Geostatistical analysis and integration of soil chemistry data with remote sensing information in the Sudbury area, Ontario.

by

Phathutshedzo Molly Nethavhani

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science (MSc) in Geology

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Office of Graduate Studies/Bureau des études supérieures

Title of Thesis
Titre de la thèse: Geostatistical analysis and integration of soil chemistry data with remote sensing information in the Sudbury area, Ontario

Name of Candidate
Nom du candidat: Nethavhani, Phathutshedzo Molly

Degree
Diplôme: Master of Science

Department/Program
Département/Programme: Geology

Date of Defence
Date de la soutenance: March 28, 2022

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Graeme Spiers
(Supervisor/Directeur(trice) de thèse)

Dr. Nadia Mykytczuk
(Committee member/Membre du comité)

Dr. Alessandro Ielpi
(Committee member/Membre du comité)

Dr. Bernhardt Saini-Eidukat
(External Examiner/Examinateur externe)

Approved for the Office of Graduate Studies
Approuvé pour le Bureau des études supérieures
Tammy Eger, PhD
Vice-President Research (Office of Graduate Studies)
Vice-rectrice à la recherche (Bureau des études supérieures)
Laurentian University / Université Laurentienne

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

Abstract

The presence of anomalous concentrations of metals within the soil profile can strongly affect its biological availability to plants, causing potential toxicity when exceeding threshold concentrations, and favoring numerous chemical exchanges. These interactions further facilitate metal dispersion in the hydrogeological and ecological systems, in response to weathering and erosion. Studies of the geospatial distribution of trace metal contaminants in Sudbury soils is thus important to unravel the dominant processes controlling dispersion patterns, contributing to sustainability of mining practice. A kriging geostatistical approach was applied to geochemical data obtained from the Sudbury Soil Survey to map multiscale geographic, enrichment trends in metal concentrations. Ordinary kriging prediction maps were developed to re-evaluate the multiscale spatial distribution of the chemicals of concern. Results show an anomalous distribution of metals centered on historical smelters, forming dominant northeast and southwest enrichment trends. The existence of these trends was validated by implementing a geostatistical Gaussian conditional simulation method, which reproduced the same spatial variability observed in the ordinary kriging maps and efficiently replicated the observed trends. The correlation analysis of the trends with remote sensing data, suggests that prevailing wind directions are likely one of the dominant driving forces controlling the trends. Integrating these results with satellite data showed improved vegetation regrowth patterns consistent with the geochemical northeast-southwest trend providing further, independent validation of the kriging results. Re-evaluation of the regional, geospatial distribution of the measured trace element concentrations will assist the monitoring and improved understanding of soil contamination trends and their impact on vegetation and other aspects of the biosphere in the Greater Sudbury area.

Keywords

Sudbury Soil Study, Kriging Interpolation, Gaussian Conditional Simulation, Remote Sensing, Normalized Difference Vegetation Index

Acknowledgments

The completion of this thesis would not have been possible without the guidance and great ideas of my supervisor, Leonardo Feltrin, and Co-supervisor, Graeme Spiers.

I would also like to thank the rest of the supervising committee for their input and corrections.

Special thanks go to IDRC, Ivan Plats Mines and Goodman School of Mines for funding my studies.

Last but not least, I would like to thank my family, in South Africa, for their undying love and support all throughout, my fellow colleagues, and friends that I met along the way.

Table of contents

List of figures

List of tables

List of equations

Chapter 1: Introduction

Introduction

The soil represents the uppermost layer of the Earth's crust, with its main components being mineral particles, organic matter, air, and water. Metals are a natural component of the soil, located within mineral particles, comprised of mineral concentrations that are dependent on soil mineral concentrations and geochemical history (CEM, 2004). However, anthropogenic activities (smelting, refining, loading, and mining) have increased the number of metals in the soil, with anomalous concentrations occurring in elevated levels at varying soil profile depths.

Human activity can lead to a variable degree of soil pollution where contamination by metals is one of the prominent contributors, especially in the Greater Sudbury area which is known to be an historically important mining district of northern Canada. Metals have been sourced from the exceptionally mineralised Sudbury Igneous Complex (SIC). Most of the soil contamination is related to early practices that involved roasting of sulphides, where combustion led to atmospheric suspension and deposition of metal-rich fumes, during smelting activity.

Soil sampling more than any other geochemical investigation, shows accurately the total historical metal pollution from the point source (Saavedra, Spiers and Dunn, 2007). Sampling of the Sudbury soils has been extensively conducted over the past years, beginning in the 1960s, to evaluate and answer questions regarding the chemicals of concern (COC), the regreening processes, and the environmental and health concerns through studying the concentration of metals in the soil and vegetation across the area (SARA, 2008). Smelting and roast yards are the major contributors of atmospheric emissions and soil contamination in the Greater Sudbury

area, with contributing factors to the emissions including changes to facilities and emission sources, changes in production levels and, addition and closure of facilities. Soil sampling, analysis and interpretation of data describing metal levels within the footprint of Sudbury smelter regions is incredibly important for future developments in ecological and human risk assessment projects (Saavedra, 2008).

This project proposes a re-evaluation of the geospatial distribution of measured trace metal concentrations to provide an up-to-date geostatistical modelling and representation to further assist the constant need of environmental monitoring and to improve understanding of soil contamination trends and their impact on vegetation regrowth and other aspects of the biosphere in the Greater Sudbury area.

The thesis will provide an introduction of the study area with the study's objectives explained. Metals and their impact on the environment will be evaluated and, a review of mining and smelting in Sudbury area will be outlined. The methods used to address the study's objectives will be documented and explained in detail. Finally, results will be analysed and discussed with the main conclusions drawn and further suggestions laid out.

1.1. Study area

1.1.1. Location

The City of Greater Sudbury is in Northern Ontario, Canada, and became a city in 2001 by the merging of multiple smaller towns and cities that made up the former municipality of Sudbury. According to the census of 2016, Greater Sudbury has a population of 161,531 and it is considered the largest municipality in Ontario by total area (3,627 km^2). City of Greater Sudbury

comprises of 330 lakes, wetlands, various habitats, and vegetation communities (City of Sudbury, 2005).

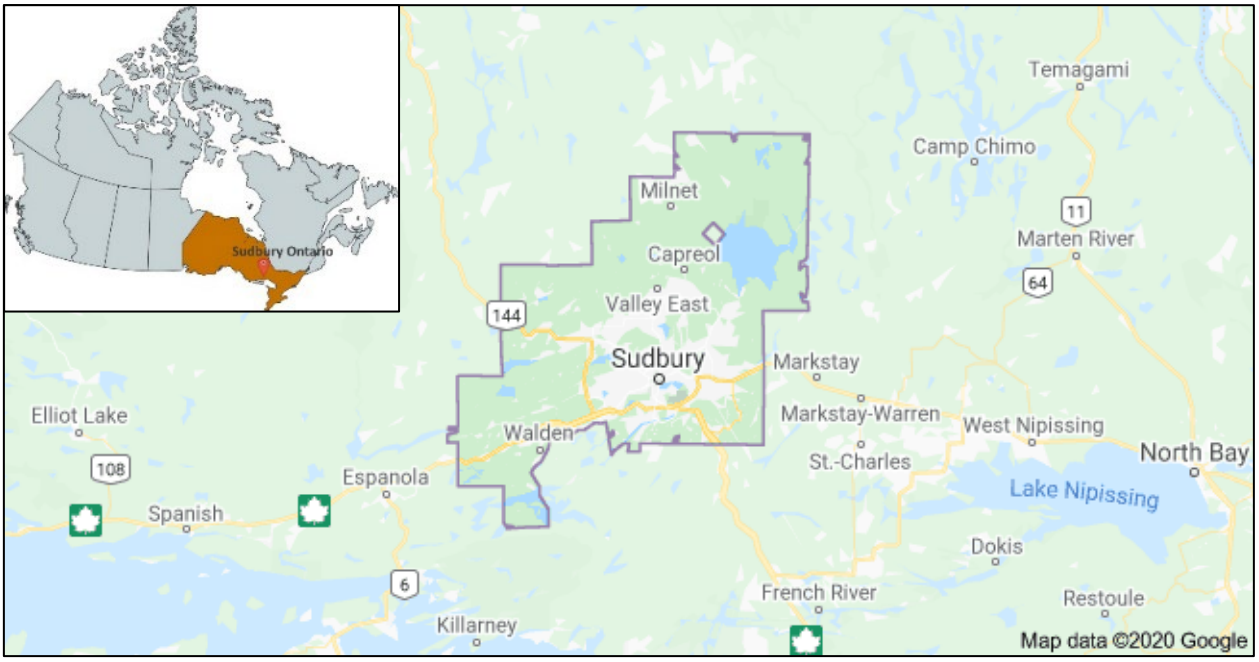


Figure 1: City of Greater Sudbury.

1.1.2. Climate

Sudbury area lies at 260 m above sea level, it is a cold and temperate region. Rains are typical all year round with the summer being mildly hot and humid while winters are usually cold and covered with snow. During the winter season in Sudbury, January lows can reach down to -18ºC (0 F) and in the summer season, July highs can be up to 25ºC (77 F) (Canadian Climate Normals, 1981-2010).

1.1.3. Geology

The Sudbury structure, formed by the Sudbury Impact Event, is an elliptical unit produced by a meteorite that collided with the southern part of the southern province of the Canadian Shield at

approximately 1.85 Ga. The structure consists of three main components: the Sudbury Basin; the SIC, which is geographically divided into the north, south and east ranges; and the Sudbury Breccia, which is an outer zone of locally brecciated footwall rocks extending as much as 80 km from the SIC (Figure 1, Rousell and Brown, 2009).

There is still disagreement about the relative contributions of crustal and mantle magmas being responsible for the formation of the SIC and its associated mineralogy, regardless of more evidence pointing and favoring an impact origin for the formation of this structure (Lightfoot and Zotov, 2005). The SIC is composed of four units, from bottom to top, contact sublayer, norite, quartz gabbro and granophyre rocks (Figure 3). The latter units form part of the main mass, with the contact sublayer forming part of the sublayer. At the top of the SIC lies the Whitewater Group comprising of four formations: Onaping, which consists of basal intrusion and fall-back breccia; Vermilion, made up of carbonate, chert and siltstone; Onwatin and Chelmsford; dominated by carbonaceous mudstone and siltstone, and muddy wackes respectively (Rousell and Brown, 2009).

The formation of the Sudbury structure, from melting of surrounding rocks, resulted in the presence of Ni-Cu and secondary PGEs mineralization embedded in the SIC resulting in the region being one of the world's largest mining camps. The deposits occur in and around the base of the SIC within the contact sublayer, footwall breccia, quartz diorite, Sudbury breccia and other footwall rocks, with pyrrhotite, pentlandite and chalcopyrite being the major ore minerals. The deposits are classified into; SIC-Footwall Contact deposit, footwall vein deposits, offset dike deposits and sheared deposits. About two thirds of the deposits are found at or below the bottom of the SIC with the majority of the remaining one third located in the offsets (Figure 3, Rousell and Brown, 2009).

Two major crosscutting breccias, which are the Sudbury Breccia (SB) and Footwall Breccia (FB) exists within the footwall of the SIC. The mineralogy of the FB is essential when discriminating anthropogenic contamination and any influences that the bedrock might have on the geochemistry of the soils (CEM, 2004).

The FB occurs in four different regions that are, discontinuous sheets, mega breccia, offset dikes and as intrusions in felsic norite. The breccia is host to most of the Sudbury Ni-Cu-PGE ore bodies, made up of quartz diorite and different other rocks, with minerals including plagioclase (dominant mineral), quartz, K-feldspar, amphibole, chlorite, biotite, epidote, pyroxene, sericite, muscovite, apatite, ilmenite, and magnetite. Pyrrhotite, chalcopyrite, pentlandite and pyrite are the main sulphide minerals in the FB, the Ni-Cu-PGE mineralization occurs as pebbly to disseminated, veinlets or lenses and valves (Figure 3, Rousell and Brown, 2009).

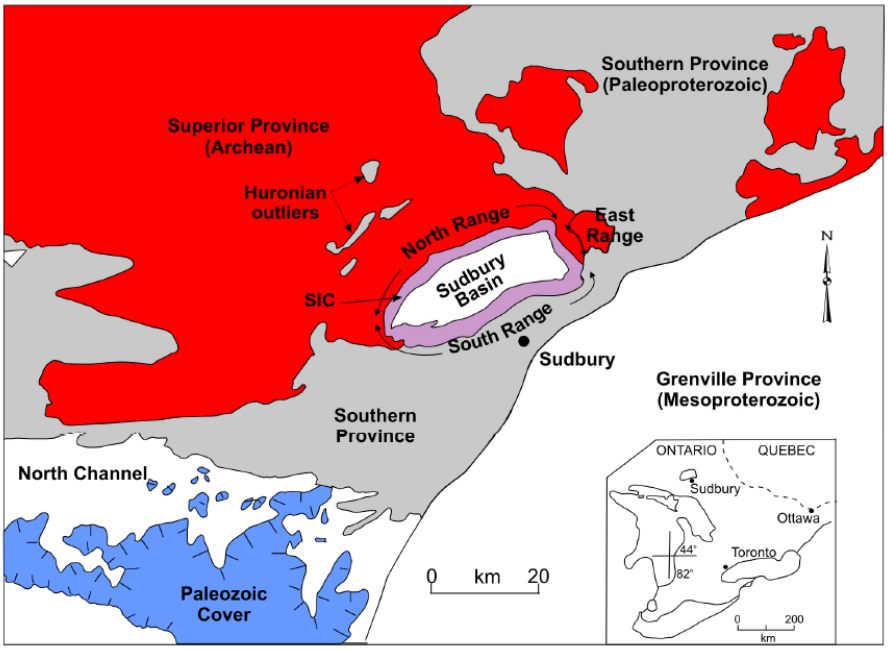


Figure 2: Regional setting of the Sudbury structure. After Rousell and Brown (2009).

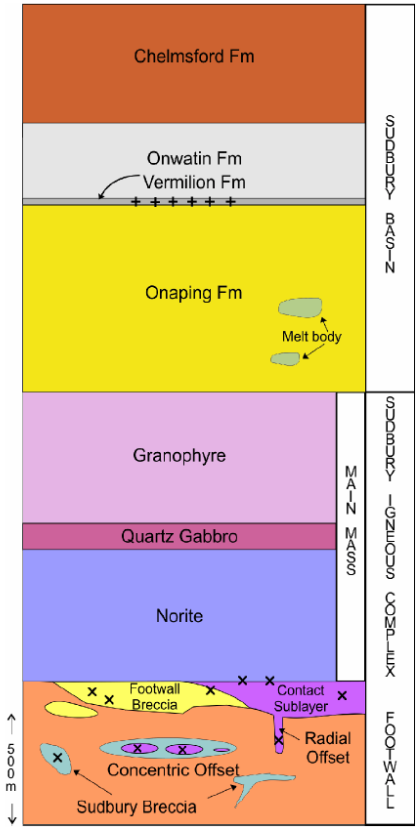


Figure 3: A stratigraphic section of the Sudbury structure, where x= Ni-Cu-PGE deposit and += Zn-Pb-Cu deposit. After Rousell and Brown (2009).

1.1.4. Glaciation and drainage

Most of the parent material in Sudbury was transported by ice forming glacial till sediments during deglaciation of the Laurentide Ice Sheet. Till refers to material that is deposited by the movement of an ice sheet and the flow direction can be determined through studying features such as striations on the rocks. The ice-flow direction in the Abitibi highlands and the Sudbury Basin was north to south, at 170˚ to 210˚, within the valley it was at 220˚and 245˚, and at 205˚ to 225˚ south of the Sudbury Basin (CEM, 2004).

The glacial sediments in the region form thin discontinuous layers above the bedrock, less than 1 m in thickness, forming thicker accumulations along the slopes of the knolls and ridges on the bedrock (CEM., 2004). The transportation of glacier sediments and contamination metals can be

influenced by the drainage pattern of an area. The Sudbury area is drained in the: 1) southward direction by rivers and streams that flow into the Georgian bay, 2) Spanish river in the westward direction draining most of the region, and 3) French River in the east draining a small part of the region (CEM, 2004). An understanding of drainage patterns and a map can assist in visualizing how they affect metal contamination levels.

1.1.5. Soil

Soil is formed by weathering of bedrock (parent) material of which the classification is based upon the mode of formation which can be ice, water, gravity, wind, lakes, and oceans. The nature of the parent rock, weathering, erosion, drainage, and ice-flow direction greatly influences the characteristics of the soil.

The soils in Sudbury area belong to five soil orders of the Canadian Soil Classification System, and they are namely: Luvisolic, Glysolic, Podzolic, Brunisolic and Organic soils (https://soilsofcanada.ca/orders/podzolic.php). The Podzolic soils are the most dominant kind found in this region. They are mainly found on well drained sandy tills in the region (Spektor, 2003). Podzolic soils are forested soils found mainly on sandy parent materials in areas that are underlain by igneous rocks mostly on the Canadian Shield, but they can also be found elsewhere on Sandy Glacio-Fluvial deposits (https://soilsofcanada.ca/orders/podzolic.php), dominant vegetation types within these soils include coniferous-dominant plant communities.

1.1.6. Vegetation

The Sudbury region is in a forest zone classified as the Great Lakes St. Lawrence Forest Region (SARA, 2008), which is a transition occurring between the boreal, coniferous forest to the north and the deciduous forest in the southern margin (Spektor, 2003). The southern margin (of the Northern Temagami Section) is characterized by large stands of red pine (*pinus resinosa*) and white pine (*Pinus strobus*). The southern margin, which is of the southern Algonquin Section, is characterized by white pine and eastern hemlock (*Tsuga canadensis*) that grew amongst sugar maple (Acer saccharum) and red oak (Quercus rubra) hardwoods in the uplands (SARA, 2008).

The pre-settlement forest could have entailed of a mixture of white (pinus strobis) and red pine (pinus resinosa) in the uplands with white cedar swamps in the lowlands (Spektor, 2003). The existence of former forests can be seen by some evidence of remnant burnt white pine stumps (Figure 4) that can be seen in the uplands and in the barren peatlands, evidence of remnant eastern white cedar stumps can be seen (SARA, 2008).

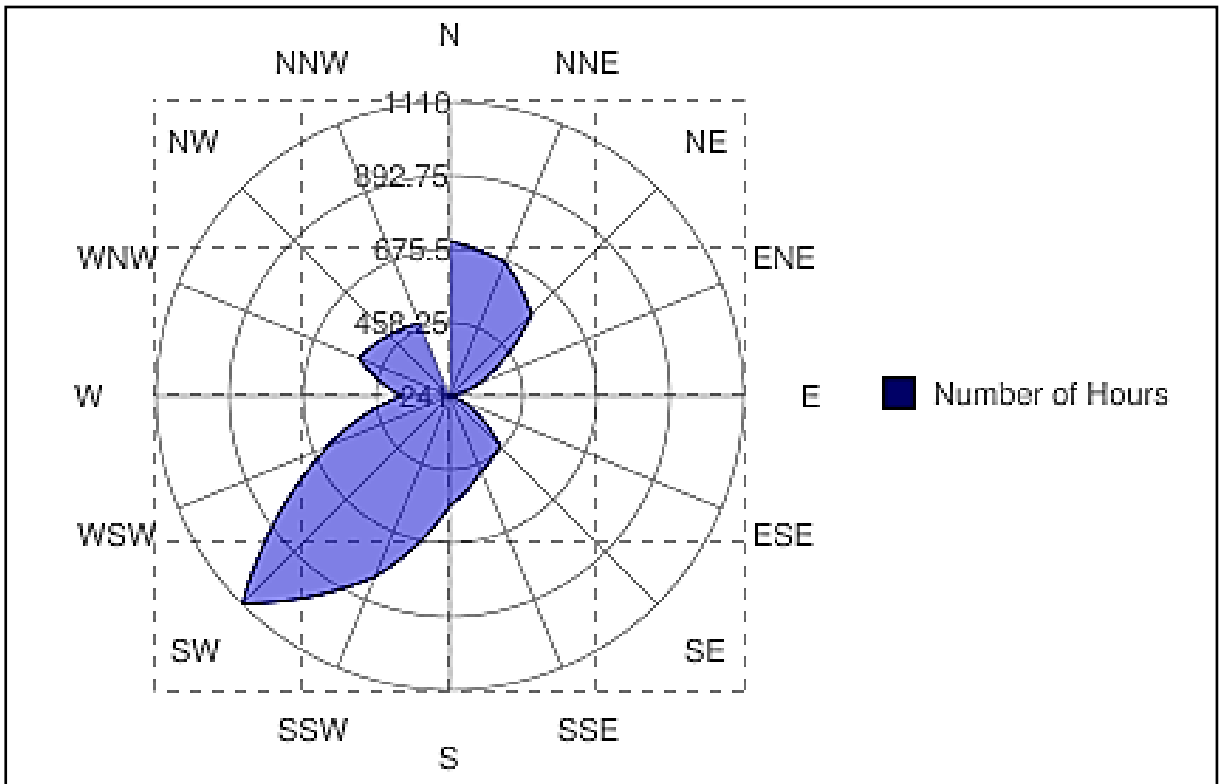
Figure 4: Remnant burnt white pine stump in Sudbury (SARA, 2008).

1.1.7. Wind patterns

The deep history of smelting in the Sudbury area made the soils to be vulnerable to atmospheric pollutants. These pollutants could have fallen through snow, rain, and dust. Atmospheric pollutants can thus be linked to prevailing wind directions as they play a vital role in determining the direction in which most of the airborne particles will be blown from smelter locations. The emissions were and are dispersed into local communities surrounding the smelter stacks as a function of the local meteorological conditions and source characteristics (SARA, 2008).

The distance at which these emissions or pollutants might travel also depends on the height of the stacks, meaning that tall stacks (e.g., Copper Cliff) might show high discrepancies as compared to shorter ones. An example is by Yang, Drohan and Yang (2020) who used seasonal wind directions

and proximity to source to determine soil contamination across an abandoned steel and iron plant, they found that the directions that had the highest pollutions were proximal to the prevailing wind direction. Winds are experienced from all directions in the Sudbury region, but the dominant directions are from the south westerly (SW) and the north to north easterly (N-NE) (Figure 5). The dominant historical wind direction has been found to be from the SW as shown in the rose plot of Figure 5. The fate of the emissions from the smelters and other operations has been determined by wind and other meteorological characteristics since the beginning of the operations and it is found that communities to the northeast and southwest would have experienced the maximum range of pollutants as compared to all the other areas in the Sudbury region (SARA, 2008).



Historical Wind Characteristics

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Speed (km/h)	16.6	16.1	17.2	17.4	15.9	14.8	13.5	13.2	14.6	16.0	16.7	16.0	15.7
Predom. Direction	SW	N	N	N	N	SW	SW	SW	S	S	SW	NW	SW

Figure 5: Quarterly wind patterns in Sudbury area (top) and yearly predominant wind directions. Wind speeds are from 1971 to 2000 indicating dominant winds from the southwest and north. Sudbury Weather Stats (2020) and SARA (2008).

1.2. Chemicals of concern (COC)

The COC are the elements found to be present in the Sudbury area that present superior potential for exposure and risk to the environment and humans. The primary source of the high levels of the COC in the Sudbury Soils Study assessment was atmospheric deposition, in addition to other

factors such as dust emitted from mine tailings and roasting beds (SARA, 2008). The measured metals from the SSS were compared with soil quality guidelines presented by the MOE (SARA, 2008). The elements that were present more than the MOE criterion were used for further risk assessments in the area. Initially, arsenic (As, metalloid), nickel (Ni), copper (Cu) and cobalt (Co) were the only elements regarded as COC in the area. Pb and Se were included at a later stage after further investigations, they were found to also have a link to smelter regions with a high positive correlation to Ni and Cu.

As, lead (Pb), Cu, Ni and selenium (Se, metalloid) were found to be distributed in the vicinity of the historic smelter regions following a NE-SW ellipsoidal trend, which is consistent with the prevailing wind directions in the Greater Sudbury area (SARA, 2008; Mantha, Schindler and Kyser, 2012). In summary, the elevated metal levels of COC were centered in the vicinity of the smelting regions of Coniston, Copper Cliff, and Falconbridge (SARA, 2008). Figure 6 presents the soil data screening process that was used in selection of the COC, starting with all the elements analyzed and the various screening parametrization that was used.

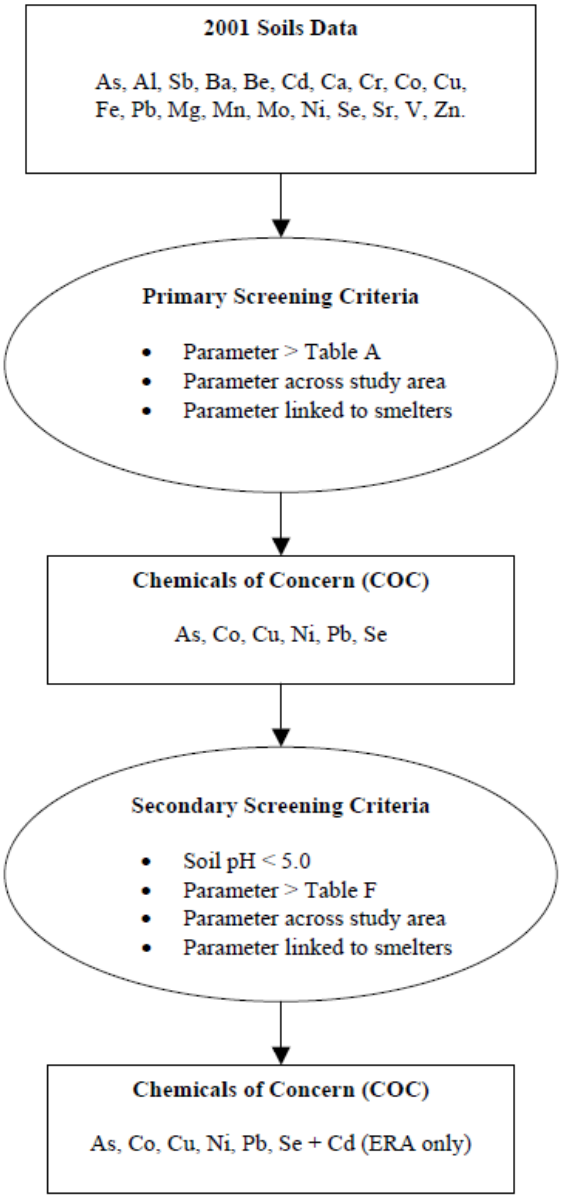


Figure 6: Soil data screening process as used in the SSS for selection of COC. Adapted from SARA (2008). Table A provides guidelines of surface soil and groundwater criteria for a potable groundwater condition while Table F is about Ontario typical range background soil concentrations. ERA refers to Environmental Risk Assessment.

