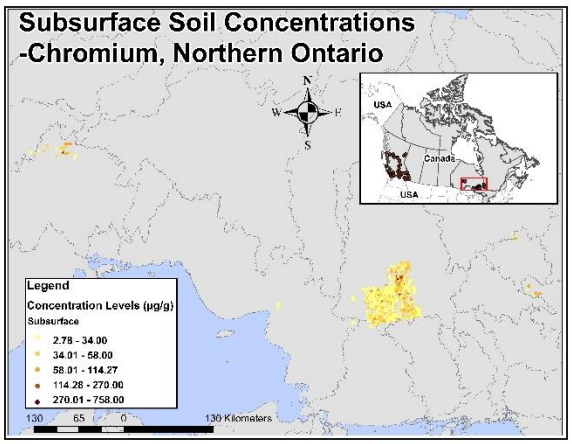


Figure 76 : Subsurface Concentration Levels for Chromium (Ontario)

Correlation analysis

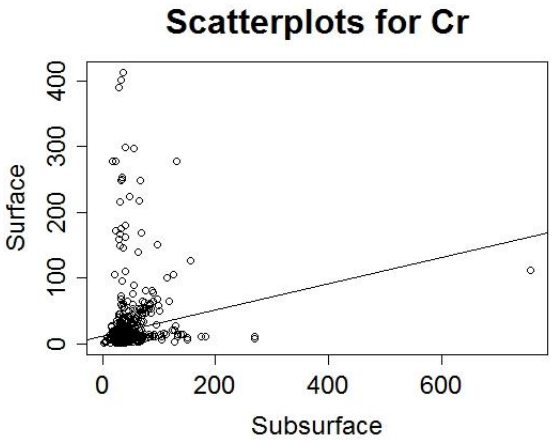


Figure 77 : Correlation between Surface and Subsurface Concentration for Cr

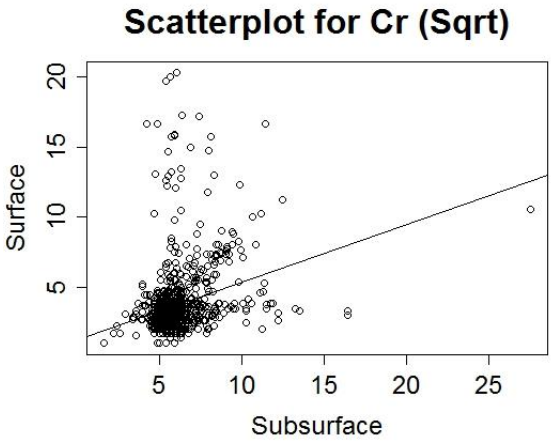


Figure 79 : Correlation between Surface and Subsurface Concentration for Cr (Sqrt)

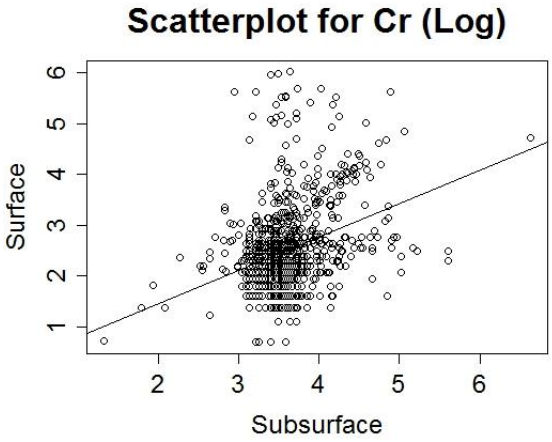


Figure 78 : Correlation between Surface and Subsurface Concentration for Cr (Log)

Table 32 : Correlation analysis between surface and subsurface Chromium concentration levels

Data	Correlation levels
Normal	0.16
Log-transformed	0.33
Squared-transformed	0.27

Ordinary Least Squares results

Table 33: OLS Analysis for Normal, Log, and Squared Transform

Chromium (n = 899)	Normal	Log	Squared
R^2 adjusted	0.02*	0.11*	0.07*
AICc	9252	2134	4129
Coefficient	0.198834*	0.655392*	0.413885*
Standard Errors	0.041218	0.061568	0.048803
t-statistic	4.823916	10.644956	8.480646
F-statistics (p-value)	23.27*	113.3*	71.92*

*** indicates significance (<0.05)**

Table 34: Residual Analysis for OLS

Chromium (n = 899)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.3723	0.9037 (<2.2e-16)	0.6464 (<2.2e-16)
Moran's I	0.067947	0.098704	0.091325
Expected Index	-0.001114	-0.001114	-0.001114
Variance	0.000059	0.000061	0.000060
z-score	8.990169	12.751745	11.888721
p-value	0.000000	0.0000000	0.0000000

Geographically Weighted Regression results

Table 35: GWR Analysis for Normal, Log, and Squared Transform

Chromium (n = 899)	Normal	Log	Squared
Neighbors	361	624	110
Residual Squares	625541.86	302.92	1854.26
Effective Number	19.898	13.574	54.448
Sigma	26.68	0.5849	1.482
AICc	8469	1597	3291
R^2 adjusted	0.60	0.51	0.65
Coefficient range	-0.0812 – 0.9555	0.2364 – 0.9696	-0.4594 – 0.9173
R^2 range	0.00 – 0.64	0.01 – 0.68	0.00 – 0.81

Table 36: Residuals Analysis for GWR

Chromium (n = 899)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.3248 (<2.2e-16)	0.9244 (<2.2e-16)	0.6319 (<2.2e-16)
Moran's I	0.003644	0.032300	0.000096
Expected Index	-0.001114	-0.001114	-0.001114
Variance	0.000054	0.000061	0.000058
z-score	0.648459	4.274606	0.158400
p-value	0.516688	0.000019	0.874142

Histograms - residuals analysis

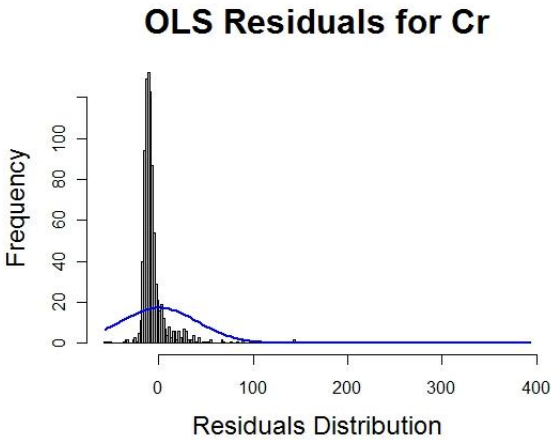


Figure 80 : Histogram for Chromium Residuals – OLS

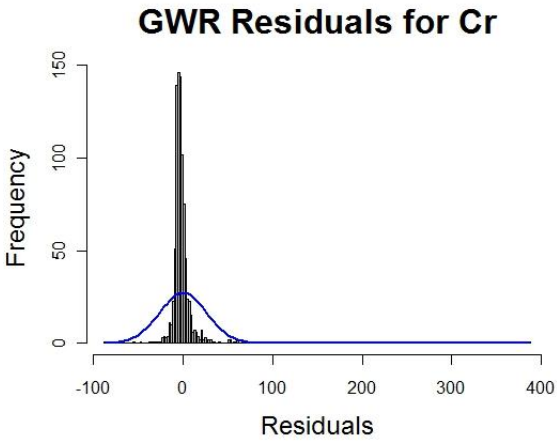


Figure 81 : Histogram for Chromium Residuals– GWR

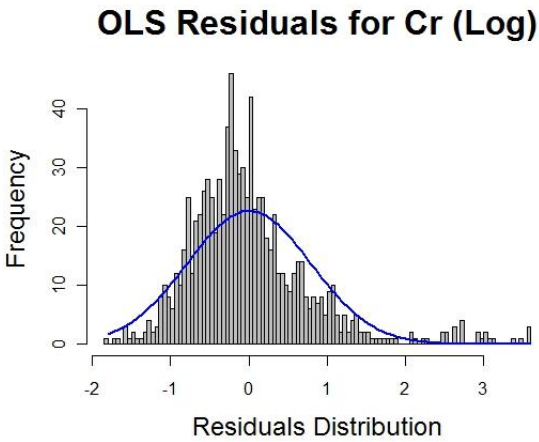


Figure 82 : Histogram for Chromium Residuals - OLS (Log)

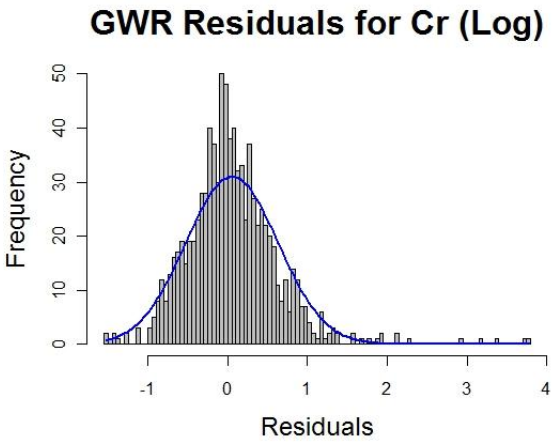


Figure 83 : Histogram for Chromium Residuals – GWR (Log)

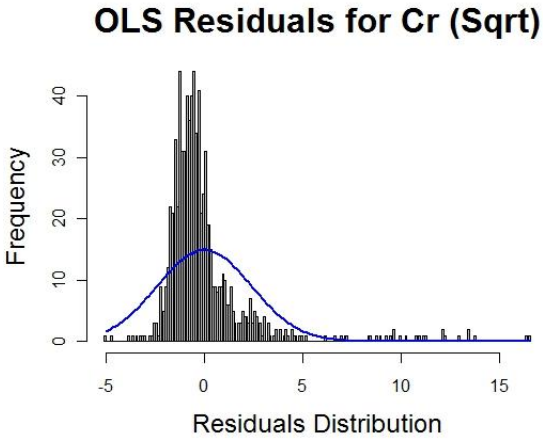


Figure 84 : Histogram for Chromium Residuals – OLS (Sqrt)

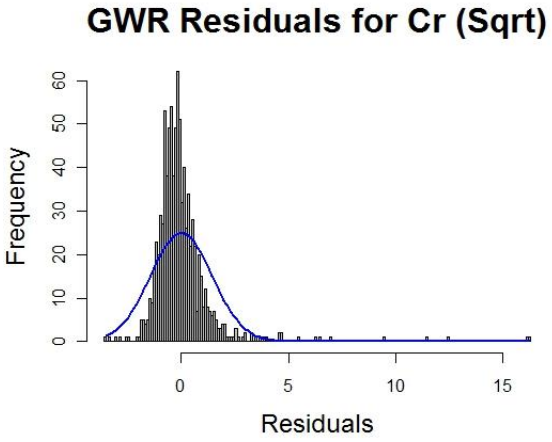


Figure 85 : Histogram for Chromium Residuals – GWR (Sqrt)

QQ plots - residuals analysis

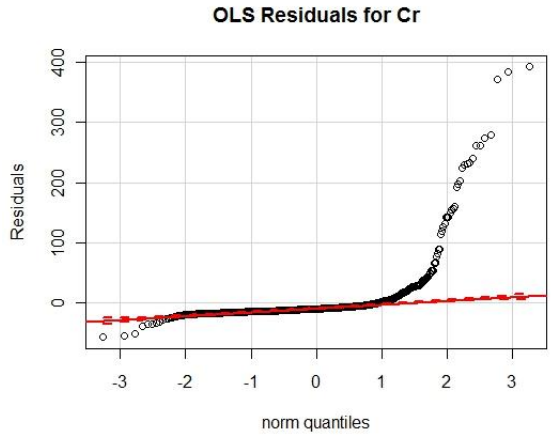


Figure 86: QQ plot for Chromium Residuals – OLS

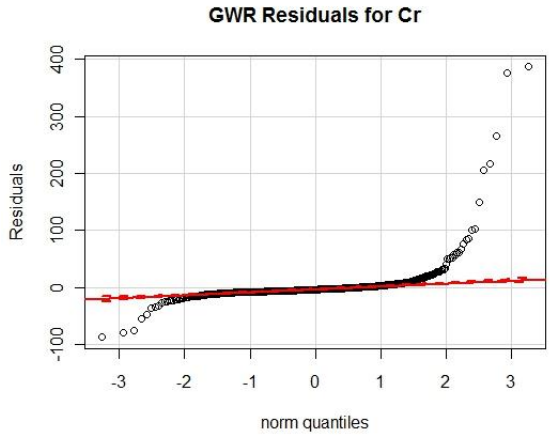


Figure 87: QQ plot for Chromium Residuals - GWR

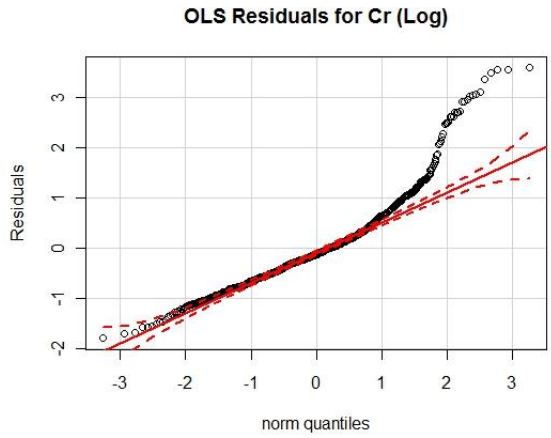


Figure 88: QQ plot for Chromium Residuals - OLS (Log)

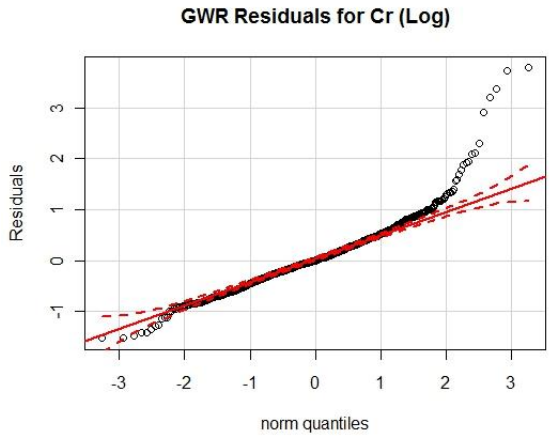


Figure 89: QQ plot for Chromium Residuals - GWR (Log)

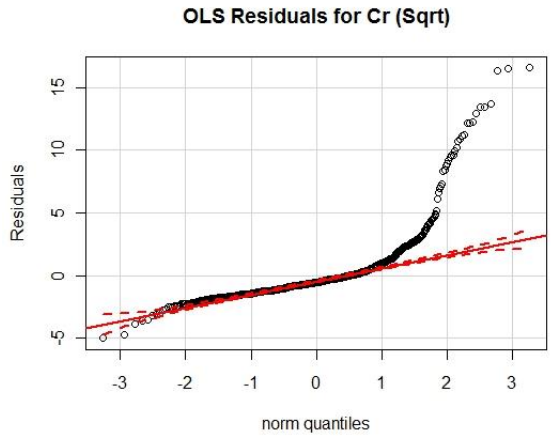


Figure 90: QQ plot for Chromium Residuals - OLS (Sqrt)

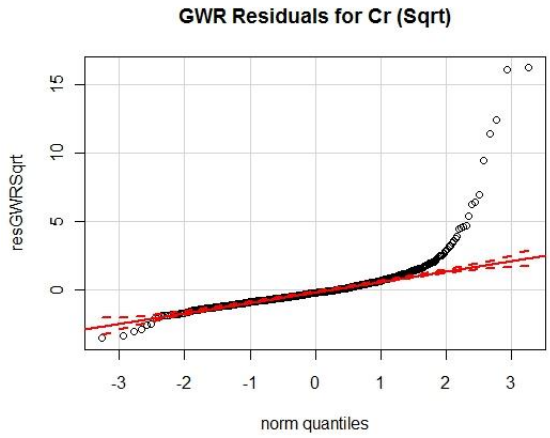


Figure 91: QQ plot for Chromium Residuals - GWR (Sqrt)

Variance of errors

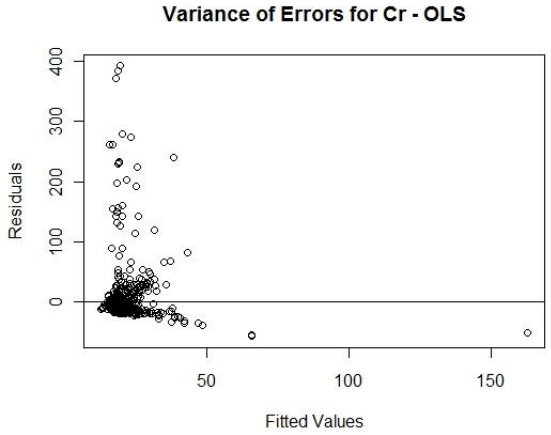


Figure 92 : Variance of Errors for Chromium - OLS

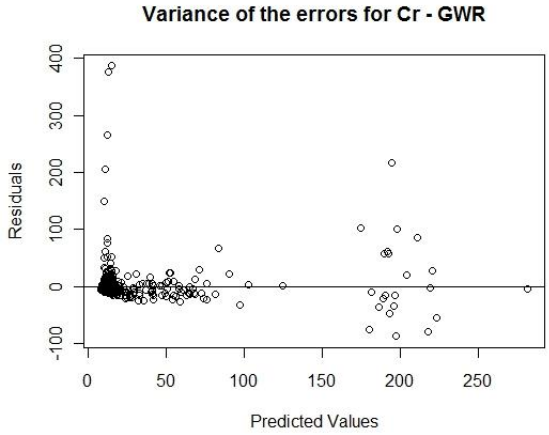


Figure 93 : Variance of Errors for Chromium - GWR

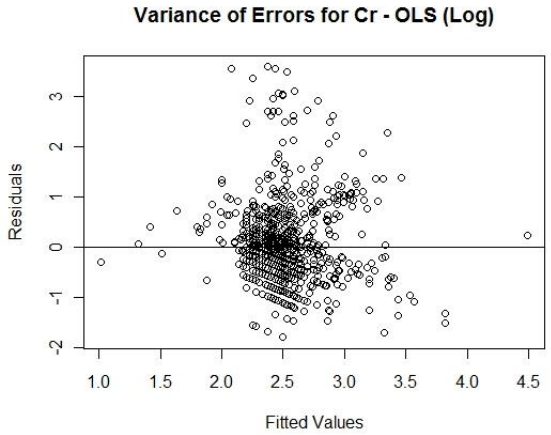


Figure 94 : Variance of Errors for Chromium - OLS (Log)

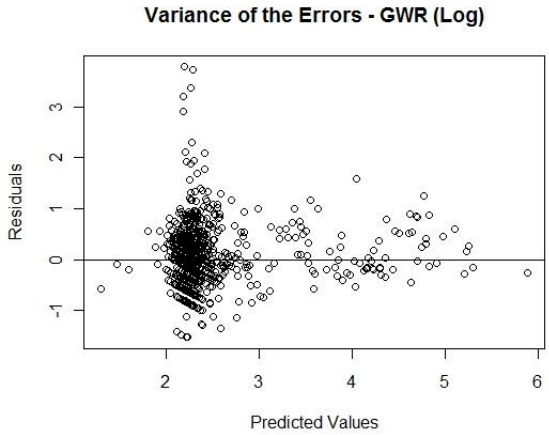


Figure 95 : Variance of Errors for Chromium - GWR (Log)

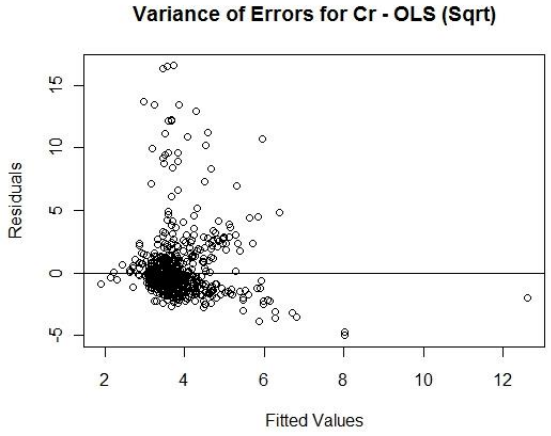


Figure 96 : Variance of Errors for Chromium - OLS (Sqrt)

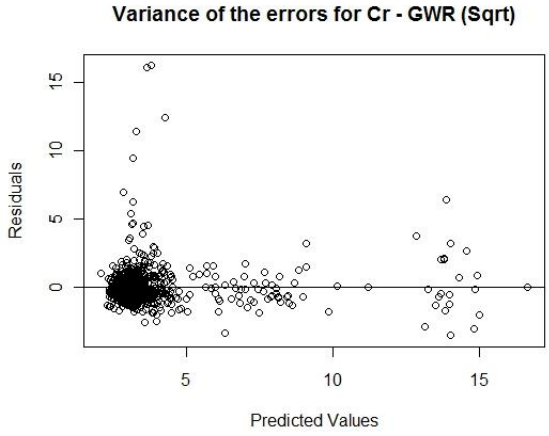


Figure 97 : Variance of Errors for Chromium - GWR (Sqrt)

Cobalt

Concentration levels and maps

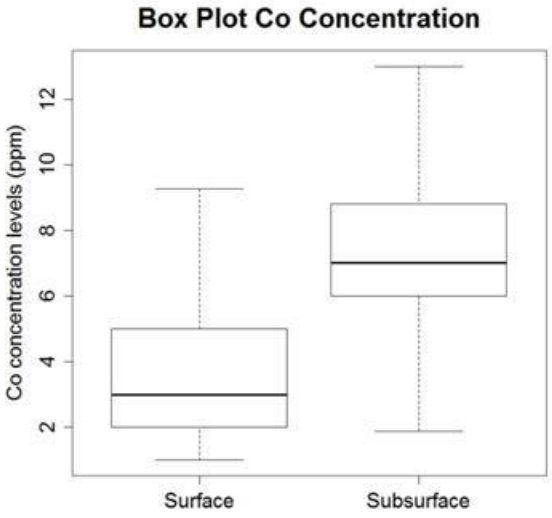


Figure 98: Box Plot of Co Concentration without outliers

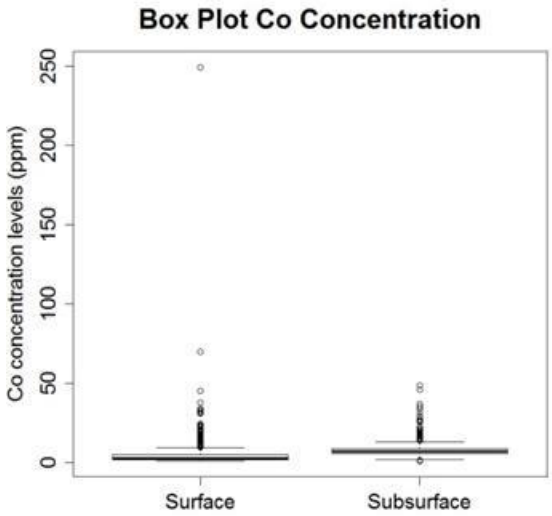


Figure 99: Box Plot of Co Concentration with outliers

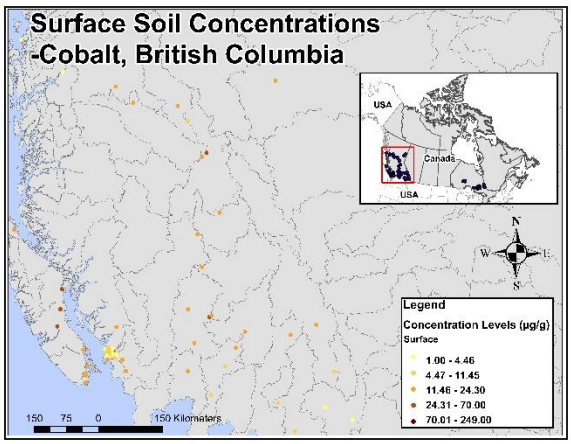


Figure 100 : Surface Concentration Levels for Cobalt (British Columbia)

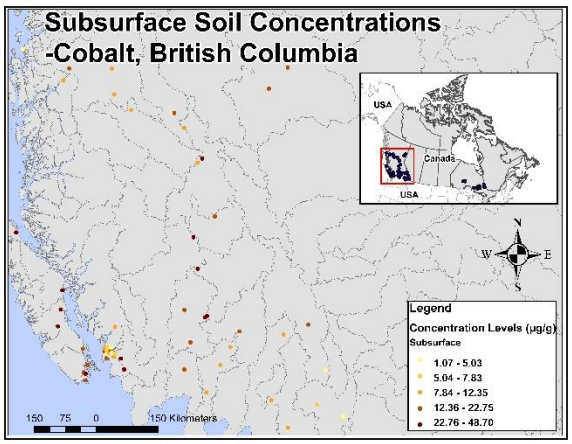


Figure 101 : Subsurface Concentration Levels for Cobalt (British Columbia)

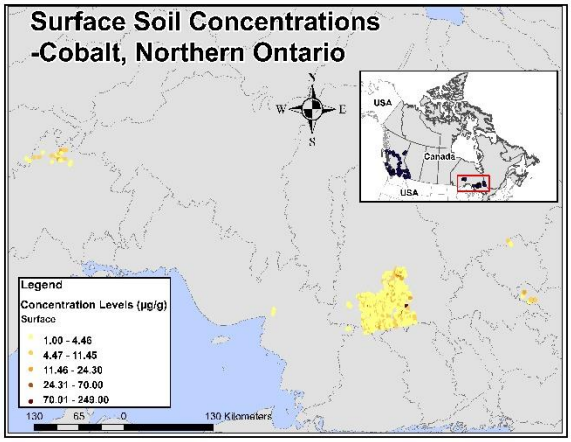


Figure 102 : Surface Concentration Levels for Cobalt (Ontario)