1.3. Regions of interest

The regions of interest in this study were centered on the historical smelters of Falconbridge, Copper Cliff and Coniston, and consider semi-barren areas surrounding them Figure 8 that were previously heavily contaminated by metals and deprived of vegetation growth.

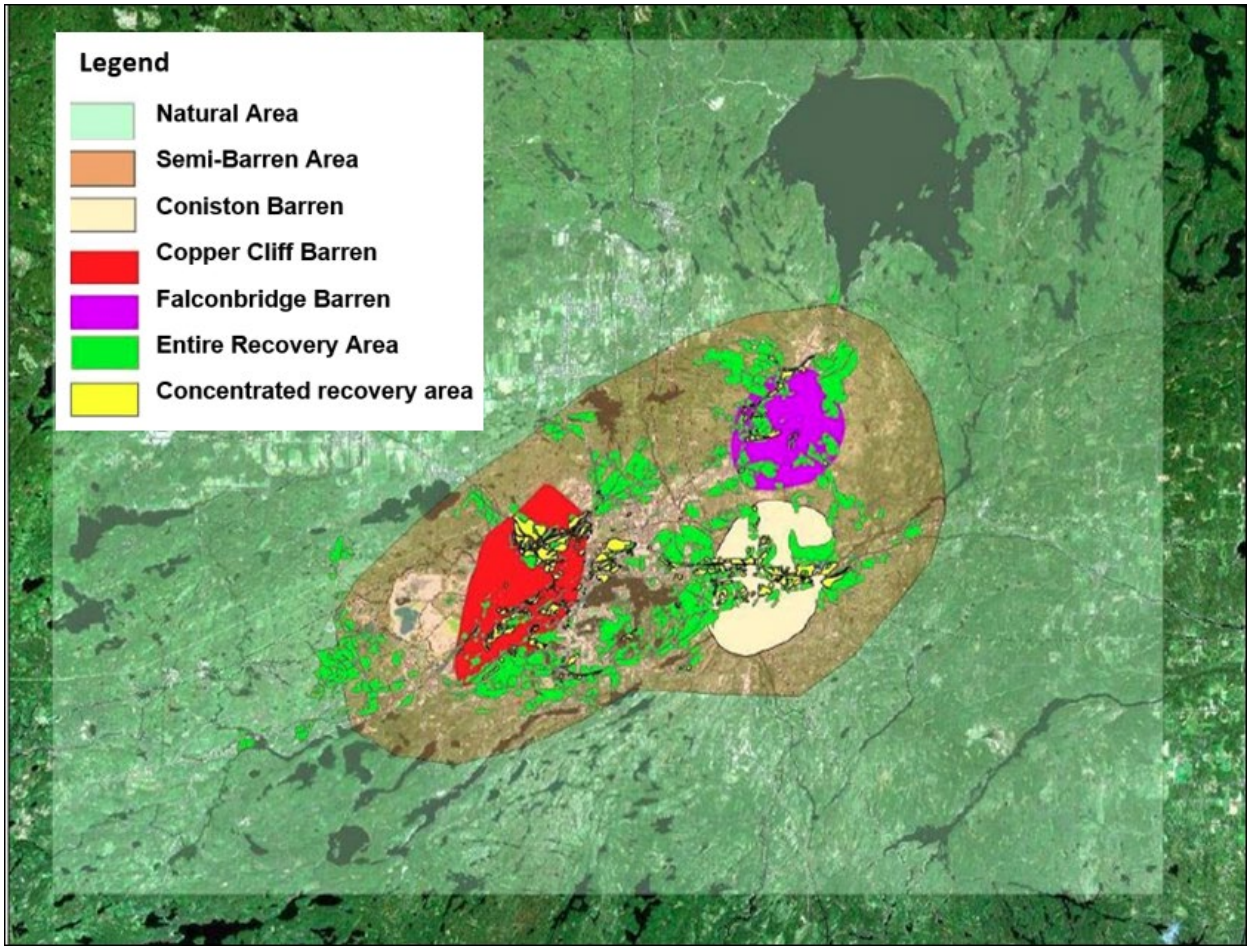


Figure 7: Regions of interest used for the SSS; the current study will also focus on some of these regions. Adapted from SARA, 2008.

1.4. Research objectives

This research project aims at conducting a detailed geostatistical study of the historical soil chemistry data and make comparisons with up-to-date Landsat Earth-Observation maps, to better understand land recovery and regreening, in the past years. The objectives of the study are to: (1) evaluate the regional, geospatial distribution of measured trace element concentrations and their dependency on the interpolation statistics of kriging and Gaussian geostatistical simulation (2) and to fulfill the broader objectives of environmental monitoring and human impact assessment.

A comparison between geostatistical and satellite data will be conducted using a Kriging spatial interpolation and conditional simulation of the multielement data and by integrating the results of these interpolations with the remote sensing information, to create comparative hybrid maps of both kriging and remote sensing map products. This will be done to improve the understanding of soil contamination trends and their impact on vegetation and other aspects of the biosphere.

1.5. Expectations

Based on the SSS results, the following is expected:

- Anomalous metal (loid) concentrations in the upper 0-5 cm soil profile depth.
- Concentration of individual metals decreasing at around 120 km from the smelters.
- Elevated concentrations centered in the vicinity of the three smelters with an asymmetric distribution resembling a NE-SW ellipsoid.
- Progressing vegetation regrowth in the smelter centroids.

The current study will aim at validating some of these results and enrich them to provide a refined interpretation and understanding, as it considers the use of different methods and new Landsat 8 data that were unavailable during the SSS study.

Chapter 2: Metals, smelting and contamination

Introduction

This section presents major key factors that will be considered during the discussion of the results. An introduction to metals, their sources, behaviour in the soil environment and how they impact the environment in terms of metal pollution or contamination is presented, to understand how the observed contamination patterns came to be. The term metals will be used when referring to both trace metals (elements) and metalloids (chemical elements with properties that are in between metals and non-metals) of concern in the study. The effects and impacts of smelting and contamination in Sudbury, as presented in this chapter, will aid as the backbone into analysing the various driving forces that could have led to the observed soil contamination patterns, this is to assist in meeting all the study's objectives. The previous soil studies conducted will be aimed at introducing the SSS in major terms, as it forms the foundation of the current study. An example from a Ni-Cu mining center in Norilsk, Russia is presented, and their findings will therefore be compared with what has been previously analysed and will be currently analysed for the Sudbury area, to be able to draw up useful conclusions.

2.1. Metals

Metals refers to elements that form positive ions by losing electrons in chemical reactions and are lustrous. They can be described and categorized as trace metals, transition metals, micronutrients, toxic metals, and heavy metals (Mclean and Bledsoe, 1992). Naturally, all soils contain trace levels of metals Table 1 meaning that their presence is not indicative of contamination, however, the concentration of metals in uncontaminated soils is basically related to the parent material geology

that the soil was formed from (Mclean and Bledsoe, 1992). Many countries in the world depend on mining for their economic growth. Metals play a vital role in our daily lives and are useful for most of the technological developments. Humans, flora, and fauna also depend on metals for survival; however, this is available in micro-nutrients, an oversaturation of these metals can cause a survival threat. Trace element concentrations exist in plant and animal tissue in amounts less that 0.01%, this increases when they become available in excess and said to be toxic. There are thirteen trace metals and metalloids, which are Ag, As, Be, Cd, Cr, Cu, Hg, Ni, Pb, Sb, Se, Tl and Zn, that are considered to be priority pollutants and can be derived by either natural or anthropogenic processes (Sparks, 2005). The Earth's crust is mostly composed of metal oxides, of which most of them are found in the soils of active mining sites or regions.

Table 1: Levels of some metals worldwide in the soils and their ranges. It should be noted that an excess of these levels is not a direct indicator of contamination as some regions also have an abundance of some metals within the parent rocks (Mclean and Bledsoe, 1992).

Metal	Selected average for soils mg/kg	Common range for soils mg/kg
Ni	40	5-500
Cu	30	2-100
Co	?	?
Pb	10	2-200
As	5	1.0-50
Se	0.3	0.1-2

2.1.1. Natural sources

The majority of trace metals occur in nature, with the main source being weathering of soil parent materials which includes both igneous, metamorphic and sedimentary rocks (Lal, 2017). The chemistry and mineralogy of soils in an area might therefore be a direct reflection of the bedrock geology. In the Sudbury basin, the bedrock is known to be highly mineralized. However this is not reflected in the higher background of soil concentrations and probably is due to the dilution with upstream rock materials as an aftermath of glaciation events (SARA, 2008). Further to this dilution effect, base metal-rich phases in sulphide-rich units of the regional bedrocks are relatively soft and might have been transported and then dissolved from the surficial materials because of glaciation and weathering.

The metal distribution in the soil parent material was collected to determine the normal background levels of the contaminants of concern (Table 2) to meet one of the Center of Environmental Monitoring's project goals. The soil samples were collected at depths greater than 80 cm (i.e., 85 to 112 cm). The analytical data obtained from these samples serves as the first attempt into determining pre-industrial levels in the regional soils, the data hence provided an accurate representation of regional background levels of *Aqua Regia* extractable metals (CEM, 2004). The mean background concentrations in the parent materials were compared with, 1) values for mean crustal abundance for granite (the dominant rock in the region), 2) values of conterminous United States, 3) and values of soils of the Canadian Shield for comparisons (CEM, 2004). Documented true natural background surface soil metal concentrations in Sudbury have been found to be similar to other regions of the Canadian Shield (SARA, 2008).

Table 2: Mean background concentrations (ppm) of the chemicals of concern in Sudbury as compared to the Canadian Shield soils and Ontario Ministry of the Environment's Table F.

Element	MOE concentration (Table F)	Canadian shield soils	CEM (mean)
As	17	n/a	1.11
Co	21	19	8.9
Cu	85	12	26.4
Ni	43	12	36.1
Pb	120	20	5.9
Se	1.9	0.18	0.06

2.1.2. Anthropogenic sources

Input of trace metals by anthropogenic sources includes agriculture, metallurgy, energy production, sewage sludge, microelectronics, and many others. Anthropogenic activities that are associated with industrial processes such as mining and smelting, manufacturing, disposal of domestic and industrial waste (e.g. slag heaps), and application of phosphorous fertilizers are found to be the major source of metal contamination in soils (Lal, 2017). Atmospheric deposition is a major catalyst for the input of trace metals in soils and plants (Sparks, 2005). In an active mining and smelting region, the soil contamination is usually due to dust fall with excess trace metals, mine waste and release of sulphur dioxide, which causes acid rain. Soils are the major recipient of metal contaminants in terrestrial ecosystems whereas in aquatic systems sediments are the major sink leading to the impact of freshwater due to runoff, drainage or disposal, and groundwater due to leaching or transport through mobile colloids (Sparks, 2005).

2.1.3. Influence on the environment

Soil contamination by trace metals might occur due to atmospheric deposition or direct deposition on the soils. Trace metals can be deposited by either way. Accumulation of trace metals in the soil might result in the degradation of soil quality, crop yield, agricultural produce, ecosystem and have negative impacts on human and fauna's health

The elevated concentrations of Ni and Cu around Sudbury smelters is one of the reasons of slow development and degradation of vegetation cover and soil (Spektor, 2003). An increase in atmospheric deposition causes an increase in metals in the air and soil causing changes in the flora and degrading species that are sensitive to metal pollution. This in turn results in only few species surviving in the area. It is a matter of survival of the fittest, flora that is resistant to metal pollution or that has acquired resistance with time will survive while their counterparts degrade. Vegetation distribution can then be used as an indicator of metal pollution and to study an ecosystems heath.

2.1.4. Metal behaviour in the soil environment.

Metal concentration in the soil solution is controlled by a variety of processes that are interconnected, ranging from inorganic to organic complexion, oxidation-reduction reactions, precipitation or dissolution reactions, to adsorption or desorption reactions (Mclean and Bledsoe, 1992). Any changes that might occur in the environmental conditions, could be man-made or natural processes, such as degradation of the soil organic waste matrix, changes in soil pH, redox potential, or soil solution composition may enhance metal mobility (Mclean and Bledsoe, 1992). The solubility, mobility, bioavailability, and toxicity of metals in the environment is controlled by biogeochemical processes at reactive natural surfaces, including clay minerals, metal oxides and

hydroxides, humic substances, plant roots and microbes surfaces (Sparks, 2005). The biogeochemical processes that metals and metalloids undergo occurs in the critical zone, which is the mixed near surface environment where the rock meets soil, water, air, and living organisms regulating the natural habitat, and can occur as described by Sparks (2005): when metal ions enter the soil solution, they may be subject to numerous pathways that can potentially overlap (Figure 8), hence the soil solution contains metals that exist as free ions (free metal ions are the most bioavailable and toxic form of metals; Mclean and Bledsoe, 1992) or as complexed to inorganic or organic ligands, these can then be;

1) Taken vertically by plants,
2) Retained on mineral surfaces, natural organic matter, and microbes,
3) Transported via the soil profile into groundwater through leaching or either by colloid-facilitated transport,
4) Precipitated as a solid phase, and
5) Diffused in porous channels such as soils.

Soils in mining regions usually have a low pH as an effect of acid rain, which increases the ability of soil to hold metals by increasing pH. The soil pH in the Sudbury area has been low for an exceptionally long time (Spektor, 2003) due to constant mining activities, resulting in the ease of metals to migrate through the soil profiles. Soil pH is inversely proportional to trace metal solubility, low pH entails high solubility and vice versa. Solid's precipitation e.g. carbonates, sulphates, oxides, and sulphides increase the limitation of metal solubility (Spektor, 2003), hence lowering the time for these solids to dissolve in the soil, the metals can then be absorbed by flora and fauna.

Not much is known about the occurrence and forms of Sudbury's main metals, which are Ni and Cu in the soils, but a study conducted using Scanning Electron Microscopy with Energy Dispersive X-ray spectroscopy (SEM/EDX) that was applied on samples collected by (Dudka, Ponce-Hernandez and Hutchinson, 1995) showed that Cu is associated with non-residual organic matter and is highly mobile if compared to Ni. Nickel on the other side forms fewer mobile sulphides and oxides. So, it is expected to find Cu traveling for longer distances as compared to Ni within the Sudbury soils. Both Ni and Cu are found as sulphides usually associated with chalcopyrite, cubanite, magnetite, ilmenite, pyrite and the most common mineral, pyrrhotite (Adamo *et al.*, 1996).

Fertilizers, land disposal, mining, atmospheric deposits

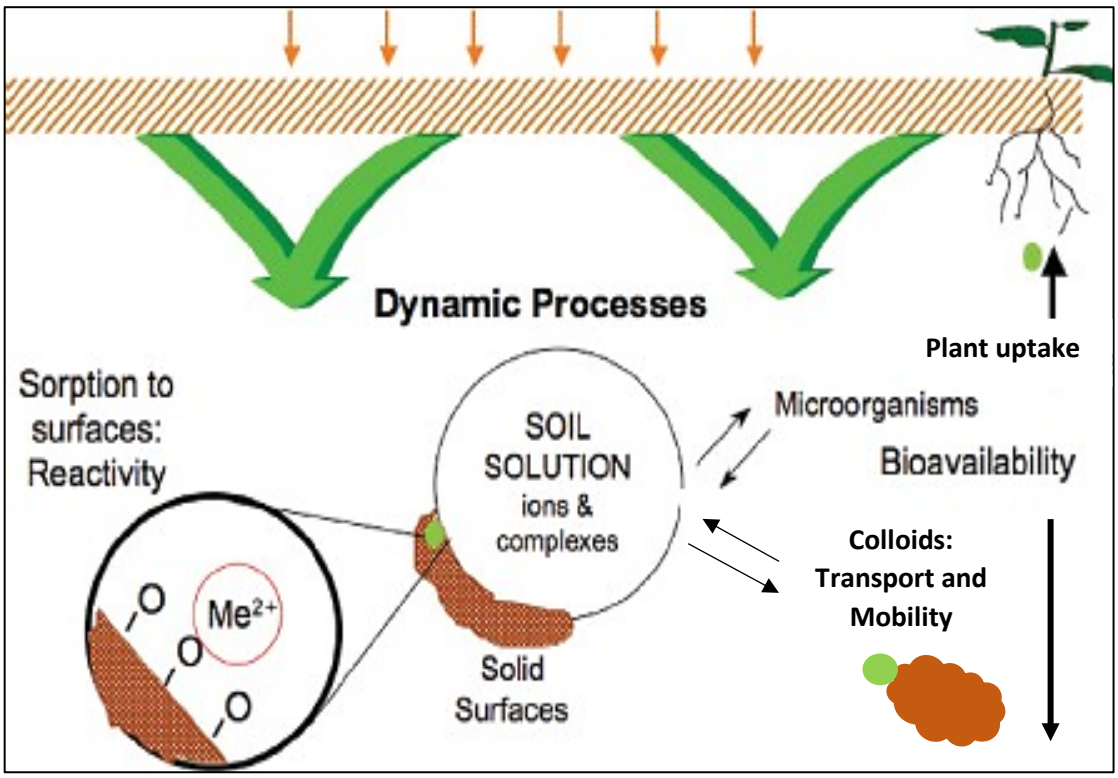


Figure 8: Biogeochemical pathways and processes of metals in the critical zone. Me = metal, Small green blobs = metal ion, big brown blobs = soil particles, and orange blob = colloid. After (Sparks, 2005).

2.2. Smelting and contamination in Sudbury

The abundance of ore minerals within the Sudbury basin has attracted mining and smelting processes, which resulted in the initiation of several smelters Figure 9 in the early 1930s and culminated with the construction of the superstack (main smelters sites: Copper Cliff facility, Coniston (closed in 1972) and Falconbridge operation; SARA, 2008).

The aim of smelting has always been to separate the ore from the rocks via ignition. Several roast yards operated around the region (Figure 9), with a blast furnace used as a production of feed in Copper Cliff (SARA, 2008). Timber was used as roast beds to ignite the ore from the rocks, this

caused deforestation and enhanced soil erosion. Roasting is an exothermic reaction, air pollution was emitted relatively close to the ground, however the emissions could not have been carried far as from tall stacks (smelters), denoting that pollution was directly carried down to the soil in the immediate vicinity, contributing to extremely high metal concentrations. By 1930, blast furnaces and roast yards were then replaced with multi-hearth roasters and reverberatory furnaces as a means of smelting (SARA, 2008).

During the peak years of smelting (between 1942-1944, as an effect of World War II), an average of 1.5 million tons of SO_2, 200 tonnes of Pb, 120 tonnes As, 50 tonnes of Se, and over 1,000 tonnes of Cu and Ni were emitted (Malcolm, 2015). Production and resources from the past exceeded 1.548 million tonnes of ore at approximately 1.2% Cu and 0.4 g/t Platinum Group Metals (PGMs), this resulted in an estimated production of 8.5 million tonnes of nickel and 8.4 million tonnes of copper over the years (SARA, 2008). However, the current stack particulate loadings have decreased significantly and are lower than historical emissions, from commencement operations to the 1950s, due to developments and implementation of highly efficient dust removals technologies.

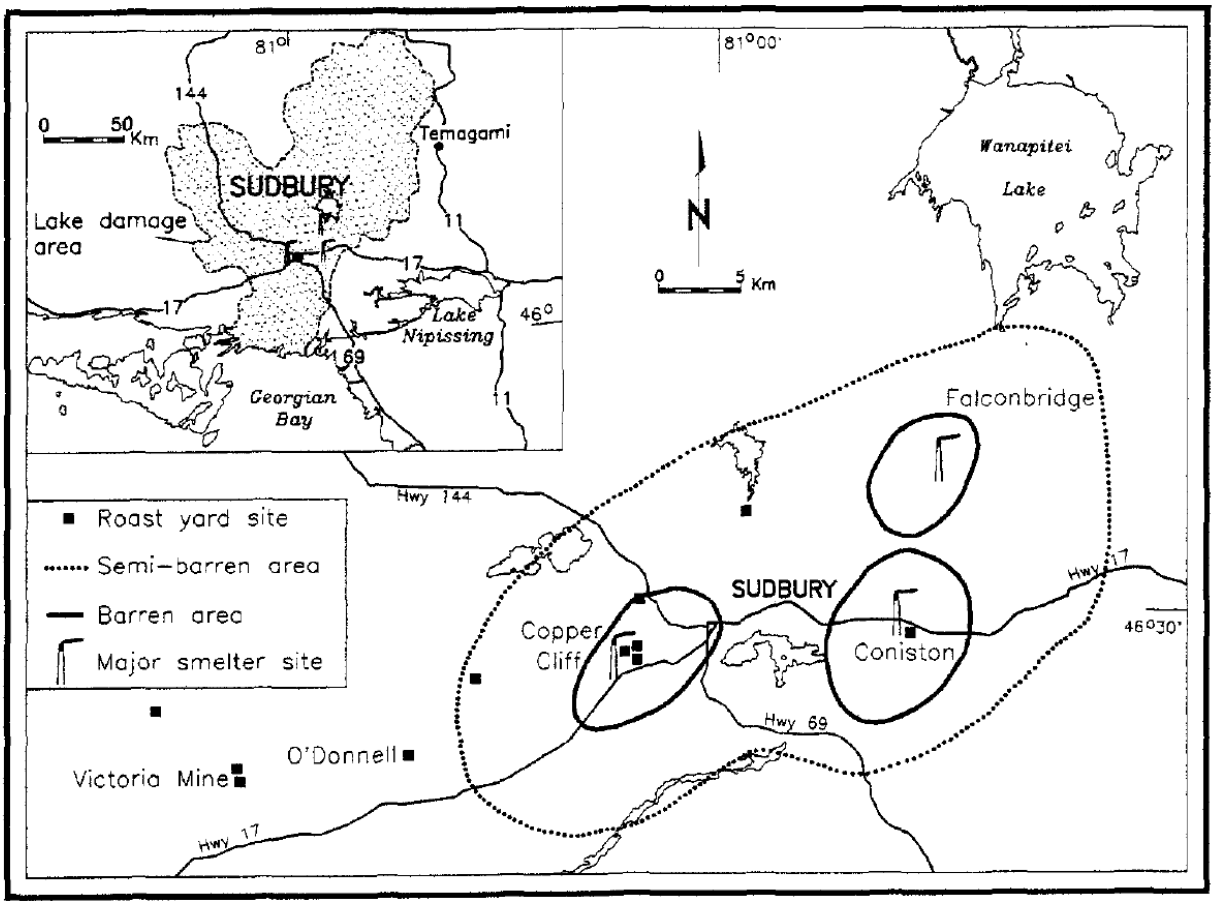


Figure 9: The location of key roasting and smelting sites in the Sudbury area, also shown are semi-barren and barren areas around the smelter centroid, with the lake damage area shown by image insert. After (Gunn et al., 1995).

The Sudbury ore is comprised of 40% iron sulphides, 4-5% Cu and Ni, and 35% sulphur, with iron in the deposit, posing difficulties in smelting as the process of separating copper and nickel from sulphur, iron and barren rock is complicated and requires expensive and advanced technology (Group, 2008e). Copper and nickel are considered as Sudbury's main metals. A study conducted by (Dudka, Ponce-Hernandez and Hutchinson, 1995) indicated that the concentrations of copper and nickel appears to be higher than other elements, with a range of Ni (5.3-2149.0 mg kg-1) and Cu (11.4-1891.0 mg kg-1, dry wt.) when compared with the world concentration of metals in uncontaminated soils (Ni , 25 mg kg-, and Cu ,12 mg kg-1), in a study conducted by Berrow and

Ure (1986). These results showed that Ni and Cu are the most elevated metals in the Sudbury region and much attention should be paid to them.

Emissions in Sudbury have decreased from the early mining years to present times. Figure 10 shows an example of COC emission rates around the Falconbridge smelter (Xstrata smelter), it shows declines around the 1980s of the COC which spikes around the 1990s.