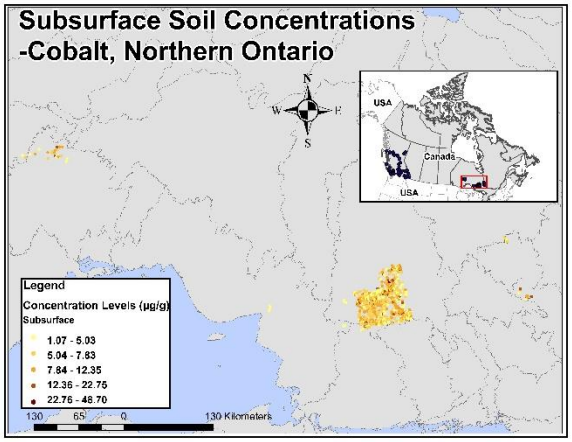


Figure 103 : Surface Concentration Levels for Cobalt (Ontario)

Scatterplot

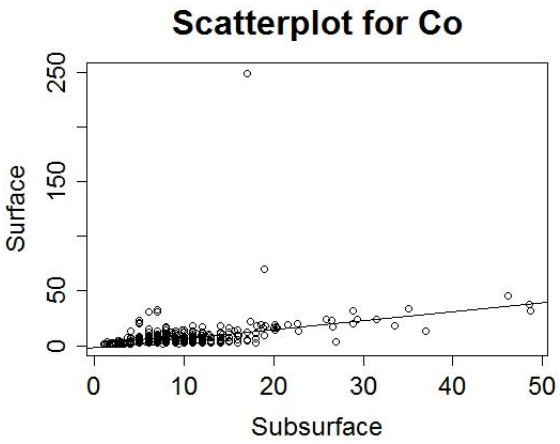


Figure 104 : Correlation between Surface and Subsurface Concentration for Cobalt

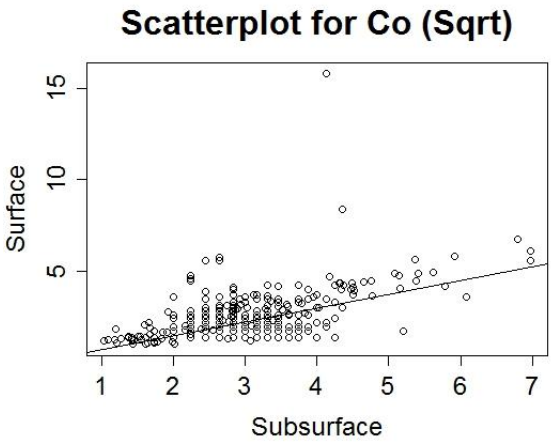


Figure 106 : Correlation between Surface and Subsurface Concentration for Cobalt (Sqrt)

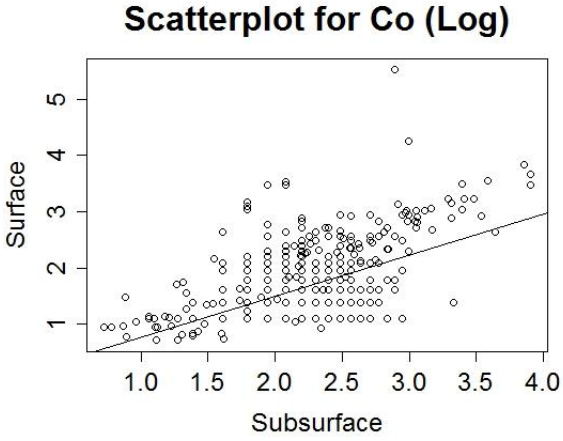


Figure 105 : Correlation between Surface and Subsurface Concentration for Cobalt (Log)

Table 37 : Correlation analysis between surface and subsurface Cobalt concentration levels

Data	Correlation levels
Normal	0.39
Log-transformed	0.52
Squared-transformed	0.54

Ordinary Least Squares results

Table 38: OLS Analysis for Normal, Logarithmic, and Squared Transform

Cobalt (n = 751)	Normal	Log	Squared
R^2 adjusted	0.15*	0.27*	0.29*
AICc	5543	1055	1870
Coefficient	0.812767*	0.728855	0.757479
Standard Errors	0.070746	0.043198	0.04317
t-statistic	11.488515	16.872558	17.546454
F-statistics (p-value)	131*	284.7*	307.9*

*** indicates significance (<0.05)**

Table 39: Residual Analysis for OLS

Cobalt (n = 751)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.2338 (<2.2e-16)	0.9367 (<2.2e-16)	0.7214 (<2.2e-16)
Moran's I	0.003774	0.41185	0.025602
Expected Index	-0.001333	-0.001333	-0.001333
Variance	0.000008	0.000019	0.000018
z-score	1.821554	9.632713	6.380705
p-value	0.068523	0.000000	0.000000

Geographically Weighted Regression results

Table 40: GWR Analysis for Normal, Logarithmic, and Squared Transform

Cobalt (n = 751)	Normal	Log	Squared
Neighbors	18	86	35
Residual Squares	6675.99	126.55	258.42
Effective Number	252.66	61.52	141.59
Sigma	3.66	0.4284	0.651
AICc	4317	896	1586
R^2 adjusted	0.89	0.44	0.57
Coefficient range	-2.735 – 19.6577	-0.299 – 1.319	-0.907 – 5.022
R^2 range	0.00 – 0.94	0.00 – 0.81	0.00 – 0.89

Table 41: Residuals Analysis for GWR

Cobalt (n = 751)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.2338 (<2.2e-16)	0.9019 (<2.2e-16)	0.6319 (<2.2e-16)
Moran's I	-0.007631	-0.006049	-0.006574
Expected Index	-0.001333	-0.001333	-0.001333
Variance	0.000019	0.000019	0.000019
z-score	-1.439174	-1.069610	-1.194029
p-value	0.150101	0.284795	0.232467

Histograms - residuals analysis

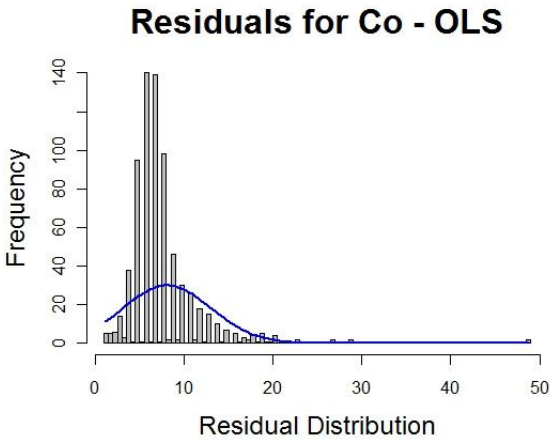


Figure 107 : Histogram for Cobalt Residuals – OLS

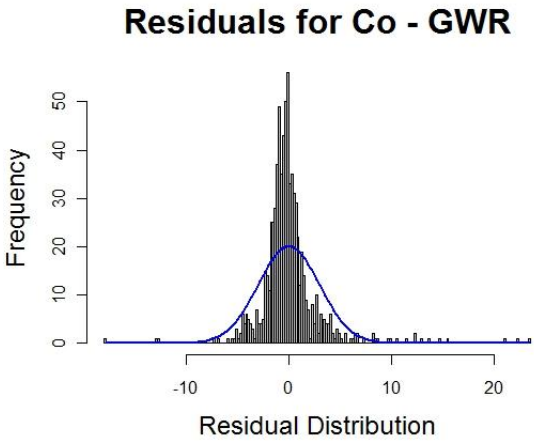


Figure 108 : Histogram for Cobalt Residuals – GWR

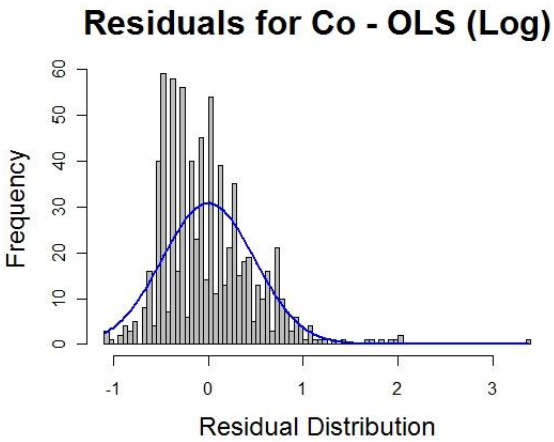


Figure 109 : Histogram for Cobalt Residuals - OLS (Log)

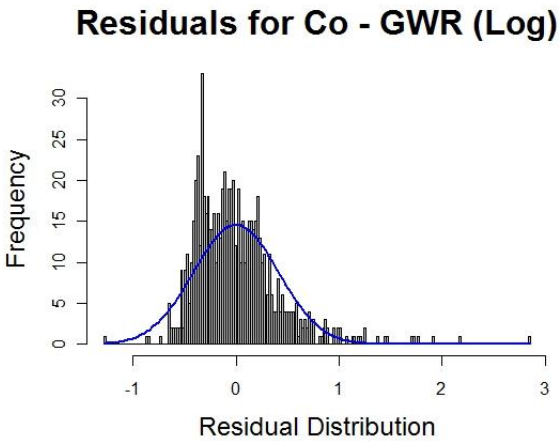


Figure 110 : Histogram for Cobalt Residuals - GWR (Log)

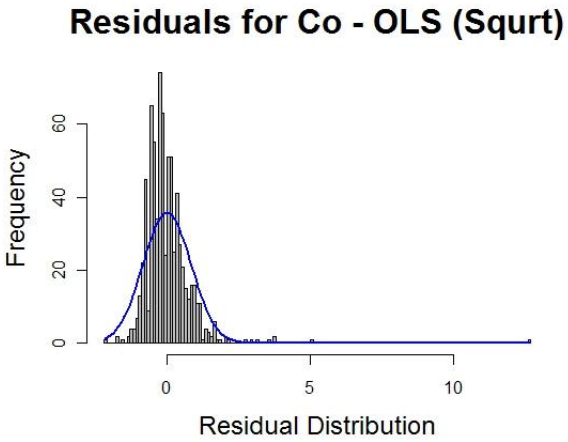


Figure 111 : Histogram for Cobalt Residuals - OLS (Sqrt)

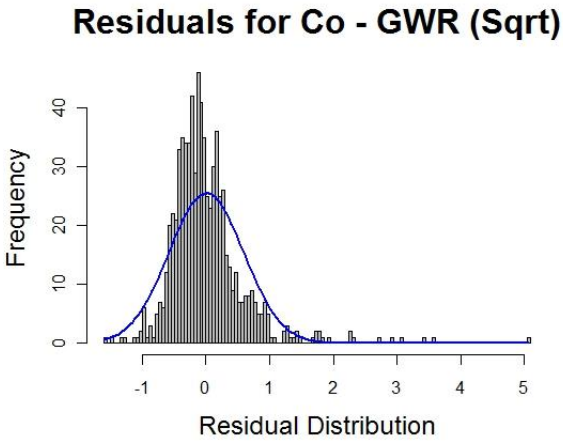


Figure 112 : Histogram for Cobalt Residuals - GWR (Sqrt)

QQ plots - residuals analysis

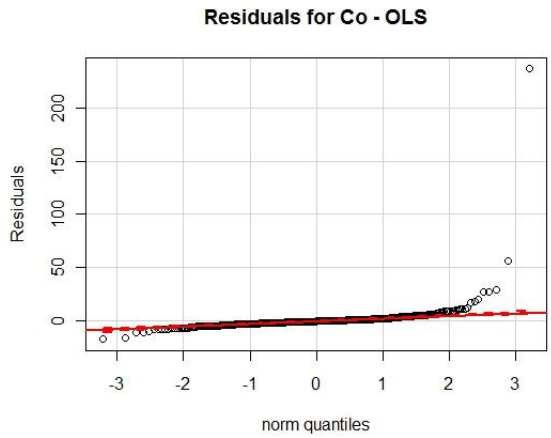


Figure 113 : QQ plot for Cobalt – OLS

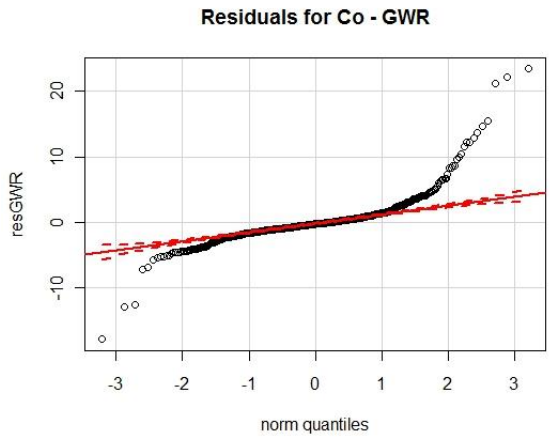


Figure 114 : QQ plot for Cobalt – GWR

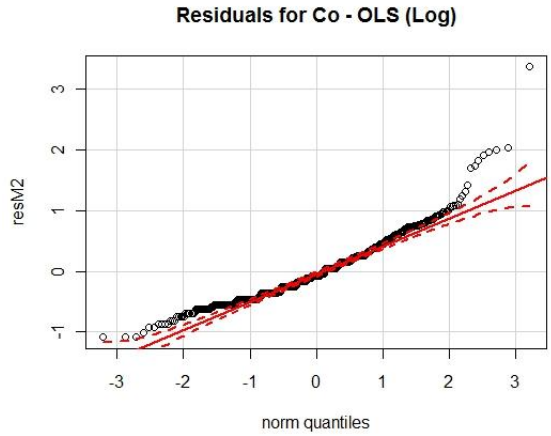


Figure 115 : QQ plot for Cobalt - OLS (Log)

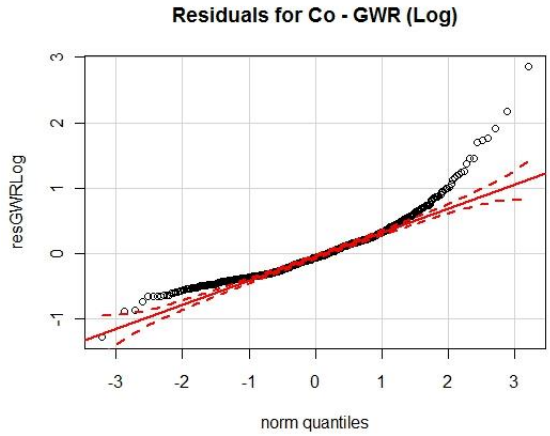


Figure 116 : QQ plot for Cobalt - GWR (Log)

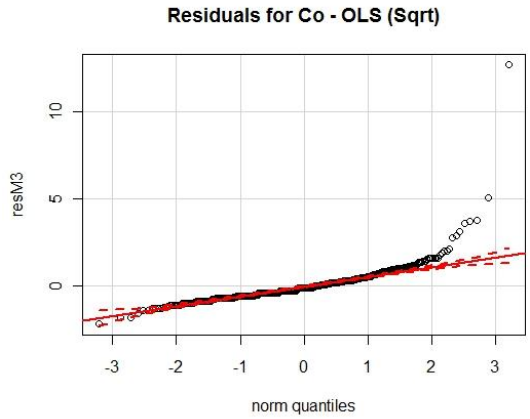


Figure 117 : QQ plot for Cobalt - OLS (Sqrt)

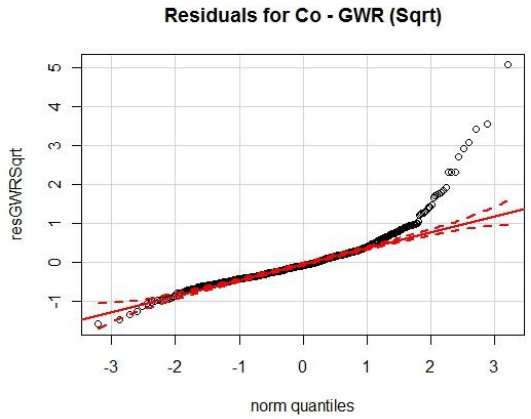


Figure 118 : QQ plot for Cobalt - GWR (Sqrt)

Variance of errors

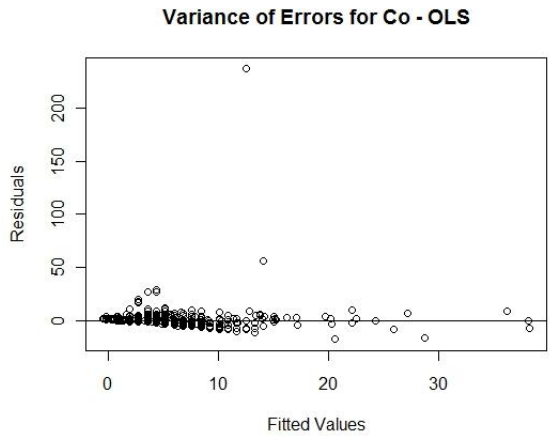


Figure 119 : Variance of Errors for Cobalt – OLS

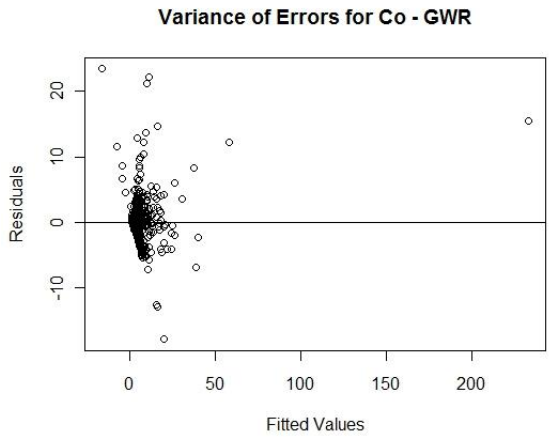


Figure 120 : Variance of Errors for Cobalt – GWR

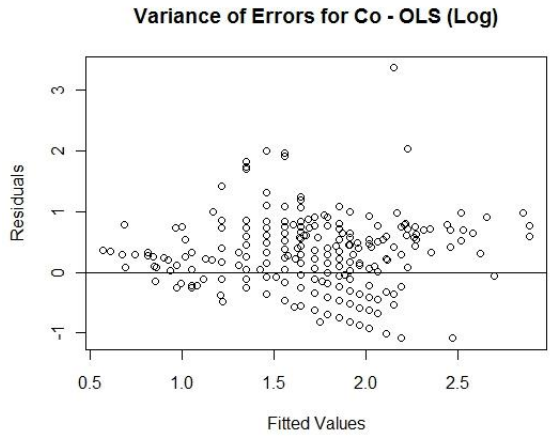


Figure 121 : Variance of Errors for Cobalt - OLS (Log)

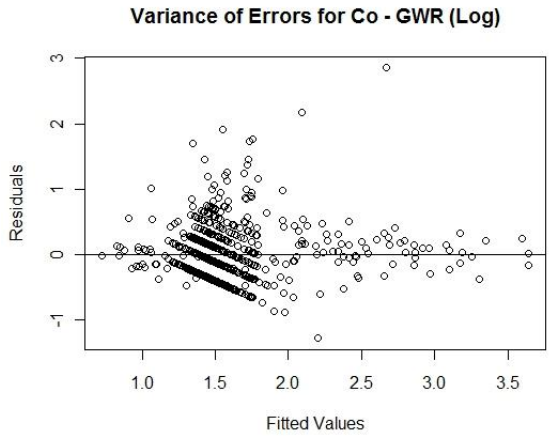


Figure 122 : Variance of Errors for Cobalt - GWR (Log)

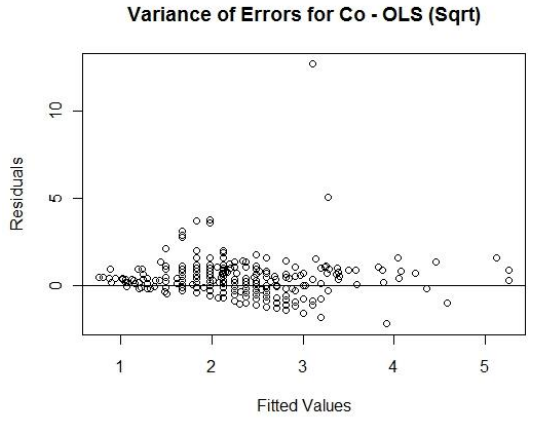


Figure 123 : Variance of Errors for Co - OLS (Sqrt)

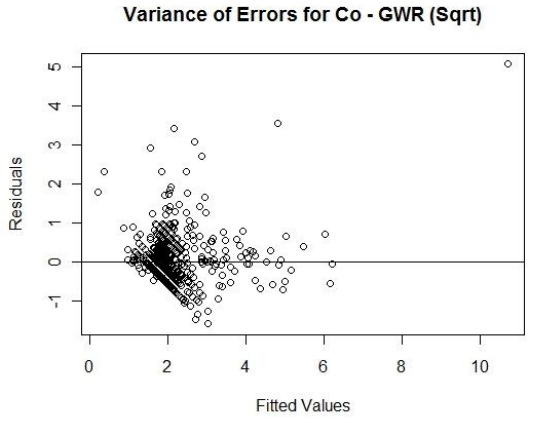


Figure 124 : Variance of Errors for Co - GWR (Sqrt)

Nickel

Concentration levels and maps

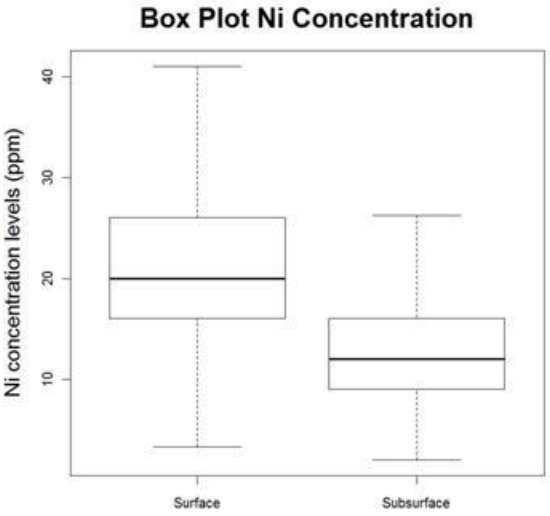


Figure 125: Box Plot of Ni Concentration without outliers

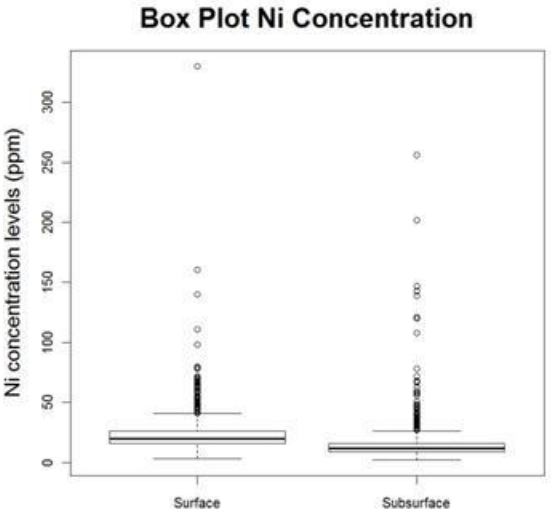


Figure 126: Box Plot of Ni Concentration with outliers

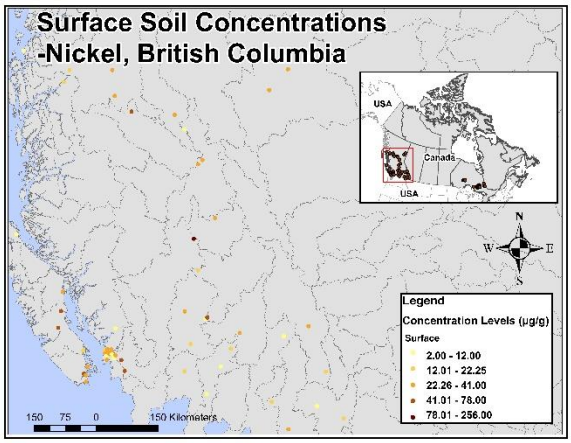


Figure 127 : Surface Concentration Levels for Nickel (British Columbia)

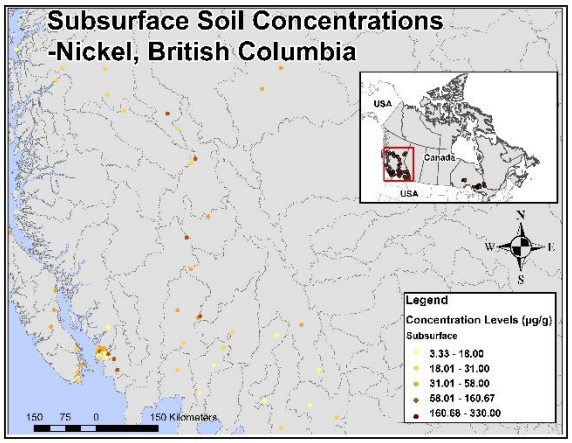


Figure 128 : Subsurface Concentration Levels for Nickel (British Columbia)

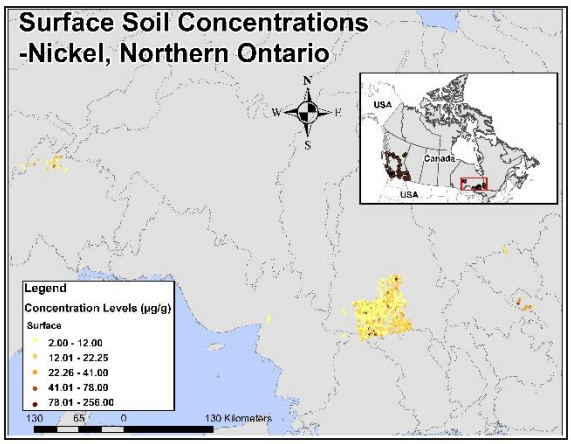


Figure 129 : Surface Concentration Levels for Nickel (Ontario)

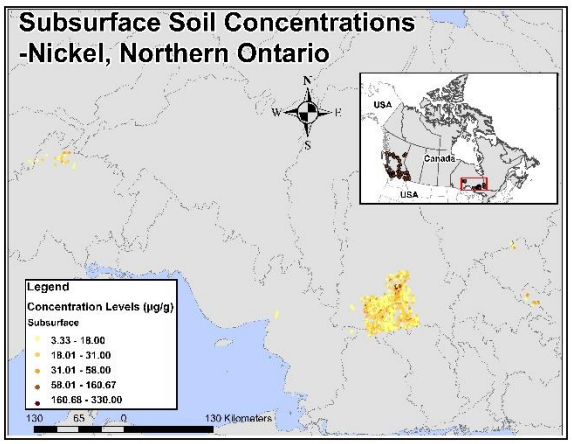


Figure 130 : Surface Concentration Levels for Nickel (Ontario)

Correlation analysis

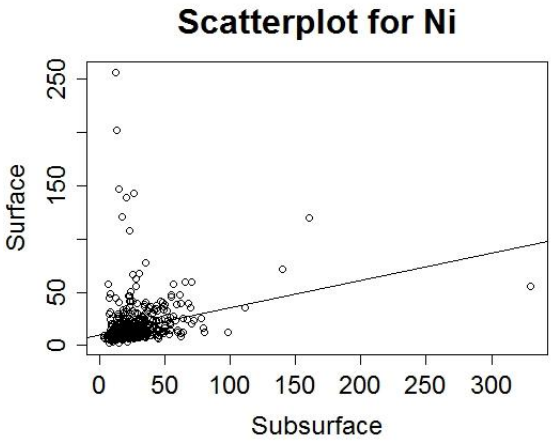


Figure 131 : Correlation between Surface and Subsurface Concentration for Ni

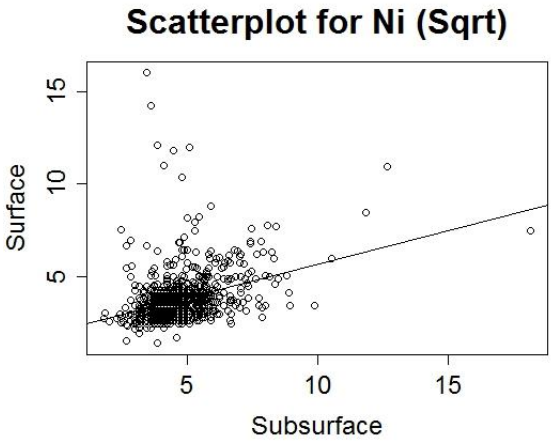


Figure 133 : Correlation between Surface and Subsurface Concentration for Ni (Sqrt)

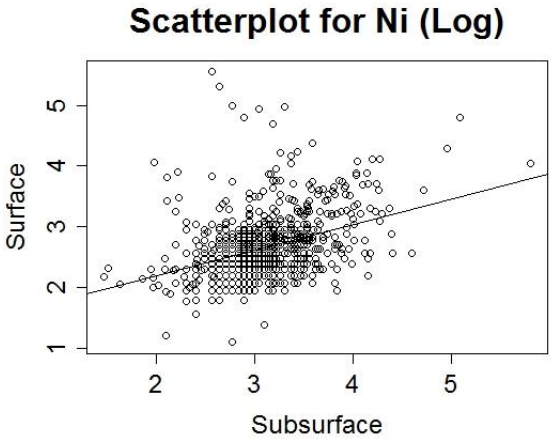


Figure 132 : Correlation between Surface and Subsurface Concentration for Ni (Log)

Table 42 : Correlation analysis between surface and subsurface Nickel concentration levels

Data	Correlation levels
Normal	0.26
Log-transformed	0.38
Squared-transformed	0.35

Ordinary Least Squares results

Table 43: OLS Analysis for Normal, Logarithmic, and Squared Transform

Nickel (n = 860)	Normal	Log	Squared
R^2 adjusted	0.07*	0.14*	0.12*
AICc	7263	1149	2812
Coefficient	0.256058*	0.419429*	0.361928*
Standard Errors	0.032746	0.035170	0.033336
t-statistic	7.819524	11.925635	10.857001
F-statistics (p-value)	61.14*	142.22*	117.87*

*** indicates significance (<0.05)**

Table 44: Residual Analysis for OLS

Nickel (n = 860)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.414 (<2.2e-16)	0.8918 (<2.2e-16)	0.8918 (<2.2e-16)
Moran's I	0.022834	0.071832	0.049827
Expected Index	-0.001164	-0.001164	-0.001164
Variance	0.000013	0.000015	0.000014
z-score	6.557755	19.012499	13.442252
p-value	0.000000	0.000000	0.0000000

Geographically Weighted Regression results

Table 45: GWR Analysis for Normal, Logarithmic, and Squared Transform

Nickel (n = 860)	Normal	Log	Squared
Neighbors	139	115	116
Residual Squares	201735.51	140.159	1032.002
Effective Number	43.09	53.34	51.746
Sigma	15.71	0.417	1.130
AICc	7205	968	2682
R^2 adjusted	0.15	0.33	0.27
Coefficient range	-0.16015 – 0.861728	-0.240749 – 0.883963	-0.212461 – 0.945109
R^2 range	0.00 – 0.75	0.00 – 0.62	0.00 – 0.68

Table 46: Residuals Analysis for GWR

Nickel (n = 860)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.3732 (<2.2e-16)	0.8624 (<2.2e-16)	0.6437 (<2.2e-16)
Moran's I	0.00323	0.007170	0.004572
Expected Index	-0.001164	-0.001164	-0.001164
Variance	0.000013	0.000015	0.000014
z-score	0.411136	2.175638	1.521692
p-value	0.680973	0.029582	0.128086

Histograms - residuals analysis

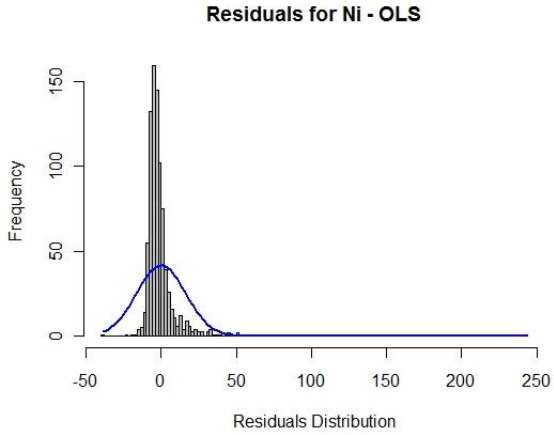


Figure 134 : Histogram for Nickel Residuals – OLS

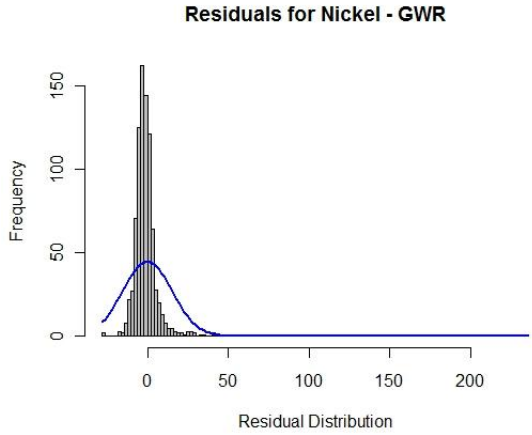


Figure 135 : Histogram for Nickel Residuals – GWR

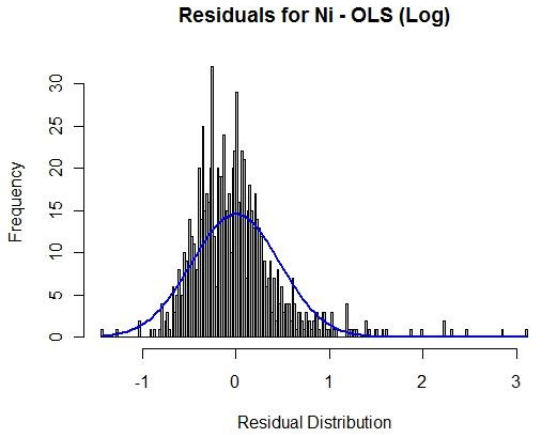


Figure 136 : Histogram for Nickel Residuals - OLS (Log)

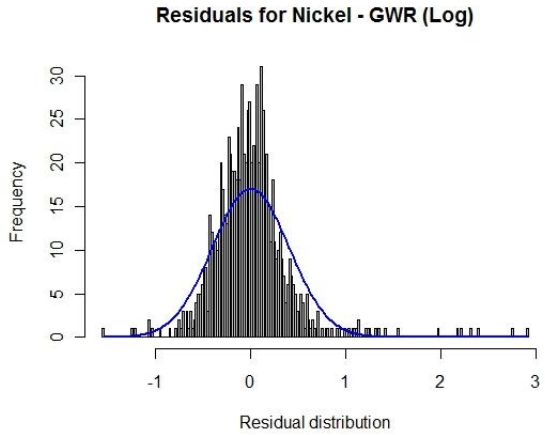


Figure 137 : Histogram for Nickel Residuals - GWR (Log)

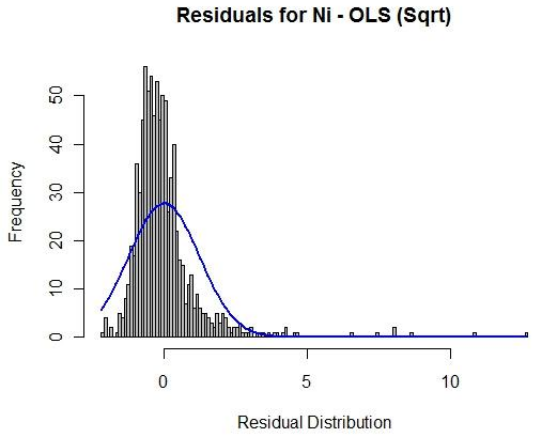


Figure 138 : Histogram for Nickel Residuals - OLS (Sqrt)

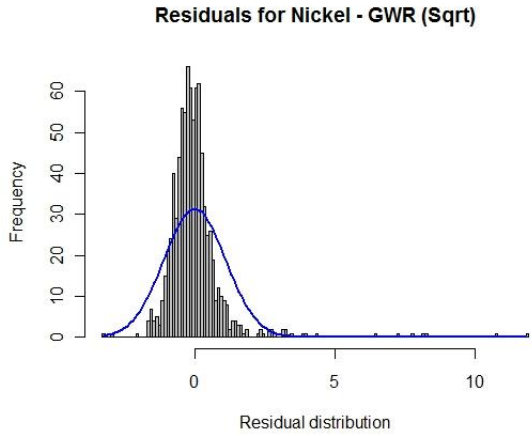


Figure 139 : Histogram for Nickel Residuals - GWR (Sqrt)

QQ plots - residuals analysis

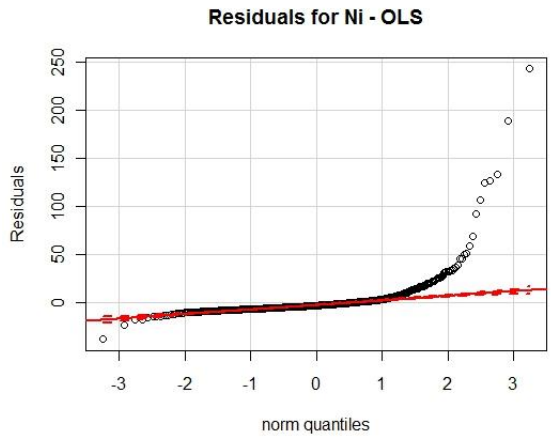


Figure 140 : QQ plot for Nickel Residuals - OLS

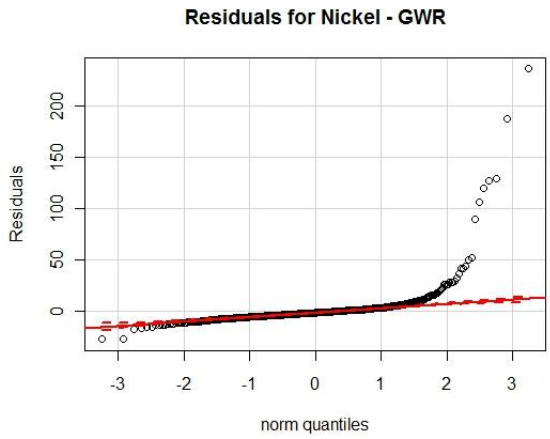


Figure 141 : QQ plot for Nickel Residuals - GWR

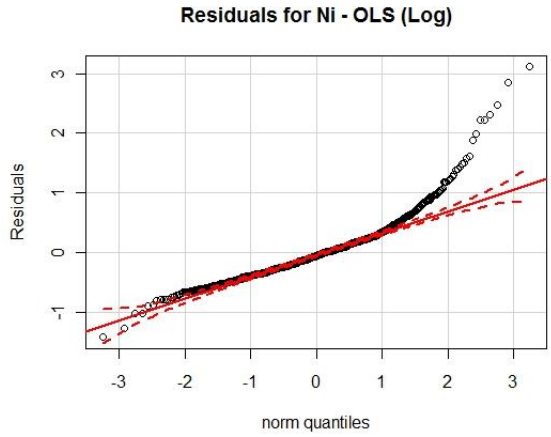


Figure 142 : QQ plot for Nickel Residuals - OLS (Log)

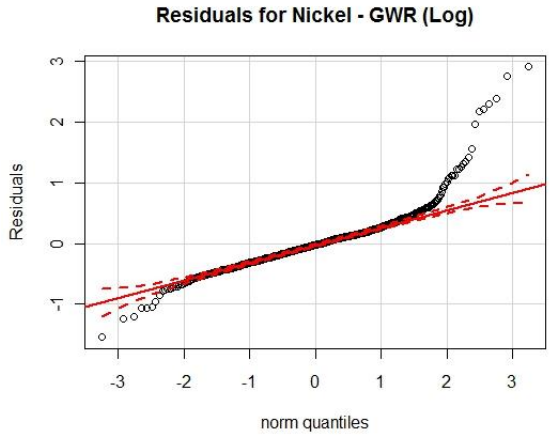


Figure 143 : QQ plot for Nickel Residuals - GWR (Log)

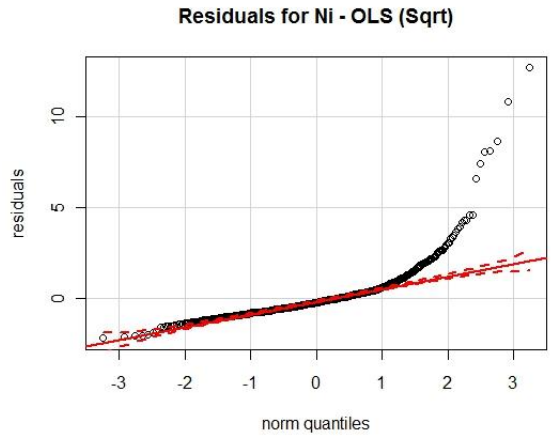


Figure 144 : QQ plot for Nickel Residuals - OLS (Sqrt)

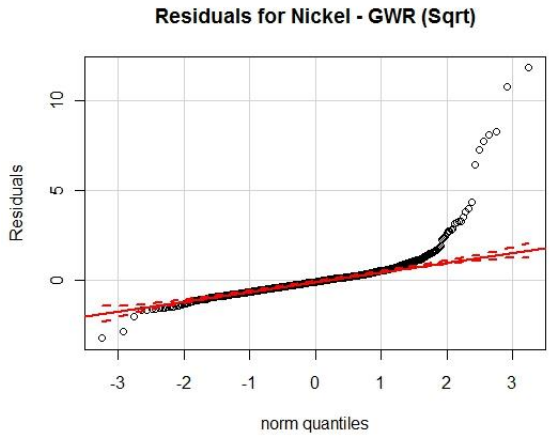


Figure 145 : QQ plot for Nickel Residuals - GWR (Sqrt)

Variance of errors

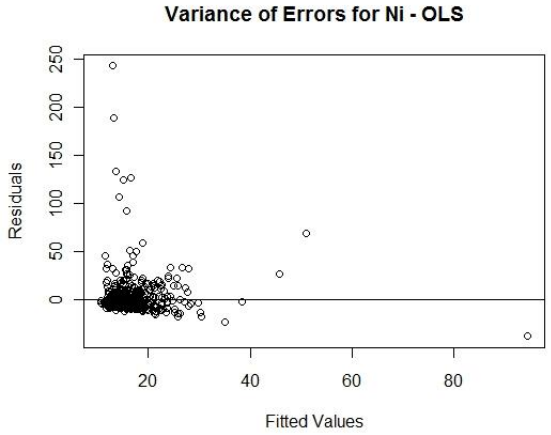


Figure 146 : Variance of Errors for Nickel - OLS

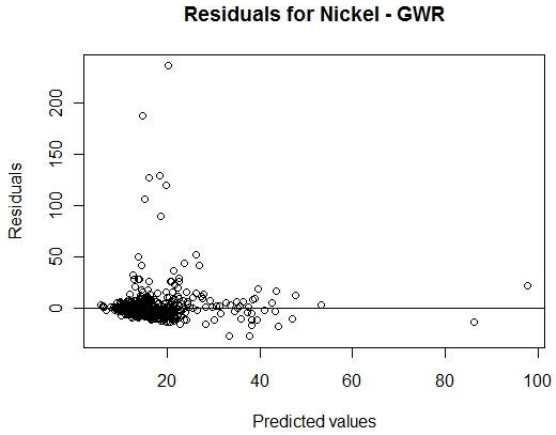


Figure 147 : Variance of Errors for Nickel - GWR

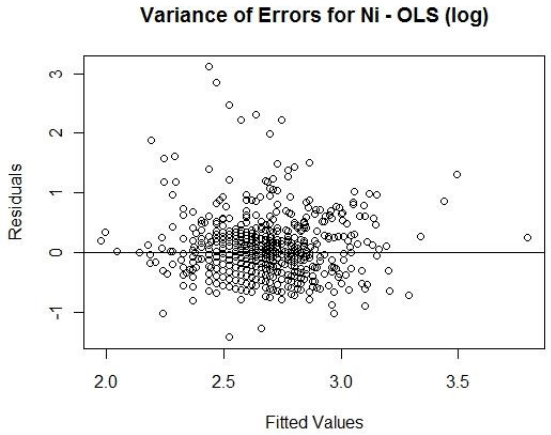


Figure 148 : Variance of Errors for Nickel - OLS (Log)

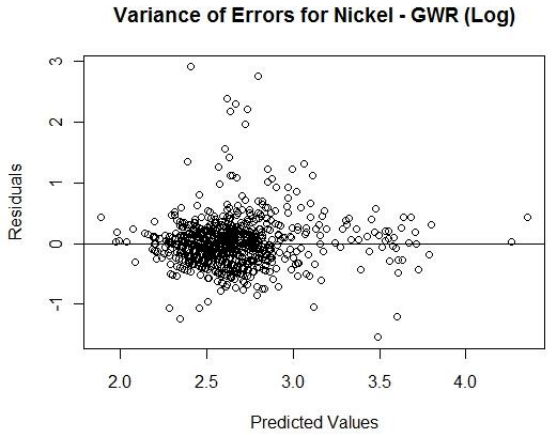


Figure 149 : Variance of Errors for Nickel - GWR (Log)

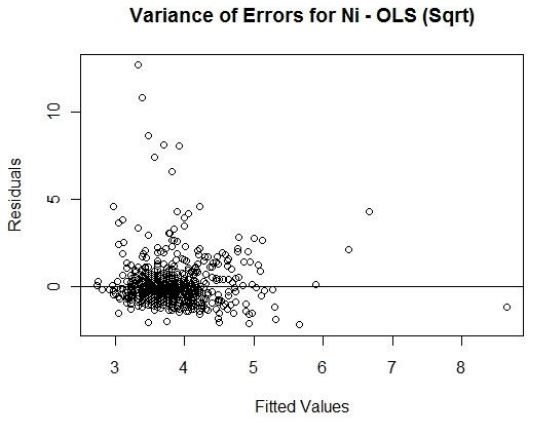


Figure 150 : Variance of Errors for Nickel - OLS (Sqrt)

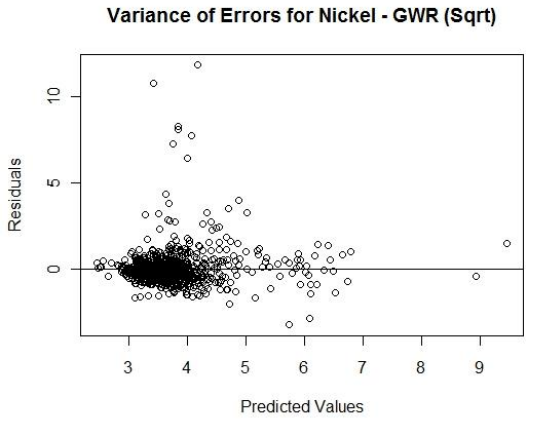


Figure 151 : Variance of Errors for Nickel - GWR (Sqrt)

Lead

Concentration levels and maps

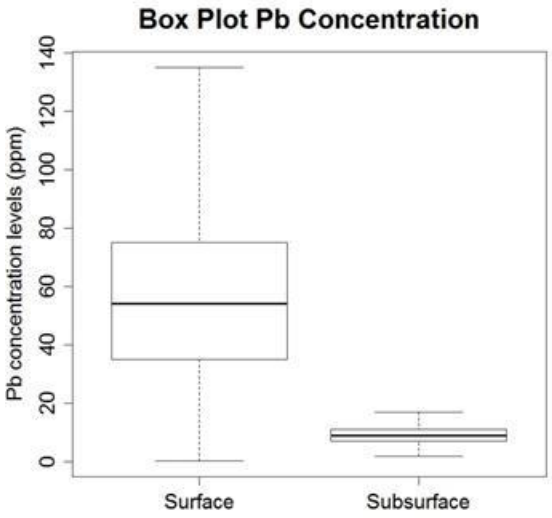


Figure 152: Box Plot of Lead Concentration without outliers

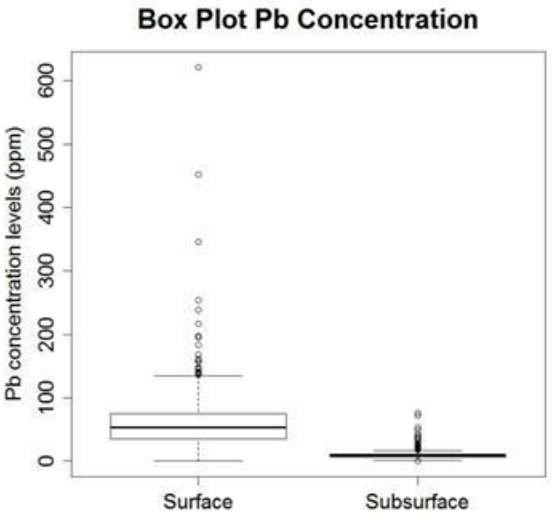


Figure 153: Box Plot of Lead Concentration with outliers

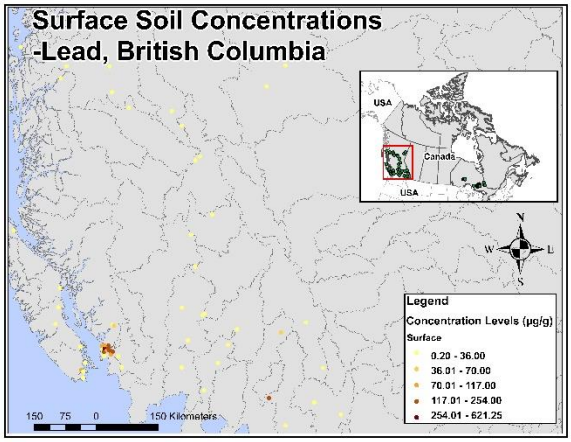


Figure 154 : Surface Concentration Levels for Lead (British Columbia)

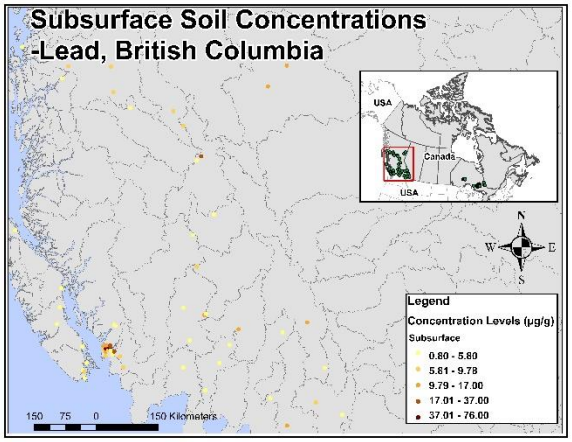


Figure 155 : Subsurface Concentration Levels for Lead (British Columbia)

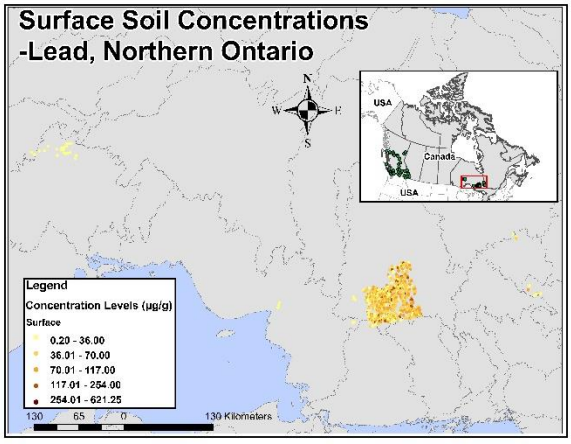


Figure 156 : Surface Concentration Levels for Lead (Ontario)