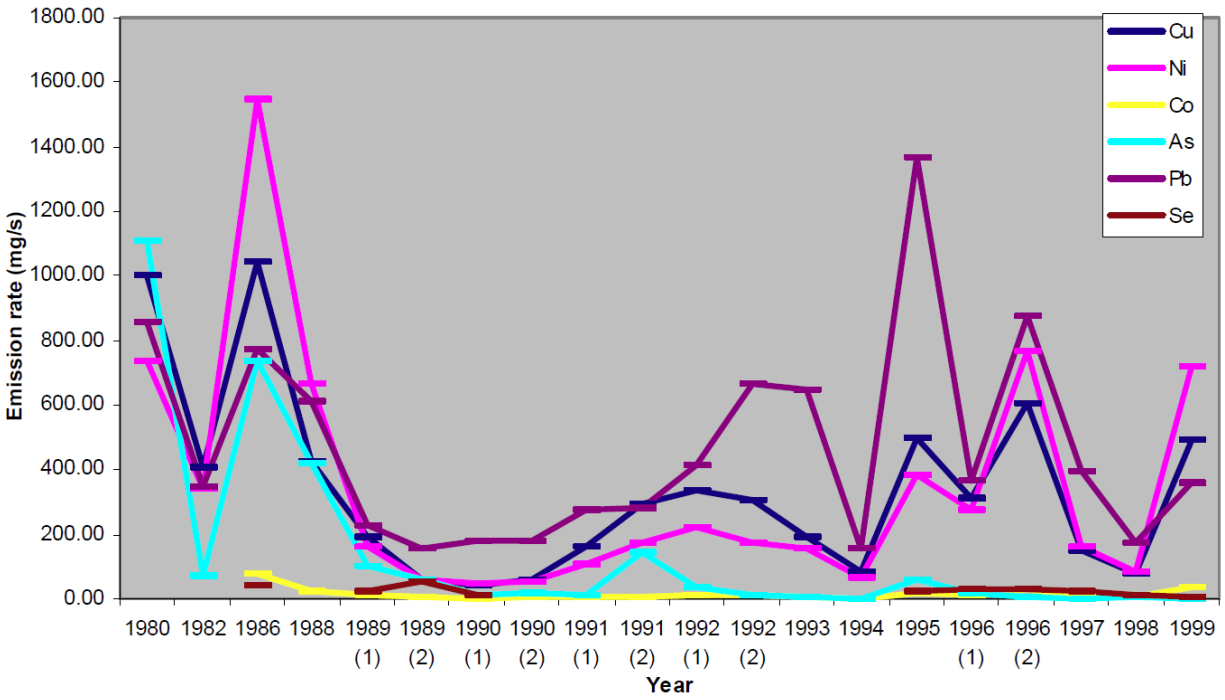


Figure 10: Emission rates of COC at Falconbridge (Xstrata) smelter. (SARA, 2008).

2.2.1. Effects on the environment

The ellipsoidal distribution of denudation of vegetation due to mining and smelting activities in this region led to high levels of soil erosion and rapid geomorphological change of land. When vegetation recovery programs started, the area had already lost most of its topsoil which contains vital nutrients as phosphorus, nitrogen, magnesium, manganese, and calcium. This could be one of the reasons for the slow land recovery. The aim of the land recovery is to increase the soil pH

where soils have been found to have exceptionally low pH in the range of 2.0 to 4.5 (Gunn *et al.*, 1995).

Much of the vegetation that is seen in Sudbury today is due to land restoration processes that started in 1917, to restore barren lands that were degraded of vegetation. These barren lands were due to slag heaps, intensive harvesting and forest fires, roasting yard impact and emission damage. Slag heaps take up a lot of space and provide an inhospitable environment for plants to grow. During precipitation, the slag can be washed off and enter the soil environment, hence contaminating it. The harvest process involves selection of certain species (e.g., red pine and white pine) for construction, conifer was also removed when building railways, sparks from the track would spark the timber left by the rail side. Fires were also caused by prospectors who lit up flora to expose the rock outcrops for exploration studies. Timber was used to ignite ore from the rocks on roast beds causing stress in the soil environment and deforestation as a lot of it was required for this process. Emissions damaged vegetation through acid rain and toxifying the surface with metals that the vegetation could not cling to for survival. The barren lands were mostly around the smelter regions (Figure 8) and can still be seen today.

The vegetation recovery in the region followed a reduction in atmospheric emissions and follows a noticeable trend. Recovery has been slow and limited in certain areas, due to that soil properties are more critical than air quality in terms of suppressing colonization of barrens (SARA, 2008). Some of the key recovery limiting factors include, loss of soil micro-organism and organic matter, physical loss of soil erosion, elevated metal levels in surface soils, soil chemistry limitations and sever micro-climate conditions. The improved growth of vegetation (Figure 12) measured since 1972 in regions that were moderately affected by pollution was the result of improved atmospheric quality (SARA, 2008). The type of metal pollutants, concentrations, accumulation, and behaviour

together with physiological system of the plants results in different reactions by the plant species. Different kinds of vegetation can be used as indicators of metal pollution levels, for example, lichens are used for biological monitoring as they have a high sensitivity to air pollution. The effect of air pollution on vegetation occurs in different stages. The first effect is the physiological level such as damaging of leaf cells, followed by visible symptoms appearing depending on level of cell degradation and under long influences the changes become more and more visible (Spektor, 2003). The study of plants can assist in giving a clear picture on the level of metal pollution in the soil environment, however, this study is not based solely on vegetation studies but rather the spatial distribution of trace metals in the soils.

Figure 11: Part of Sudbury showing the changing landscape (Greater Sudbury.ca, 2009).

2.2.2. Soil studies

Previously, environmental studies had only focused on sulphur emissions, it was only until the 1960s when environmental concerns expanded to include metal levels and acidification of the soils, that is when the soil studies began (SARA, 2008). The studies' results indicated that the

soil acidity and concentration of copper and nickel were elevated in the same areas where sulphur dioxide had been measured, making soils studies a pre-requisite.

In September 2001, the Ontario Ministry of the Environment (MOE) released a report summarizing previously unreported extensive soil and vegetation chemistry data that was conducted in the Greater Sudbury area during 1971 to 2000, from the Ministry's Sudbury Regular Survey, the Sudbury Special Survey, and the year 2000 Surface Soil Survey (MOE, 2001). The recommendations of the report were the following: (1) a more detailed study should be undertaken to fill data gaps and (2) a Human Health Risk Assessment (HHRA) and Ecological Risk Assessment (ERA) should be undertaken (MOE, 2001). These recommendations gave rise to the Sudbury Soil Study (SSS), with the underlying objective of determining whether Sudbury soils containing metal and metalloid levels (e.g., Arsenic) above generic guidelines posed an unacceptable ecological and human health risk.

In 2001, two mining companies Vale Inco and Xstrata Nickel funded the SSS of which the report was released in 2008, this study is the most extensive soil study to ever be conducted for the Sudbury area (SARA, 2008). Three comprehensive soil surveys were conducted, with the aim of quantifying anthropogenic metals in the local soils and to determine pre-industrial metal levels in local soils and around smelter regions. The studies are namely, the Sudbury Urban Soil Study (SUSS), Sudbury Regional Soils Study (SRSS) and the Falconbridge Soil Survey (FSS). Data from these studies was combined, forming the 2001 SSS. The SUSS aimed at investigating metal concentrations in local urban soils and compared them to MOE guidelines, while the SRSS analyzed the distribution of metal species in undisturbed soils on a regional scale of the 0-5, 5-10, and 10-20 cm soil profile depths (Malcolm, 2015). On the other hand, the FSS focused on sampling properties that are owned by Falconbridge.

The SSS identified six COC; Ni, Cu, Co, Pb, As and Se, to be distributed at approximately 20 km from the smelter regions. The SRSS showed that samples that are within the 0-5 cm soil profile depth are highly enriched in metals than the deeper soil depths, and that their distribution is consistent with atmospheric deposition being the driving force and historical industrial activities of the Sudbury area (Malcolm, 2015). The Center of Environmental Monitoring (CEM) in Laurentian University also did a report on the SSS that indicated the concentrations of these metals (COC) along concentration gradients indicating the effects of smelter emissions as being relative to regional backgrounds at approximately 120 km from downtown Sudbury. The detailed maps of these can be found in the CEM (2004) report.

2.2.3. Correlations with other Ni-Cu mining centers

2.2.3.1. Russia-Noril'sk

Noril'sk region sits at about 300 km north of the arctic circle with a population of approximately 170,000 people. Located on the Tunguska basin in the center of the 3000*4000 km Siberian continental flood basalt (Habashi, 2012), the Noril'sk mining district is one of the world's largest mining camps being principal producers of copper, nickel, and PGEs. Noril'sk history is no different from that of Sudbury with regards to environmental pollution by metal contaminants.

This area is one of the most environmentally damaged in Russia, with vegetation degradation due to mining smelters around the region. Smelting in this region began around 1935 with the aim of producing copper and nickel at a much lower rate (Spektor, 2003). The extensive smelting led to surrounding ecosystem degradation destroying forests at 120 kms south-south-west from the

smelter source. Just like Sudbury, nickel smelting in this region is the main source of pollution in the form of smog and acid rain by sulphur dioxide emissions and metal contamination. The forest degradation was only recorded in 1968, but during this time not enough finances where available to switch over and start implementing current technologies of lower emissions. The soils in this area lack a top soil horizon with soil profiles having 3000-4000 mg/kg and 500-2000 mg/kg of copper and nickel respectively mostly in the 5-10 km zone from the emission point source (Spektor, 2003).

The acquired information from this mining region will be used later for comparisons with what is already known and will be found for the Sudbury region.

Chapter 3: Methods

Introduction

Characterization of spatial variability of trace metals in soil is essential in the identification of pollution sources and the potential risk that this might cause to humans and the environment. Trace metal concentrations might depend on the geochemical nature of the bedrock, abiotic and biotic factors (Guagliardi, Cicchella and De Rosa, 2012). Geostatistics allows for the description and modelling of available spatial patterns, prediction (kriging) at unsampled locations and the assessment of the associated uncertainty by means of stochastic simulation (Saavedra, Spiers and Dunn, 2007). Geostatistical methods of ordinary kriging and simulation are being applied in this study to facilitate predictions of spatial distribution of trace metals through geospatial interpolation.

This chapter introduces theories and methods that were used to investigate the distribution of spatial variation of soil pollutants and vegetation using respectively geostatistical methods and remote sensing techniques. The methods applied are introduced using brief theories, with important equations (equation derivations are not shown) where applicable and their use in the study. Kriging geostatistical interpolation methods were applied to estimate trends and the spatial distribution of metals, as the soil sampling that was conducted provides limited insights on trends, due to the sparsity of the information. On the other hand, remote sensing techniques were applied to look at vegetation distribution from a satellite point of view and further compare the analysis to what is depicted from geostatistical models. The aim of integrating the two methods is to compare the soil contamination trends obtained from kriging maps with remote sensing applications to assess any correlation to make precise decisions regarding the effect that these contaminants are having on the vegetation distribution of the study region. The chapter is divided into three main

sections (Figure 12), firstly, statistical, and geostatistical interpolation methods that were applied are introduced, then the application of remote sensing techniques with major focus on vegetation analysis is also explained and lastly, an illustration of the integration methodology will be presented. The use of ESRI ArcGIS 10.6.1. and its *geostatistical analyst* tool, excel spreadsheet, and an R package was implemented to meet the study's objectives.

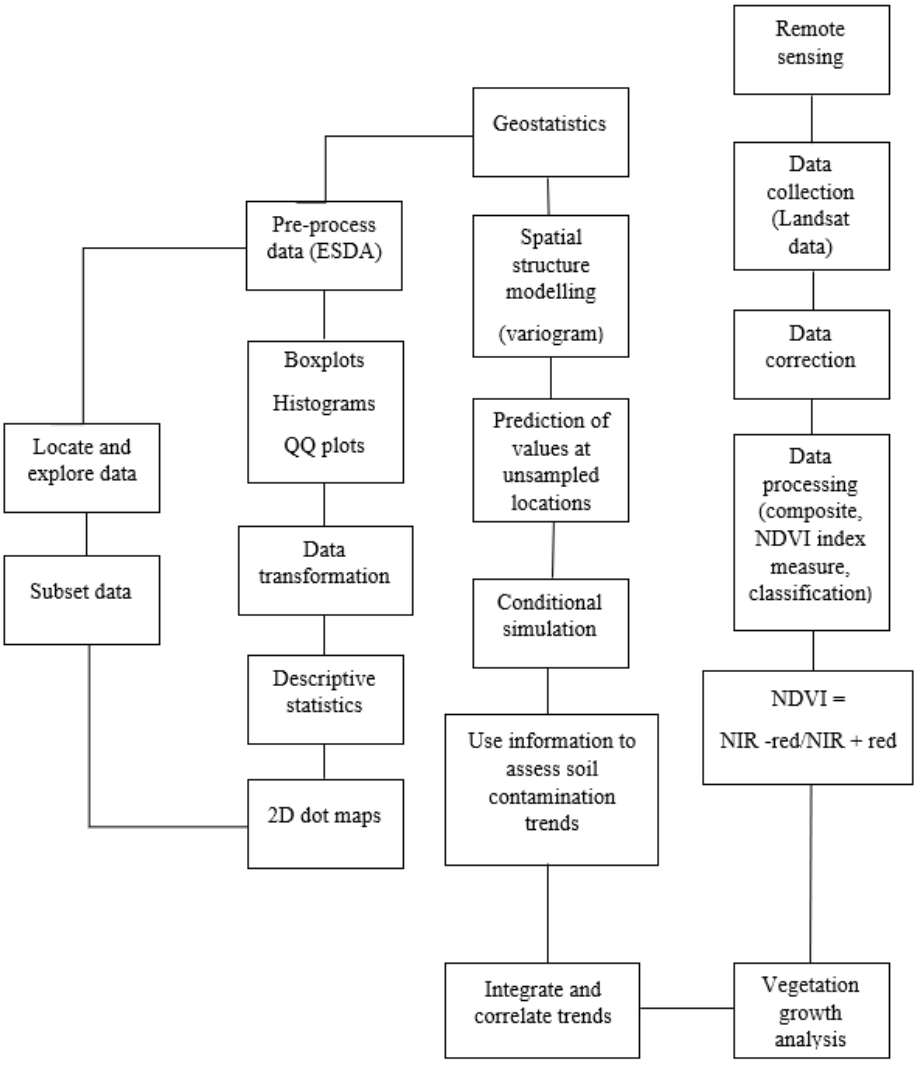


Figure 12: Method workflow implemented to meet the study's objectives.

3.1. Data

Sampling is usually the first step in determining the concentration of pollutants at site locations. Data obtained from the 2001 Sudbury Soil Study (SSS) was acquired from an open-source dataset in a Microsoft excel format, with the data comprising of approximately 8000 samples. The soil samples were analysed for 20 key elements at soil profile depths of 0-5 cm, 5-10 cm and 10-20 cm, the analysis procedure can be found on the (SARA, 2008) report. The concentrations of the metals are given in µg/g (ppm). The data was prepared in an Excel spreadsheet and exported into ArcGIS 10.6.1. and mapped (Figure 14).

The data was subset by soil profile depth in ArcGIS and mapped as 2D dot maps for each COC. Data sub setting can be advantageous as this can help limit time in fitting a model and avoiding numerical errors by using a uniform subset. This assists in saving time for acquiring an accurate model, validation can be easily conducted, effects of clusters can be reduced, and the robustness of the data can be improved (Rennen, 2009). The general intuition is that using more data in geostatistics produces a better model and a large dataset is regarded as a good starting point in building a model (Rennen, 2009), therefore the entire sub-dataset was used for analysis.

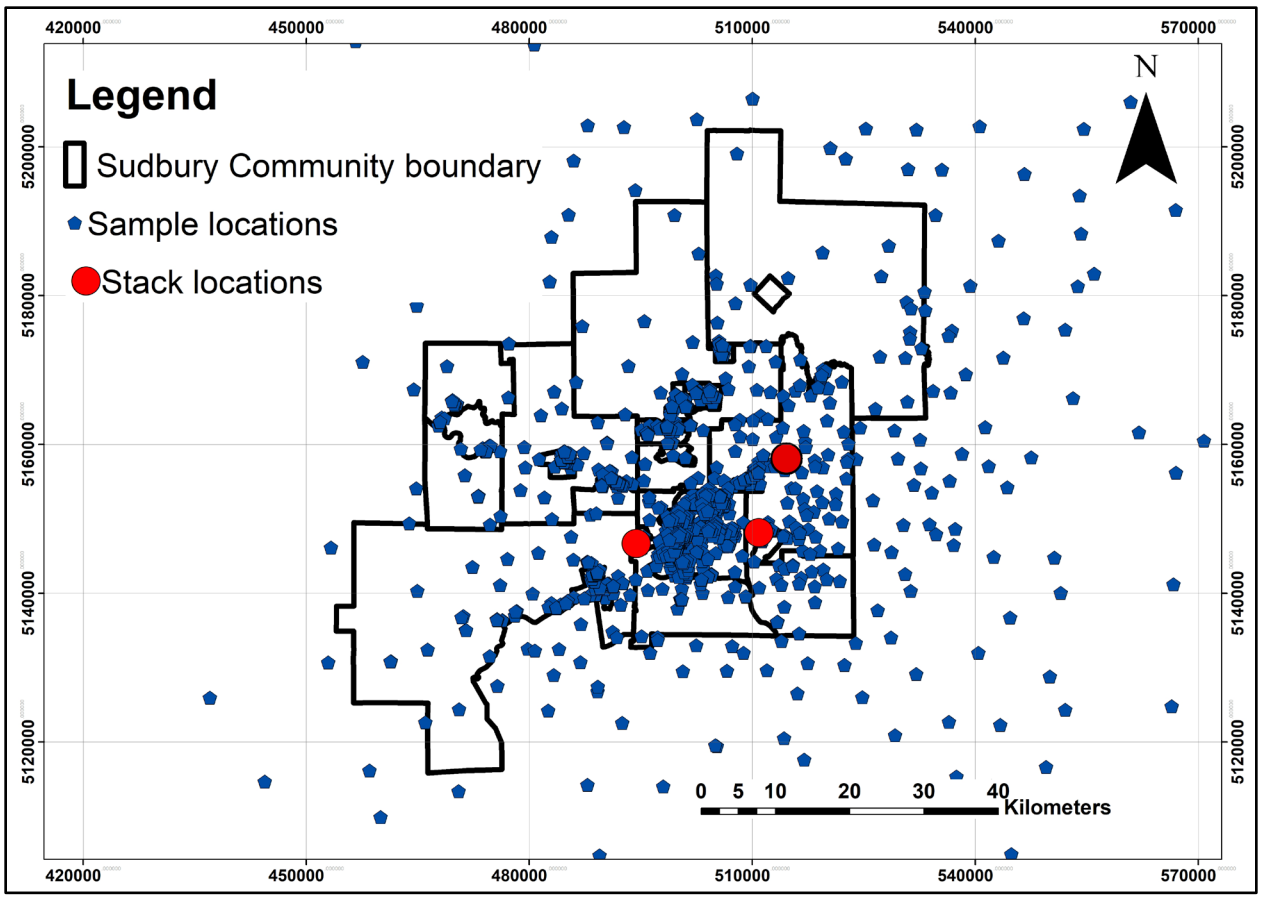


Figure 13: Data point map showing locations where samples were collected.

3.2. Exploratory spatial data analysis

After mapping the data, exploratory spatial data analysis was performed for data exploration to understand the data behaviour and distribution. Box plots, histograms and normal Q-Q plots were produced for all the COC to look at data outliers, assess data distribution, and measure normality. Data outliers that were identified using boxplots were removed as they can vigorously affect the interpolation. Data that was skewed was log transformed, and normal Q-Q plot was used to assess normality. In the normal Q-Q plot, the closer the data points are to the 45-degree line, the closer

the data is to a normal distribution. Scatterplots were used to understand the relationship existing within these metals. This was done to assess similarities in the trends and behavior of the metals.

The trend analysis tool of ArGIS was implemented to study global trends of the attributes (COC). A surface might comprise of different main components, which are a global trend (e.g., topographic influence) that is fixed and short-range variation (also known as short range error, e.g., precipitation); the latter can be modelled in spatial autocorrelation and the nugget effect (defined in section 3.3.1.4.). An existing global trend in the data can be removed using polynomials of different orders (e.g., 1^{st}, 2^{nd}, or 3^{rd} order) or modelled depending on knowledge of the phenomena, with justifications of doing so. This study deconstructed the data into the global trend plus short-range variation, meaning that it was assumed that the global trend is fixed, and the short-range variation is randomly autocorrelated depending on its location in space. The final surfaces produced will be a combination of the global trend and short-range variation. This was done as removing the trend comes with a lot of complications (e.g., choosing the right polynomial, knowledge of the global trend) and due to that modelling sensitive values as soil contamination can often be difficult, so caution should be applied. However, data detrending was also tested as removing the global trend can lead to better visualization of local variations since kriging is carried out on the residuals, although the geostatistical analyst tool adds the trend back to the kriging results obtained from the residuals to rescale the data and produces different results.

3.3. Spatial prediction

Spatial interpolation or prediction refers to the use of values that are at sampled locations to estimate values at unsampled locations. All spatial interpolation methods are based on Tobler's first law of geography that states that 'Everything is related to everything else, but near things

are more related than distant things' (Sui, 2004), this is documented as spatial autocorrelation. Spatial autocorrelation can be positive or negative. A positive spatial autocorrelation entails that the closer that observational units are to each other the more similar they are, the opposite is true for negative spatial autocorrelation. Spatial prediction models are used to measure spatial autocorrelation of a dataset, which is important in understanding how attributes are distributed in space. These models can be classified as (Hengl, 2007): mechanical (deterministic) models (e.g. theissen polygons, inverse distance interpolation, regression on coordinates, natural neighbors, and splines), linear statistical (probability) models (e.g. kriging (plain geostatistics), regression-based, Bayesian- based, regression-kriging, and hybrid models) and expert-based systems (e.g. machine-learning algorithms, hand-drawn maps, and data driven expert based on neural network algorithms). This study makes use of kriging linear statistical models to look at spatial autocorrelation trends of the COC. In depth details on the above mentioned different kinds of interpolation models can be found in Hengl (2007).

3.3.1. Geostatistics

Geostatistics is a linear statistical modeling technique that specializes in the analysis and interpretation of geographically referenced data. It is based on the theory of regionalized variables developed by G.Matheron in honour of the pioneering work of mining engineer D. Krige and statistician H.S. Sichel as applied in the South African mining industry. This formed the first generation of geostatistics known as linear geostatistics. Geostatistics is used with two different meanings, as a collection of all statistical and probabilistic methods that are applied in geosciences and as an alternative name for the theory of regionalized variables (Bàrdossy, 1997). The theory of regionalized variables is based on an introduction of random functions (*RF*).

Introduction of *RF* was such that for each point x there exists a random variable $Z(x)$ such that two points x and y with variables $Z(x)$ and $Z(y)$ are different but not independent (Matheron, G. Kleingeld, 1987). The theory provides a brief and essential methodology for describing and analysing data that is spatially distributed.

Geostatistical methods are based on statistical models that include autocorrelation, which are statistical relationships among the sample data, on top of their capability of producing a prediction surface, they can also provide an accurate measure of the predictions (Johnston *et al.*, 2003). Geostatistical analysis of data occurs in two main ways (1) the variogram or covariance is modelled to examine and model spatial autocorrelation of the data, (2) kriging is applied to predict values at unsampled locations. This means that the data is used twice, (1) to estimate the spatial autocorrelation and (2) to make the predictions (Johnston *et al.*, 2003).

3.3.1.1. Kriging interpolation

Kriging is the most used linear geostatistical method. It is a general term used as a synonym for geostatistical interpolation. Kriging is referred to as the best unbiased linear estimator (*B.L.U.E.*). This is because the estimates are unbiased and have a minimum variance as it is an optimal interpolator, this feature makes it more creditable as compared to other estimation methods (Oliver *et al.*, 1989). Kriging has been compared with various other interpolation methods (e.g., Isaaks and Srivastava, 1989) and has proven to have higher credibility and to be reliable.

Ordinary kriging (*OK*) is a standard version of kriging methods (Hengl, 2007) that was used in this study, as it usually proves to be robust for estimating unknown values at unsampled locations (Mousavifard *et al.*, 2013). *OK* is by far the most implemented kriging type as it provides an

estimate for the entire region around measured data points (Mousavifard *et al.*, 2013) as compared to other methods (e.g. *deterministic methods*). Isaaks and Srivastava (1989) looked at estimates produced by polygonal deterministic and *OK* methods, looking at different statistics (such as true value vs. estimate plots) they found that *OK* estimates had a lower overall and conditional bias as compared to the polygonal ones. It is for this reason, in addition to that kriging can calculate error estimates, and is able to interpolate data sources allowing a better and accurate representation of trends and overall spatial distribution of elements, that kriging is employed in this study. The *OK* weights are estimated by assuming a constant mean (μ) for the data (no trend) using the equation:

$$z(s) = \mu(s) + \varepsilon(s) \qquad \text{Equation 1}$$

Where $z(s)$ is the unknown variable of interest at location s that is decomposed into a deterministic trend $\mu(s)$ and $\varepsilon(s)$ is the form of the random autocorrelated errors with spatial dependence (Johnston *et al.*, 2003). The predictor is developed as a weighted sum of the data defined by:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \qquad \text{Equation 2}$$

Where $Z(s_i)$ is the measured value at location *i,* the unknown weight for the value measured at location *i* is given by λ_i and, s_0 is the location of prediction (Johnston *et al.*, 2003). The weights are calculated by solving a system of linear equations comprising of coefficients that depend only on the variogram and they are not selected based on any rules but on the variation of the function in space (Johnston *et al.*, 2003; Floch, 2013). These weights are also proportional to the

correlation between the estimated and measured points and inversely proportional to the intercorrelation existing between the data points. It is expected that some of the predictions might be above or below the actual values, the difference between these values should be zero on average. Making the predictions unbiased (Johnston *et al.*, 2003).

Some other types of commonly used kriging methods are the *universal* (*UK*) and *simple* (*OK*) kriging methods. Universal kriging assumes a polynomial trend model which can be a linear model, of the type:

$$z(s) = \mu(s) + \varepsilon(s) \qquad \text{Equation 3}$$

Where $z(s)$ is the unknown variable of interest at location s, $\mu(s)$ is some sort of deterministic function and $\varepsilon(s)$ is the form of the random autocorrelated errors with spatial dependence.

Simple kriging assumes stationarity of the first moment within the entire study domain with a known mean, using the model:

$$z(s) = \mu + \varepsilon(s) \qquad \text{Equation 4}$$

Where $z(s)$ is the unknown variable of interest at location s, μ is a known constant mean and $\varepsilon(s)$ is the form of the random autocorrelated errors with spatial dependence. Due to that μ is assumed and known fully, it is also assumed that $\varepsilon(s)$ is also exactly known at the data locations (ESRI, 2020).