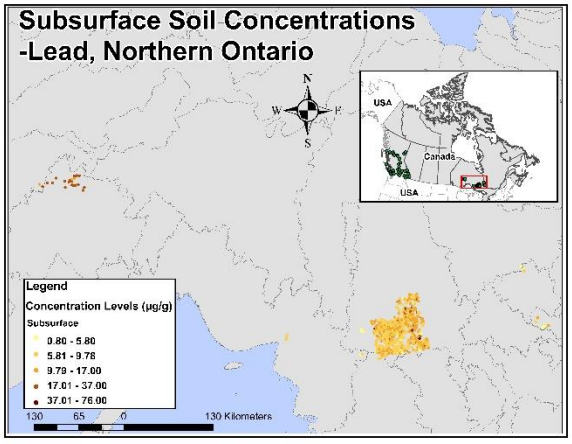


Figure 157 : Subsurface Concentration Levels for Lead (Ontario)

Correlation analysis

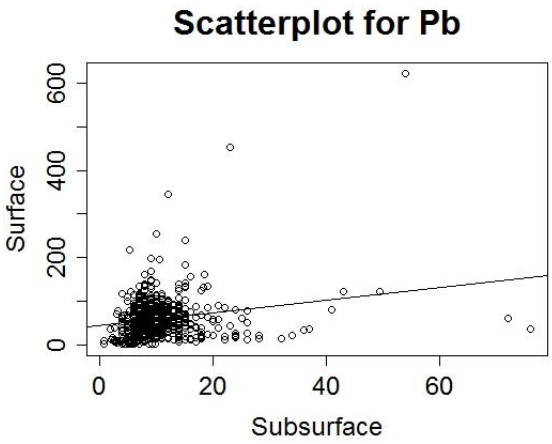


Figure 158 : Correlation between Surface and Subsurface Concentration for PB

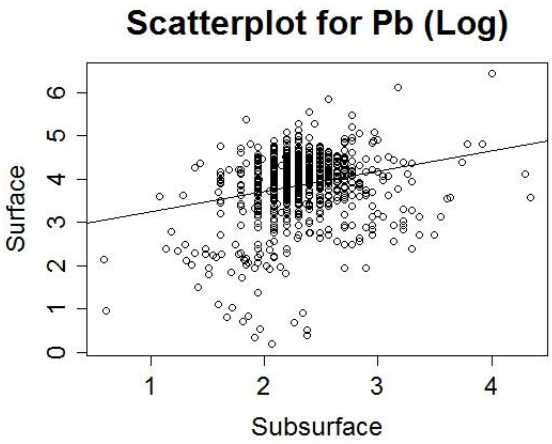


Figure 159 : Correlation between Surface and Subsurface Concentration for Lead (Log)

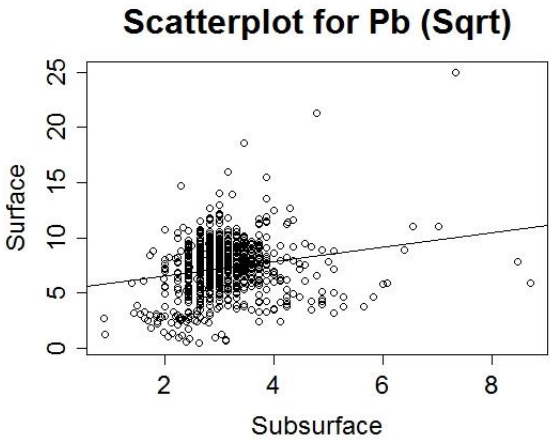


Figure 160 : Correlation between Surface and Subsurface Concentration for Lead (Sqrt)

Table 47 : Correlation analysis between surface and subsurface Lead concentration levels

Data	Correlation levels
Normal	0.20
Log-transformed	0.23
Squared-transformed	0.20

Ordinary Least Squares results

Table 48: OLS Analysis for Normal, Logarithmic, and Squared Transform

Lead (n = 846)	Normal	Log	Squared
R^2 adjusted	0.04*	0.05*	0.04*
AICc	8690	1938	3890
Coefficient	1.432874*	0.465382*	0.654042*
Standard Errors	0.239640	0.066900	0.112910
t-statistic	5.979281	6.956389	5.792621
F-statistics (p-value)	35.75*	48.39*	33.55*

*** indicates significance (<0.05)**

Table 49: Residual Analysis for OLS

Lead (n = 846)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.7827 (<2.2e-16)	0.9037 (<2.2e-16)	0.9037 (<2.2e-16)
Moran's I	0.021279	0.070487	0.046926
Expected Index	-0.001183	-0.001183	-0.001183
Variance	0.000014	0.000014	0.000014
z-score	6.031233	18.879970	12.672602
p-value	0.000000	0.000000	0.000000

Geographically Weighted Regression results

Table 50: GWR Analysis for Normal, Logarithmic, and Squared Transform

Lead (n = 846)	Normal	Log	Squared
Neighbours	27	55	49
Residual Squares	740444.258	234.38	2847
Effective Number	198.82	107.29	117.51
Sigma	33.82	0.56	1.977
AICc	8509	1497	3629
R^2 adjusted	0.35	0.48	0.35
Coefficient range	-6.303045 – 25.252047	-0.618096 – 4.062231	-1.452569 – 5.407172
R^2 range	0.00 – 0.80	0.00 – 0.64	0.00 – 0.57

Table 51: Residuals Analysis for GWR

Lead (n = 846)	Normal	Log	Squared
Shapiro-Wilk test (p-value)	0.8239 (<2.2e-16)	0.9719 (1.076e-11)	0.9539 (1.252e-15)
Moran's I	-0.006774	0.008467	-0.000378
Expected Index	-0.001183	-0.001183	-0.001183
Variance	0.000014	0.000014	0.000014
z-score	-1.484405	2.540950	0.212247
p-value	0.137701	0.011055	0.831915

Histograms - residuals analysis

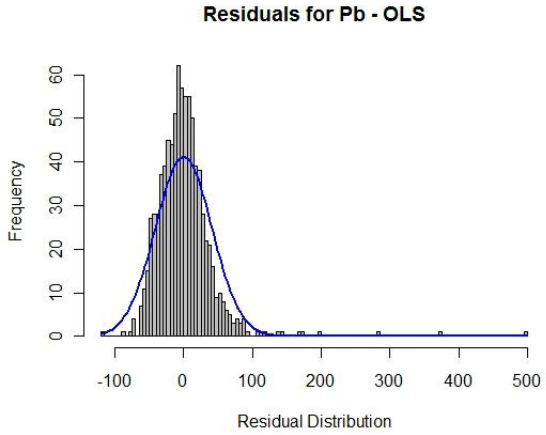


Figure 161 : Histogram for Lead Residuals - OLS

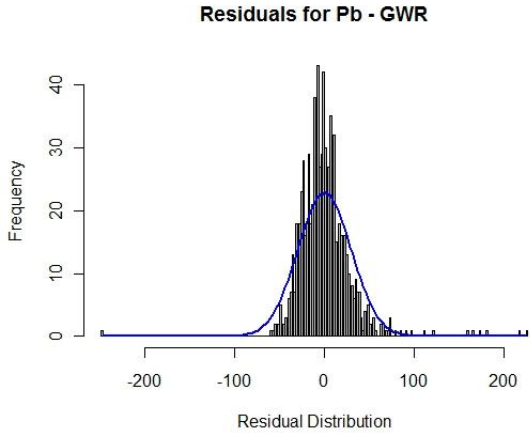


Figure 162 : Histogram for Lead Residuals - GWR

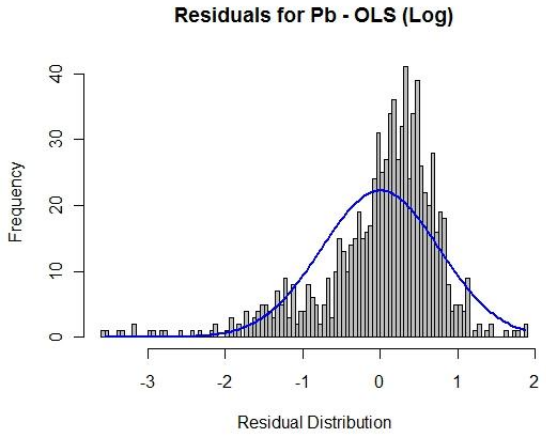


Figure 163 : Histogram for Lead Residuals - OLS (Log)

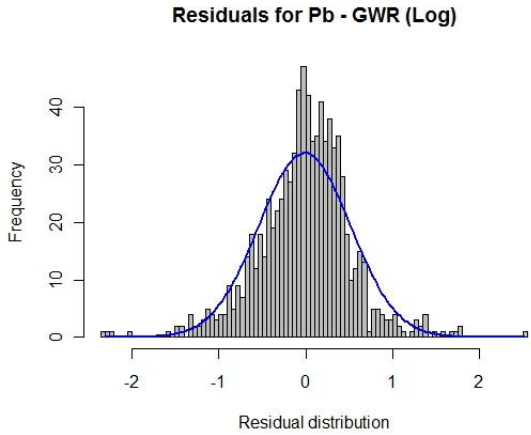


Figure 164 : Histogram for Lead Residuals - GWR (Log)

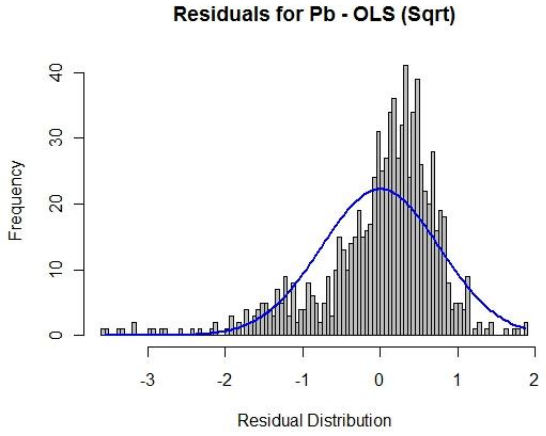


Figure 165 : Histogram for Lead Residuals - OLS (Sqrt)

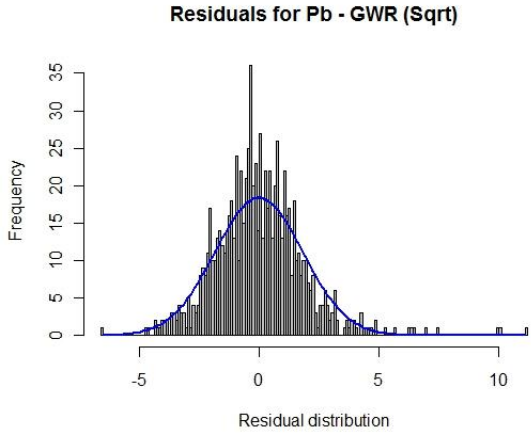


Figure 166 : Histogram for Lead Residuals - GWR (Sqrt)

QQ plots - residuals analysis

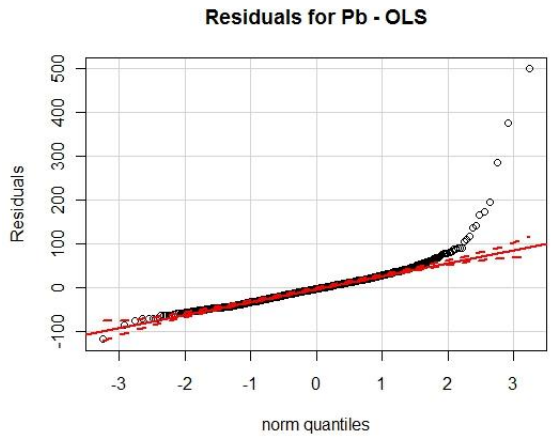


Figure 167 : QQ plot for Lead residuals - OLS

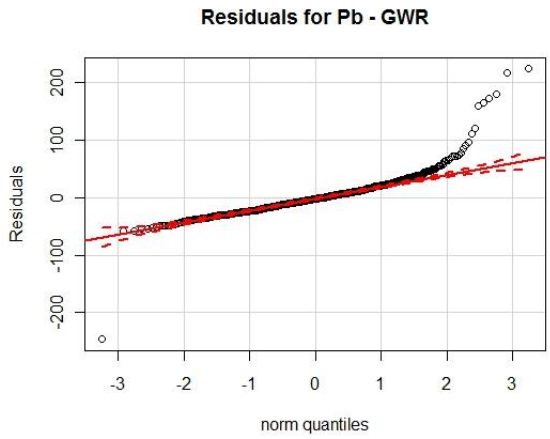


Figure 168 : QQ plot for Lead residuals - GWR

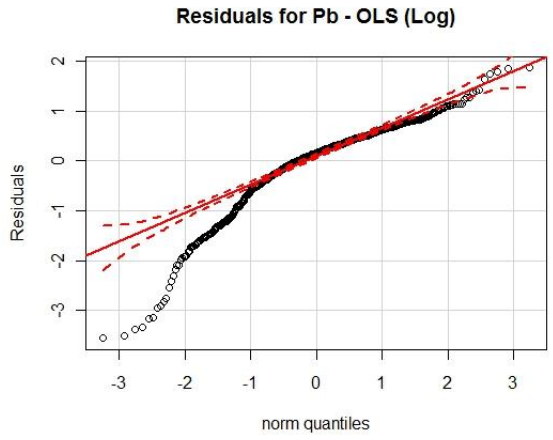


Figure 169 : QQ plot for Lead residuals - OLS (Log)

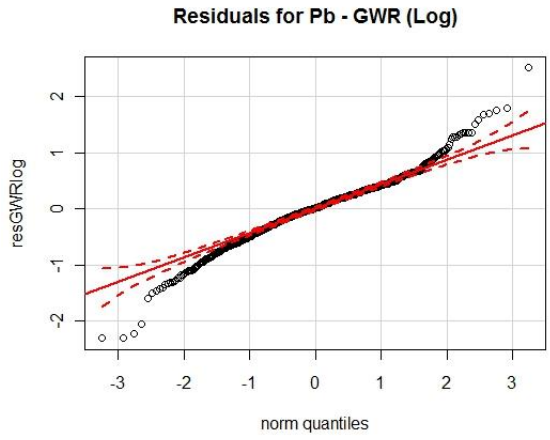


Figure 170 : QQ plot for Lead residuals - GWR (Log)

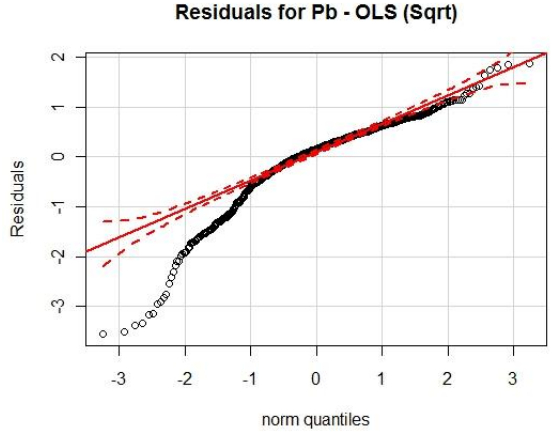


Figure 171 : QQ plot for Lead residuals - OLS (Sqrt)

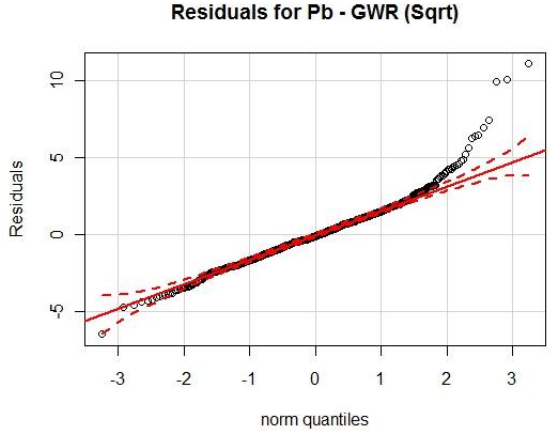


Figure 172 : QQ plot for Lead residuals - GWR (Sqrt)

Variance of errors

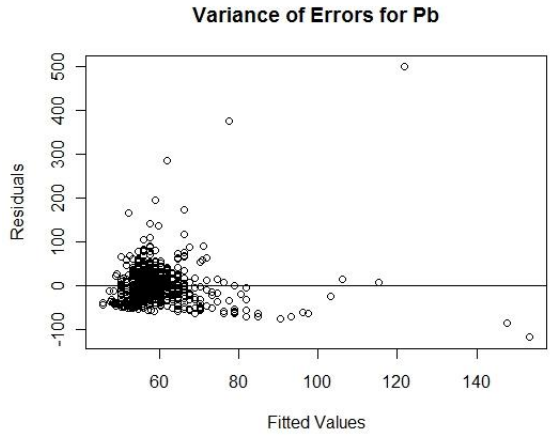


Figure 173 : Variance of Errors for Lead – OLS

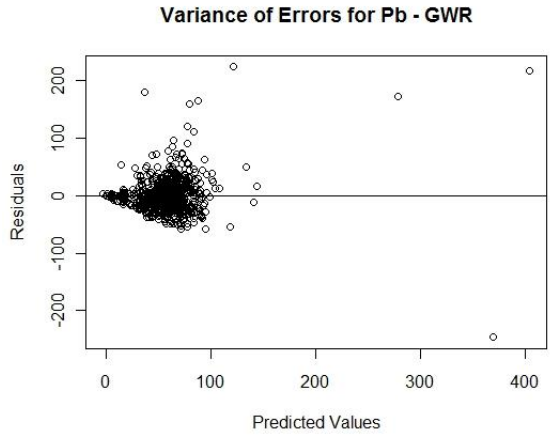


Figure 174 : Variance of Errors for Lead - GWR

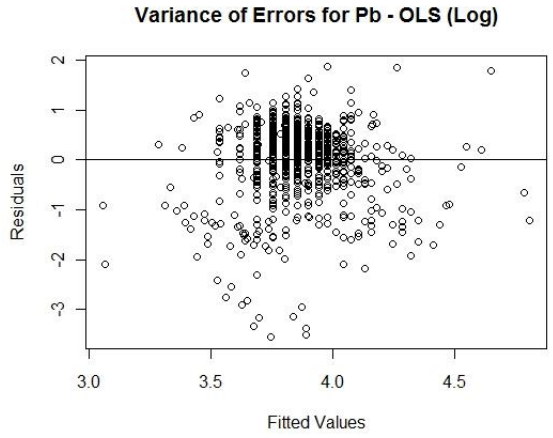


Figure 175 : Variance of Errors for Lead - OLS (Log)

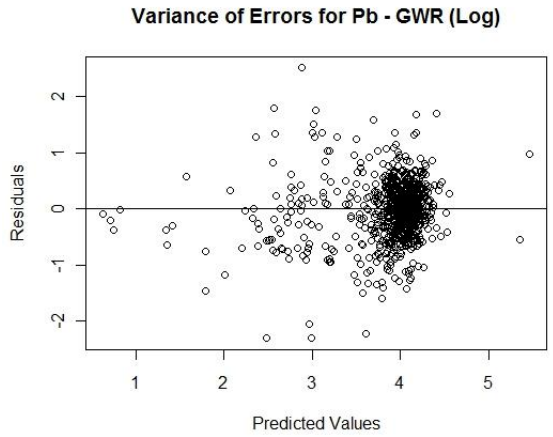


Figure 176 : Variance of Errors for Lead - GWR (Log)

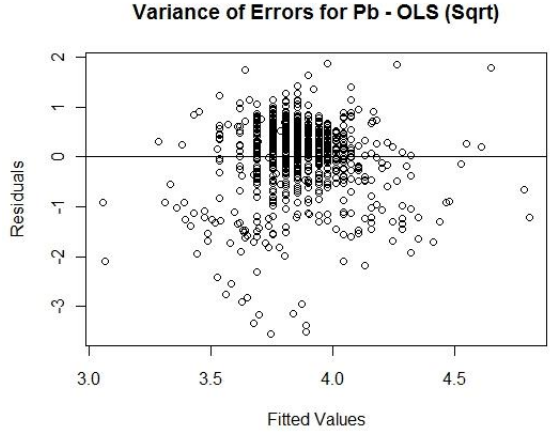


Figure 177 : Variance of Errors for Lead - OLS (Sqrt)

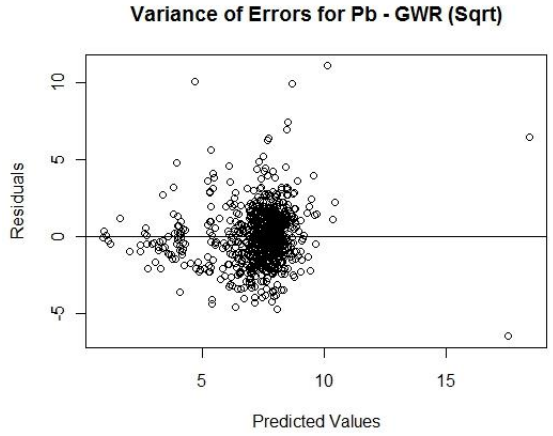


Figure 178 : Variance of Errors for Lead - GWR (Sqrt)

-Coefficient Distribution

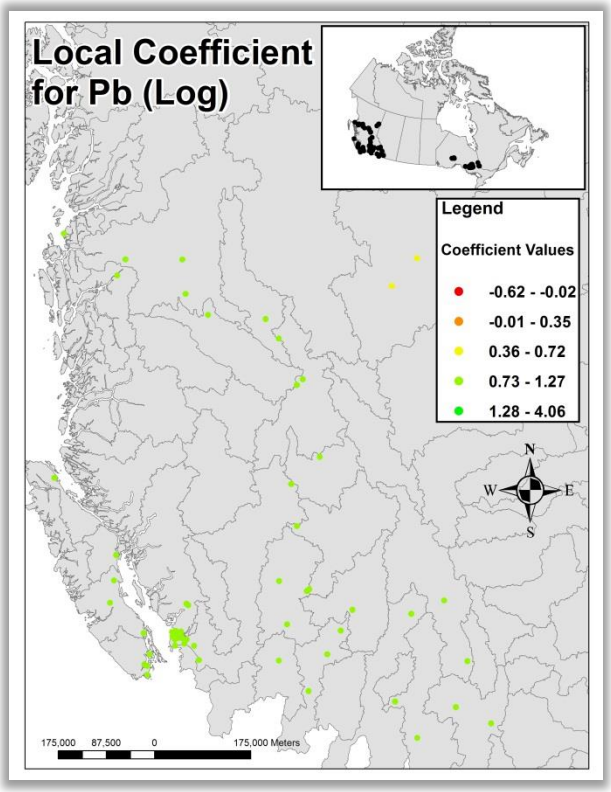


Figure 179 : Spatial distribution of the regression coefficients for BC (Log)

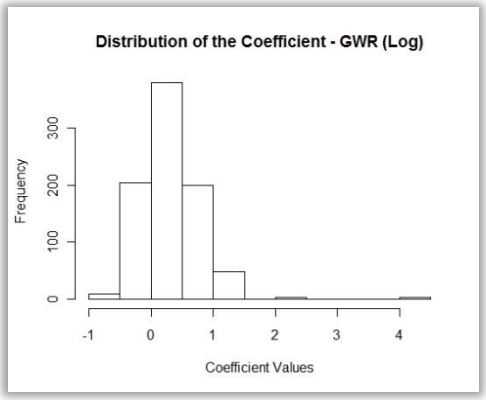


Figure 180 : Histogram of the coefficient (Log)

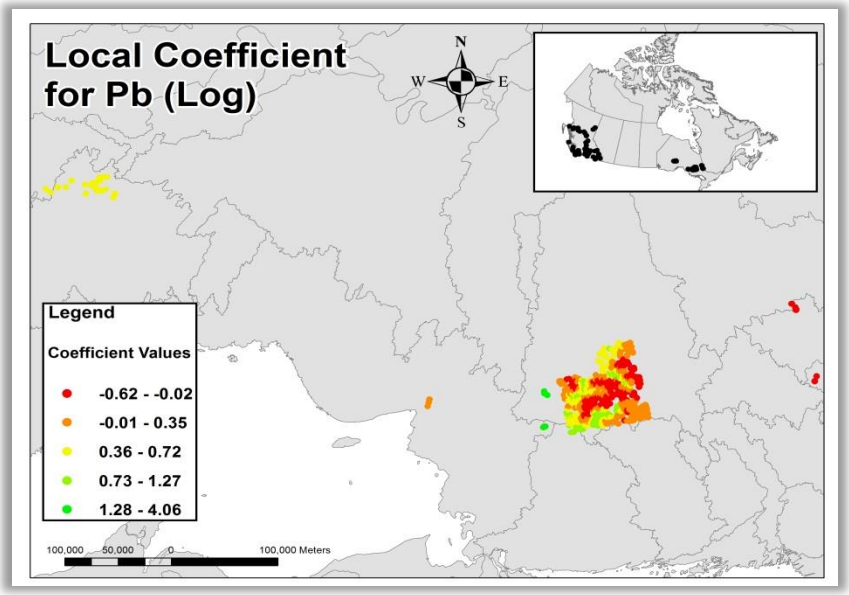


Figure 181 : Spatial distribution of the regression coefficients for Ontario (Log)