3.3.1.2. Spatial dependency, second order stationarity and the intrinsic hypothesis

Geostatistics is solely based on random processes with dependence (Johnston *et al.*, 2003). Spatial dependency (also called spatial autocorrelation) entails that it is impossible to predict

values between data values that are spatially independent (Macleod, 2006). Dependency rules are usually unknown but can become apparent through continued observation of several samples (Johnston *et al.*, 2003). Overlooking dependency might result in incorrect analysis and poor decisions when dealing with spatially dependent data (Macleod, 2006). Statistics relies on the means of the replication to estimate the dependency rules. To obtain the required replication in a spatial setting, stationarity is used which is an assumption required for spatial data wherein it is believed that estimates can be derived with the variation and uncertainty of the estimates understood from repeated observations (Johnston *et al.*, 2003). This replication is provided by similar distances between any two data points (Johnston *et al.*, 2003).

Second order stationarity assumes that the expected value of a random function is constant all through out the area (Bàrdossy, 1997) and the covariance is the same between any two points that are the same distance apart no matter their location, the covariance is dependent on the distance separation between points and not on their coordinate positions. Stationarity does not directly propose second order stationarity, stationarity random functions may sometimes not be second-order stationarity as their first moments might be undefined (Myers, 1989).

The intrinsic hypothesis assumption is slightly weaker than the second order stationarity (Bàrdossy, 1997). It requires the mean and variance to depend strictly on the separation distance between samples and not on the coordinate position of the data, hence distances that are alike between different pairs of data points provide statistical replication (Macleod, 2006).

A *RF* can satisfy the intrinsic hypothesis and not second order stationarity (Myers, 1989). This means that a *RF* can enable a variogram to be calculated but this is not the case for the covariance (Floch, 2013). (The relationship between the covariance function and the variogram is shown in

Figure 15. The *variogram* function is the most frequently applied in the analysis of soils and will be considered for this study. Stationarity is what the *variogram* assumes and is dependent on, if data is in a non stationarity form, data detrending and data transformation should be applied to enable a normal Gaussian distribution (Macleod, 2006). Kriging can then be applied once the dependency and stationarity assumptions are known.

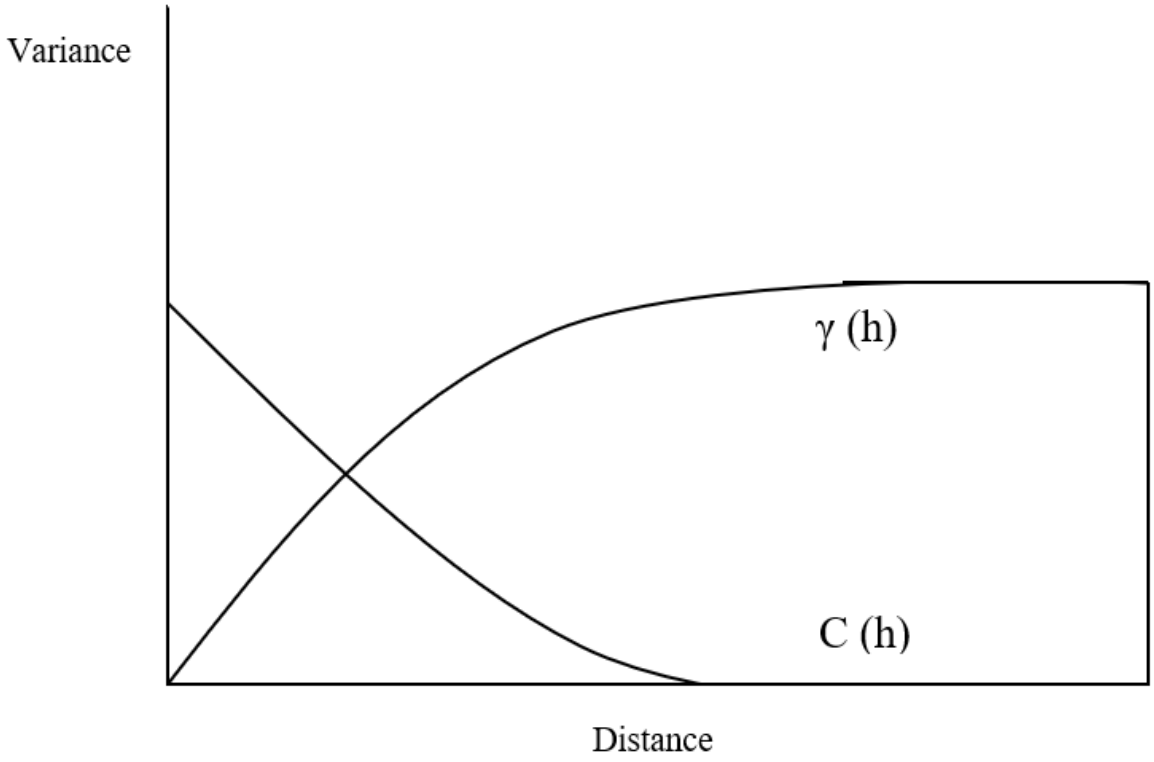


Figure 14: Relationship between the covariance, C (h), and variogram, γ (h), function.

3.3.1.3. Data transformation and normality

The kriging estimator is a weighted average and hence sensitive to few exceptionally large variables and if the data is skewed, modeling of the variogram becomes hard (Macleod, 2006).

The main goal of log transforming data is to have a frequency distribution that is or mimics a gaussian normal distribution, hence decreasing skewness. By doing so, the variance of the data is reduced, and descriptive statistics are improved. If data transformation is not applied, low values might be overestimated while high values might be underestimated, which is one of the draw backs of kriging. Sample data that was heavily skewed (with a skewness of >5) was transformed using a natural logarithm to simplify variography and fulfill the kriging assumptions of data normality.

3.3.1.4. Variogram modelling

The *variogram* is the core of geostatistics, from it the form of the model that is applicable to the phenomena under study, the kriging weights, and standard errors of estimation by kriging can be gathered (Cressie and Hawkins, 1980). The kriging estimates, Gaussian simulation and indicator methods all need a *variogram* for each variable in each domain (Deutsch, 2014). It should be noted that some authors refer to this as the *variogram* (e.g. Leuangthong, McLennan and Deutsch, 2004; Gringarten and Deutsch, 2001; Bàrdossy, 1997, Myers and Journel, 1990; Myers, 1989; Journel, 1989 and 2003, Chilès, 2012; Compositing, Mory and Deutsch, 1989; Matheron and Kleingel, 1987; Floch, 2013) while some refer to it as the *semi-variogram* (e.g. Cambardella *et al.*, 1994; Cressie and Hawkins, 1980; Yamamoto, 2005; Klauberg *et al.*, 2018; Johnston *et al.*, 2003; and Macleod, 2006). The term can also be used interchangeably. For argument's sake, this study will not dwell much on the effective use of the terminology but will use *'variogram'* as the preferred term. The *variogram* is a measure of variability that increases as values become more different (Gringarten and Deutsch, 2001), which is used to quantify spatial autocorrelation

patterns of spatial locations. To further elaborate Equation 5, the *semi-variance* of the *variogram* is used for analysing temporal and spatial correlations representing a measure of the degree of differences between different points or observations in a variogram and is defined by:

$$\hat{\gamma}(h) = \frac{1}{2} \times \frac{1}{n(h)} \sum_{i=1}^{n(h)} (z(x_i + h) - z(x_i))^2 \qquad \text{Equation 5}$$

Where $\hat{\gamma}(h)$ is the variance at lag distance h, $n(h)$ is the number of observation pairs separated by h, $z(x_i)$ is a measured variable at spatial location i, and $z(x_i + h)$ is a measured variable at spatial location $i + h$ (Johnston *et al.*, 2003). The variogram plot (Figure 16) is a graph that shows how the variance $\hat{\gamma}(h)$ varies as the observation distances between points increase (MuhaimeedAmal, 2013) giving estimates of γ(.) at observed values of h (Compositing, Mory and Deutsch, 1989). Only the pairs of lags $\|h\|$ that are apart are needed if the process is isotropic and if the process is anisotropic variogram estimates can be made at different directions through selecting certain directions and averaging the pairs of the lag distances $\|h\|$ that are apart in the chosen direction (Compositing, Mory and Deutsch, 1989).

The variogram allows to estimate graphically various parameters, sill, nugget, range, and partial sill (Figure 16) which are helpful in the comparison of autocorrelation trends. The sill is the value on the *y-axis* at which the graph reaches its asymptote, beyond this distance (which is the range on the *x-axis*), observations are spatially uncorrelated. The nugget is the discontinuity from the origin of the graph (*y-intercept*), a high nugget effect often leads to smoother estimates excluding closer to data locations where artifact discontinuities become more pronounced (Compositing, Mory and Deutsch, 1989). Therefore, the nugget effect value holds more influence when modelling the variogram and can affect the outcome of kriging. The partial sill exists when there

is a nugget effect and is given by the sill value minus the nugget value. If a variogram point cloud shows a random distribution or does not show an increasing $\hat{\gamma}(h)$ on the *y-axis* with increasing separation distance h on the *x-axis*, it shows that there is no positive spatial autocorrelation amongst the data points (MacLeod, 2006), the resulting variogram will be a horizontal line.

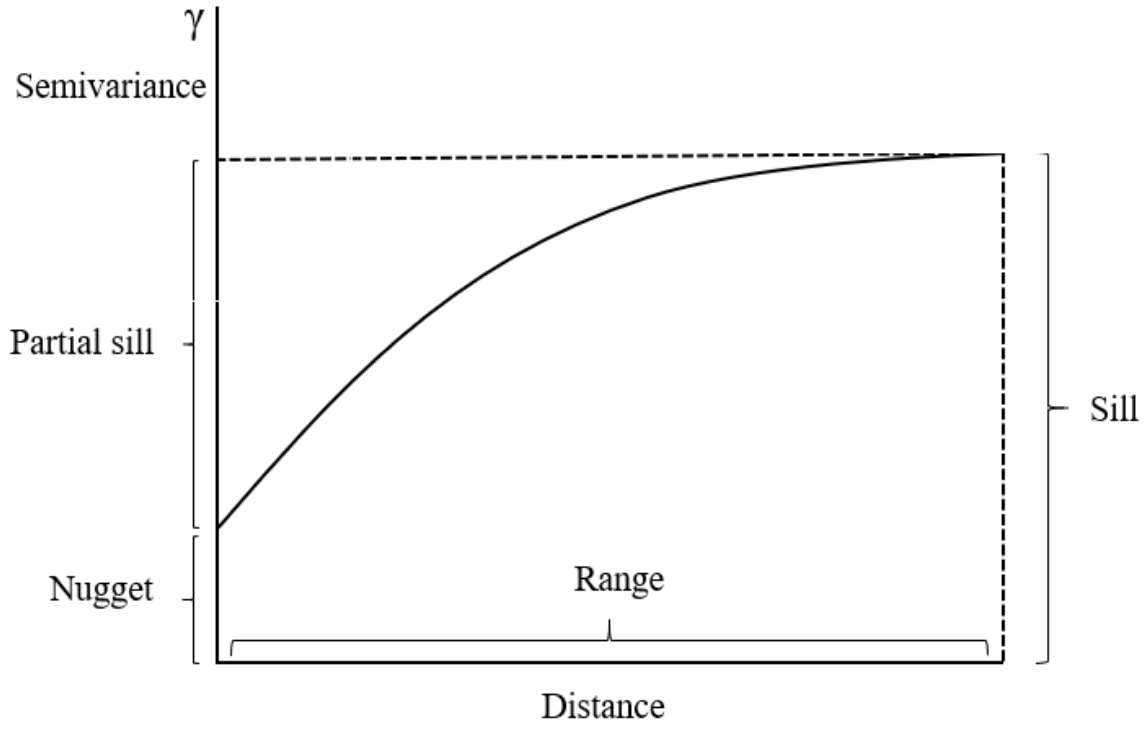


Figure 15: Characteristics of the variogram.

The variogram comprises all of the information on the spatial variation of the attribute under study when the intrinsic hypothesis is obeyed. This allows the *semi-variance* to be estimated from a single realization of the underlying sample (Oliver *et al.*, 2009). If the second order stationarity conditions are all met (i.e., for every distant point the corresponding random variables are

independent) this will produce a variogram that is constant after the sill, known as sill variograms (Floch, 2013). However, if the second order stationarity is not obeyed whilst the intrinsic hypothesis is, variograms without a sill are produced. Such models however still qualify to be called variograms. Some of the important widely used theoretical variogram models are *spherical*, *exponential* and *gaussian*, these are variograms with a sill and can be defined by the following equations (Floch, 2013):

Spherical model:

$$\gamma(\mathrm{h}) = \begin{cases} c_0 + c_s \left[\begin{matrix} 0 \\ \frac{3}{2}\left(\frac{h}{a}\right) - \frac{1}{2}\left(\frac{h}{a}\right)^3 \\ c_0 + c_s \end{matrix} \right] & \begin{matrix} h = 0 \\ 0 < h \leq a \\ h > 0 \end{matrix} \end{cases} \qquad \text{Equation 6}$$

Where: c_0 is the nugget effect, c_s is the partial sill, $c_0 + c_s$ is the sill and a is the range.

The spherical model is the commonly used. This model has a linear behaviour at small separation distances near the origin and flattens out at a larger distance and reaches the sill $c_0 + c_s$ at range a (Isaaks and Srivastava, 1989).

Exponential model:

$$\gamma(\mathrm{h}) = \begin{cases} c_0 + c_s \left[\begin{matrix} 0 \\ 1 - exp\left(-\frac{h}{a}\right) \end{matrix} \right] & \begin{matrix} h = 0 \\ h > 0 \end{matrix} \end{cases} \qquad \text{Equation 7}$$

As the spherical model, the exponential model is also linear at noticeably short distances near the origin but rises more steeply and then flattens out graphically, reaching the sill $c_0 + c_s$ asymptotically at a range a (Isaaks and Srivastava, 1989).

Gaussian model:

$$\gamma(\mathrm{h}) = \begin{Bmatrix} c_0 + c_s \left[1 - exp\left(-\left(\frac{h}{a}\right)^2\right)\right] & h > 0 \\ 0 & h = 0 \end{Bmatrix} \quad \text{Equation 8}$$

The gaussian model is also like the exponential in the sense that it also reaches the sill asymptotically with a range *a,* but different in that it has a parabolic behaviour at the origin and often used to model phenomena that is extremely continuous (Isaaks and Srivastava, 1989).

However, choosing a suitable theoretical model is often challenging, the choice is always a concession. The choice of the variogram model to be used depends wholly on the spatial variability of the analysed natural phenomena which is greatly important in the analyses of soils since they are powerfully variable in space, this permits the determination of uncertainty of the results (Borkowski and Kwiatkowska-Malina, 2017). All three models were tested to determine which model best represents the data by adjusting the empirical variogram to theoretical models with well defined analytical forms through looking at the range, sill, nugget, and partial sill values. This was done to assess models that will better define the kriging equations.

3.3.1.4.1. Directional influences

During variogram modelling, the dependency of the phenomena on direction can also be interpreted, the variogram models can be omnidirectional (isotropic) or directional (anisotropic). The empirical variogram depends only on the separation distance and not on direction in the isotropic case and hence all the directional theoretical variograms will be the same (Isaaks and Srivastava, 1989) for the same attribute under study. In the anisotropic case, the variogram reveals

major changes in the range or sill direction (Isaaks and Srivastava, 1989). A knowledge of the phenomena under study is necessary when deciding on what kind of variogram to use. The current study considered using omnidirectional and directional variograms, as they are easy to control, and directional variograms as atmospheric pollution and metal mobility might be continuous in specific directions, due to impact of natural phenomena such as wind patterns. Directional influences and anisotropy were accounted for when analysing the data as they affect the predictions in the output kriging surface.

3.3.1.4.2. Parameter selection

After data preparations and specific transformations, an empirical variogram is calculated in various directions, which is a graph of the averaged variogram values (*y-axis*) and distance (*x-axis*), and different parameters are selected for modeling. The direction of the variogram (*figure 18*A) was defined by directional trends based on the knowledge of trends previously reported for the elements under study (e.g. SARA Group, 2008). The lag parameters and tolerance (*figure 18*b) were assigned based on the average of a nearest neighbour estimate of sample pair distances calculated in ArcGIS. The angle tolerances and bandwidths consider variability that exists in calculating directions and the sample configuration (Figure 17C) (Deutsch, 2014), these parameters were selected based on the knowledge of sampling space between samples. The final variogram (Figure 17D) shows a combination of all the above listed parameters.

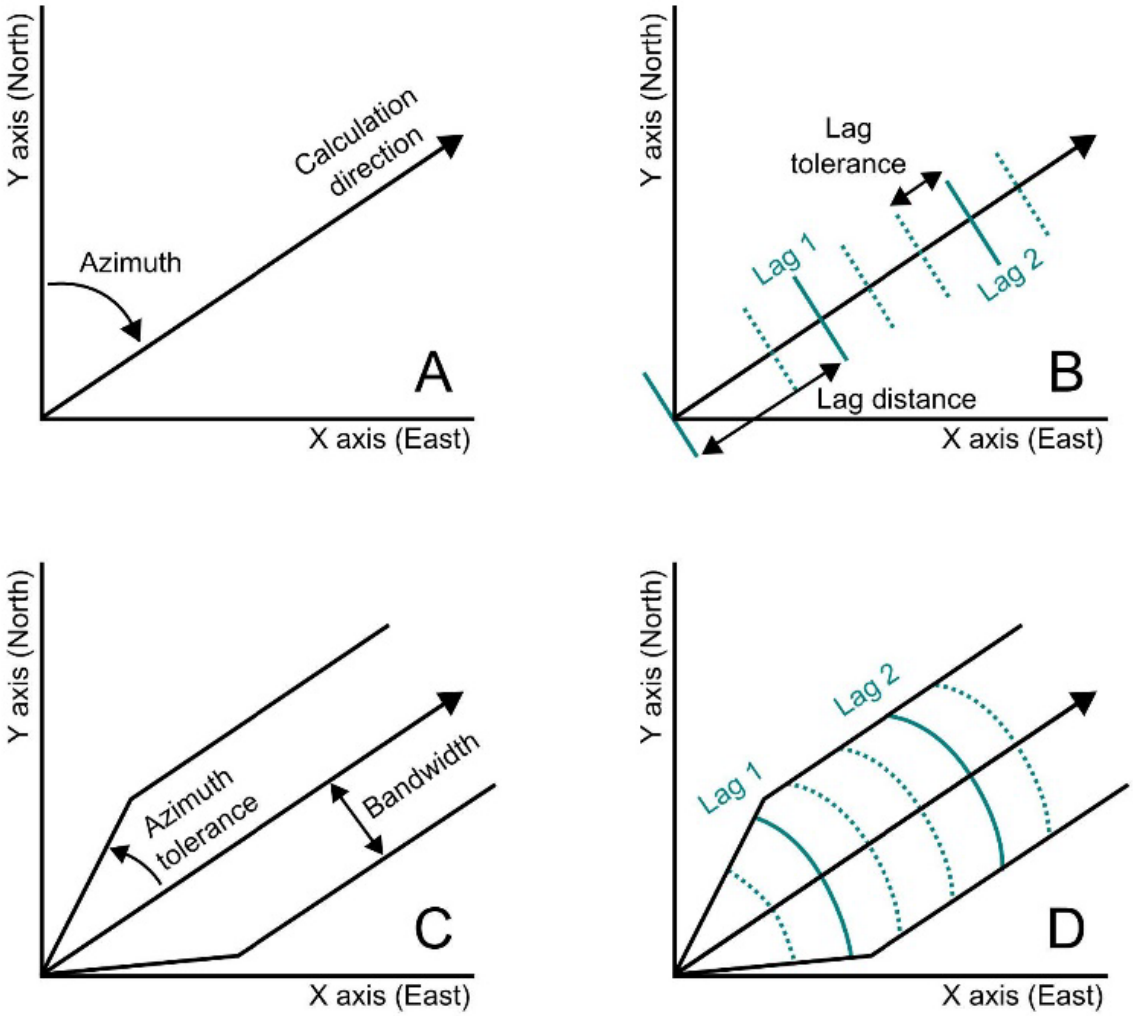


Figure 16: A sketch showing different tolerance parameters of the variogram. The parameters are sketched in 2D but the angle of tolerance and bandwidth are reported in 3D from Deutsch (2014).

3.3.1.5. Data validation

A cross validation technique was used to check the results of different modeling approaches to choose the one that performed better. This technique allows the comparison of estimated and true values by assisting one to choose between different weighting procedures, neighborhood search strategies and variogram models (Isaaks and Srivastava, 1989). The method makes use of all the data set, to estimate the trend and autocorrelation models of the kriging estimates by removing

data points one at a time and using the rest of the data to make an estimate of that vacant data value.

Sample cross-validation residuals have important spatial information that help in making informed decisions on the type of model providing the best *OK* predictions, it was used to assess the accuracy of the resulting surfaces produced and to make corrections and adjust variogram parameters. A careful study of the spatial distribution of cross validation residuals with the focus on the entire goals of the estimation exercise can give insights leading to case specific improvements of the estimation procedure, making cross validation a useful preliminary step before final estimates can be calculated (Isaaks and Srivastava, 1989).

Validation is done using calculated statistics and follows Johnston et al. (2003) recommendations: for an accurate method the Mean Error (*ME*) must be close to zero, Root Mean Square Error (*RMSE*) and Average Standardised Error (*ASE*) should be as small as possible and Root Mean Square Standardised Error (*RMSSE*) should be close to 1; if the predictions are unbiased, the *ME* should be close to zero and if standard errors are accurate the *RMSSE* should be close to 1, with the *RMSSE* small when the predictions are close to the measured values. The summary statistics are obtained using the ArcGIS geostatistical analyst tool. They are mathematically written as (Johnston *et al.*, 2003):

$$\text{ME} = \frac{\sum_{i=1}^{n}(\hat{Z}(s_i) - z(s_i))}{n} \qquad \text{Equation 9}$$

$$\text{RMSE} = \sqrt{\frac{\left(\sum_{i=1}^{n}(\hat{Z}(s_i) - z(s_i)\right)^2}{n}} \qquad \text{Equation 10}$$

$$\text{ASE} = \sqrt{\frac{\sum_{i=1}^{n} \hat{\sigma}\ (s_i)}{n}} \qquad \text{Equation 11}$$

$$\text{MSE} = \frac{\sum_{i=1}^{n} \left(\hat{Z}(s_i) - z(s_i)\right)/\hat{\sigma}\ (s_i)}{n} \qquad \text{Equation 12}$$

$$\text{RMSSE} = \sqrt{\frac{\sum_{i=1}^{n} [\left(\hat{Z}(s_i) - z(s_i)\right)/\hat{\sigma}\ (s_i)]^2}{n}} \qquad \text{Equation 13}$$

Where: $\hat{Z}(s_i)$ is the predicted value obtained from cross validation, $z(s_i)$ is the observed value and $\hat{\sigma}\ (s_i)$ is the prediction standard error for location s_i.

Cross validation also has its own limitations. An example is that pairs of true and estimated values can be generated only at sample locations, resulting in an inaccurate reflection of the actual performance of an estimation method as the estimation at sample locations is typically not representative of estimation at all of the unsampled locations (Isaaks and Srivastava, 1989).

3.3.1.6. Gaussian conditional simulation

Kriging algorithms are preferred due to their ability of estimating data accounting for its spatial variability; however, such methods sometimes overly smoothen the data resulting in a misrepresentation of the exact behavior of the attribute under study (Leuangthong, McLennan and Deutsch, 2004). It is for this reason that interpolated maps should not be used when applied to cases sensitive extreme values and patterns of continuity such as soil pollution data and soil physical properties (Goovaerts, 1999). Simulation can assist in overcoming such a problem due its ability to improve heterogeneity characterization, while reproducing the spatial variability that was

modelled from the data. Geostatistical simulation uses the basis of kriging algorithms (Equations 14, 15) and retains most of the positive attributes of kriging, which are, reproducing the exact data and use of spatial correlations existing between data (Leuangthong, McLennan and Deutsch, 2004).

Accounting for variability between the simulated locations assists in overcoming the kriging smoothing effect. The drawback of conditional simulation is that it provides multiple equiprobable images of the phenomena or attribute under study and choosing a single representative image becomes a problem (Yamamoto, 2005). The least attractive feature about conditional simulation is that the stochastic realizations are not error free, with the errors being larger than those of kriging estimates, however, the best representative image can still be chosen based on certain criteria (outlined below).

The ability to condition the simulation assists in selecting the best possible realizations of the simulation that reproduces the experimental data set (Torcal *et al.*, 1999). To achieve this reproduction, the kriging equation is incorporated (Torcal *et al.*, 1999):

$$y_{cs} = y_{cs}(x) + [y_k^*(x) - y_{sk}^*(x)], \quad \text{Equation 14}$$

$$y_{cs}(x) = y_s(x) + \sum_{i=1}^{N} \lambda_i [y(x_i) - y_s(x_i)\ ], \quad \text{Equation 15}$$

Where:

$y_{cs}(x)$ = the conditionally simulated datum,

$y_s(x)$ = the non-conditionally simulated datum,

$y(x)$ = the normalized experimental datum,

$y_k^*(x)$ = the value estimated by kriging from the datum $y(x)$,

$y_{sk}^*(x)$ = the value estimated by kriging from the datum $y_s(x)$,

λ_i = the measured weights of the kriging system, and

N = the number of nearby points that will be considered in the kriging.

The Gaussian conditional geostatistical simulation was applied to the log-transformed dataset using the workflow modified from (Klauberg *et al.*, 2018):

(1) The data distribution of the attribute under study was transformed to normal, the trend was removed, data was declustered and normal score transformation was applied
(2) Simple kriging was then applied to the data set in which a variogram model was fitted with the search neighborhood defined
(3) The conditional simulation was performed 100 times using the simple kriging as the input raster and the original data was used as the conditioning field.
(4) When the conditional simulations were applied the output rasters were generated cell by cell as mean and standard deviation that are calculated across all the 100 realizations.

3.3.1.6.1. Accuracy assessment

Once the conditioned Gaussian values are obtained, the model was checked for accuracy as this is an especially important factor in the decision-making process. Accuracy of the simulated maps refers to their ability to replicate data and statistics from the original input dataset with reasonable bias based on the knowledge of the phenomena under study. The accuracy of the simulated maps was compared with the original kriging maps and tested for truthfulness. Stochastic simulation

does not focus on minimization of the error variance unlike kriging, its focus is solely on the reproduction of statistics as the sample histogram, variogram model and honoring the data values (Goovaerts, 1999). To confirm the statistic reproduction, consistency and accuracy of the models can be checked by assessing that the simulated models (Leuangthong, McLennan and Deutsch, 2004):

(1) Reproduce data values at their exact location,

(2) The distribution of the attribute of interest must be reproduced, and

(3) The spatial continuity characterized by the variogram model must be the same as that of the original data.

Original input data was compared with gaussian conditional simulation results of the realizations by assessing scatterplots of the original vs. simulated data to verify data reproduction, reproduction of the histogram and Q-Q plot shape by correlating summary statistics (mean, median and variance) to check for the reproduction of the distribution of the data and the variogram reproduction. This checks must only be applied in normal or transformed space for gaussian simulation as only the normal scores variogram is imposed directly (Leuangthong, McLennan and Deutsch, 2004). The variogram was calculated for multiple realizations and compared to the input variogram in similar directions. Any departures from these expectations may be a cause for concern (Leuangthong, McLennan and Deutsch, 2004); however, any ergodic fluctuations should reasonable and unbiased. The values that are obtained from the geostatistical simulation should be aligned with the experimental (original) information and should reproduce the observed variability (Torcal *et al.*, 1999).

3.4. Remote sensing applications

This study made use of Landsat images collected from heavy industrial years (1970s) to present times of limited heavy emissions (2019) using ArcGIS 10.6.1. software. The Landsat data was acquired from the USGS Earth Explorer and GloVIS open-source websites. The Landsat satellites provide easy access to a large chain of continuous records of satellite-based observations of the Earth. Landsat serves as an invaluable resource for monitoring global change and is also a primary source of medium spatial resolution Earth observations used in decision making processes (Chander, Markham and Helder, 2009). The Landsat program began around 1972 and now has 4 satellite groups which are: Group 1; Landsat 1 (L1) and Landsat 2 (L2) carrying the Multispectral Sensor (MSS); group 2, Landsat 5 (L5) carrying the Thematic Mapper Sensor (TM); group 3, Landsat 6 that carried the Thematic Mapper sensor (TM) (Landsat 6 failed upon launch, Chander, Markham and Helder, 2009), Landsat 7 (L7) carrying the Enhanced Thematic Mapper plus sensor (ETM+) and group 4, which is Landsat 8 (L8) carrying the Operational Land Imager (OLI) and the Thermal Infrared sensor (TIRS). Landsat 9 is to be launched in mid-2021, with a higher image capacity than past the Landsat's. A total of four cloud free (or low cloud) Landsat images covering the study area, or some parts of the study area were available for the period between 1979 and 2019 corresponding to the areas previously heavily affected by smelting activities. The images were used to calculate the NDVI and classify the land observation around the smelting regions and away from them.

Table 3: Different Landsat satellites used for image processing. The image acquisition date, scene ID and Image path are reported.

Satellite and sensor	Acquisition date	Scene ID	WRS path	WRS row
L2-MSS	1979-06-03	LM20210281979154PAC03	21	28
L7-ETM+	2000-08-27	LE70190282000240EDC00	19	28
L8-OLI	2013-09-24	LC80190282013267LGN01	19	28
L8-OLI	2019-07-07	LC80190282019188LGN00	19	28

3.4.1. Image pre-processing

The ease of detecting and quantifying changes in the Earth's environment highly depends on the sensors that can provide known accuracy, precision and consistence in measurements of the Earth through time (Chander, Markham and Helder, 2009). This makes radiometric calibration of data a pre-requisite for creating data that has high quality and is consistent with no artifacts. The digital raw numbers of the Landsat Level 1 images were converted to spectral radiance and spectral reflectance, to overcome product artifacts. These conversions were conducted as the images received were uncalibrated and to correct for the sun angle and atmospheric effects to have a better representation of the spectral response of objects. The values used in the conversion are

found in the MTL document of the Landsat data and the referenced authors. Conversion of DN values to radians and reflectance for the MSS, TM and ETM+ sensors was calculated based on the methods outlined by Chander, Markham and Helder (2009):

DN values to radiance

$$L_\lambda = (\frac{L_{MAX_\lambda} - L_{MIN_\lambda}}{Q_{Cal_{max}} - Q_{Cal_{min}}})(Q_{cal} - Q_{Cal_{min}} + L_{MIN_\lambda}) \quad \text{Equation 16}$$

Where:

L_λ = Spectral radiance at the sensor's aperture $[W/(m^2 Sr\ \mu m)]$

Q_{cal} = Quantized calibrated pixel value [DN]

$Q_{Cal_{min}}$ = Minimum quantized calibrated pixel value corresponding to L_{MIN_λ}

$Q_{Cal_{max}}$ = Maximum quantized calibrated pixel value corresponding to L_{MAX_λ}

L_{MIN_λ} = Spectral at-sensor radiance that is scaled to $Q_{Cal_{min}}[W/(m^2 Sr\ \mu m)]$

L_{MAX_λ} = Spectral at-sensor radiance that is scaled to $Q_{Cal_{max}}[W/(m^2 Sr\ \mu m)]$

DN values to reflectance

$$\rho_\lambda = \frac{\pi * L_\lambda * d^2}{E_{sun_\lambda} * cos\theta_s} \quad \text{Equation 17}$$

Where:

ρ_λ= Planetary TOA reflectance [unitless]

π = Mathematical constant ~3.14159 [unitless]

L_λ = Spectral radiance at the sensor's aperture $[W/(m^2 Sr\ \mu m)]$

d = Earth-Sun distance [astronomical units]

The OLI sensor DN values were directly converted to top of atmospheric reflectance without calculating the radiance based on equations from (Department of the Interior U.S. Geological Survey, 2016) Landsat 8 handbook:

$$\rho_\lambda = \frac{M\rho * Q_{cal} + A_p}{Sin\theta_{SE}} \qquad \text{Equation 18}$$

Where:

ρ_λ = TOA planetary reflectance

$M\rho$ = Reflectance multiplicative scaling factor for the band (REFLECTANCE_MULT_BAND_n from the metadata).

A_ρ = Reflectance additive scaling factor for the band (REFLECTANCE_ADD_BAND_N from the meta data)

Q_{cal}= Level 1-pixel value in DN

$Sin\theta_{SE}$ = Local sun elevation angle

3.4.2. Image processing

3.4.2.1. Band composite

Multiple bands were used to create single raster datasets. Different band combinations are used for different sensors. Various composites are commonly viewed to assess vegetation distribution,

barren regions, geology, and water bodies. An example is the 4, 3, 2 band combination for natural colour image of L8.

3.4.2.2. Band ratioing

Band ratioing techniques are used to modify spectral differences that exist within bands to reduce any shadow effects that might be caused by topographic artifacts (Rajendran, Al-Sayigh and Al-Awadhi, 2016). Band ratio images were constructed by dividing the NIR band by the RED and the GREEN band. The NIR band is highly reflective such that the acquired band ratios show vegetated regions by bright pixels because the shadow effects are now greatly reduced on topography. The spatial distribution of the vegetation was then assessed, taking note of regions with little-moderate or no vegetation into consideration.

3.4.2.3. NDVI

Vegetation cover analysis was applied as a biophysical indication to locate regions with low and high vegetation cover, hence giving an idea of how metal contamination levels have or are affecting vegetation recovery trends. NDVI is a simple and straightforward numerical indicator that defines vegetation distribution based on characteristics of reflectance patterns of green vegetation (Gandhi *et al.*, 2015). The values obtained are used to assess and differentiate between regions with live vegetation, semi-barren and barren. Meaning that, NDVI measures the level of chlorophyll that is being reflected off the surface or region of interest by calculating the difference between the NIR and the RED band. Larger values are indicative of greater chlorophyll density,

such that the level of greenness that is shown is equal to the chlorophyll concentration (Gandhi *et al.*, 2015). NDVI values closest to zero are indicative of bare soil or water bodies, while values that are 0.1 and below generally correspond to barren areas of rock outcrops, snow, or sand. Values that are moderate (0.2-0.3) indicate low to moderate vegetation cover, representing shrub and grassland. High values that are between 0.6 and 0.8 indicates high vegetation density which may be temperate or rain forests (Gandhi *et al.*, 2015).

The NDVI is calculated as:

$$\text{NDVI} = (\text{NIR} - \text{RED})/(\text{NIR} + \text{RED}), \text{ the resulting values are between -1 and 1} \qquad \text{Equation 19}$$

Where RED is the visible red reflectance, with a wavelength of 600-700 nm, and NIR is the near infrared reflectance, with a wavelength of 750-1300 nm. NDVI takes the difference between the NIR and RED band and normalizes it to balance out uneven illumination effects that might occur, such as clouds or hills (Gandhi *et al.*, 2015).

2.2.4. Integration of kriging with remote sensing techniques

Integrating geostatistical and remote sensing approaches is an effective way in incorporating spatial information into mapping, to improve and better understand prediction quality of estimates. A study by López-Granados *et al.* (2005) proved this to be true. They made a comparison of various prediction methods (linear regression, ordinary kriging, and simple kriging) for mapping various soil properties (texture, organic matter, pH, phosphorus and potassium) by a means of incorporating secondary spatial information into mapping for precision

farming in Caracol (Seville, Andalusia, southern Spain). The primary information was from intensive grid sampling and the secondary spatial information from an aerial colour photograph of bare soil. To improve the linear regression prediction maps, the soil variables were interpolated using the ordinary kriging method, simple kriging was also employed on some specific soils. The prediction performances were later compared using the Mean Square Error (MSE). The poorest predictions were depicted for simple linear regression techniques as there was a low correlation between soil attributes and spectral values as compared to the kriging methods. The study results showed that when sparse and expensive soil measurement methods are combined with secondary information such as remotely sensed data from aerial photographs the geostatistical techniques become adequate to accurately map soil properties. The obtained results could also be used to suggest sampling intensity for future surveys in the area. The integration of kriged estimates and digital data from aerial photography were found to be accurate enough to improve the soil management zones. It was also suggested that the methodology can be used when mapping more permanent soil physical properties due to the effectives found.

This kind of approach will assist when integrating geostatistical and remote sensing techniques in this study for soil contamination and vegetation change, although the workflow adopted varies significantly.

Chapter 4: Results and discussion

Introduction

The overall results of the present study are being presented. Firstly, the data distribution and variograms are discussed. The metal variability with soil profile depth (0-5, 5-10 and 10-20 cm) is being presented in the form of geostatistical prediction and simulation maps to show observations and validations. Statistical variability of the metals is also being presented. NDVI images from pre-industrial years to present times is being used to assess vegetation distribution and further integrated with geochemical maps. To correlate and understand the uncertainty of the metals, wind patterns, vegetation, bedrock geology and metal mobility is briefly assessed and discussed.

4.1. Exploratory analyses and data transformation

The data obtained from the 2001 SSS comprised of 3870, 2670, and 2592 samples for 0-5, 5-10 and 10-20 cm soil depth profiles, respectively. The descriptive statistics (Table 4) were calculated for each depth to detect and analyze patterns of regularity. The table shows essential statistical characteristics that are vital in understanding the metal distribution.

Observing the descriptive statistics, all elements are not normally distributed, which is a common case for contaminated soils. These is indicated by the fact that the mean, mode, and median are not all equal, which is one characteristic feature of a normal distribution. However, the mean and median for Co, As, Pb and Se are relatively close to each other. The coefficients of skewness and kurtosis for all the elements is higher than that of a normal distribution. The metals are skewed to the right, with a skewness greater than 0.5. Extreme values also exist within the entire dataset. The maximum values of all elements at all three depths are greater than their mean values. This makes it necessary to exercise caution when calculating the average value used in the site assessment, as

any additional high-concentration samples may push the mean to the right hence misleading decision makers (Ersoy, Yunsel and Atici; 2006). The maximum values also show some sort of contamination in these soils, as the mean of all the elements is greater than what was found for the Sudbury area by the MOE, with exceptions of Se (Table 5).

The histogram, boxplot, and normal Q-Q plot of Ni, as a chosen example for presentation purposes, are illustrated in figure 29 (the rest of the elements are in the appendix). The raw data histogram confirms further what was observed earlier by the descriptive statistics, that the data is skewed to the right, this is observed for all the elements at all depths. This can be clearly seen by the departure from the symmetrical shape of a normal distribution. The histograms also further confirms that all elements contain some level of contamination. The normal QQ plots confirms this kind of distribution whereby the values can be seen deviating from the 45˚ line.

As a result of the data skewness, normal logarithmic transformation was applied to the data (Figure 18). The transformed data shows an improvement in normality. Ni, Cu, Pb and Co show the greatest improvement and similar patterns, with their mean and median being relatively close to each other. As and Se still represent a positively skewed distribution even after transformation. This could be due to that some of their values are extremely low. Although Ordinary kriging is referred to as the Best Linear Unbiased Estimator, even if the data is not normally distributed, the transformation was necessary to make it the best predictor and provide the best estimates, while making variogram modeling easy.

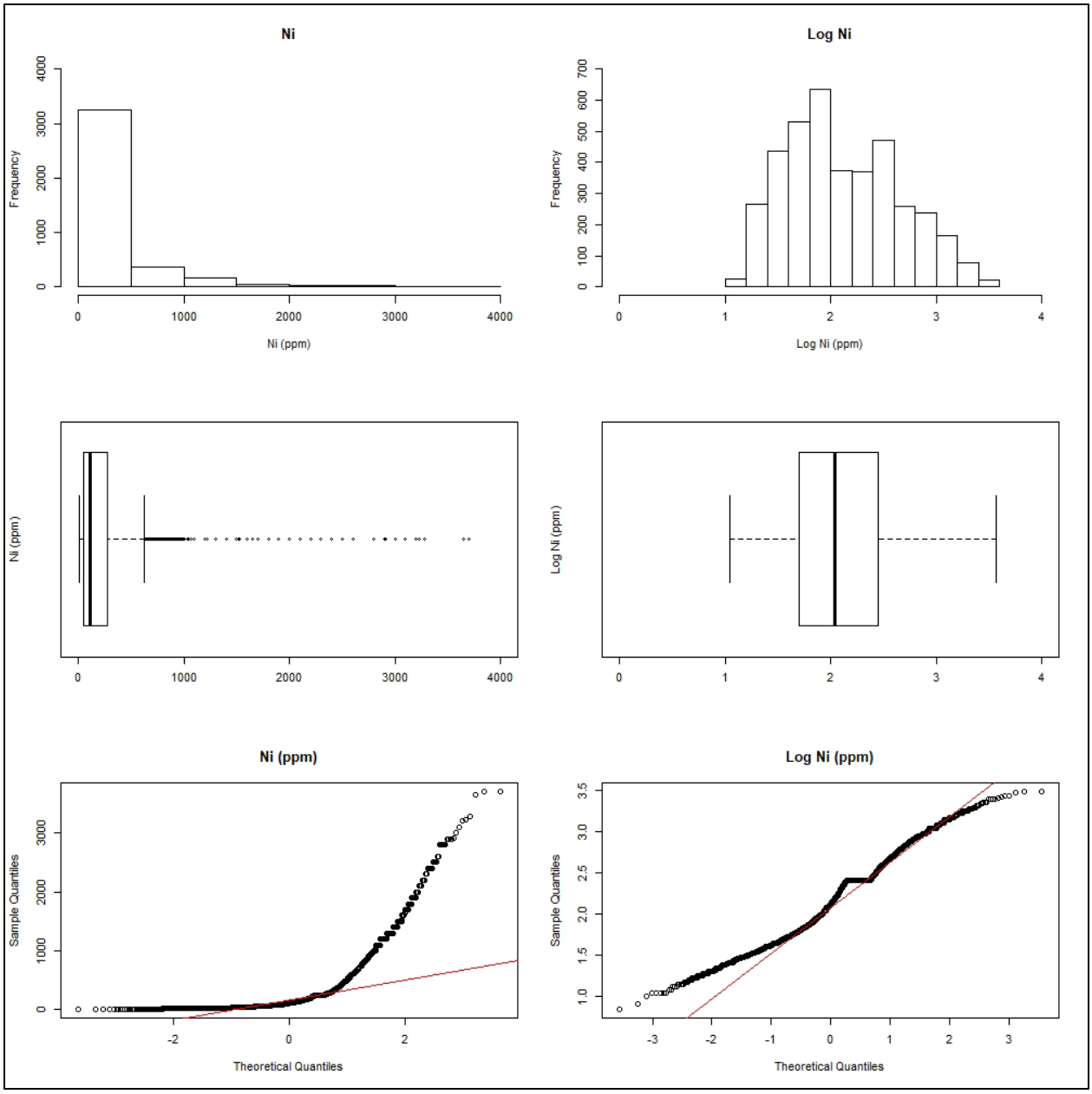


Figure 17: Histogram, boxplot and normal QQ plot for Ni at 0-5 cm.

Table 4: Descriptive statistics of the COC at 0-5, 5-10 and, 10-20 cm. All concentrations are given in ppm. The values on the left of the variable are the raw data and those on the right are the log transformed data. MOE represents the mean background concentrations that have been measured for the Sudbury region by the Ontario Ministry of the Environment.

	Variable	As		Co		Cu		Ni		Pb		Se	
0-5 cm													
	Mean	13.22	0.82	15.36	1.03	297.00	2.05	285.17	2.11	37.38	1.33	1.56	0.02
	Standard Error	0.42	0.01	0.28	0.01	8.53	0.01	7.05	0.01	0.71	0.01	0.05	0.01
	Median	6.00	0.78	9.00	0.95	89.00	1.95	110.00	2.04	22.00	1.34	0.00	-0.30
	Mode	2.00	0.40	6.00	0.78	250.00	2.40	255.00	2.41	34.00	1.53	0.00	-0.30
	Standard Deviation	25.88	0.46	17.35	0.33	530.69	0.59	438.83	0.53	44.19	0.48	2.98	0.40
	Sample Variance	669.63	0.21	301.14	0.11	281636.41	0.35	192569.25	0.28	1952.65	0.23	8.86	0.16
	Kurtosis	48.30	0.14	14.81	0.61	21.59	-0.60	12.41	-0.63	13.36	-0.64	41.99	0.22
	Skewness	5.85	0.90	3.36	0.86	3.97	0.42	3.14	0.41	2.97	-0.07	4.89	1.05
	Range	398.00	2.20	188.00	1.98	5595.00	3.03	3689.00	2.53	409.00	2.61	49.00	1.99
	Minimum	2.00	0.40	2.00	0.30	5.00	0.72	11.00	1.04	1.00	0.00	0.00	-0.30
	Maximum	400.00	2.60	190.00	2.28	5600.00	3.75	3700.00	3.57	410.00	2.61	49.00	1.69
	Count	3870.00	3870.00	3870.00	3870.00	3870.00	3870.00	3870.00	3870.00	3870.00	3870.00	3870.00	3870.00
5-10 cm													
	Mean	19.36	0.95	13.97	1.01	260.45	2.10	274.11	2.15	34.07	1.32	1.30	0.01
	Standard Error	0.83	0.01	0.28	0.01	7.03	0.01	7.26	0.01	0.81	0.01	0.03	0.01
	Median	9.00	0.95	10.00	1.00	120.00	2.08	130.00	2.11	18.00	1.26	1.00	0.00
	Mode	2.00	0.40	14.00	1.14	250.00	2.40	255.00	2.41	34.00	1.53	0.00	-0.30
	Standard Deviation	42.83	0.48	14.60	0.32	363.21	0.54	375.22	0.50	42.04	0.40	1.73	0.34
	Sample Variance	1834.25	0.23	213.09	0.10	131922.06	0.29	140790.59	0.25	1767.40	0.16	2.98	0.12
	Kurtosis	47.45	0.48	22.37	0.62	11.09	-0.70	12.22	-0.68	11.16	-0.26	6.67	-0.32
	Skewness	6.10	0.72	3.97	0.51	2.95	0.11	3.06	0.24	3.01	0.47	2.13	0.79
	Range	568.00	2.36	149.00	2.18	2995.00	2.81	3093.00	2.65	339.00	2.53	14.00	1.45
	Minimum	2.00	0.40	1.00	0.00	5.00	0.67	7.00	0.85	1.00	0.00	0.00	-0.30
	Maximum	570.00	2.76	150.00	2.18	3000.00	3.48	3100.00	3.49	340.00	2.53	14.00	1.15
	Count	2670.00	2670.00	2670.00	2670.00	2670.00	2670.00	2670.00	2670.00	2670.00	2670.00	2670.00	2670.00
10-20 cm													
	Mean	16.95	0.91	11.98	1.00	199.34	2.00	219.33	2.10	32.57	1.27	1.03	-0.04
	Standard Error	0.71	0.01	0.17	0.00	4.80	0.01	4.98	0.01	0.93	0.01	0.03	0.01
	Median	8.00	0.90	10.00	1.00	110.00	2.04	130.00	2.11	18.00	1.26	0.00	-0.30
	Mode	2.00	0.40	14.00	1.14	250.00	2.40	255.00	2.41	34.00	1.53	0.00	-0.30
	Standard Deviation	36.32	0.47	8.72	0.25	244.21	0.54	253.47	0.48	47.21	0.44	1.38	0.31
	Sample Variance	1319.14	0.22	75.97	0.06	59639.30	0.29	64246.64	0.23	2228.39	0.19	1.89	0.09
	Kurtosis	68.41	0.19	21.16	0.10	8.87	-0.73	10.30	-0.94	48.40	-0.38	4.91	-0.30
	Skewness	6.92	0.71	3.32	0.37	2.57	-0.15	2.62	0.06	5.22	0.42	1.79	0.85
	Range	618.00	2.39	108.00	1.74	1997.00	2.87	2490.00	2.40	788.00	2.60	11.00	1.34
	Minimum	2.00	0.40	2.00	0.30	3.00	0.43	10.00	1.00	2.00	0.30	0.00	-0.30
	Maximum	620.00	2.79	110.00	2.04	2000.00	3.30	2500.00	3.40	790.00	2.90	11.00	1.04
	Count	2592.00	2592.00	2592.00	2592.00	2592.00	2592.00	2592.00	2592.00	2592.00	2592.00	2592.00	2592.00
MOE	Mean	17.00		21.00		85.00		43.00		120.00		1.90	

4.2. Variogram analysis

Variogram analyses was performed on the transformed data of all elements. The variogram graphs show binned points as red circles (●) and are generated through binning empirical variogram points by making use of square cells that are one lag wide. The average points (✚) are produced through grouping empirical variogram points falling within angular sectors, while the fitted theoretical variogram model is shown by the blue line. Visual inspection of the empirical semivariograms showed that the gaussian (Figure 26, 30) and spherical models would be suitable models to be fitted. The various features and parameters of the variogram are presented in Table 5.

The theoretical omnidirectional variogram models of the variables poses an exceptionally low nugget effect of less than 1 (Figure 21). The lower nugget effect implies that the spatial model or the variograms can influence the estimation of the variables and assist in limiting data over smoothing, the sampling density is satisfactory to reveal spatial characteristics of the dataset. The existence of the nugget in the data set indicates that short scale randomness or noise in the variables of interest have been accounted for. There is no exact pattern in this nugget effect at various depths. The variograms show positive spatial autocorrelation which is represented by the existence of the sill in all of them. The range of influence for As and Co are the highest being 12,770.5 m and 10,420.654 m for the 0-5 cm, respectively, but decreases with an increase in soil profile depth. Pb shows the lowest range of influence of 80.663 m at 0-5 cm depth increasing with an increase in depth to 6,774.037 m at 10-20 cm. The value of the range is dependent on the lag distance applied. The correlation between the predicted and measured values is also shown (Figure 19, 20), with a correlation of greater than 0.5, this shows that the kriging method is performing accurately.

Directional sample variograms were also tested out (Figure 21*)*. These models revealed changes of the range with direction implying the existence of geometric anisotropy. The directional variograms revealed that the data is anisotropic, as it varies with direction. Directional variograms plotted for the direction of maximum continuity and minimum continuity, which are 6˚ and 96˚ respectively, for Ni at 0-5 cm are shown in figure 32. At 6˚, the variogram slowly levels off indicating positive spatial autocorrelation, while at 96˚the variogram rapidly levels off showing weak or negative spatial autocorrelation. For the other elements, the directions of maximum and minimum continuity are: Cu, 2˚ and 88˚; As, 89˚ and 1˚; Co, 78˚ and 12˚; Se, 94˚ and 4˚; Pb, 74˚ and 16˚; respectively. The values are all similar for all the other depths.

In summary, all the variogram models showed positive spatial autocorrelation trends as they all showed an increasing semivariance, $\hat{Y}(h)$, with an increase in the distance separation, (h). The second order stationarity is obeyed, the variograms tend to remain constant after the sill value is reached.

These variograms assist in providing a good measure of the goodness of the sampling method that is being used, the spatial variability of samples, and assists in correlating the simulation variogram with the measured spatial variability (Ersoy, Yunsel, and Atici, 2008). But these models cannot tell us about pattern characterization and regions with high and low concentrations, which is why it is required to present the estimates in a form of geochemical maps.