R-squared Distribution

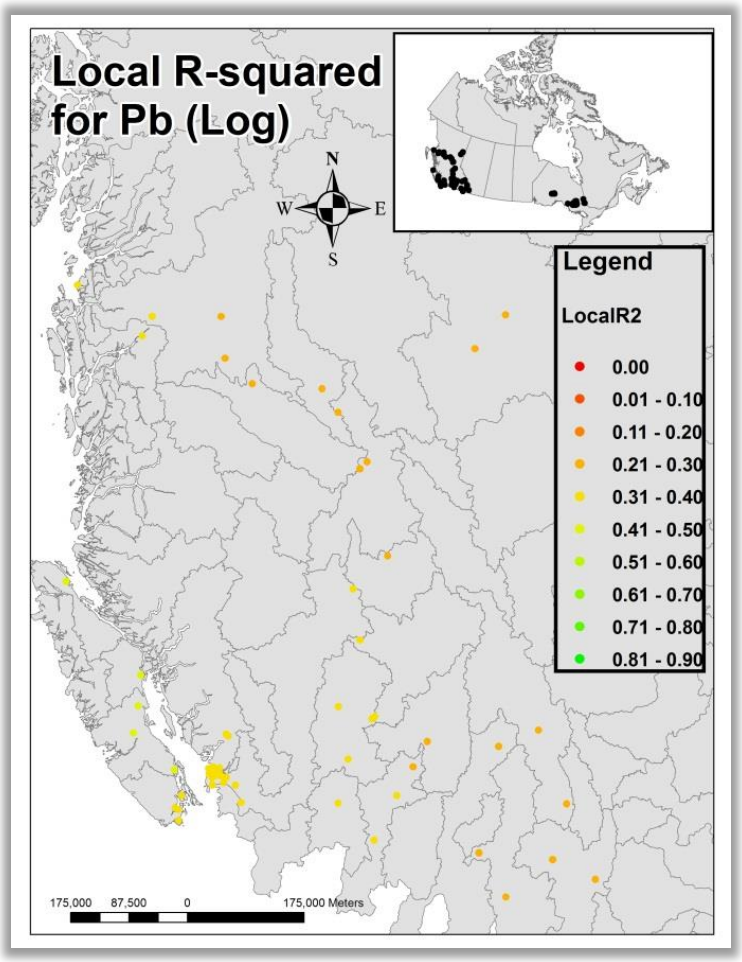


Figure 182 : Spatial distribution of the local R-squared for BC (Log)

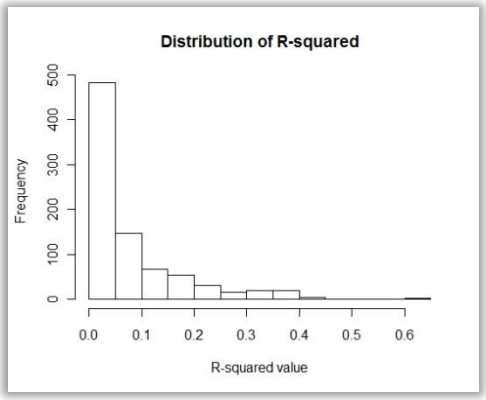


Figure 183 : Distribution of the R-squared for Lead

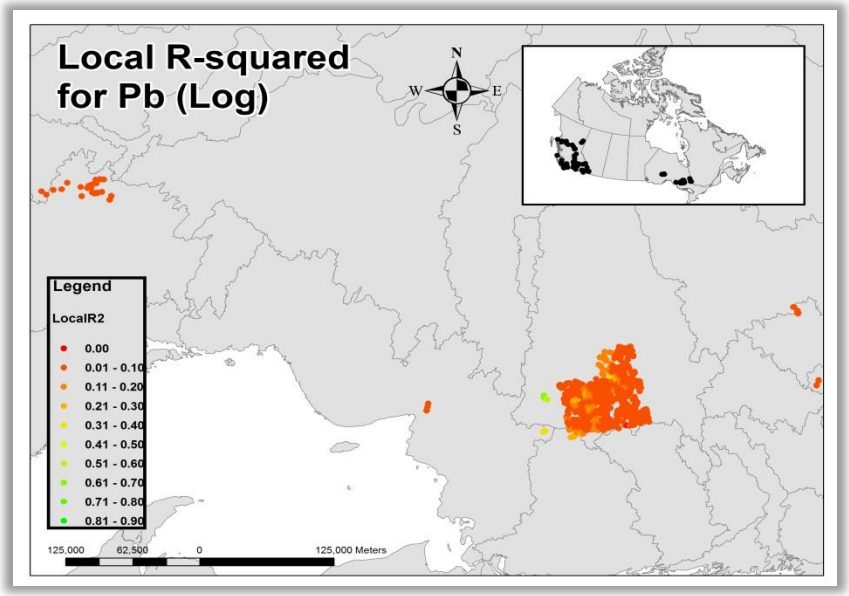


Figure 184 : Spatial distribution of the R-squared for Ontario (Log)

Appendix C – LUR results

This Appendix presents information about the concentration data and the results of the Arsenic and Lead LUR models, which includes the distribution of concentration levels, correlation analysis between the dependent and the independent variables, regression results, model adequacy, and bootstrap analysis.

Figures 188 to 199 (Arsenic) and 207 to 224 (Lead) show some level of correlation between the measured concentrations and the independent variables used to develop the LUR models. The graphs include the correlation levels between the variables without transformation (blue lines), with the concentration log-transformed (red line), and both the concentration levels and the independent variables log-transformed (green line).

Table 52 (Arsenic) and Table 57 show the LUR analysis results, including R-squared and the variables significance levels. Table 53 (Arsenic) and Table 58 show the 95% confidence interval for the two LUR models coefficients. Table 54 (Arsenic) and Table 59 (Lead) show the normal distribution of the residuals for both LUR models. Table 55 (Arsenic) and Table 60 (Lead) show the analysis of variance significance for the variables, no spatial autocorrelation, and the explanatory values for each of the variables in the LUR models. Figures 202 to 204 (Arsenic) and 225 to 227 (Lead) show that the residuals for both models are normally distributed and are suitable for linear regression analysis.

Finally, Table 56 (Arsenic) and Tables 61 (Lead) show the distribution of the R-squared from the bootstrap analysis. Figures 230 to 233 show the consistant distribution of the Pb model coefficients after the bootstrap analysis.

List of Tables and Figures

Land Use Regression analysis – Arsenic model

Concentration levels

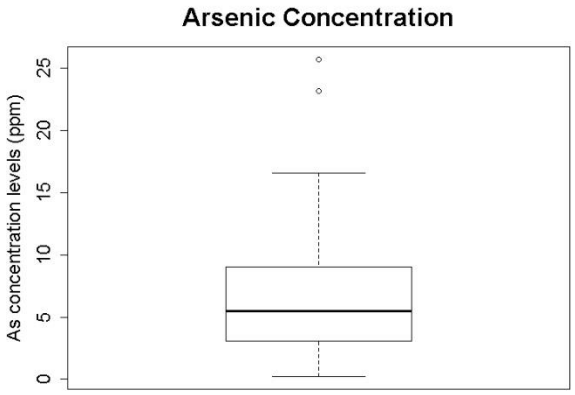


Figure 185 : Concentration levels for Arsenic

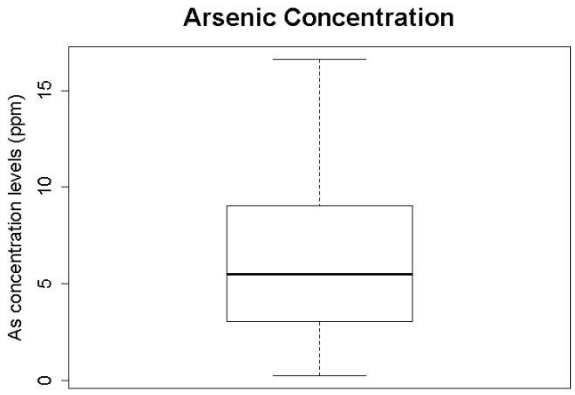


Figure 186 : Concentration levels for Arsenic (no outlier)