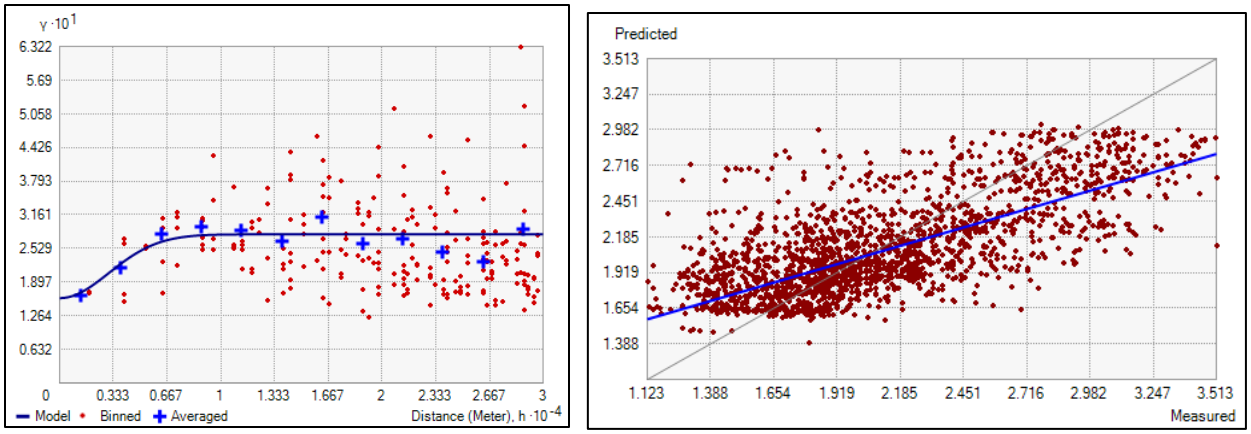


Figure 18: Gaussian variogram model (left) of Ni with a nugget effect, sill, and range of influence of 0.16, 0.119 and 7,465.555 m, respectively. The correlation between the predicted and measured values is on the right image.

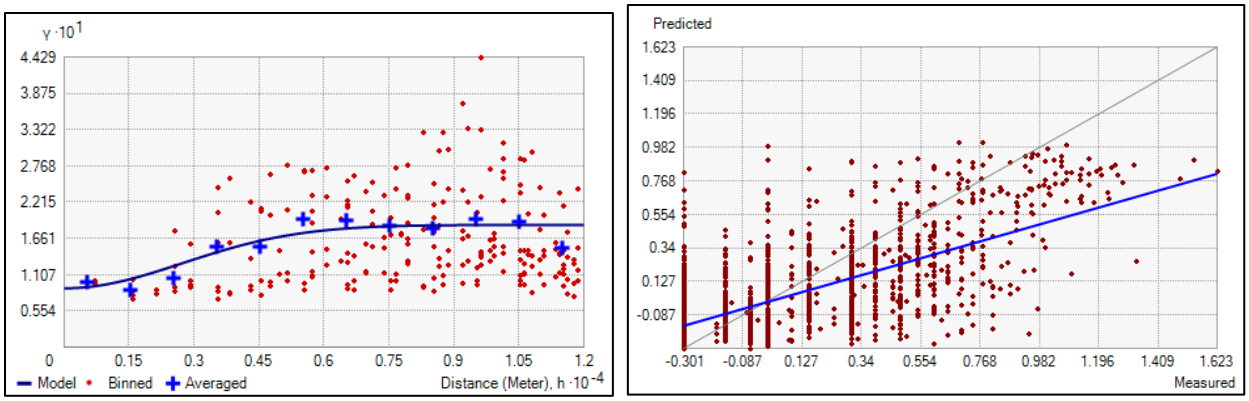


Figure 19: Gaussian variogram model (left) for Se with a nugget effect, partial sill, and range of influence of 0.0891, 0.097, and 6653.33 m, respectively. The correlation between the predicted and measured values is on the right image.

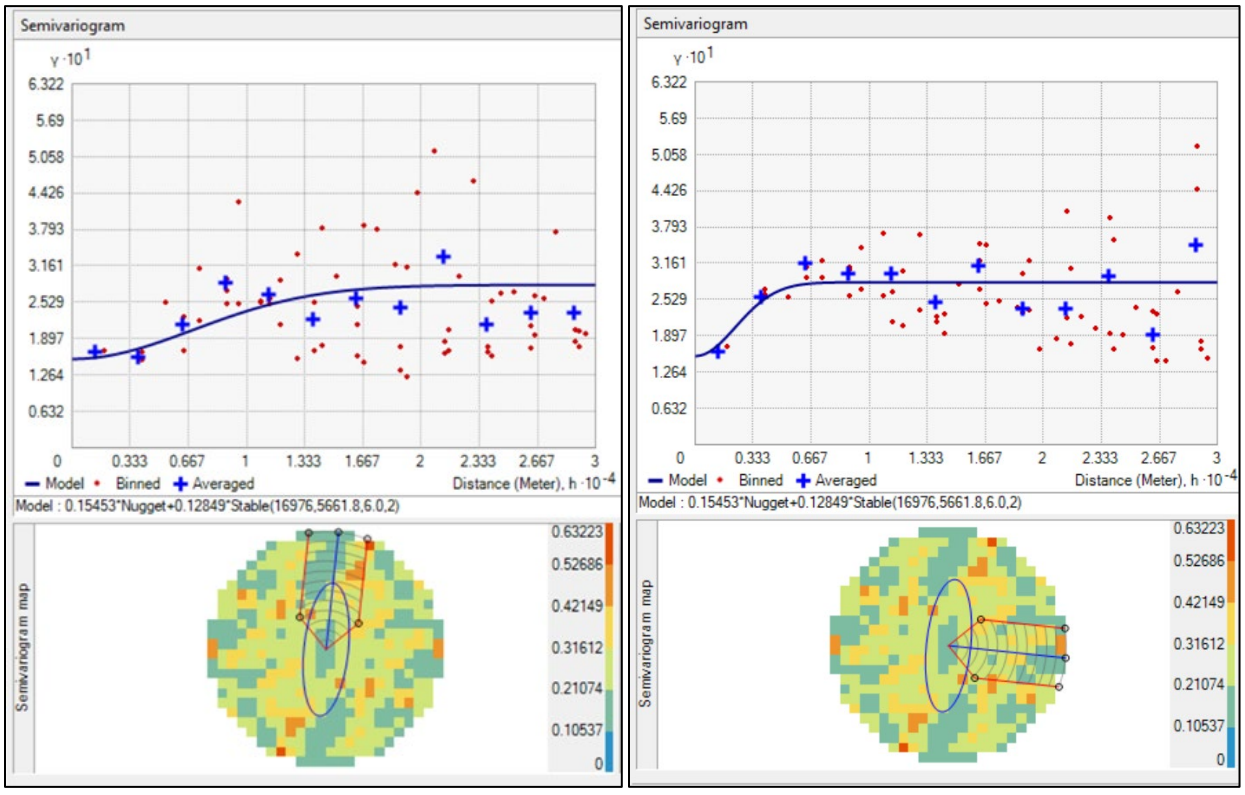


Figure 20: Directional gaussian variogram models for Ni at 0-5 cm soil profile depth. The image on the left is at the direction of maximum continuity at an angle of 6˚, while the right image is the direction of minimum continuity at 96˚. The tolerance used is 45˚ with a bandwidth of 3 m.

Table 5: Parameters of the variogram models. The range of influence is given in meters.

Variable	As	Co	Cu	Ni	Se	Pb
0-5 cm						
Variogram model type	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Number of lags	12	12	12	12	12	12
Lag size	2500	2500	2500	2500	1000	13
Nugget	0.117	0.066	0.176	0.16	0.089	0.049
Range	12770.495	10420.654	7226.567	7465.555	6656.33	80.663
Anisotropy	No	No	No	No	No	No
Partial Sill	0.097	0.038	0.167	0.119	0.097	0.134
Sill	0.214	0.104	0.343	0.279	0.186	0.183
Neigbhours to include	10	10	10	10	5	5
Include at least	5	5	2	5	2	2
Sector type	Four and 45 degree	Four and 45 degree	Four and 45 degree	Four and 45 degree	Four and 45 degree	Four and 45 degree
5-10 cm						
Variogram model type	Gaussian	Gaussian	Spherical	Gaussian	Spherical	Gaussian
Number of lags	12	12	12	12	12	12
Lag size	2000	2000	2000	2000	30	40
Nugget	0.138	0.067	0.059	0.147	0.026	0.083
Range	12450.406	11287.38	135.728	7740.769	114.016	192.28
Anisotropy	No	No	No	No	No	No
Partial Sill	0.123	0.039	0.155	0.107	0.07	0.079
Sill	0.261	0.106	0.214	0.254	0.096	0.162
Neigbhours to include	5	5	5	5	5	5
Include at least	5	2	2	2	2	2
Sector type	Four and 45 degrees	Four and 45 degrees	Four and 45 degrees	Four and 45 degrees	Four and 45 degrees	Four and 45 degrees
10-20 cm						
Variogram model type	Gaussian	Gaussian	Gaussian	Gaussian	Spherical	Gaussian
Number of lags	12	12	12	12	12	12
Lag size	2000	3000	1000	1000	50	1000
Nugget	0.13	0.046	0.116	0.114	0.036	0.153
Range	10813.968	7856.052	7245.416	7100.658	224.033	6774.037
Anisotropy	No	No	No	No	No	No
Partial Sill	0.118	0.021	0.169	0.123	0.048	0.068
Sill	0.248	0.067	0.285	0.237	0.084	
Neigbhours to include	5	5	5	5	5	5
Include at least	2	2	2	2	2	2
Sector type	Four and 45 degrees	Four and 45 degrees	Four and 45 degrees	Four and 45 degrees	Four and 45 degrees	Four and 45 degrees

4.3. Data validation

Prior to examining the interpolated surfaces, cross-validation statistics of the variograms for the kriging models, and data reproduction of the simulations were closely examined.

4.3.1. Cross validation

The statistics for the cross-validation parameters are presented in Table 6. The statistics show that the data is robust and efficient enough to be used to predict values, however, some models

did not meet all the standard requirements for a good estimation. The presented statistics are for the best models selected.

Table 6: Cross validation summary statistics for ordinary kriging.

Variable	Mean Error (ME)	Root Mean Square Error (RMSE)	Mean Standardised Error (MSE)	Root Mean Square Standardised Error (RMSSE)	Average Standard Error (ASE)	Variogram model
0-5 cm						
As	0.007	0.324	0.021	0.919	0.359	Gaussian
Co	0.001	0.23	0.004	0.868	0.268	Gaussian
Cu	0.002	0.386	0.007	0.882	0.453	Gaussian
Ni	0.002	0.363	0.003	0.877	0.425	Gaussian
Se	0.003	0.274	0.014	0.853	0.329	Gaussian
Pb	-0.006	0.333	-0.002	0.889	0.393	Gaussian
5-10 cm						
As	-0.004	0.334	-0.009	0.849	0.399	Gaussian
Co	-0.007	0.233	-0.025	0.841	0.278	Gaussian
Cu	-0.006	0.39	-0.013	0.903	0.428	Spherical
Ni	-0.008	0.366	-0.02	0.869	0.419	Gaussian
Se	0.003	0.224	0.007	0.784	0.298	Spherical
Pb	0.002	0.318	0.001	0.832	0.392	Gaussian
10-20 cm						
As	-0.009	0.327	-0.022	0.839	0.393	Gaussian
Co	-0.006	0.183	-0.024	0.803	0.229	Gaussian
Cu	-0.009	0.347	-0.025	0.901	0.379	Gaussian
Ni	-0.009	0.327	-0.022	0.839	0.393	Gaussian
Se	0.002	0.209	0.005	0.784	0.281	Spherical
Pb	-0.01	0.335	-0.025	0.804	0.422	Gaussian

The results of ME and MSE showed that the models of Co, Se and Pb at 0-5 cm, As, Co, Ni and Cu at 5-10 cm and, As, Co, Cu, Ni and Pb at 10-20 cm displayed the least bias. Models of As and Cu at 0-5 cm and, Se at 5-10 cm shows the most underestimation. This was expected for Se and As given that they did not resemble a normal distribution even after transformation. A startling

observation is that although As and Se were not close to a normal distribution for all the profile depths, the underestimation was only on particular depths i.e., 0-5 cm for As and 5-10 cm for Se. The ASE and RMSE of the models were relatively close to each other and small, ranging from 0.209 to 0.428 illustrating a good measure of variability, with the exceptions of Cu and Ni at 0-5 and 5-10 cm. This makes the models of Cu and Ni less reliable and trustworthy at these depths due to the slight difference in their ASE and RMSE values. Models with RMSSE deviating from 1 and showing higher prediction errors are that of Se at both 5-10 and 10-20 cm, which is not a surprise given that the transformed histograms of Se were not close to a normal distribution (positively skewed) and has sparsely distributed data. The rest of the models show lower prediction errors with Cu standing out the most at 0.9.

4.3.2. Data reproduction

Various tests were conducted to check the reproduction of the original data by the gaussian simulation models.

Visual observation- The simulations were only produced for the upper soil profile depth of 0-5 cm as no trends of interest were observed for the other depths. Visual inspection of the simulation means of 100 realizations indicated that they closely resemble the ordinary kriging estimations and proved to be robust. The low and high values are both located in areas where it was expected, based on the visualization of the ordinary kriging maps. The previously observed trends are also positively replicated by the simulations (Figure 22, 23). The variability is clearly reasonable and not very much questionable, it reproduces what we are interested in meeting the study's first objective.

Descriptive statistics reproduction- The statistics of the simulation mean were used to compare with that of the raw data, shown in Table 7. The mean of the mean of the simulations is relatively close to the mean of the raw dataset, with Co, Cu and Ni closer than the other metals, with a difference therefore, the data statistics were closely reproduced.

Correlation scatter plots- Scatterplots were implemented to accurately check the reproduction of the data Figure 22. The R-squared value was calculated to check for the goodness of fit for the linear regression models. The values were low showing a weak effect of the fitted model on the simulated values, meaning that not all the variances have been accounted for as most of the data points are falling far away from the regression line. The lowest R-squared value is that of Pb at 0.37 with the highest being Se at 0.58, the rest are, As, 0.53; Co, 0.53; Cu, 0.52 and, Ni 0.49. Much attention should be paid to the simulated model of Pb due to this low coefficient of determination. However, the R-squared value does not really illustrate if a regression model provides adequate fit for the data and these models can still be used to draw conclusions. The coefficient of correlation (r) shows that the simulated values and the original data have a strong positive relationship. The highest correlation is that of Se at 0.76 with Pb being the lowest at 0.61, the values for Cu, Co, Ni and As are 0.72, 0.73, 0.70 and 0.7, respectively, which are relatively satisfying.

Histogram reproduction- Comparisons of the raw data histograms and simulated values are shown in Figure 23, 24. The mean and standard deviation of the original data and the simulated values are relatively close; hence the simulations were able to reproduce the original histograms.

QQ plot reproduction- The QQ plots were analyzed to further assess the histogram reproduction, as the histogram might have masked some information due to binning Figure *25.* Most of the elements show great histogram reproduction such as Cu, Ni, Pb, Se and to a lesser extent Co. This

was also seen on the histograms where all the elements reproduced the original histogram on average, with As being an exception.

Variogram reproduction- Variogram models were calculated for the mean of the 100 realizations and their reproduction compared with the sample variograms. The models are illustrated in Figure 26 and 27. The models fitted to the variogram of the simulated realization mean are relatively close to the sample or original data in terms of the variogram shape. The range of influence for most of the elements was accurately reproduced, but the partial sill and nugget effect are not being reproduced. This could mean that the nugget effect that was previously depicted for the sample data was due to measurement error, and the variation that is associated with the nugget should not be reproduced by the simulations if it corresponds to measurement error (Marcotte, 1995). This great departure in the nugget effect could also be due to some sort of trend within the data which can be related to regions with elevated metal levels, this effect might not have an impact in the spatial structure of the data.

Table 7: Descriptive statistics of the mean of the simulations.

Variable	As	Co	Cu	Ni	Pb	Se
Mean	0.789	1.010	2.031	2.094	1.276	-0.026
Standard Error	0.005	0.004	0.007	0.006	0.005	0.005
Median	0.701	0.971	1.917	2.002	1.204	-0.173
Mode	1.515	1.505	2.636	2.664	1.635	0.231
Standard Deviation	0.330	0.264	0.433	0.370	0.288	0.300
Sample Variance	0.109	0.070	0.188	0.137	0.083	0.090
Kurtosis	0.295	0.485	-0.309	-0.642	-0.916	0.997
Skewness	0.981	0.164	0.792	0.636	0.321	1.378
Range	1.699	1.510	1.969	1.440	1.483	1.216
Minimum	0.205	0.162	1.071	1.439	0.433	-0.302
Maximum	1.904	1.672	3.040	2.879	1.916	0.914
Sum	3054.738	3907.424	7859.705	8103.023	4939.709	-101.376
Count	3870	3870	3870	3870	3870	3870

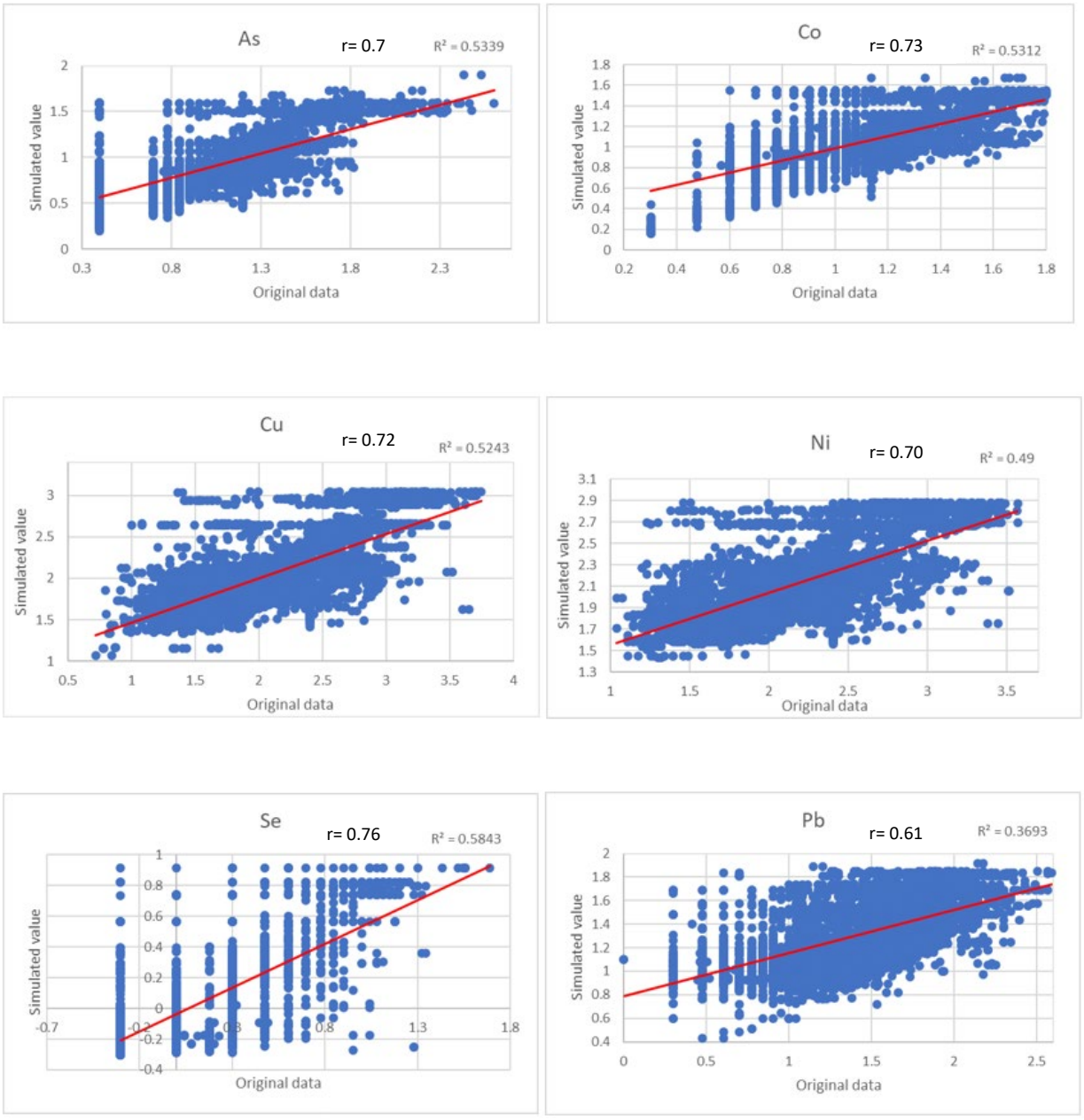


Figure 21: Correlation plots of the original data vs. simulated values, with the R-squared and r value on the top right corner of each plot.

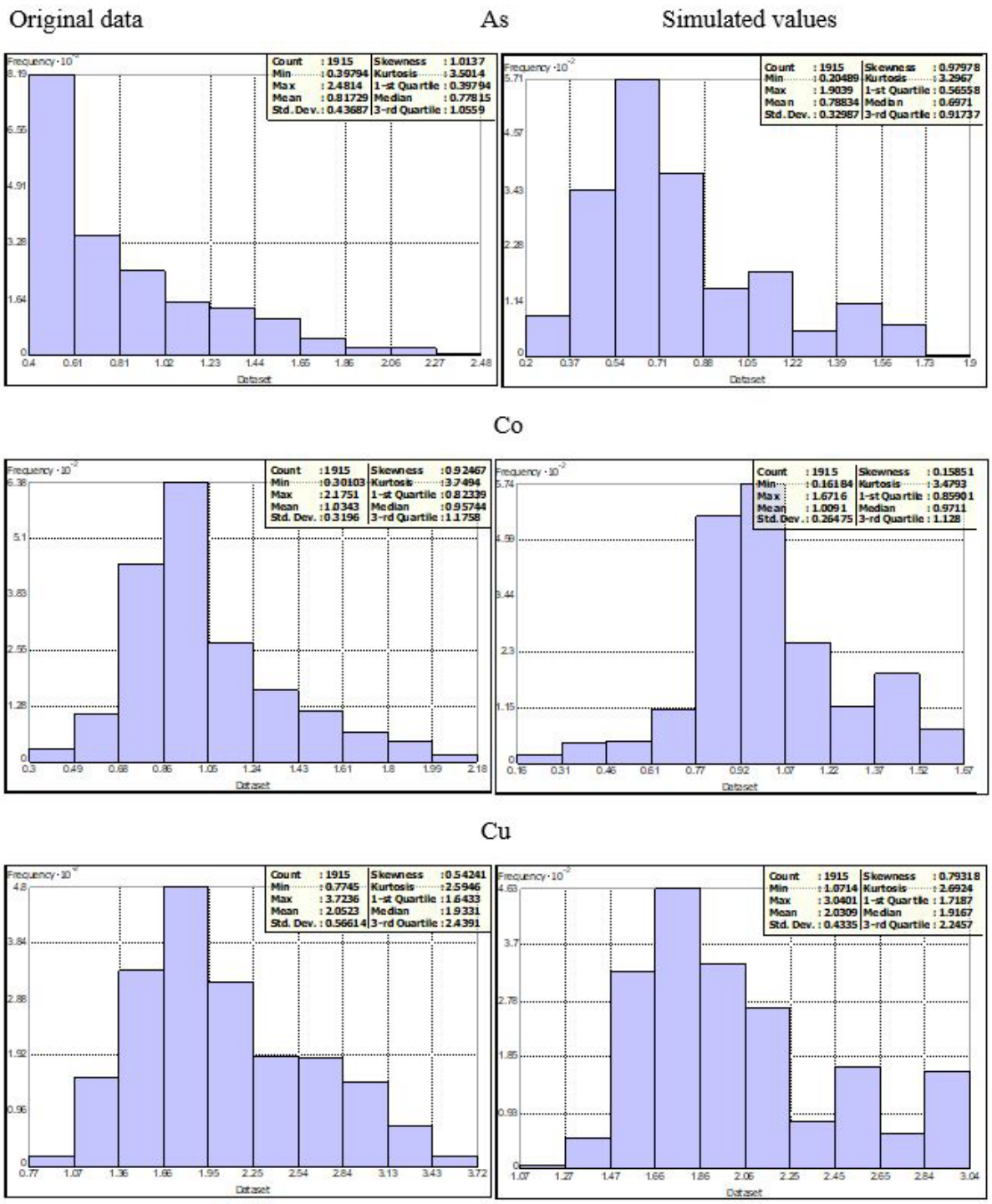


Figure 22: Histogram reproduction check for As, Co and Cu, the mean and standard deviations of the plots are shown on the top right corner of each plot.

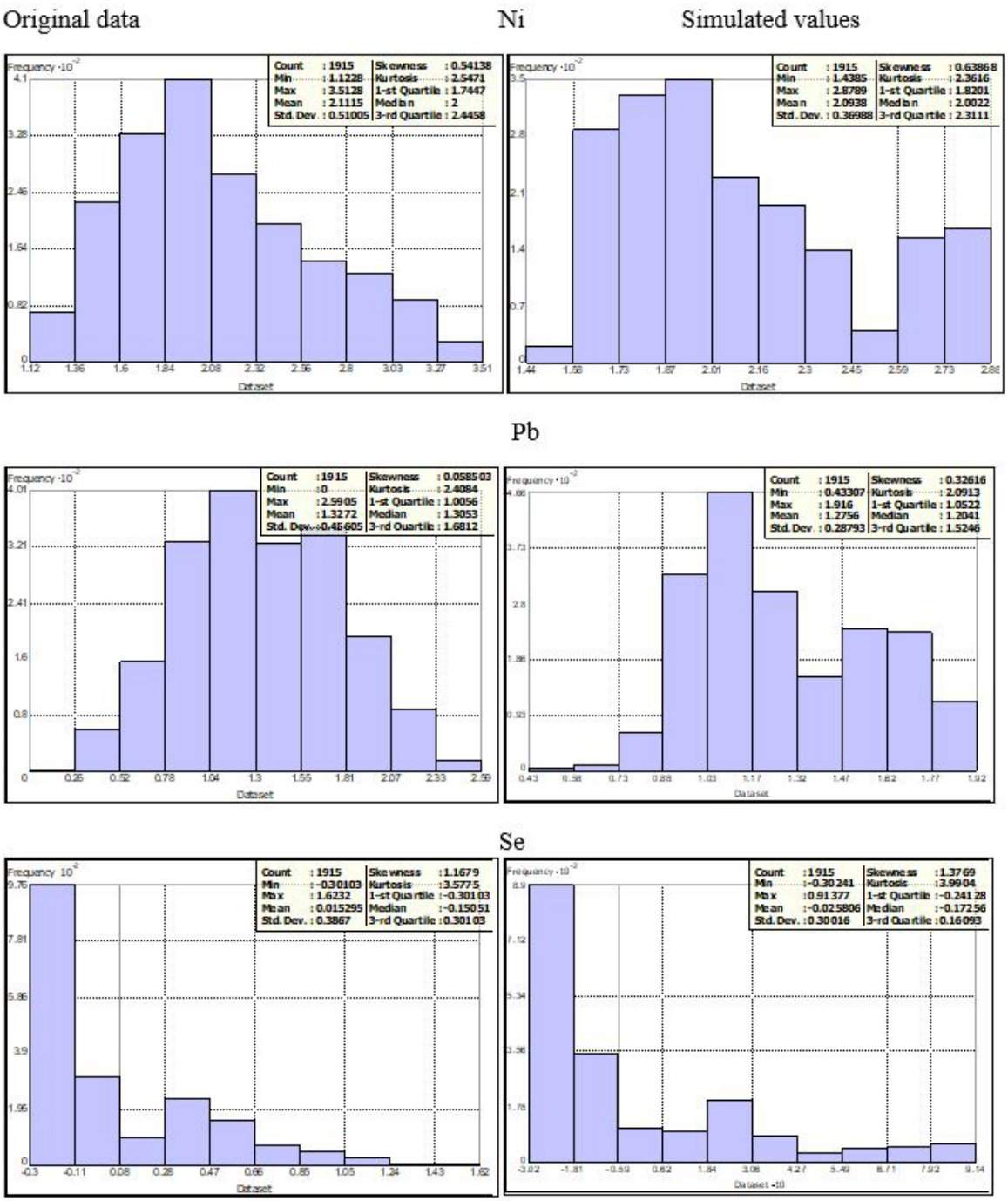


Figure 23: Histogram reproduction check of Ni, Pb and Se, the mean and standard deviations of the plots are shown on the top right corner of each plot.

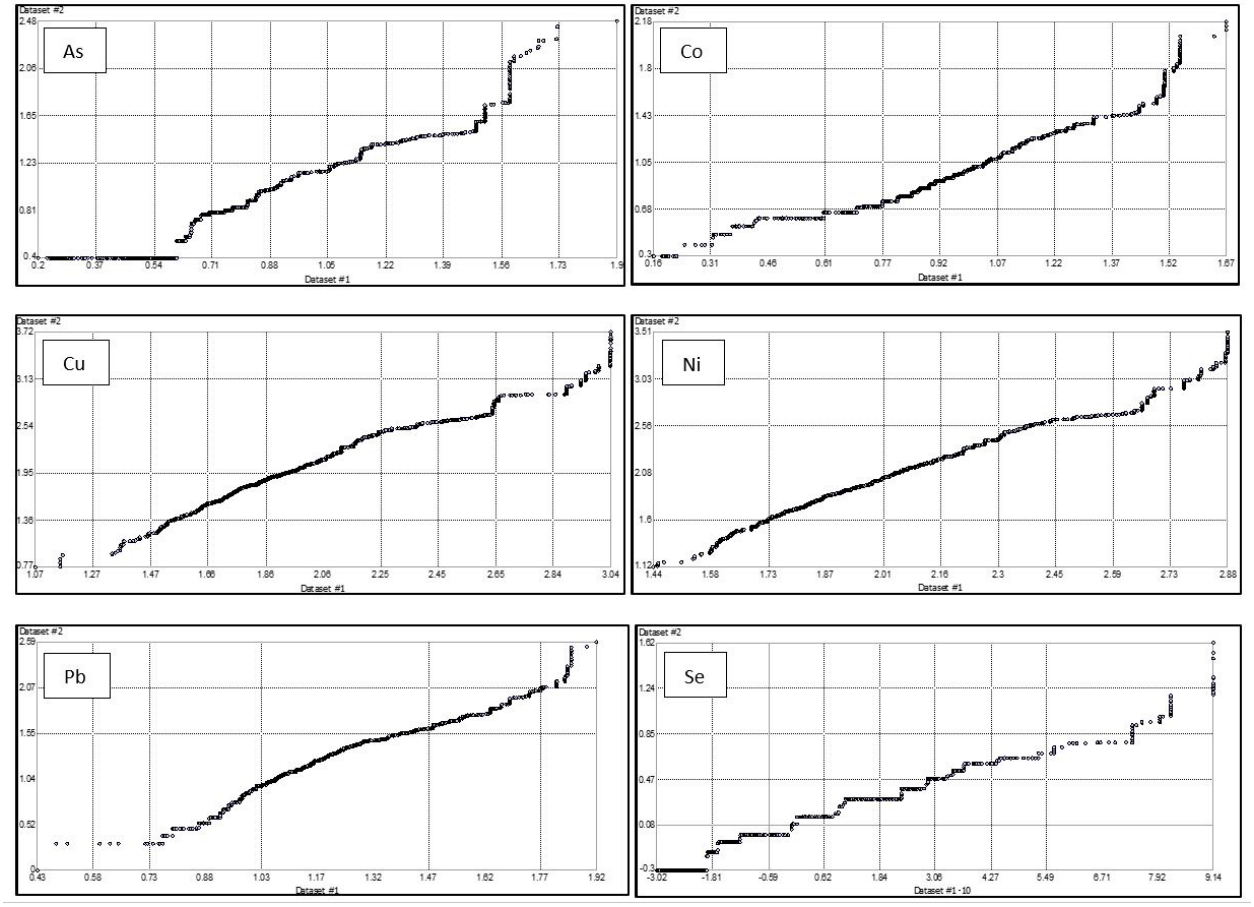
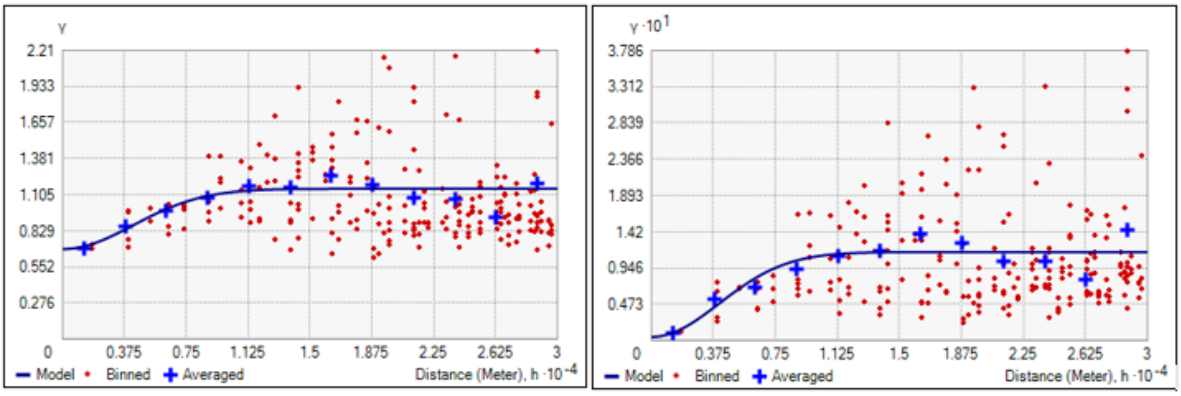


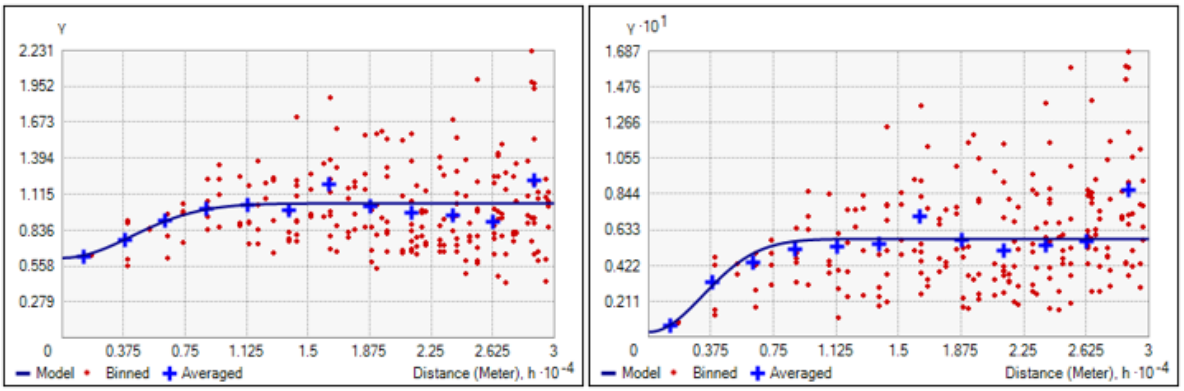
Figure 24: General QQ plots for the simulated vs. original data to assess histogram reproduction.

Dataset #1 represents simulated values while Dataset #2 is the original value.

As



Co



Cu

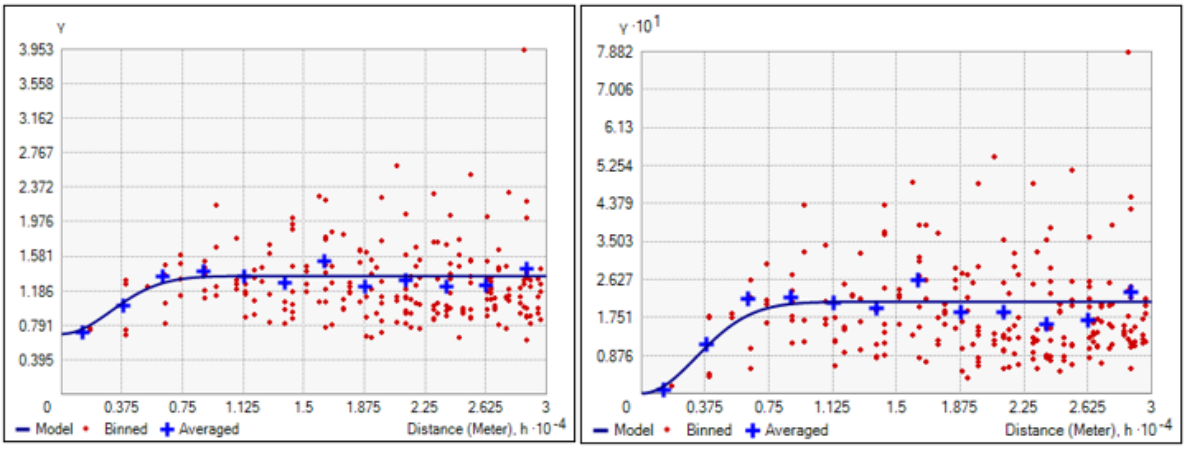
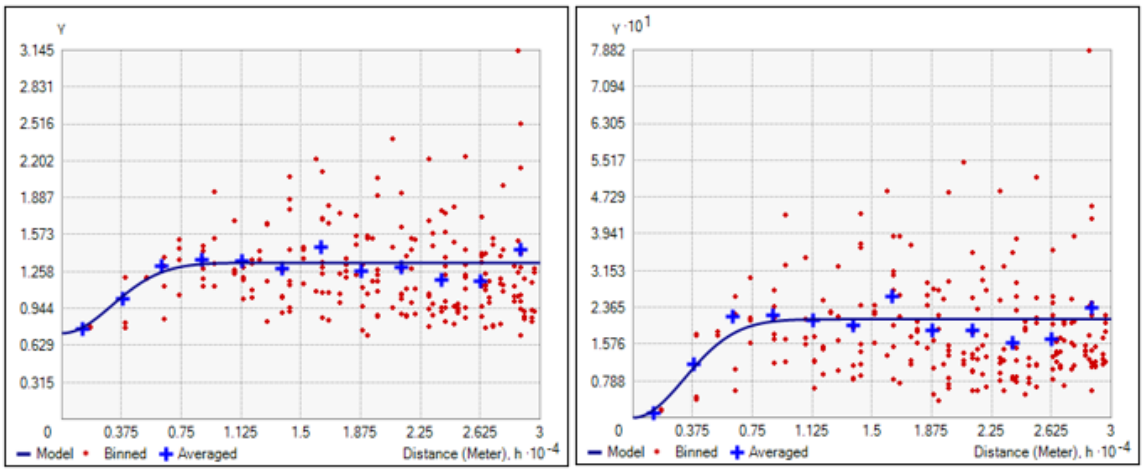
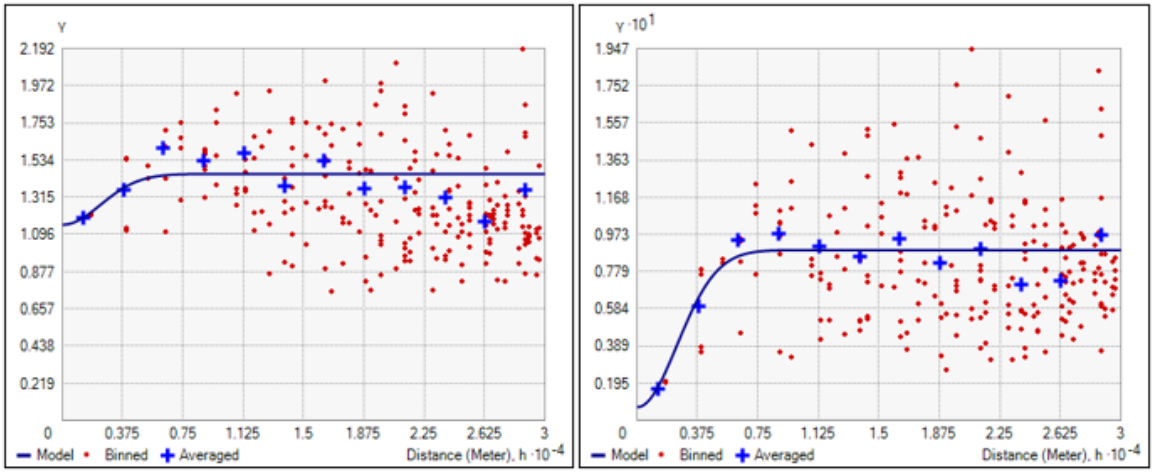


Figure 25: Sample (left) and simulated variograms (right) of the elements, showing the nugget, partial sill, and range of influence for the data.

Ni



Pb



Se

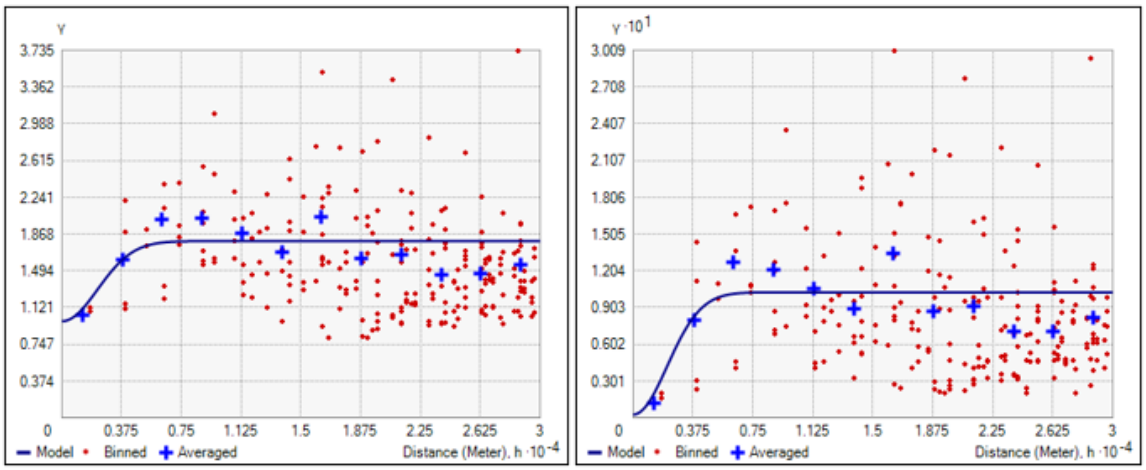


Figure 26: Sample (left) and simulated variograms (right) of the elements, showing the nugget, partial sill, and range of influence for the data.

4.4. Kriging and simulation models

To accurately analyze the patterns that the elements are making in the soil environment, the data is presented in the form of geochemical maps, this meets the first objective of the study. This will enable the detection of regions with positive concentration anomalies and possibly relate the observed patterns to different driving forces. Mapping the distribution of the COC will enable the identification of spatial trends and assist in subsequent risk assessment projects.

Ordinary kriging maps of the elements As, Co, Cu, Ni, Pb and Se are presented alongside their corresponding gaussian simulation models. The maps are presented using a logarithmic scale with filled contours (connecting areas with the same values) representing concentrations given the same color. The log transformed kriging models produced showed that higher concentrations of Cu, Ni, As, Co and Se are centered within the vicinity of the three historic smelters, with the highest values around Copper Cliff and Falconbridge. According to SARA (2008), the highest concentrations of nickel (3200 mg/kg) and copper (5600 mg/kg) were detected within Copper Cliff, this is evident from the kriging models currently produced (Figure 37, 38). A fascinating factor is that the spatial distribution of Ni closely resembles that of Cu, which is in line with the findings of the SSS.

A broad-scale gradient is readily observed approximately along the northwest to southeast direction (NW-SE), with the highest values of Cu (5.92-6.97 µg/g) and Ni (5.92-6.79 µg/g) progressively decreasing away from the smelters. A secondary trend is also reported, marking a narrow and elongated zone with elevated values of both Cu and Ni (northeast to southwest (NE-SW) direction). The kriging model of Ni shows local discrepancies (< 3-5 kms) if compared to the Cu interpolation results. The kriging estimates of As, Co and Se also show similar observations as Cu and Ni, but in varying variabilities. Arsenic shows the same trends but in lower concentrations,

the highest concentrations are seen around the Falconbridge smelter, which is a similar observation with Se showing high concentrations in Copper Cliff. Cobalt on the other hand, shows similar concentrations distributed almost equally around the three smelters in similar patterns as that of Cu and Ni. The odd element in terms of patterns is Pb, it does not show any clear patterns and its estimates are not easy to interpret as no trends can be depicted. But however, the models show evidence of smoothing as the discrepancy between regions of high and low values cannot easily be made, but the trends can be seen with ease. The variability in patterns of these elements could be attributed to various soil types within the study area partly explaining the incongruence as well as a higher magnitude of control on air dispersion processes vs later mobilization in the soils.

The simulations were produced using cell by cell declustered data and simple kriging models as outlined in Chapter 3 (3.1.1.6). The goal of using simulated models was to refine the kriging models as some of them showed evidence of over and underestimation with smoothing effects. The simulated maps of the elements show that much of the uncertainty is in preferred areas that were highly sampled, which are areas in the smelter centroid. The simulations were able to overcome the drawbacks that were observed in the kriging estimates (i.e., over and underestimation, and smoothing effect). The observed patterns from the estimates can now be clearly seen for Cu, Ni and Se, but however kriging seems to have accurately mapped the trends better as compared to the simulation models. The high values are at the regions were the distribution shows to occur most and so are the lower values. Although the simulations failed to replicate (on average) the variograms of the sample data, the maps can still be used for soil contamination interpretations. These geochemical maps can be used in land contamination characterization, site inspection of contaminated zones and to plan future sampling methods.

The middle (5-10 cm) and lower (10-20 cm) soil profile depths were also mapped (Figure 30). These ordinary kriging estimates did not show any pronounced trends. The maps show that the concentrations around the smelter centroid have decreased with an increase in the soil profile depth. At 5-10 cm minimum anomalies can still be seen around the smelters for the elements and appears to have decreased at 10-20 cm. The uncertainty that existed around these smelters can be said to decrease with an increase in soil profile depth. This is in line with what has been depicted already in earlier studies.

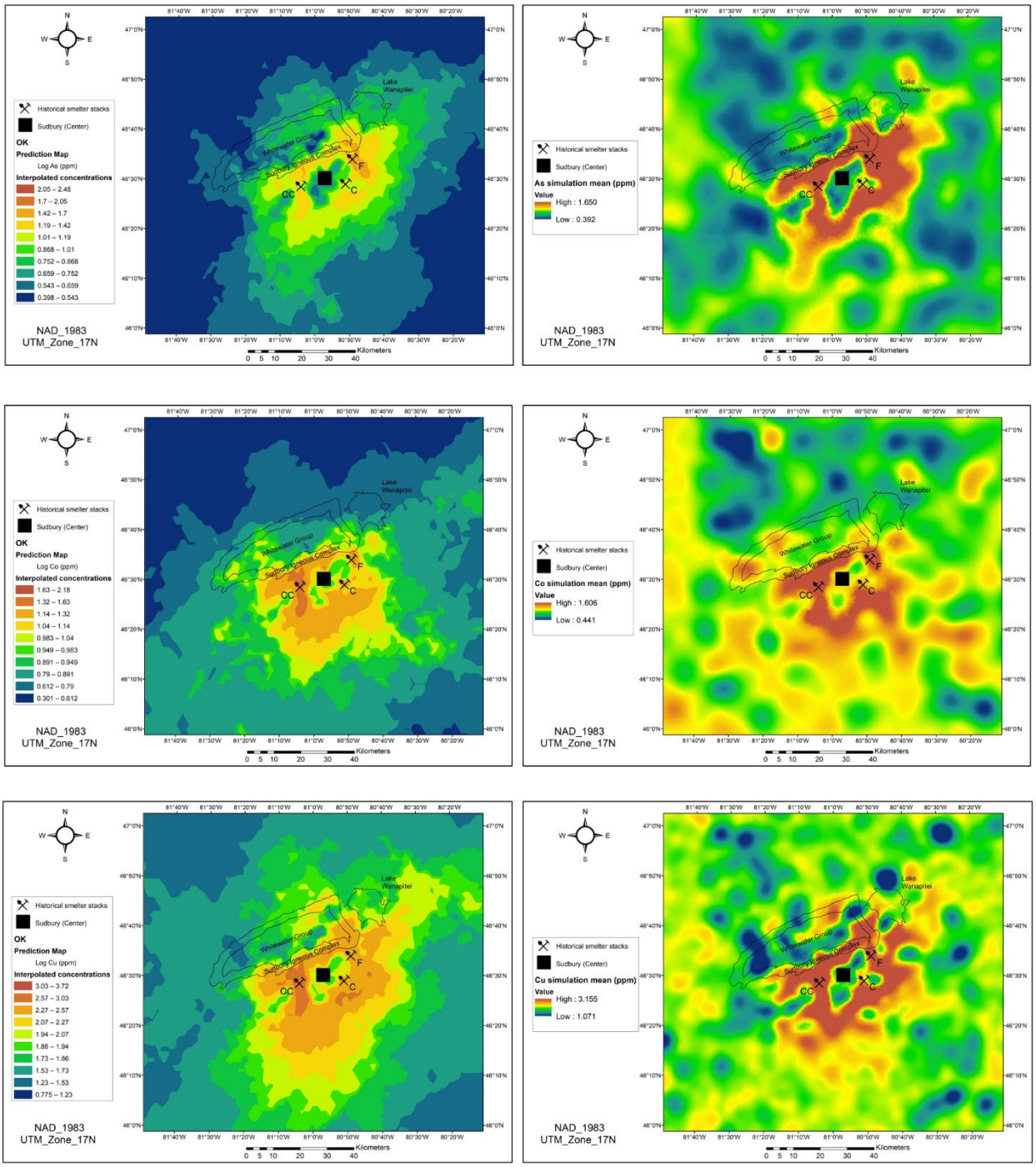


Figure 27: Ordinary kriging estimates (on the left) of As, Co, and Cu and the gaussian simulated values (on the right). All the values are log transformed and given in ppm. Historical smelter stacks are labeled as CC, Copper Cliff; C, Coniston; and F, Falconbridge. Common features in the region i.e., Sudbury Igneous Complex and Lake Wanapitei are inserted.