Correlation analysis

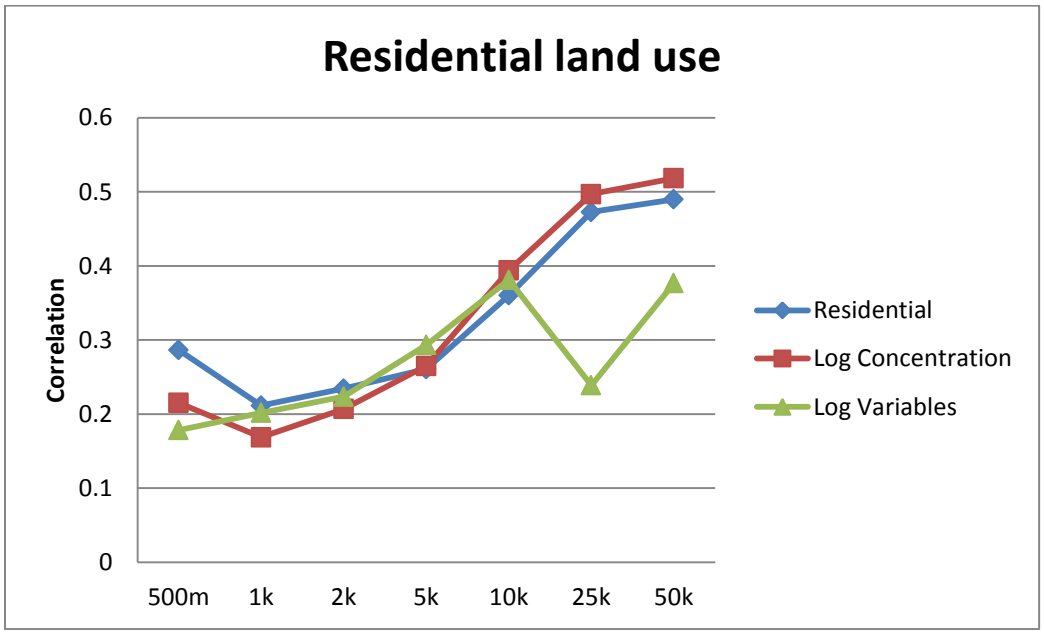


Figure 187 : Correlation between residential land use and concentration levels

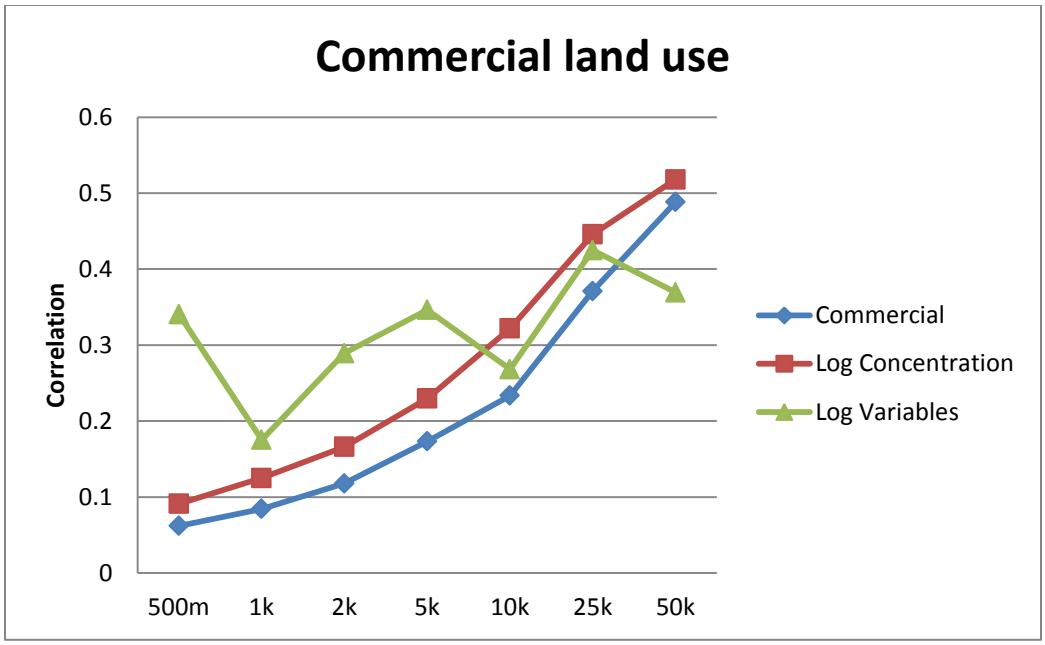


Figure 188 : Correlation between commercial land use and concentration levels

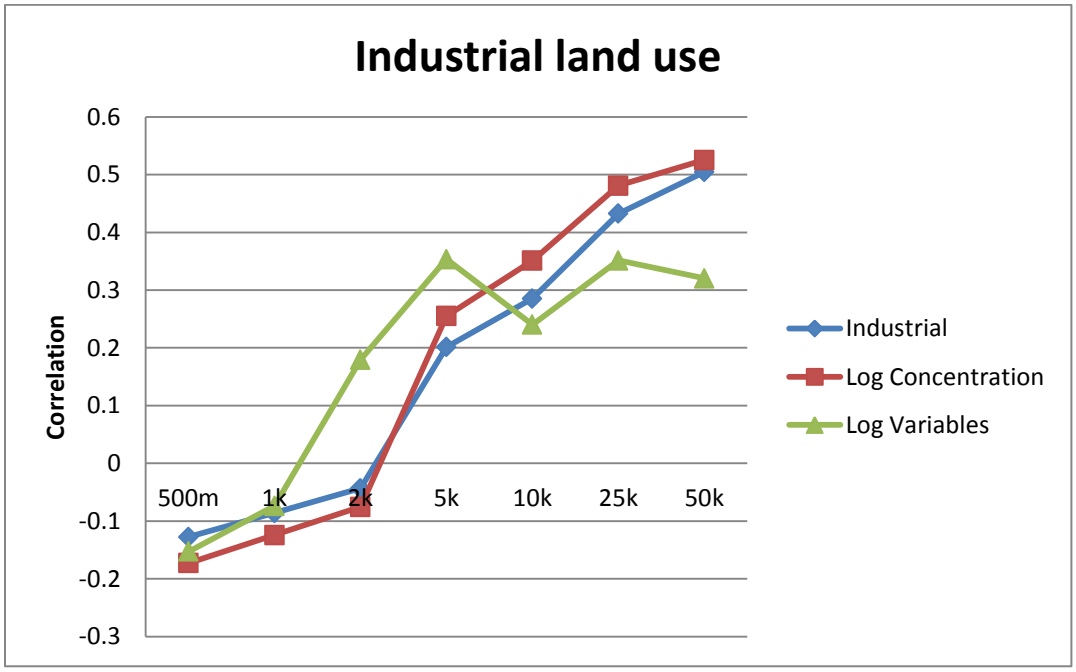


Figure 189 : Correlation between industrial land use and concentration levels

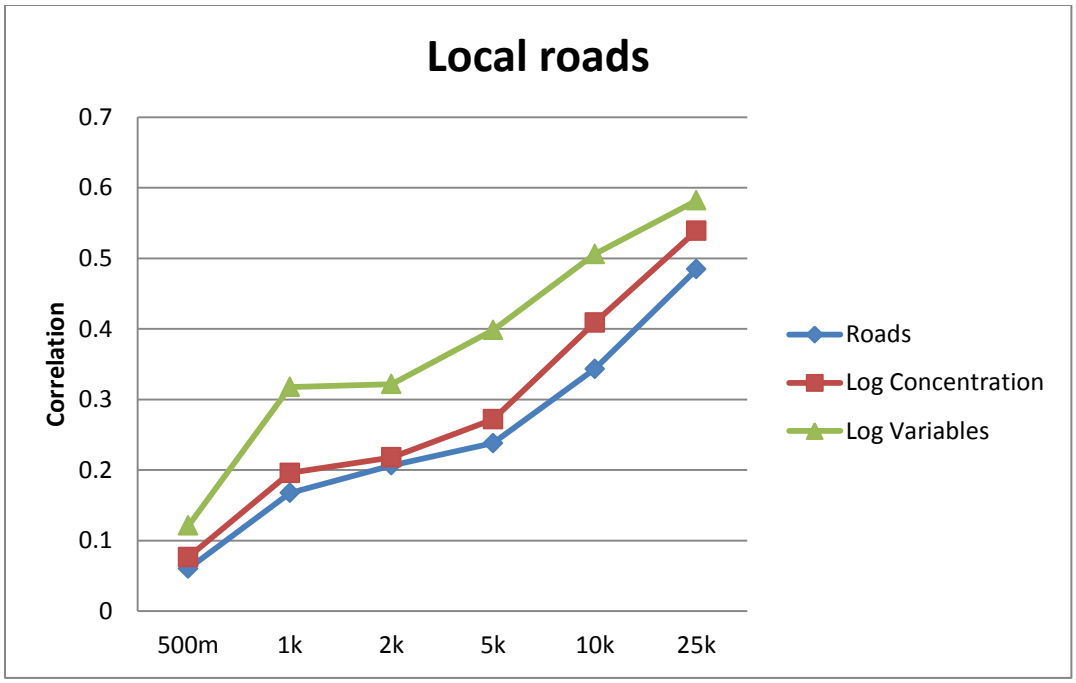


Figure 190 : Correlation between local roads and concentration levels

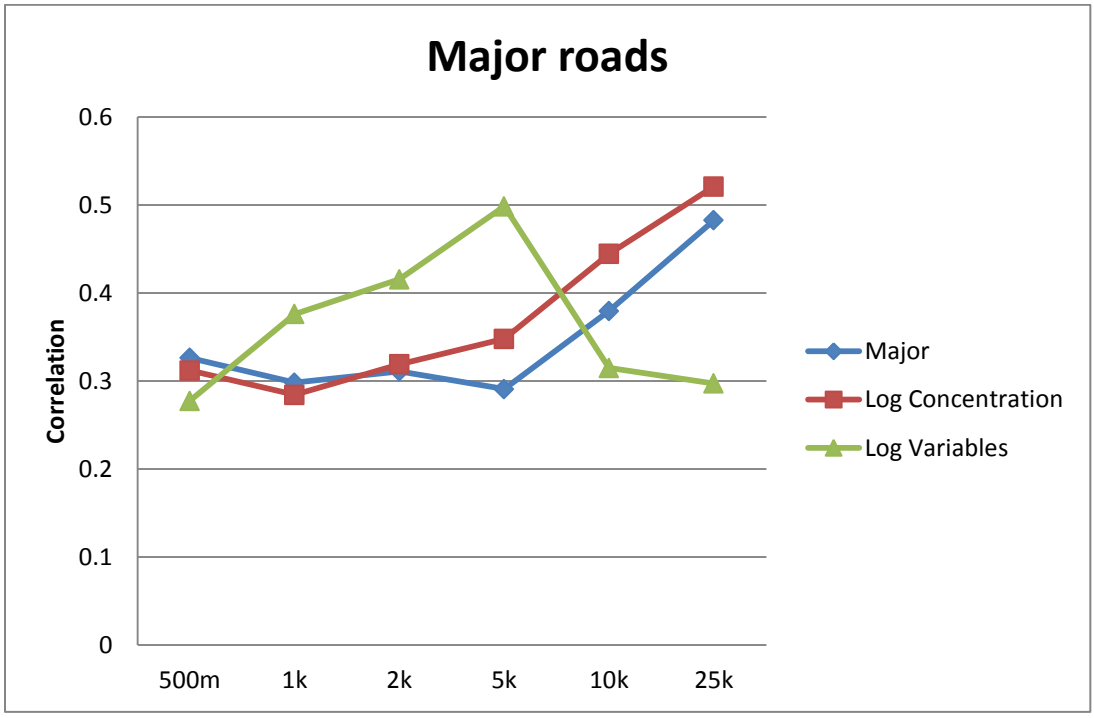


Figure 191 : Correlation between major roads and concentration levels

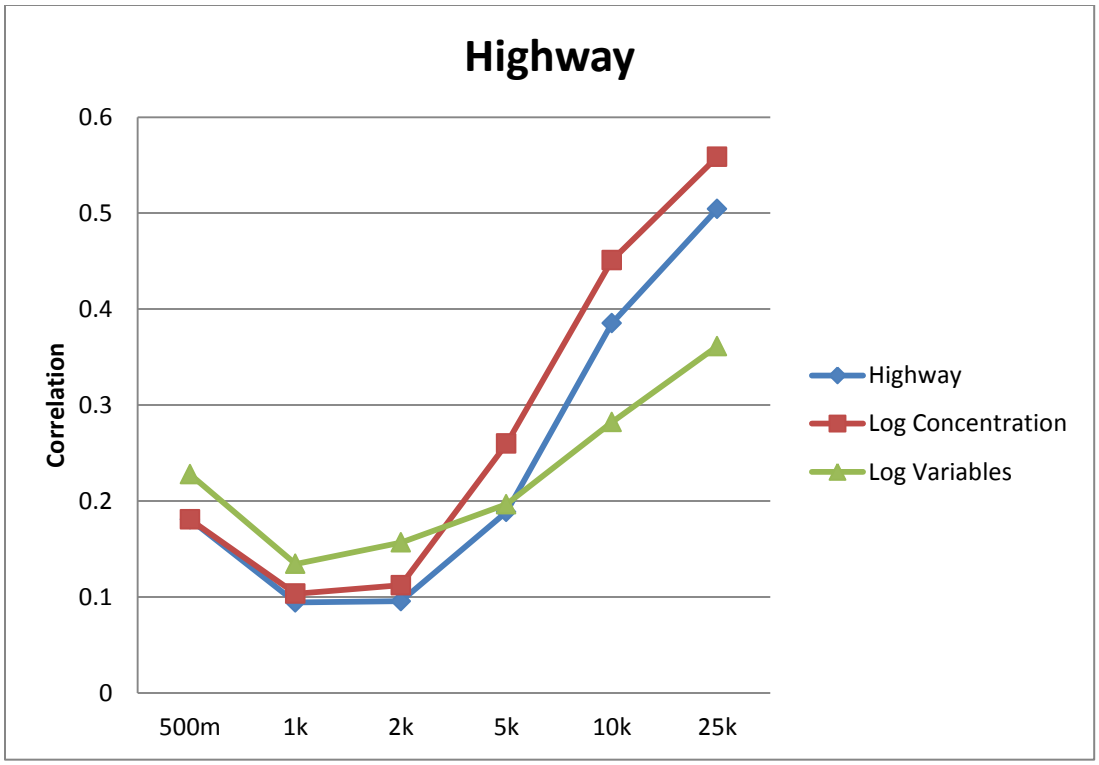


Figure 192 : Correlation between highway and concentration levels

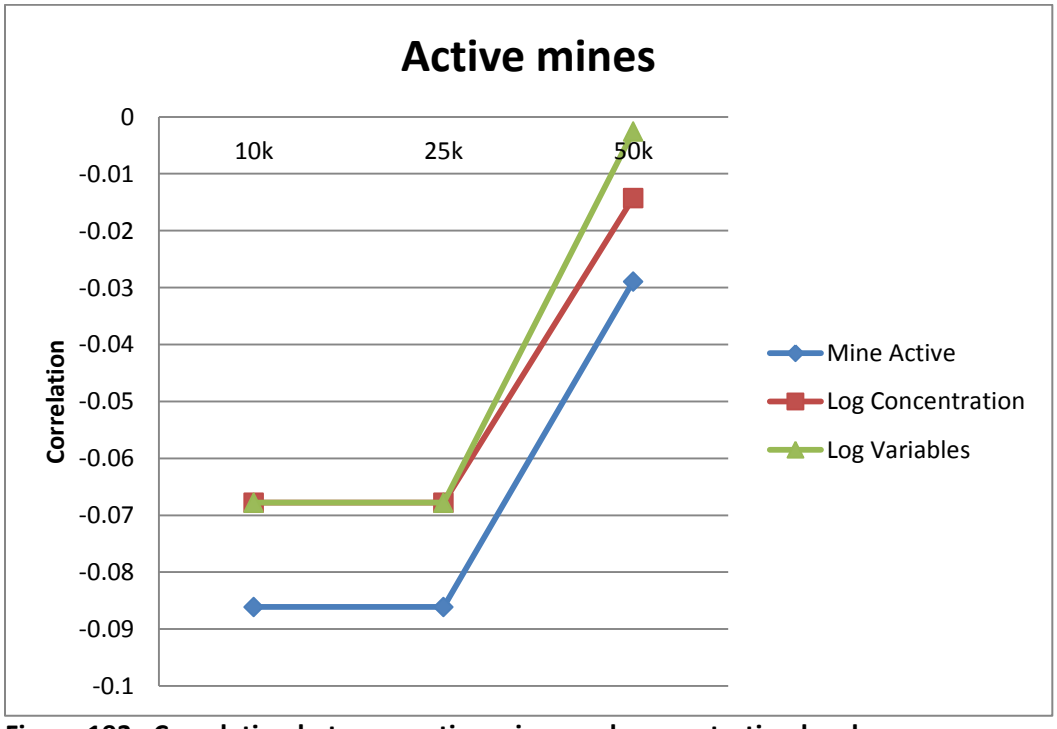


Figure 193 : Correlation between active mines and concentration levels

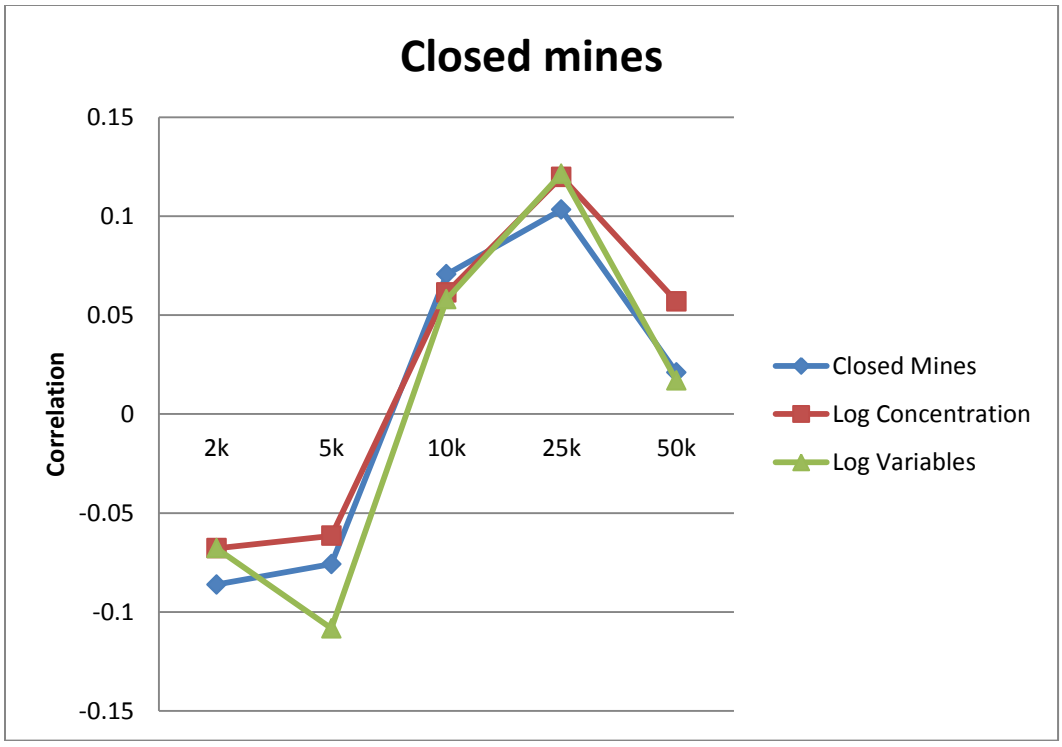


Figure 194 : Correlation between closed mines and concentration levels

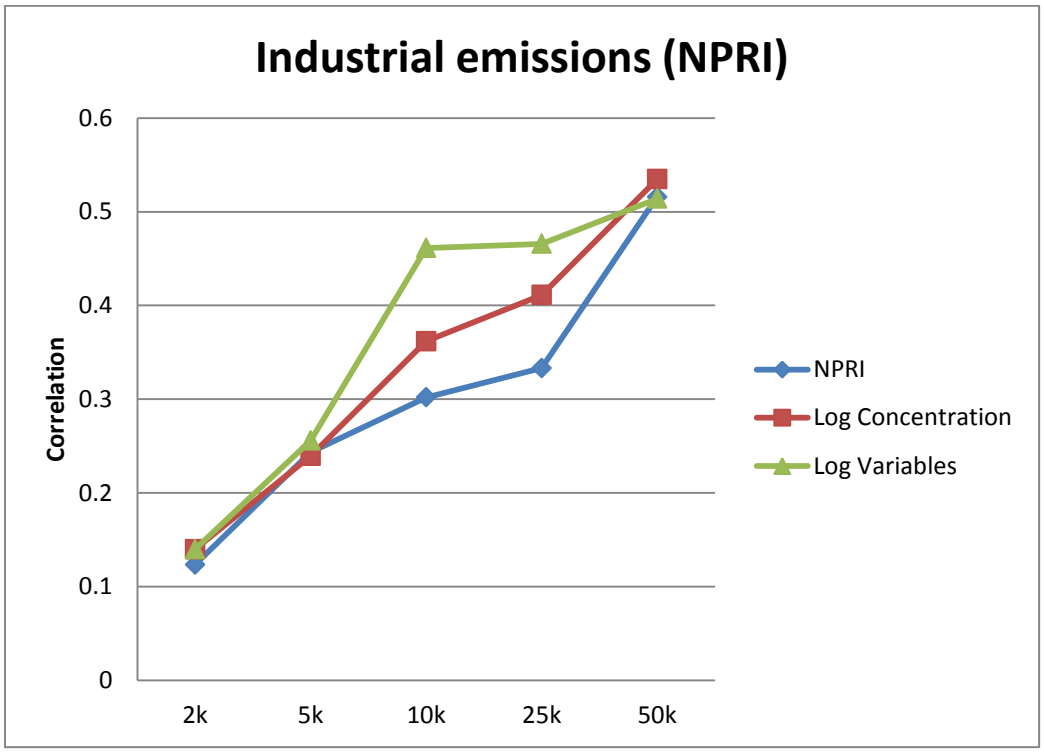


Figure 195 : Correlation between industrial emission (NPRI) and concentration levels

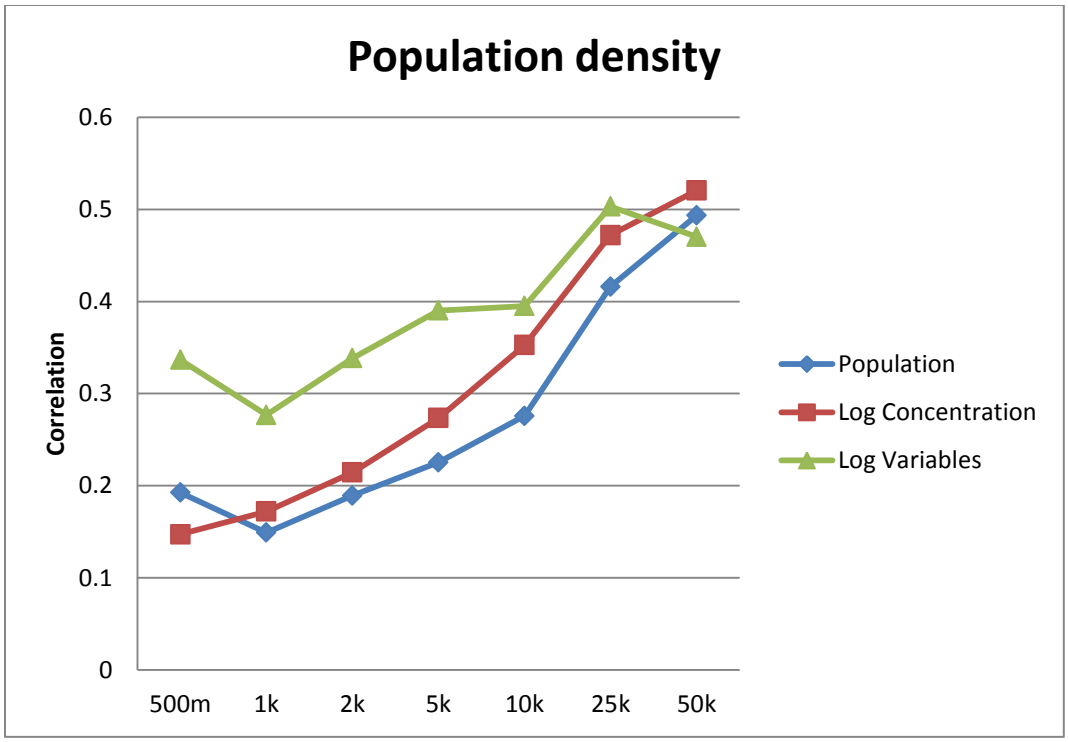


Figure 196: Correlation between population density and concentration levels

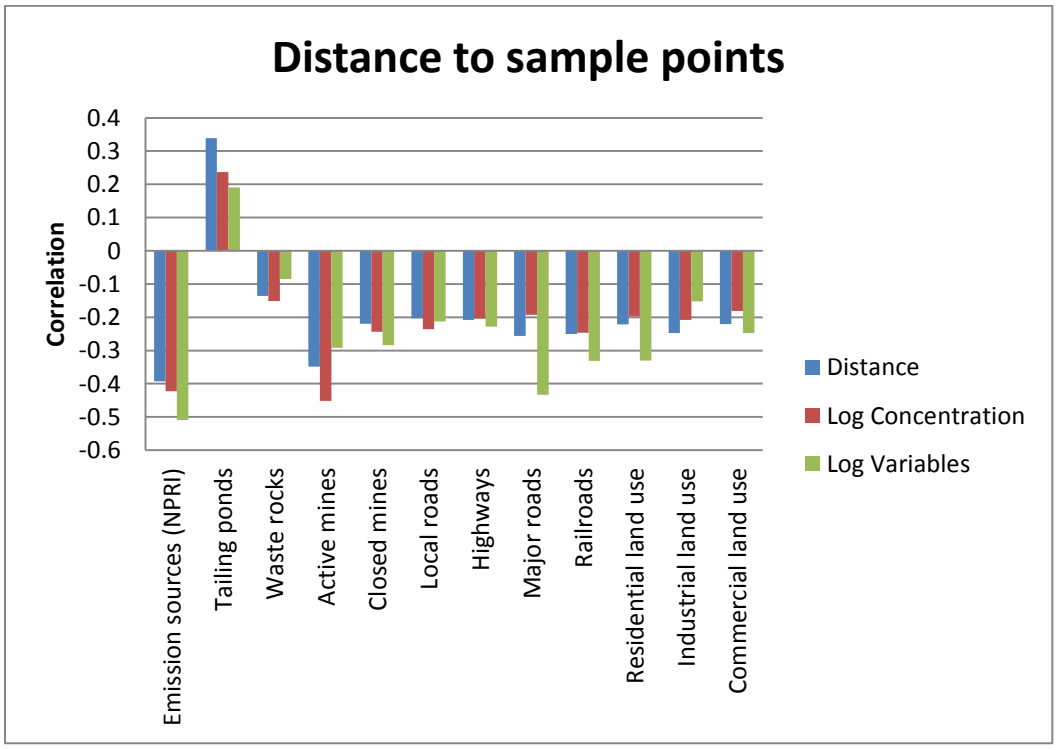


Figure 197: Correlation between distance to sample points and concentration levels

Linear regression

Table 52 : Land Use Regression analysis – Arsenic model

Variables	Coefficient	Standard Errors	t-Value	Probability
Intercept	-5.47857	1.36517	-4.013	0.000235
Length of roads within 25 km (log)	0.49839	0.09166	5.437	2.39e-6
Bimodal volcanic rocks	0.63752	0.53936	1.182	0.243704
Granodiorite, Quartz Diorite	0.19868	0.41314	0.481	0.633023
Intermediate volcanic rocks	0.67861	0.65334	1.039	0.304758
Mafic intrusive rocks - Diorite, Gabbro	-1.36899	0.65123	-2.102	0.041434
Mafic volcanic rocks	-0.28572	0.46129	-0.619	0.538922
Non-marine sedimentary rocks	0.33297	0.53214	0.626	0.534799
Off-shelf mio-geoclinal rocks	0.30759	0.65315	0.471	0.640074
Sedimentary, mafic volcanic rocks	-0.03304	0.65345	-0.051	0.959913
Undivided sedimentary rocks	0.10135	0.38985	0.260	0.796124
Undivided volcanic rocks	0.66303	0.44543	1.489	0.143913

Residual standard error: 0.531 on 43 degrees of freedom
Multiple R-squared: 0.5414,
Adjusted R-squared: 0.424
F-statistic: 4.614 on 11 and 43 degrees of freedom
p-value: 0.0001284

Confidence interval

Table 53 : Confidence interval of Arsenic model coefficients

Variables	2.5%	97.5%
Intercept	-8.231684	-2.72544695
Length of roads within 25 km (log)	0.3135388	0.6832403
Bimodal volcanic rocks	-0.4501992	1.72522946
Granodiorite, Quartz Diorite	-0.6344928	1.03185171
Intermediate volcanic	-0.638972	1.99619537
Mafic intrusive rocks – Diorite, Gabbro	-2.6823305	-0.05565163
Mafic volcanic rocks	-1.2160047	0.64455851
Non-marine sedimentary rocks	-0.7401803	1.4061269
Off-shelf mio-geoclinal rocks	-1.0096185	1.62478984
Sedimentary, mafic volcanic rocks	-1.3508513	1.28477826
Undivided sedimentary rocks	-0.6848496	0.88754962
Undivided volcanic rocks	-0.2352655	1.56132125

Residual distribution

Table 54 : Residual distribution (standardized residuals)

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
-0.8822	-0.3546	0.0000	0.0000	0.3109	1.029
(-1.862)	(-0.7482)	(0.0000)	(0.0000)	(0.656)	(2.171)

Analysis of variance

Table 55 : Analysis of variance for Arsenic

Variables	Degree of Freedom	Sum of Squares	Mean of Squares	F-Value	Probability
Length of roads within 25 km (log)	1	8.9556	8.9556	31.7576	1.237e-6
Bedrock geology	10	5.3575	0.5357	1.8998	0.07171
Residuals	43	12.126	0.282		

Shapiro-Wilk normality test

W: 0.9837

p-value: 0.6572

Spatial Autocorrelation

Neighbourhood: 179206

Moran's Index: 0.04295

Expected Index: -0.018519

Variance: 0.005316

z-score: 0.843081

p-value: 0.399183

Variable Variance Explanation

Length of roads within 25 km (log): 31.53%

Bedrock geology: 20.26%

Graphs

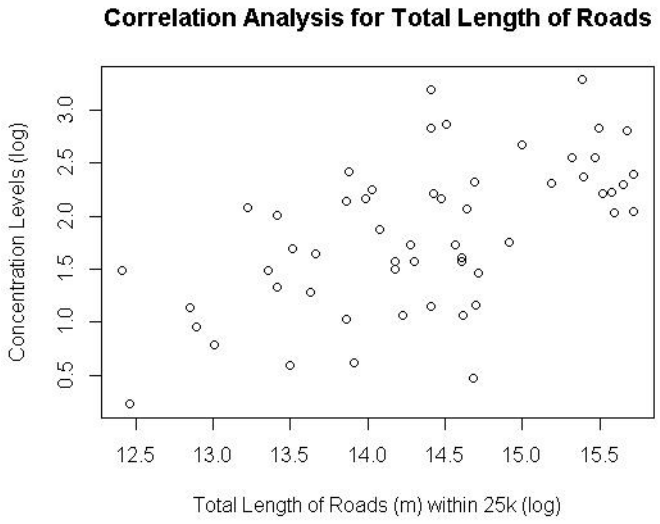


Figure 198 : Correlation between total road length (m) and concentration levels (log)

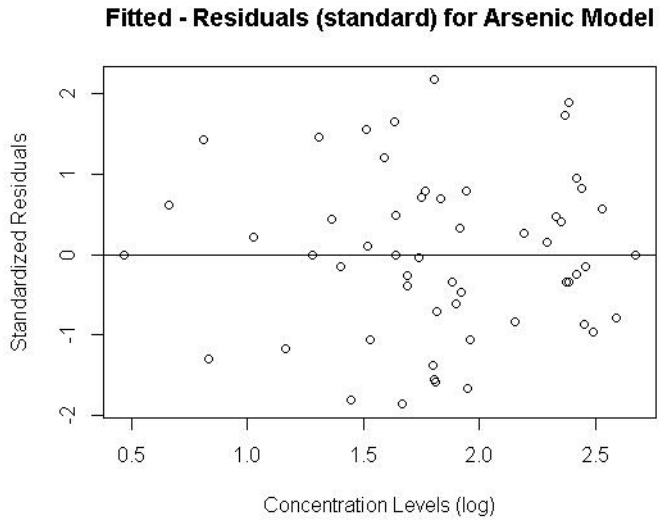


Figure 199 : Fitted standardized residuals for Arsenic model

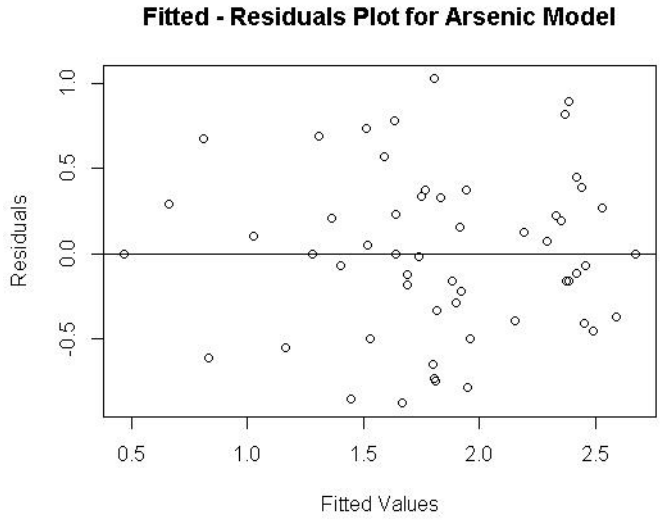


Figure 200 : Fitted residuals for Arsenic model

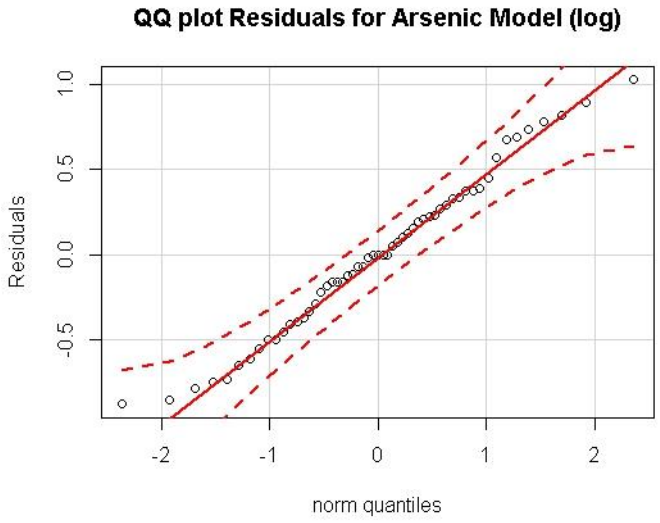


Figure 201 : QQ plot residuals for Arsenic model

Figure 202 : QQ plot standardized residuals for Arsenic model

Bootstrap analysis

Distribution of the R-squared

Table 56 : Distribution of the R-squared after bootstrap analysis

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0.3074	0.5779	0.6327	0.6289	0.6840	0.8939

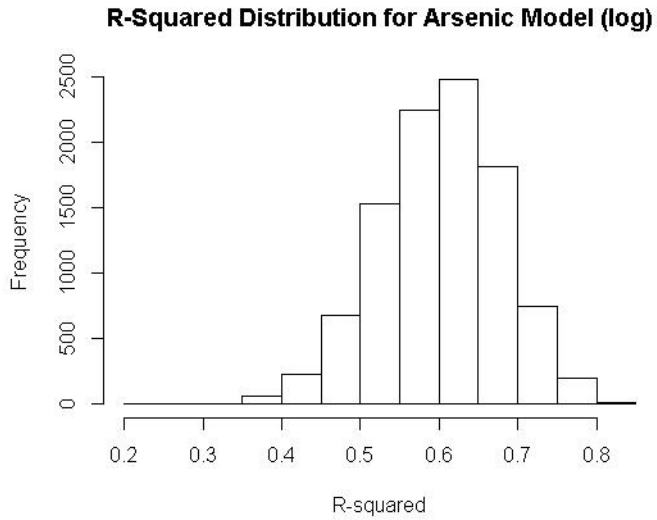


Figure 203 : Bootstrap R-squared distribution for Arsenic model

Land Use Regression analysis – Lead model

Concentration levels

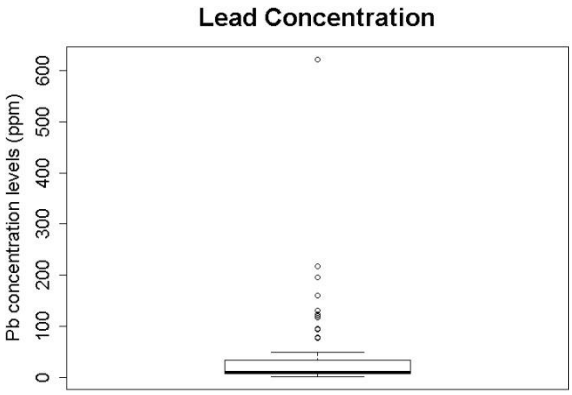


Figure 204 : Concentration levels for Lead

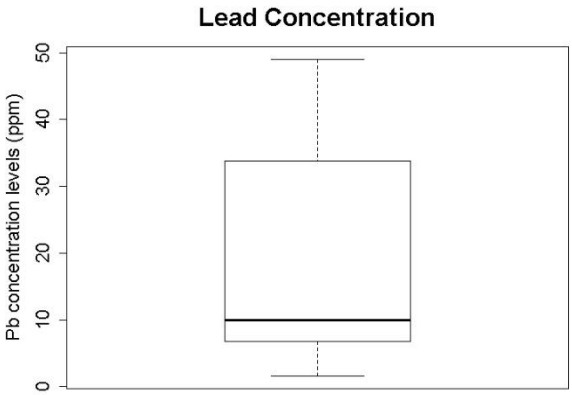


Figure 205 : Concentration levels for Lead (no outlier)

Correlation analysis

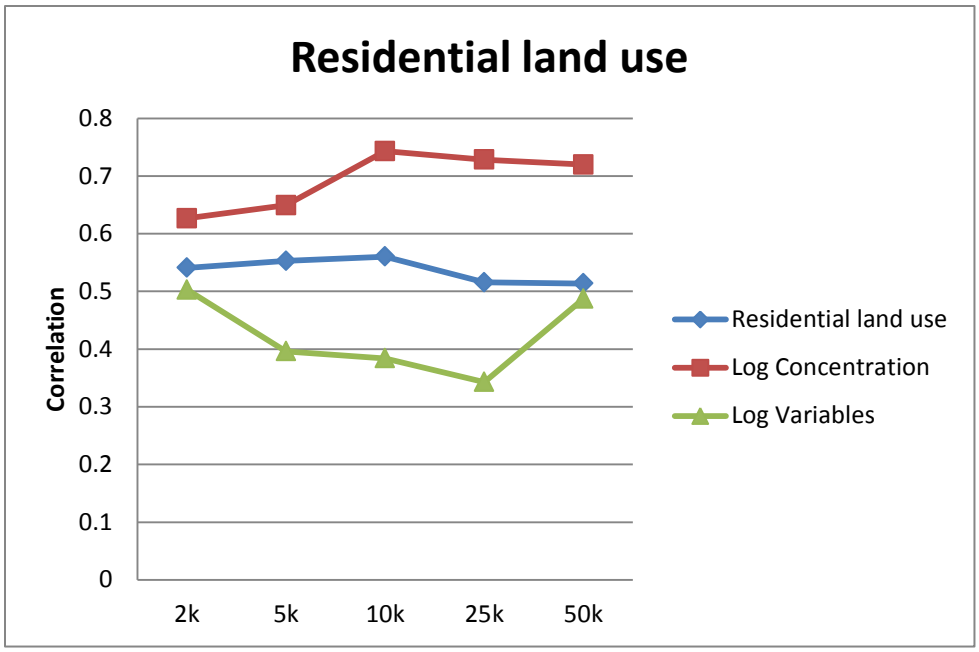


Figure 206 : Correlation between residential land use and concentration levels

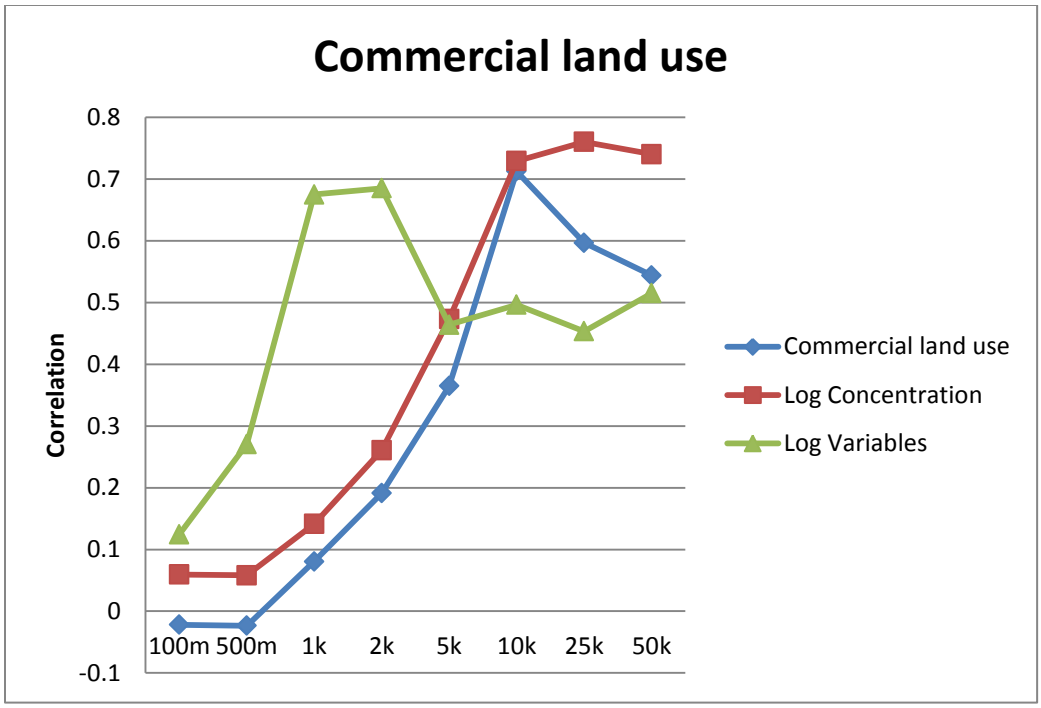


Figure 207 : Correlation between commercial land use and concentration levels

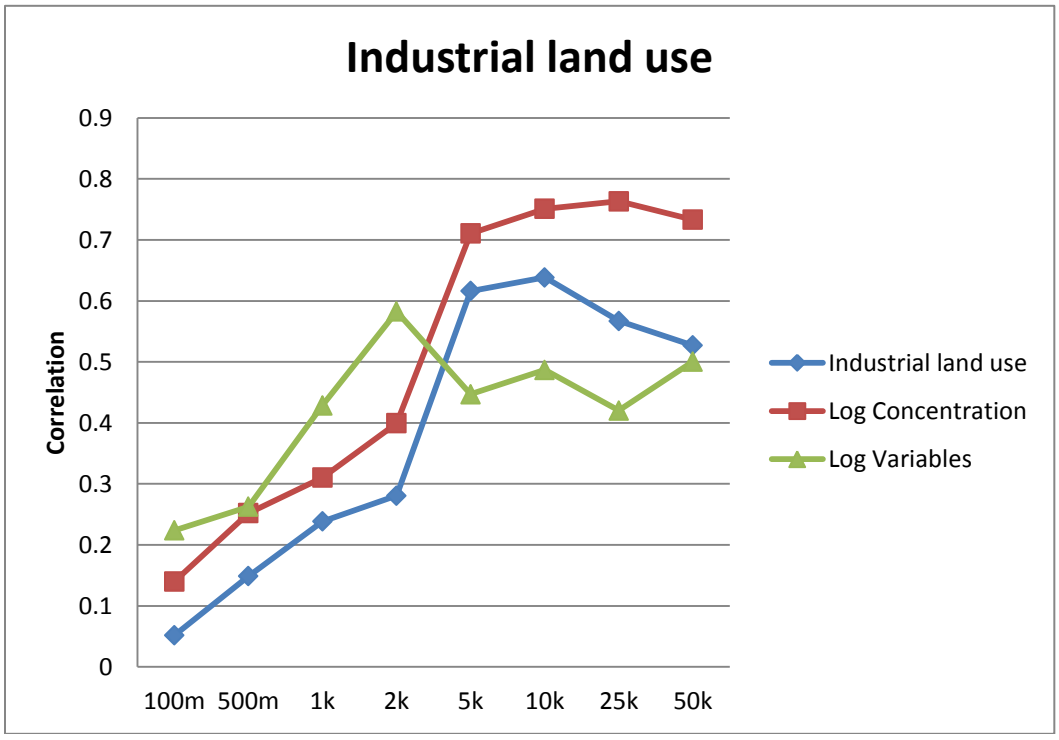


Figure 208 : Correlation between industrial land use and concentration levels

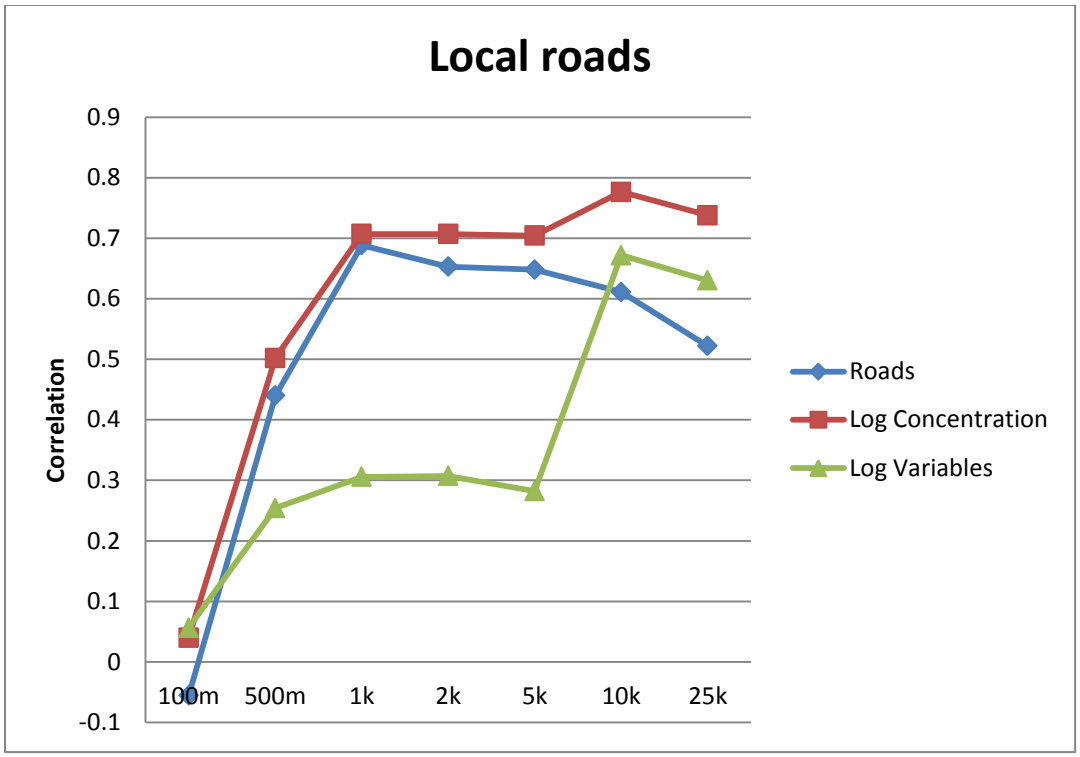


Figure 209 : Correlation between local roads and concentration levels

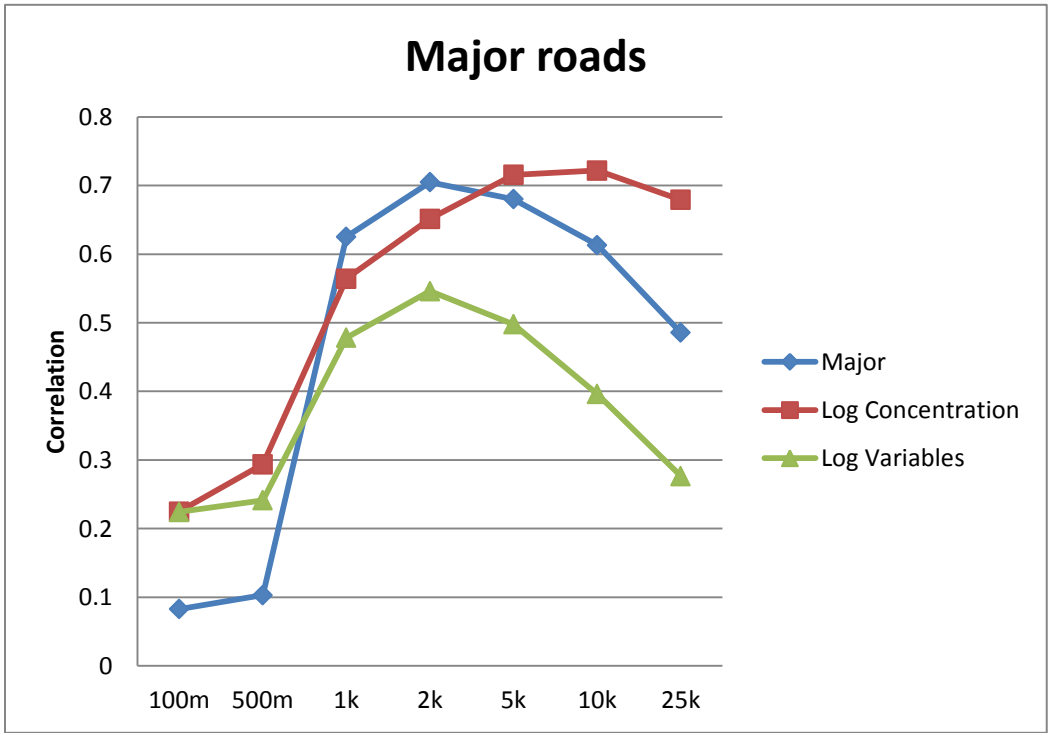


Figure 210 : Correlation between major roads and concentration levels

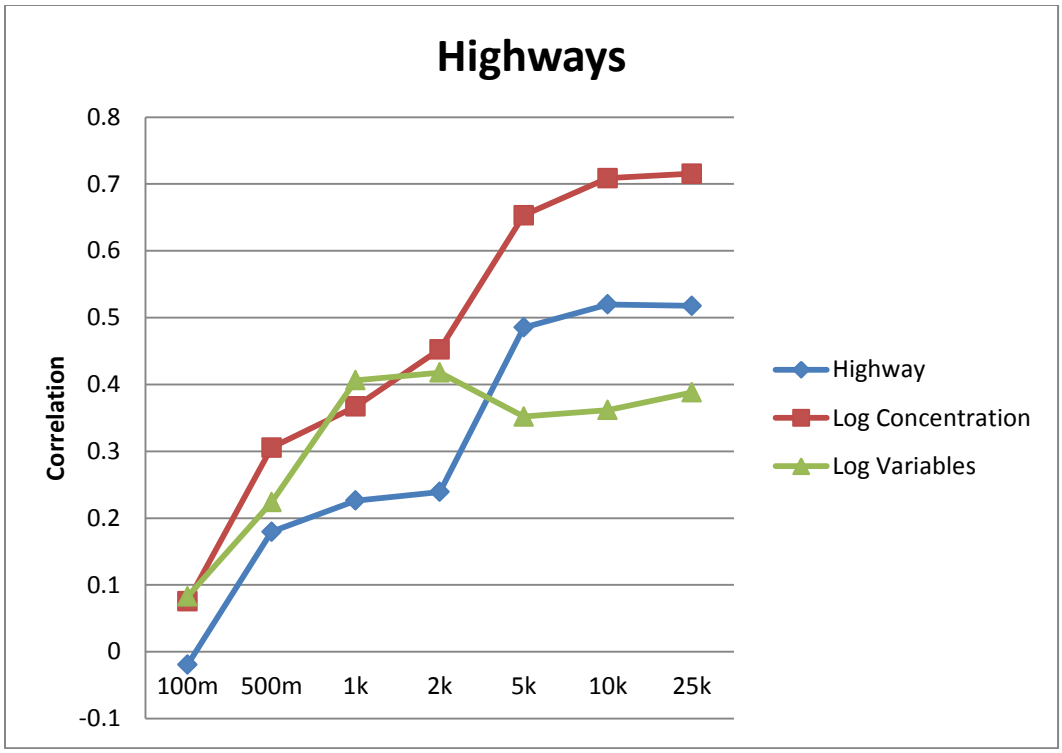


Figure 211 : Correlation between highways and concentration levels

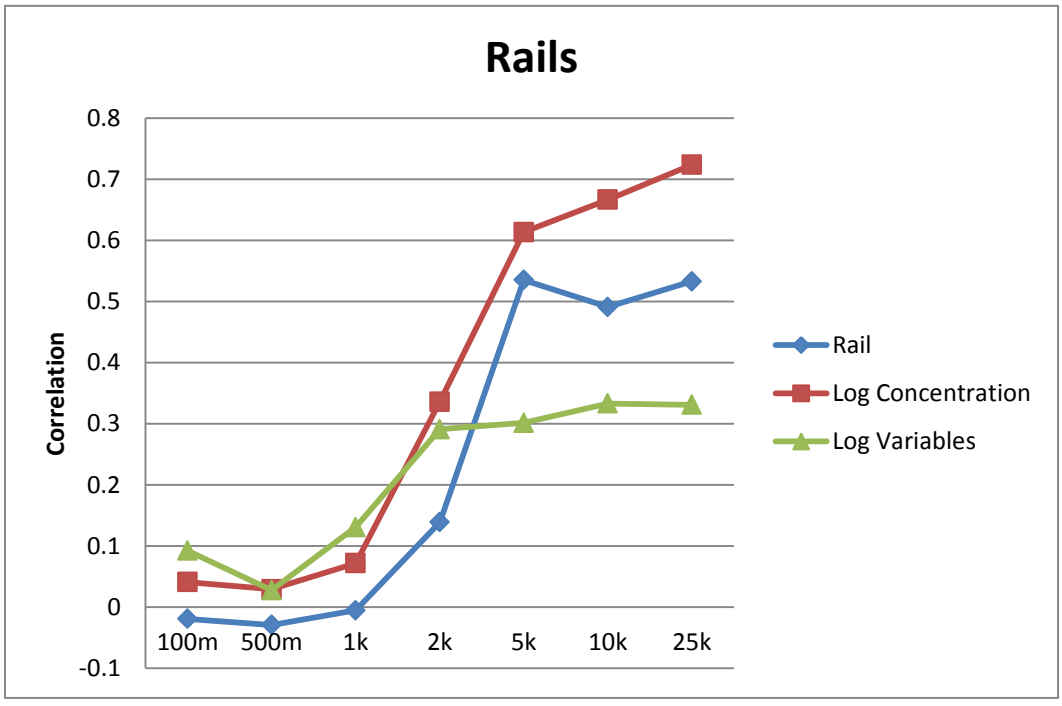


Figure 212 : Correlation between rails and concentration levels

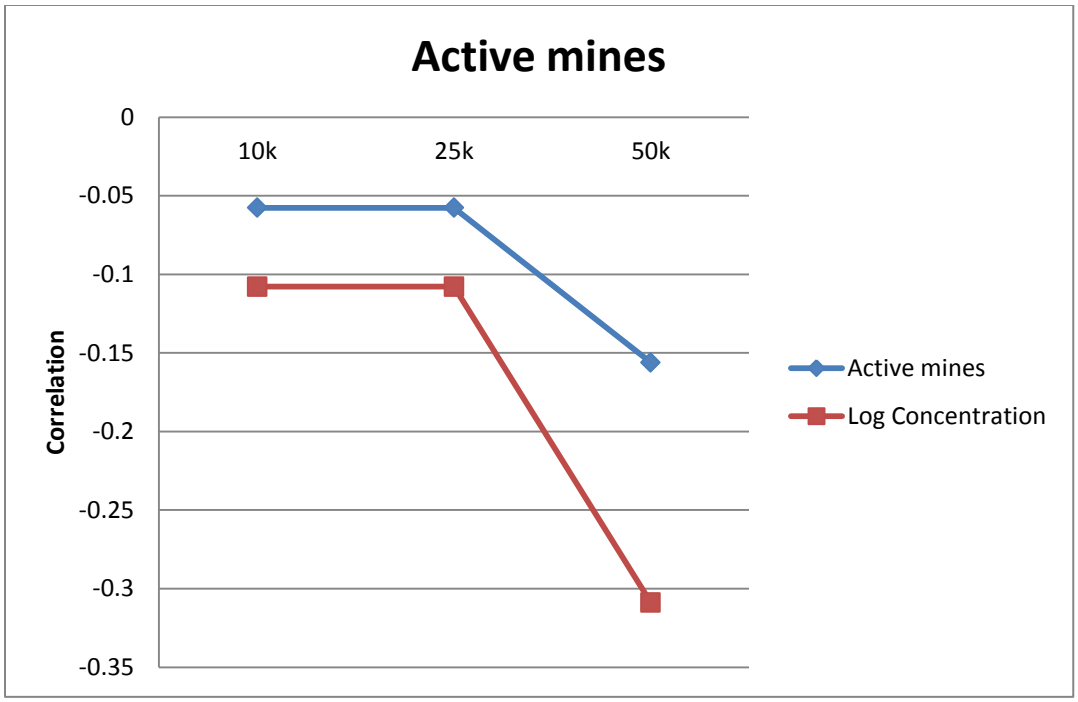


Figure 213 : Correlation between active mines and concentration levels

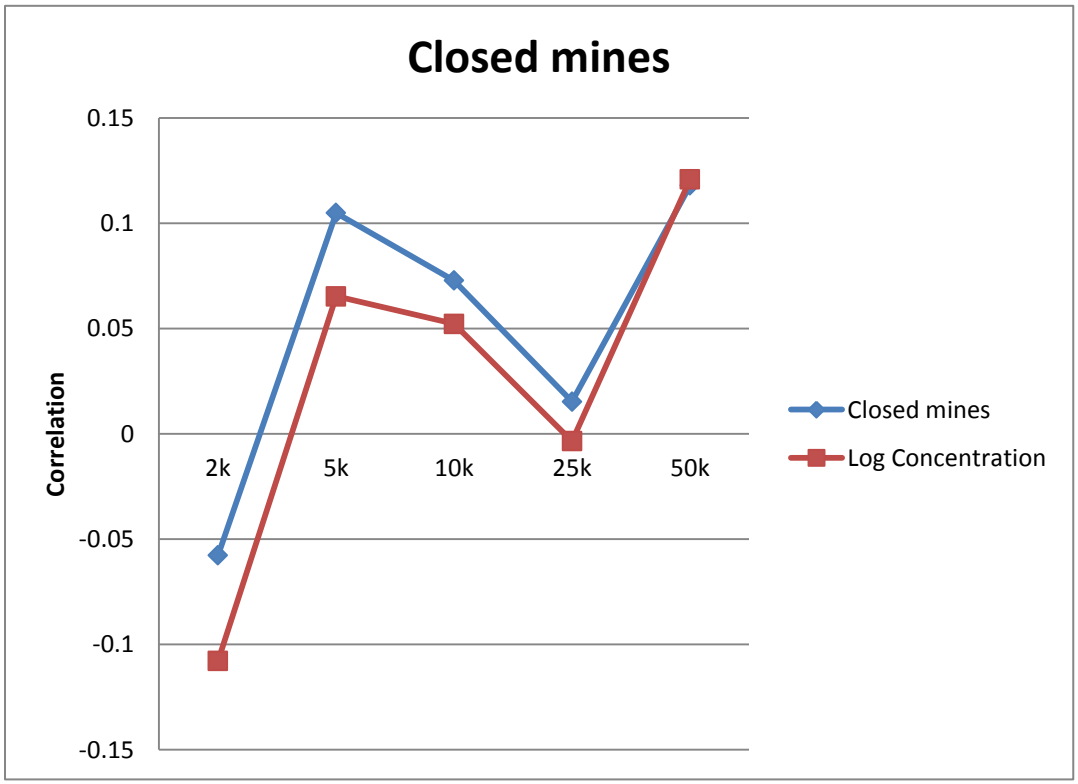


Figure 214 : Correlation between closed mines and concentration levels

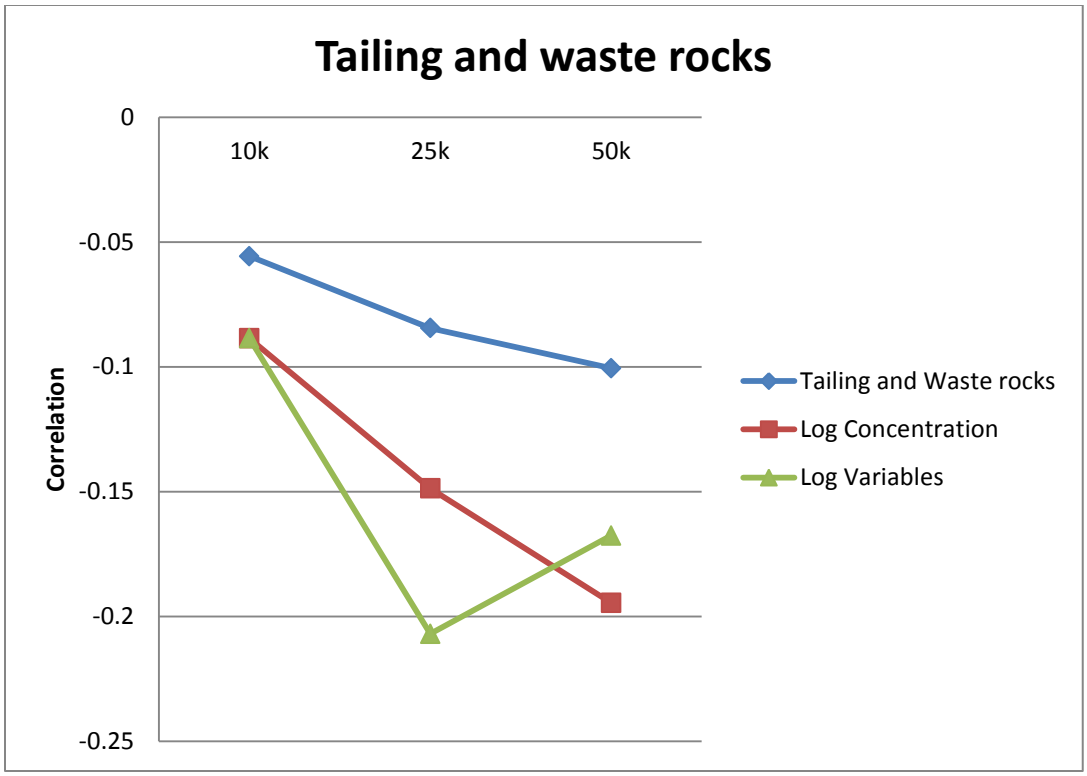


Figure 215 : Correlation between tailing and waste rocks and concentration levels

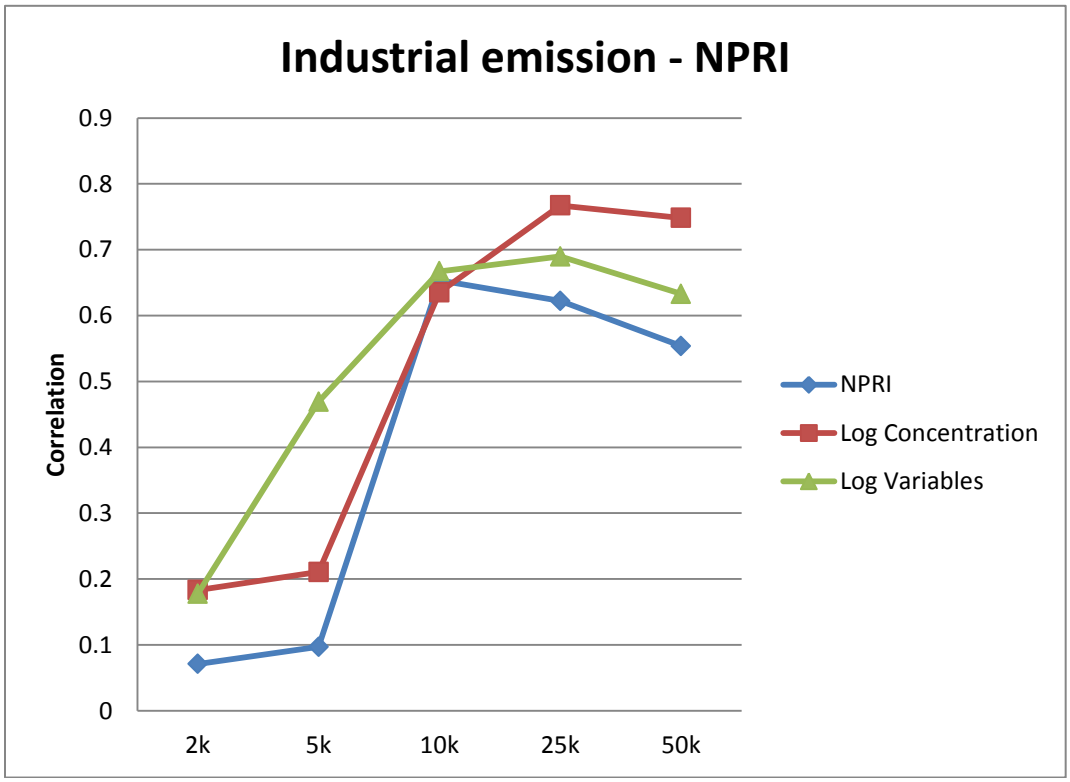


Figure 216 : Correlation between industrial emission (NPRI) and concentration levels

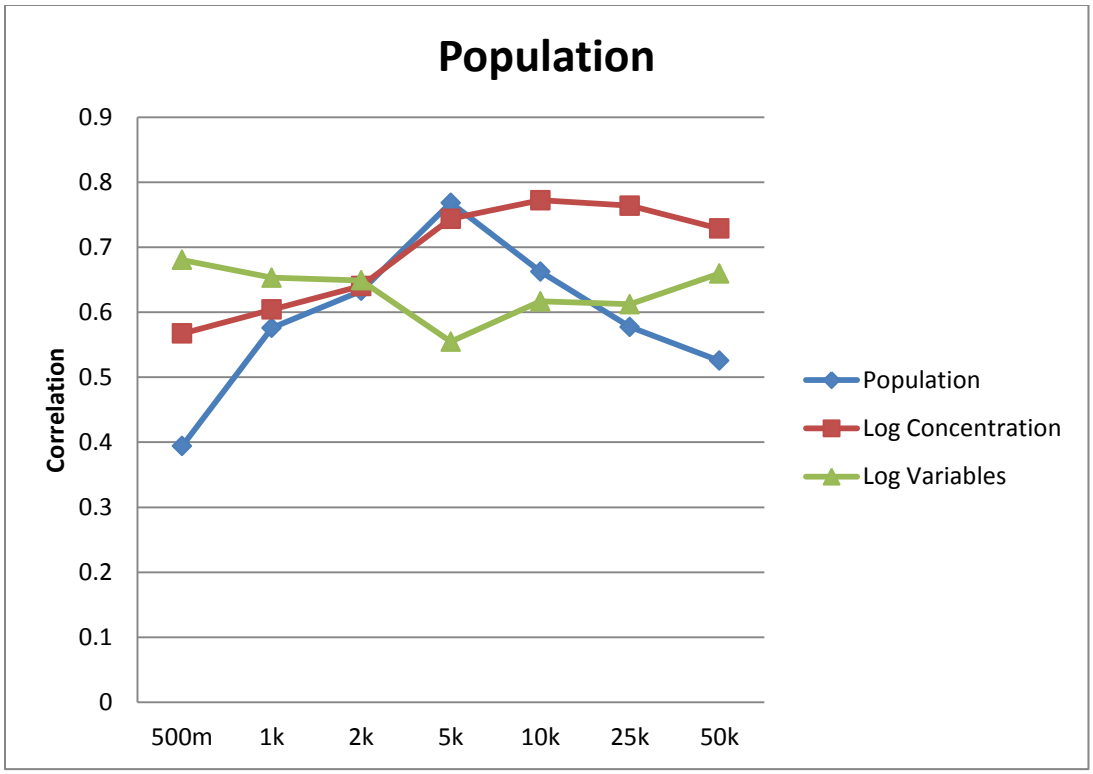


Figure 217 : Correlation between population density and concentration levels

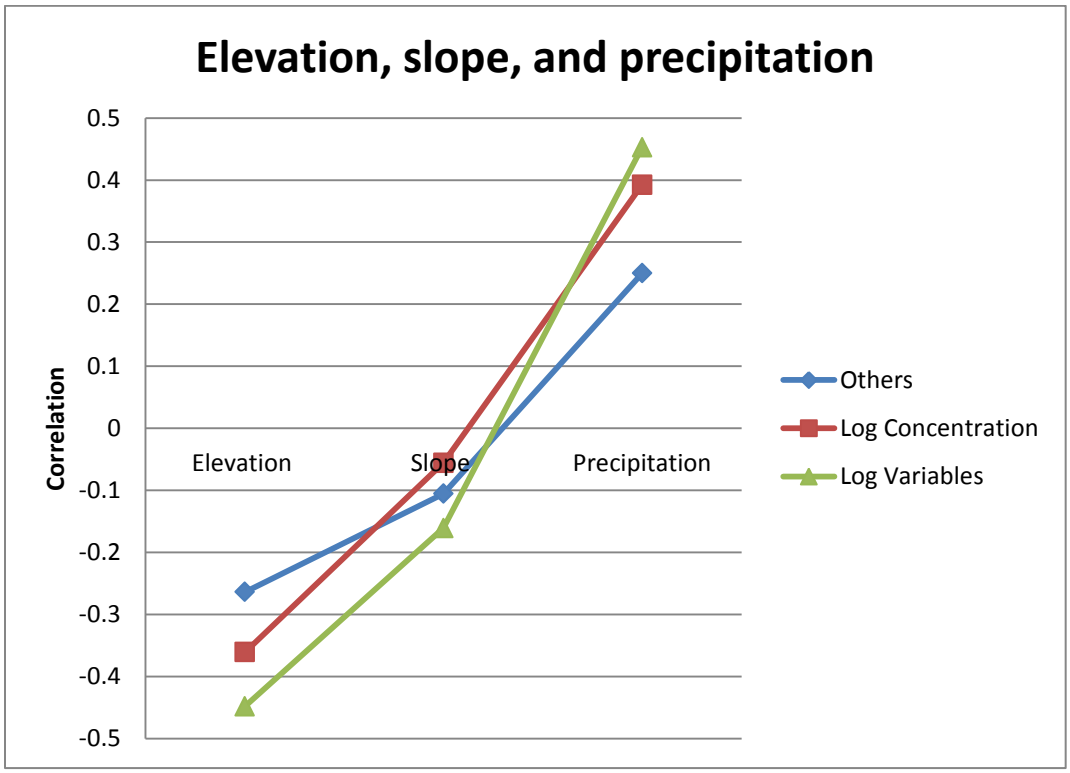


Figure 218 : Correlation between elevation, slope, and precipitation and concentration levels

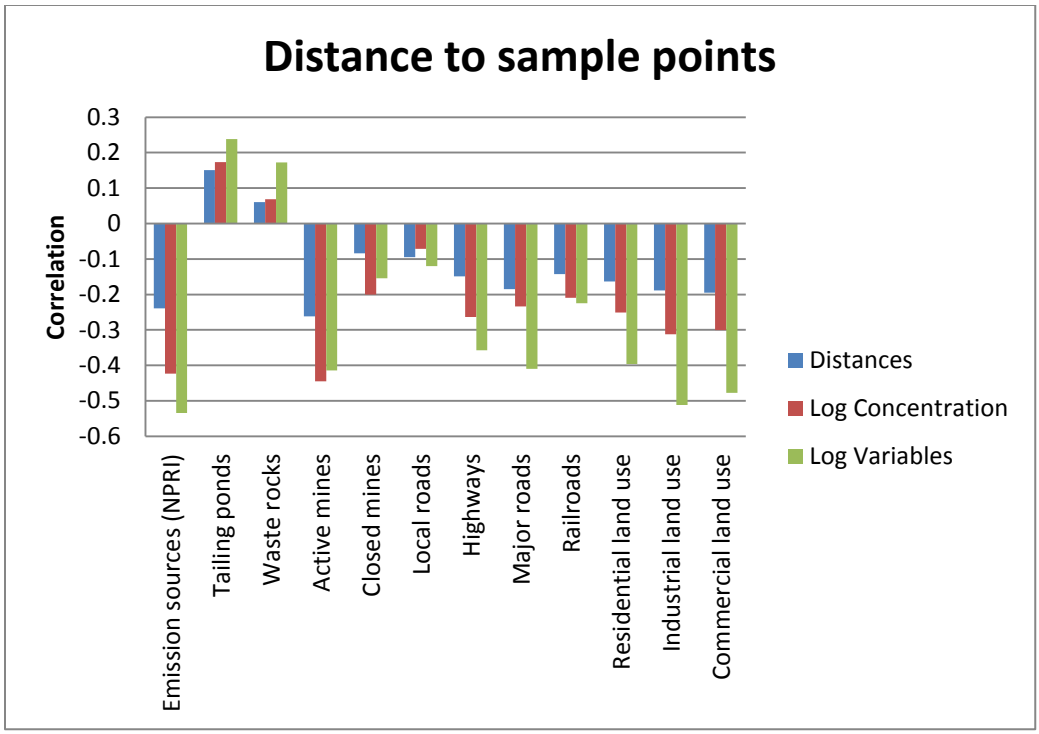


Figure 219 : Correlation between distance to sample points and concentration levels

Linear regression

Table 57 : Land Use Regression analysis – Lead model

Variables	Coefficient	Standard Errors	t-value	Probability
Intercept	1.820e+00	1.352e-01	13.464	< 2e-16
Industrial land use within 5 km	1.489e-07	2.937e-08	5.069	5.62e-06
Industrial emission within 25 km (log)	9.761e-02	2.955e-02	3.303	0.00175
Industrial emission within 10 km	5.257e-05	1.210e-05	4.344	6.67e-05
Presence of past mines within 50 km	1.199e-02	2.180e-03	5.501	1.23e-06
Residual standard error: 0.5762 on 51 degrees of freedom				
Multiple R-squared: 0.7919				
Adjusted R-squared: 0.7756				
F-statistic: 48.52 on 4 and 51 degrees of freedom				
p-value: < 2.2e-16				

Confidence interval

Table 58 : Confidence interval

Variables	2.5%	97.5%
Intercept	1.588547e+00	2.136248e+00
Industrial land use within 5 km	9.037017e-08	2.118802e-07
Industrial emission within 25 km (log)	3.497405e-02	1.510159e-01
Industrial emission within 10 km	2.553661e-05	7.814455e-05
Presence of past mines within 50 km	4.687555e-03	1.578716e-02

Residual distribution

Table 59 : Residual distribution (standardized residuals)

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
-0.9361 (-1.68700)	-0.3507 (-0.63210)	-0.0529 (-0.09535)	0.00000 (0.00000)	0.3838 (0.69170)	1.4910 (2.68800)

Analysis of variance

Table 60 : Analysis of variance for Lead

Variables	Degree of Freedom	Sum of Squares	Mean of Squares	F-Value	Probability	VIF
Industrial land use within 5 km	1	41.065	41.065	123.705	3.033e-15	1.832618
Industrial emission within 25 km (log)	1	6.655	6.655	20.046	4.272e-05	2.342749
Industrial emission within 10 km	1	6.665	6.665	20.077	4.223e-05	1.459620
Presence of past mines within 50 km	1	10.046	10.046	30.262	1.226e-06	1.109414
Residuals	51	16.930	0.332			

Shapiro-Wilk normality test

W: 0.9658

p-value: 0.1122

Spatial Autocorrelation

Neighbourhood: 179206

Moran's Index: -0.099722

Expected Index: -0.018182

Variance: 0.005253

z-score: -1.125045

p-value: 0.260570

Variable Variance Explanation

Industrial land use within 5 km: 10.48%

Industrial emission within 25 km (log):4.45%

Industrial emission within 10 km: 7.70%

Presence of past mines within 50 km: 12.35%

Graphs

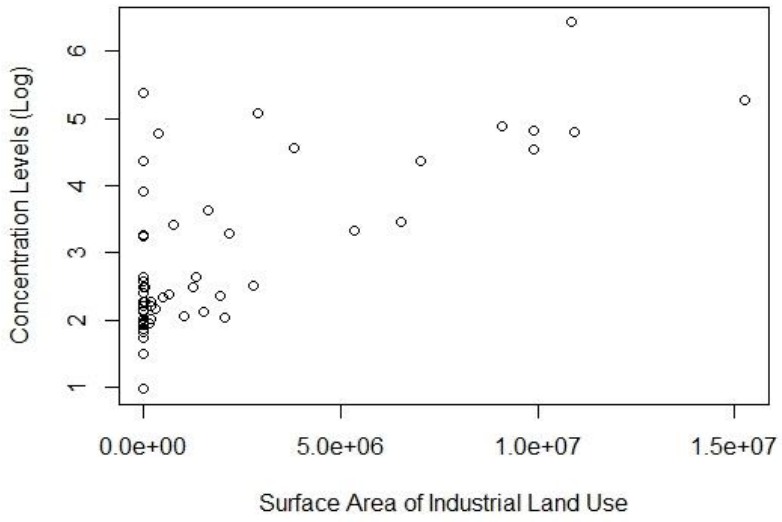


Figure 220 : Correlation between Lead concentration levels (log) and Industrial land use within 5 km

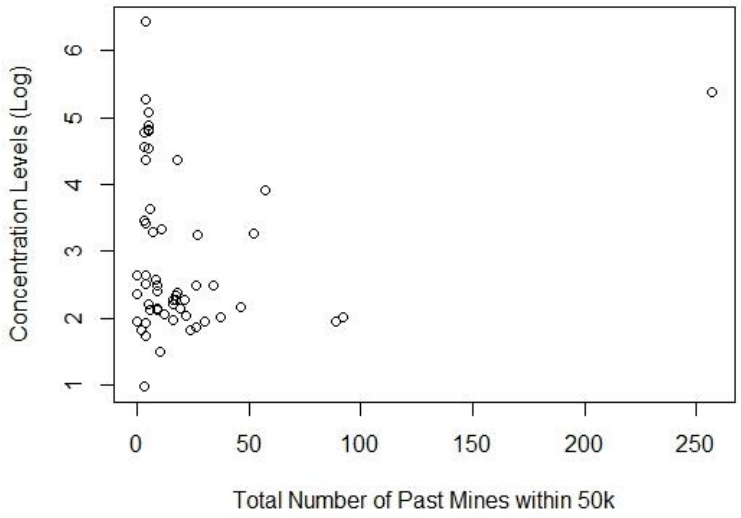


Figure 221 : Correlation between Lead concentration levels (log) and Past mines within 50 km

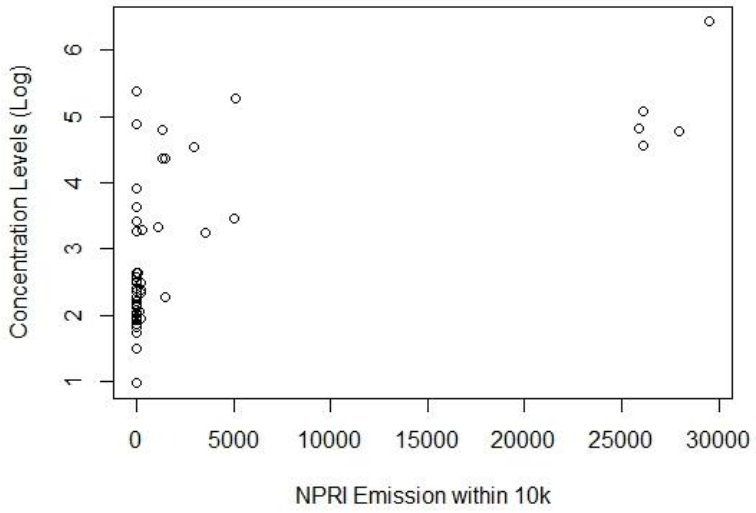


Figure 222 : Correlation between Lead concentration levels (log) and Industrial emission within 10 km

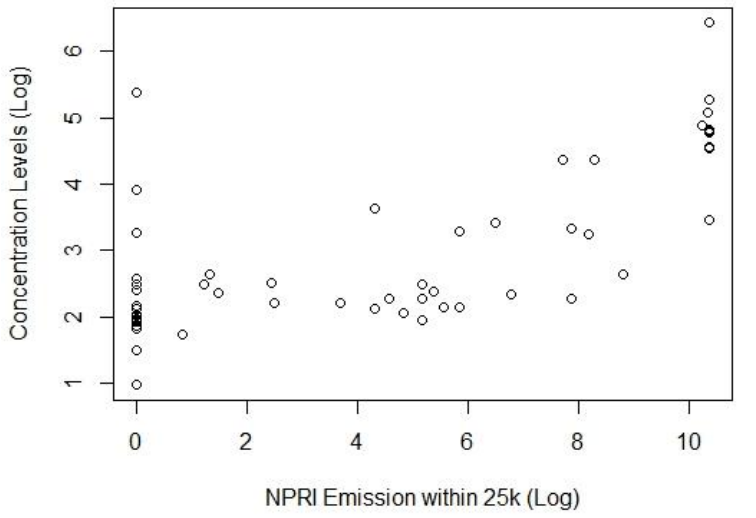


Figure 223 : Correlation between Lead concentration levels (log) and Industrial emission within 25 km (log)

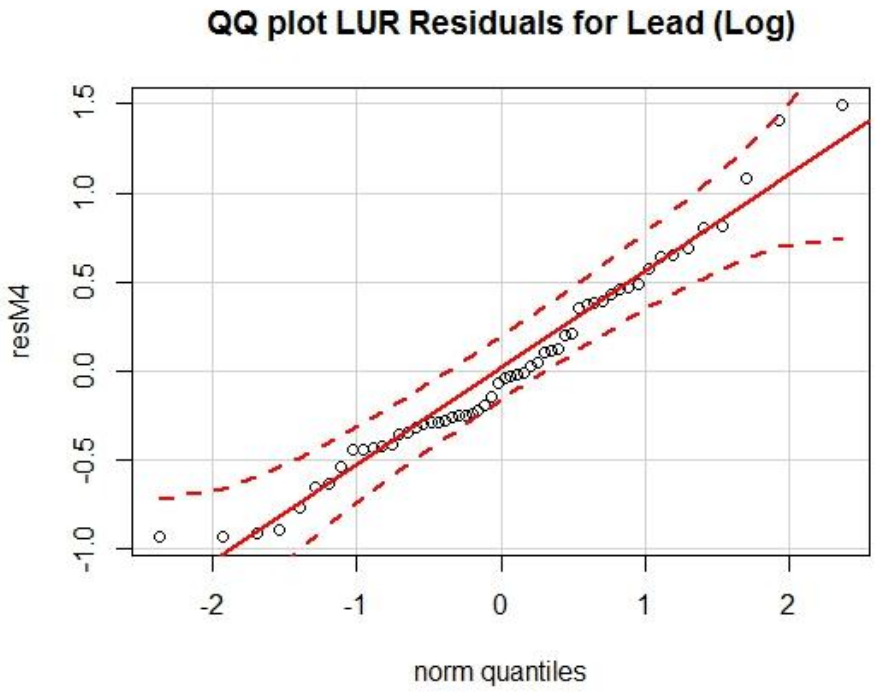


Figure 224 : QQ plot residuals for Lead model

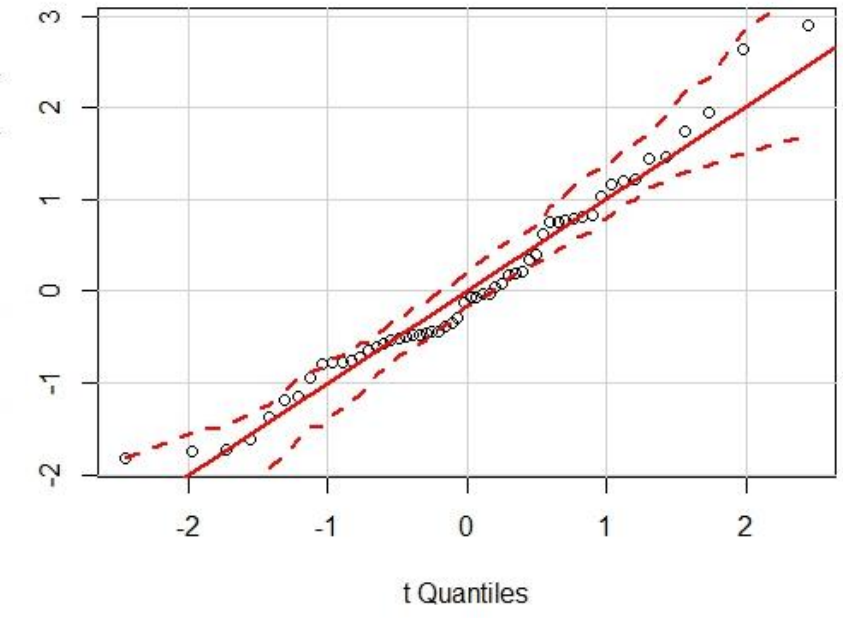


Figure 225 : QQ plot residuals (standard) for Lead model

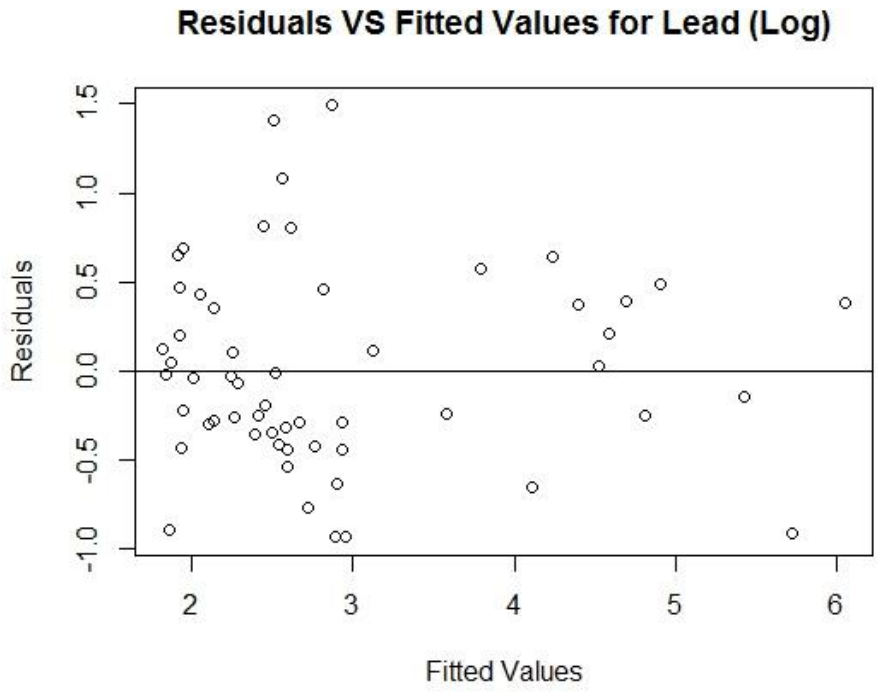


Figure 226 : Fitted residuals for Lead model

Bootstrap analysis

Distribution of the R-squared and the coefficients

Table 61 : Distribution of the R-squared of the Lead model after the bootstrap analysis

	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
R-squared	0.4900	0.7684	0.8087	0.8021	0.8424	0.9518

Graphs

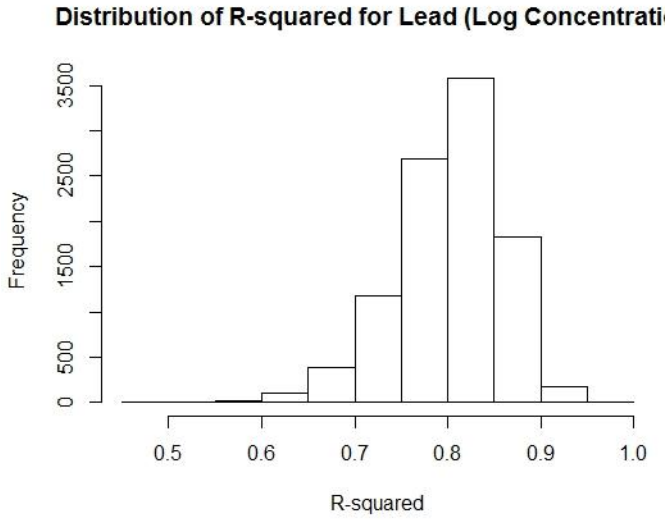


Figure 227 : R-squared distribution from the bootstrap analysis

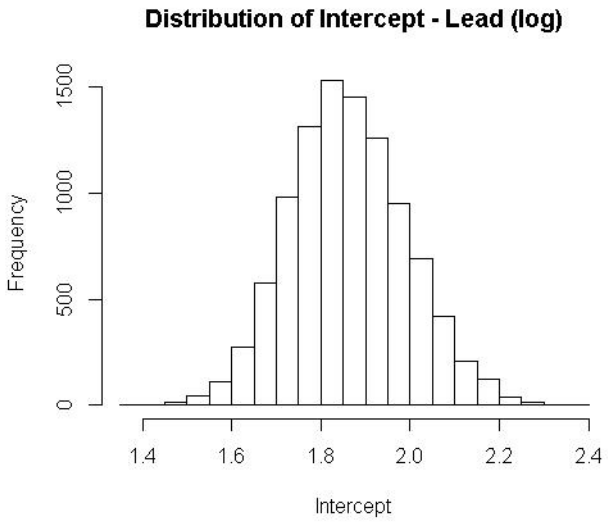


Figure 228 : Intercept distribution for Lead model

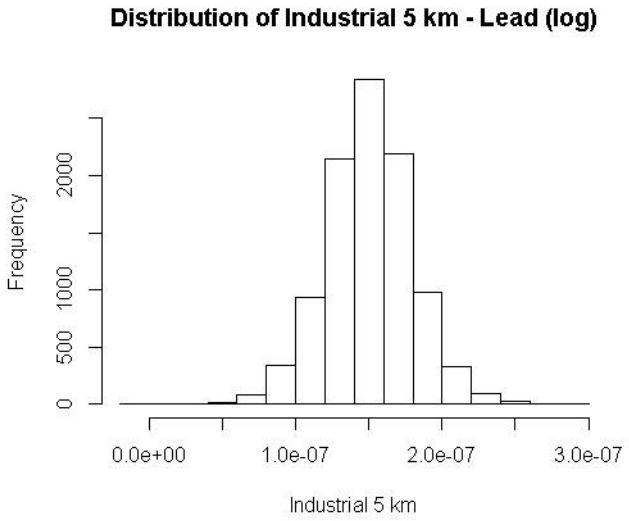


Figure 229 : Coefficient distribution for industrial land use within 5 km from the bootstrap analysis

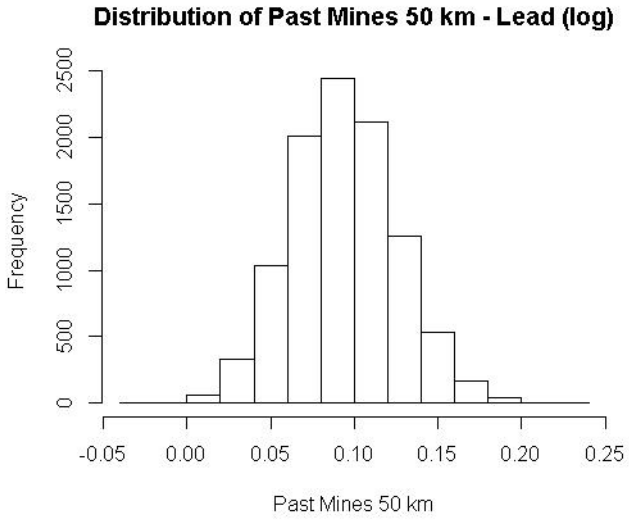


Figure 230 : Coefficient distribution from bootstrap analysis for industrial past mines within 50 km

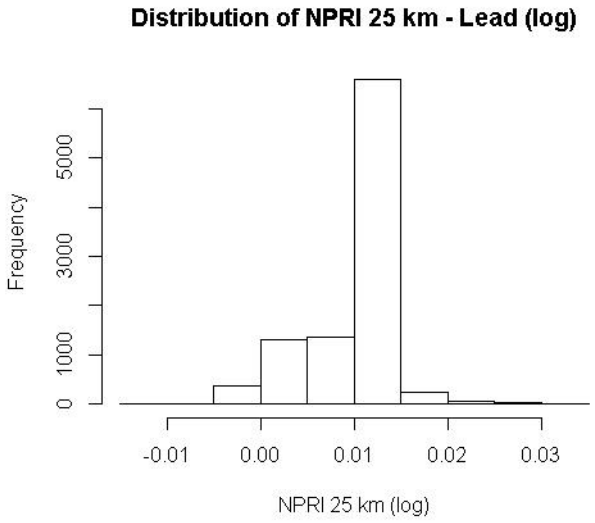


Figure 231 : Coefficient distribution from bootstrap analysis for industrial emission within 25 km (log)

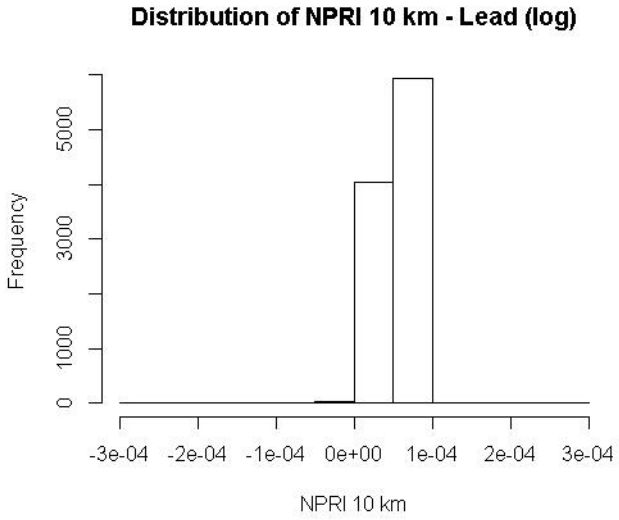


Figure 232 : Coefficient distribution from bootstrap analysis for industrial emission within 10 km