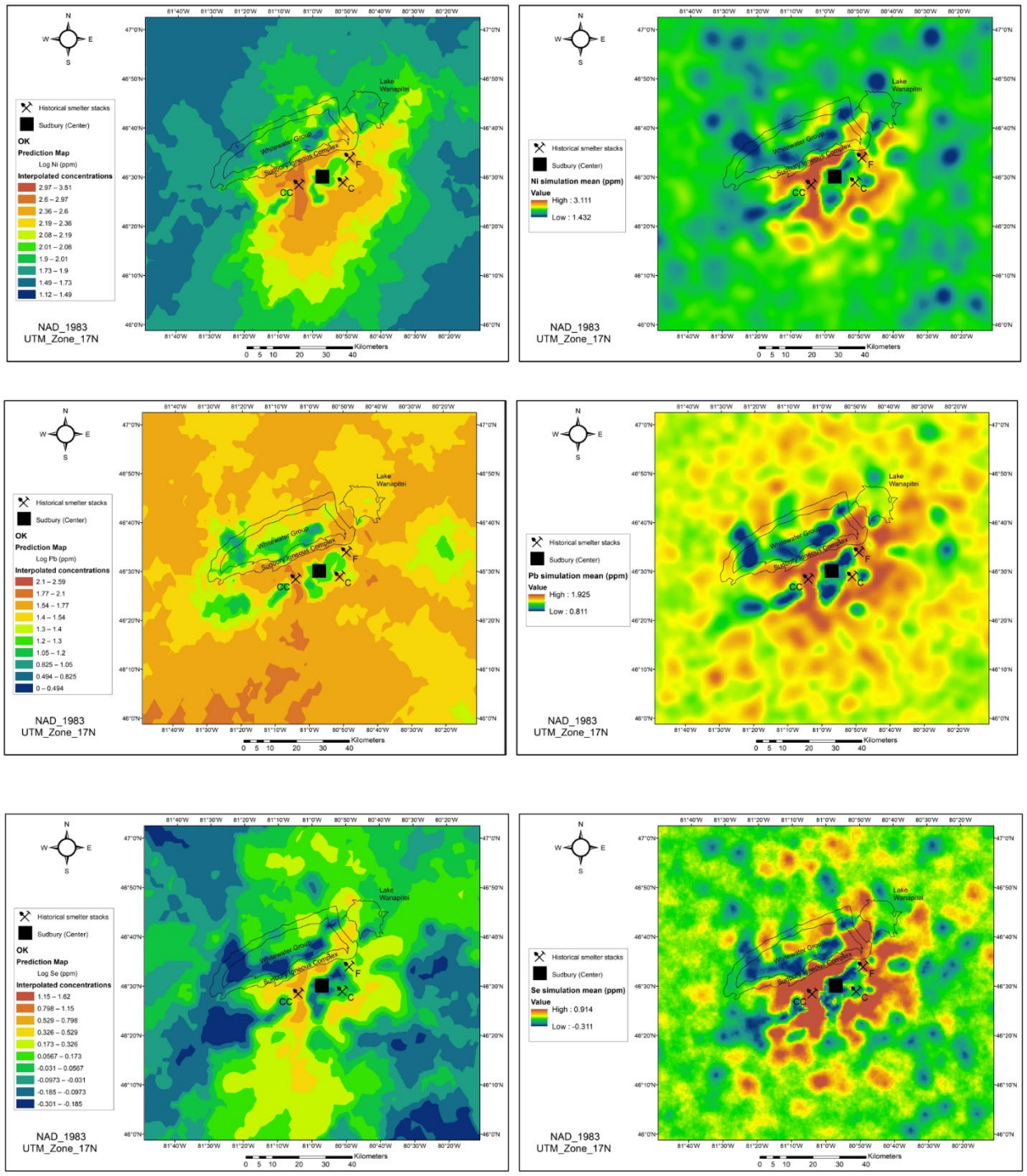
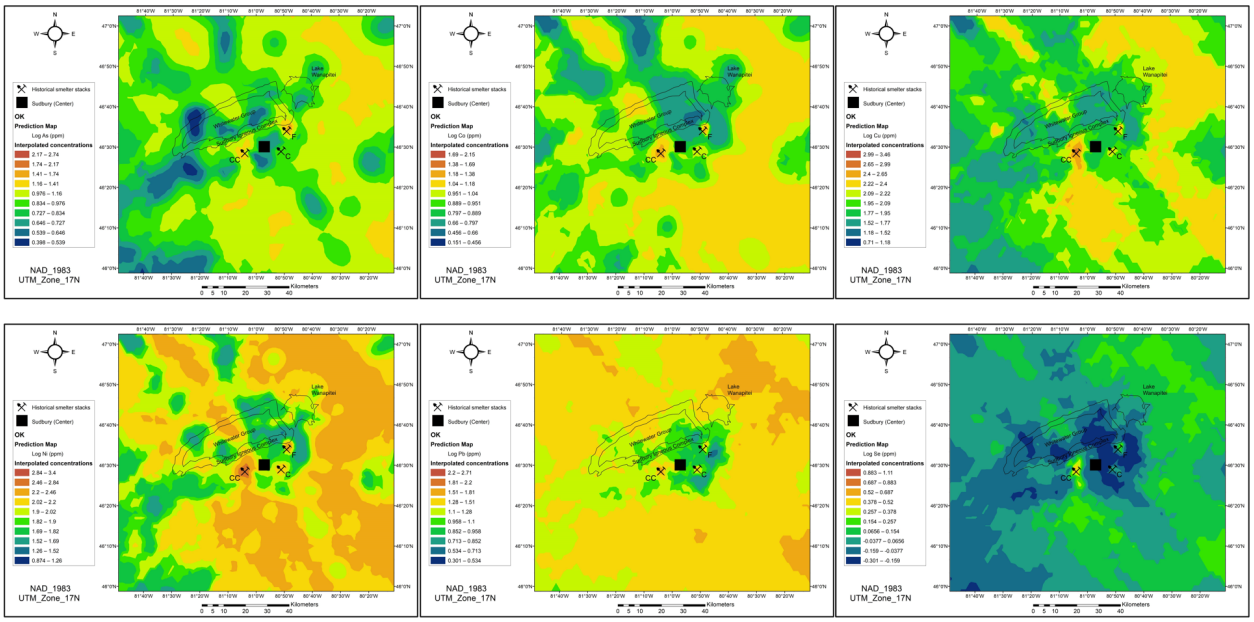


Figure 28: Ordinary kriging estimates (on the left) of Ni, Pb, and Se, and the gaussian simulated values (on the right). All the values are log transformed and given in ppm. Historical smelter stacks are labeled as CC, Copper Cliff; C, Coniston; and F, Falconbridge. Common features in the region i.e., Sudbury Igneous Complex and Lake Wanapitei are inserted.

5-10 cm



10-20 cm

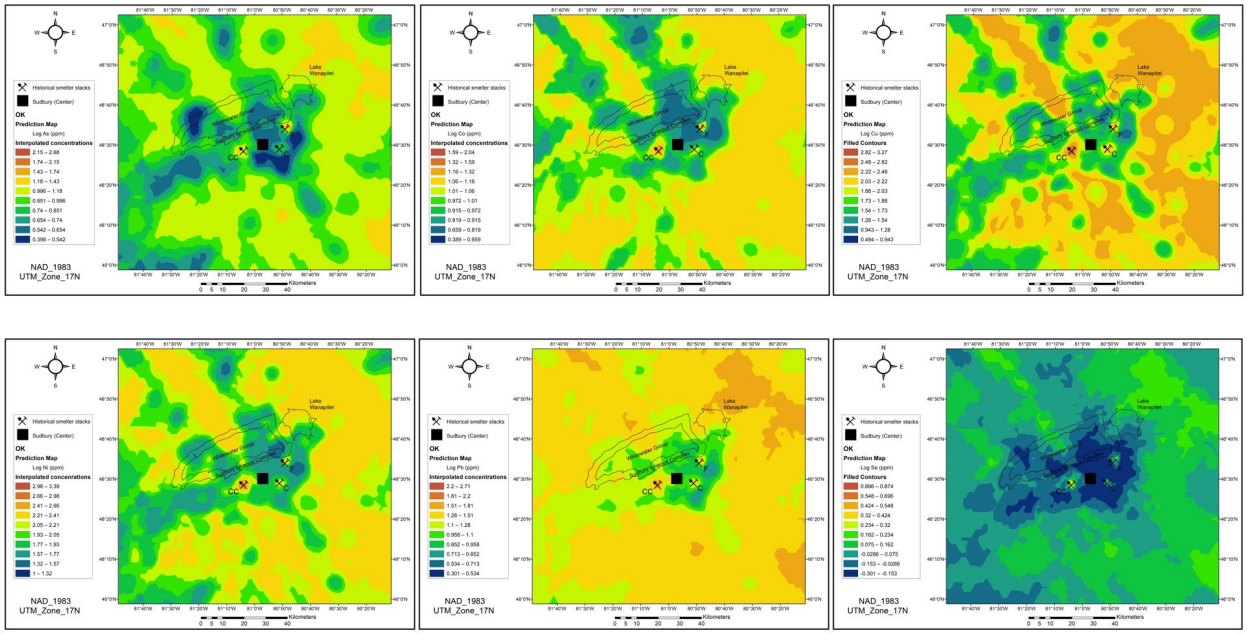


Figure 29: Ordinary kriging estimates for soil profile depth 5-10 (top) and 10-20 cm (bottom) for As, Co, Cu, Ni, Pb, and Se. The interpolated values are given in ppm.

4.5. Vegetation analysis

Figure 30 shows band composite images from the year 2019, this is to give a picture of the real-life images that the spectral variabilities and vegetation change will be based upon. The images show real features of the region such as vegetated, non-vegetated regions, and previous barren areas of concern. The historic smelters are labelled as CC, Copper Cliff; C, Coniston and F, Falconbridge, for all the images under this section.

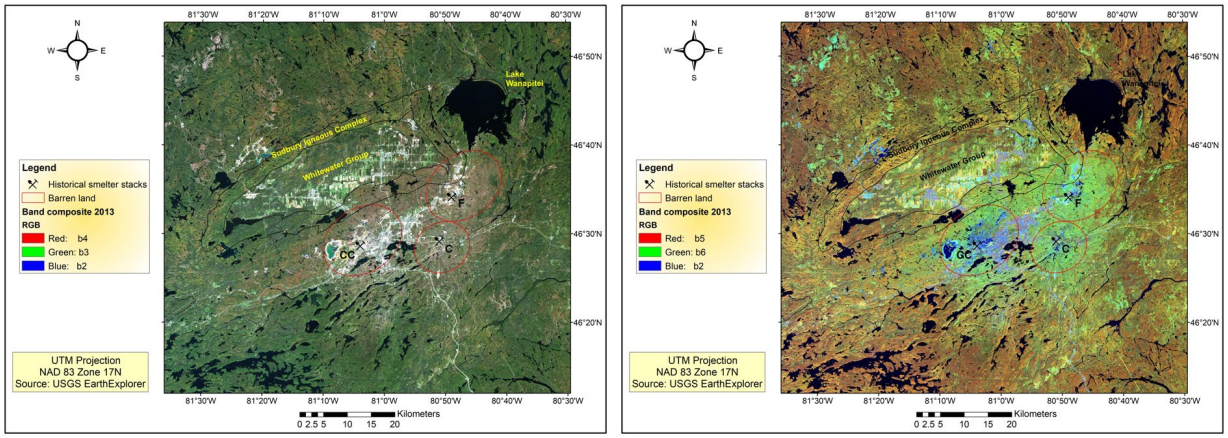


Figure 30: RGB 4,3,2 natural colour (left) and 5,6,2 vegetation and soil cover band composites for Landsat 8 of the year 2019.

The green colour on the left image and rusty orange to yellow colour on the right image indicates regions that are vegetated while the brownish colour (left image) and greenish colour (right image) indicates non-vegetated areas, infrastructure and roads are represented by white and blue-purple colour on the left and right images, respectively. Water bodies are in black in both images.

Band ratioing was applied to the Landsat images, figure 31 shows this for the year 2019. This gave an indication of regions that are expected to be vegetated and non-vegetated when assessing vegetation growth patterns. The image shows vegetated regions by bright pixels represented by the colour green, as the shadow effects are now reduced, and can be seen all around the area with a mild green surrounding the smelters.

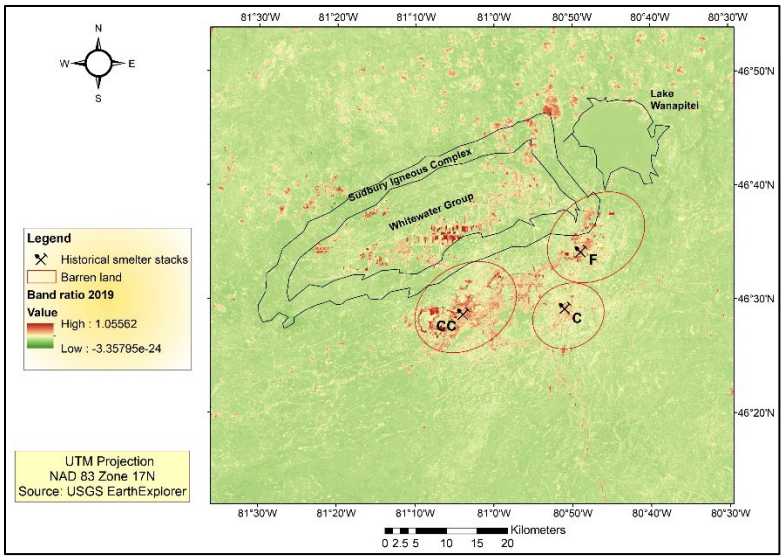


Figure 31: Band ratio image from 2019 Landsat data.

NDVI was used to determine the production of green vegetation and vegetation changes in and around regions that were previously damaged due to excess mining and smelting activities around the Sudbury region. The calculated NDVI images are presented in Figure 32. The main features of the images show that barren areas existed during the pre-industrial years (1979), indicated by NDVI values of <0.1 and has since decreased following the post-industrial years. The barren areas formed a northeast-southwest pattern, these could be following the geographic distribution of the smelters or some other driving forces. The lowest NDVI values <0.1 are observed on non-vegetated soil and water bodies, this is due to the high reflection that is given off by these areas, which in turn produce low values in the NIR band and high values in the RED, hence the values are low, the opposite is true for the vegetated regions. Visual inspection shows that the vegetation of the area has since recovered and most of the previous barren areas are now vegetated, this is shown by the NDVI classified images (Figure 33). The data was classified using unsupervised classification algorithms and the classes were made based on visual inspection of the band

composites in comparison with Google Earth images. The scars on the land in and around the smelter centroid that were captured in 1979 can be seen to have gradually decreased during the following mapped years. This can be attributed to improved dust removal and mining technologies, and regreening programs in the region. The semi-barren and barren patterns are like what was mapped by Gunn et al. (1995) in Figure 9. The 2019 image shows a lot of clouds at the top left and right, cloud masking was not effective for this data. The clouds should not be interpreted as non-vegetated areas in Figure 33.

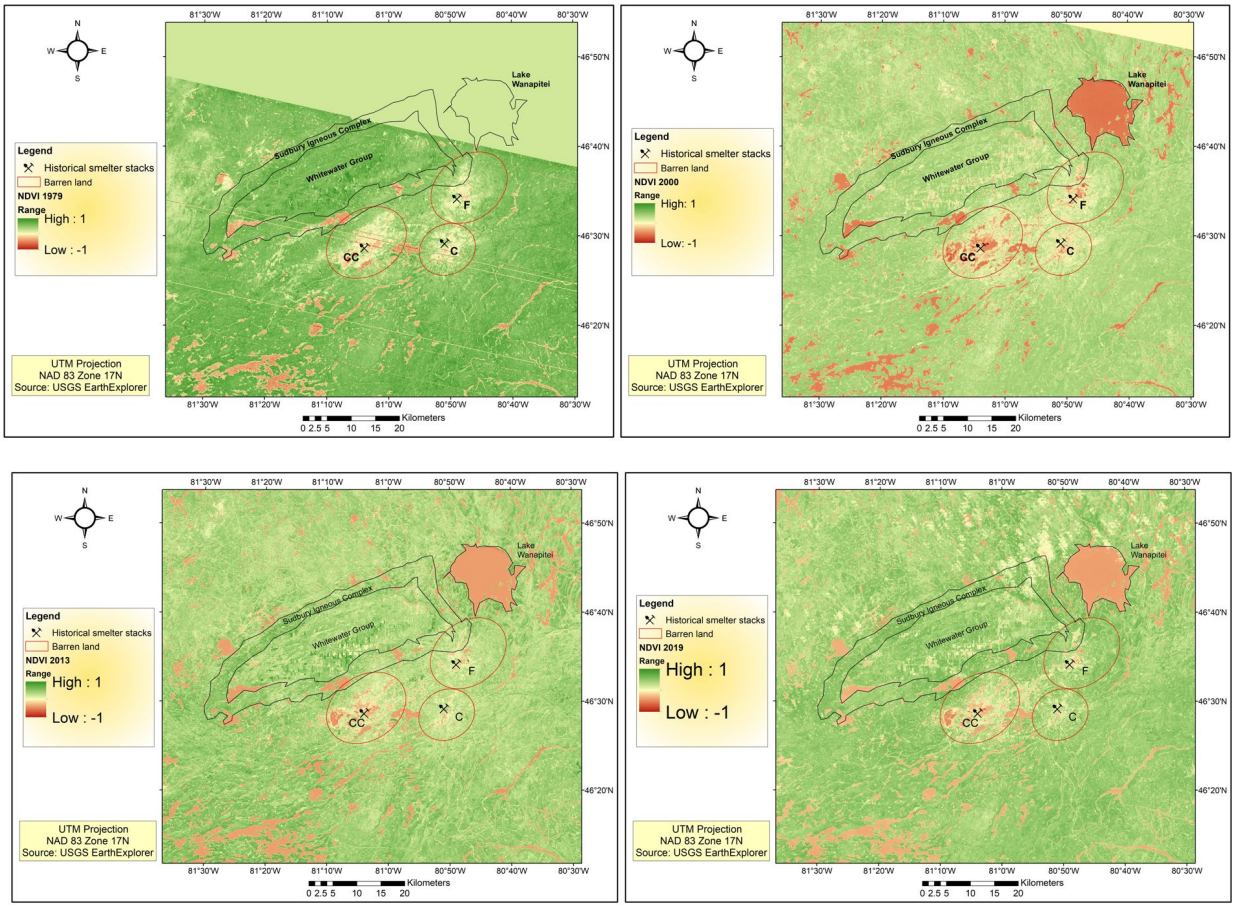


Figure 32: NDVI images of the years 1979, 2000, 2013 and 2019. Vegetation is indicated by a green colour while non-vegetated regions and water bodies are in red-pale yellow colour.

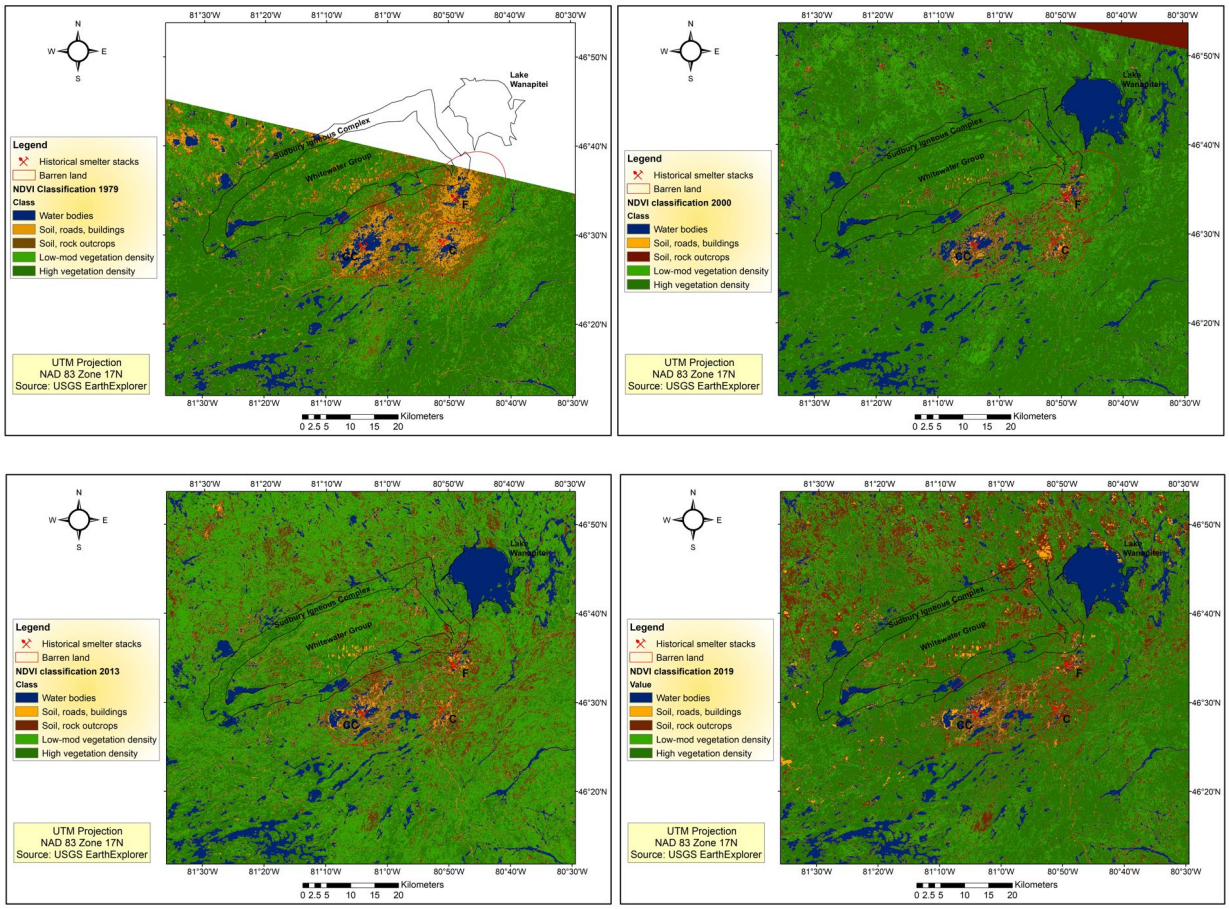


Figure 33: Unsupervised classified NDVI images.

4.6. Integration

The kriging estimates and simulation trends were integrated with NDVI observed trends to map and create hybrid maps for precision interpretation of the observed patterns. The patterns observed from the geostatistical analysis conducted were directly combined with the NDVI and the trends were traced (Figure 34). The 1979 NDVI image was chosen for this as it clearly shows where the barren areas were located which makes it easy for integration. The printed patterns coincide with regions that have been mapped to be barren in 1979 elongated in the same directions. The occurrence of the observed distribution of elements according to wind patterns,

geology of the region, vegetation cover and metal deposition and mobility will be discussed in the following sections.

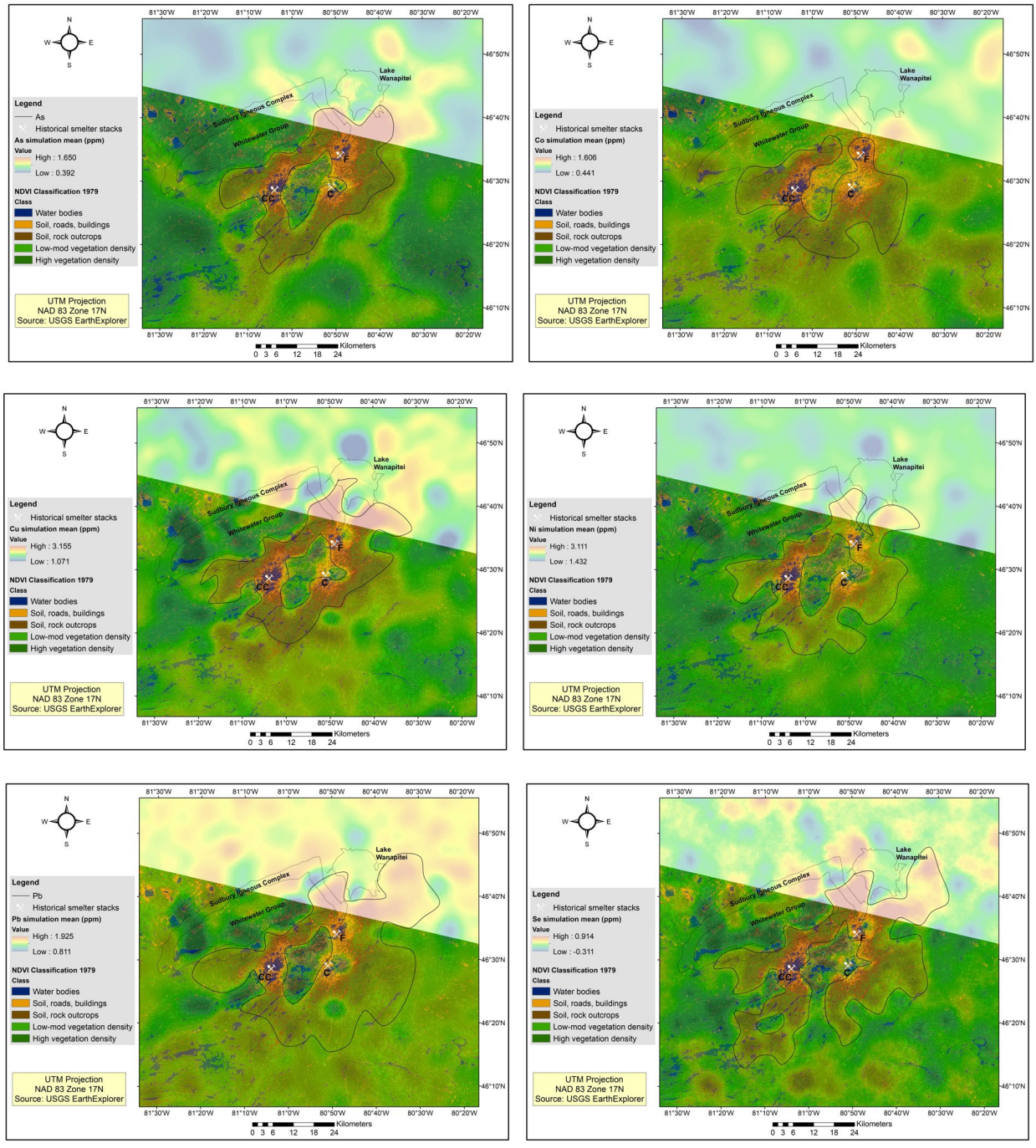


Figure 34: Observed patterns from simulated maps printed onto the 1979 NDVI classification map.

4.7. Discussion

4.7.1. Wind patterns

Most of the observed trends are causally linked to the predominant wind direction that is observed annually for the Sudbury region. The NE-SW trend that is depicted for the elements As, Cu, Ni, Co and Se, and to a lesser extent Pb, is in alignment with the prevailing wind direction (figure 5, 35). In this case, the pollutant values are observed to change more slowly in the NE-SW direction which is aligned with the wind than perpendicular to the wind (Figure 35). This is the same observation that was predicted by geostatistics, more especially looking at the anisotropic variogram examples, and its existence further confirmed by NDVI mapping. The observation comes as no surprise given that most of the contamination by trace elements in the study area was due to atmospheric deposition. A study conducted by Hutchinson and Freeman (1980) on Sudbury soils, showed that the metals, Ni and Cu in particular, appeared to have enhanced and sometimes reduced concentrations depending on the period. They placed deposition collectors at 10, 20, 30 and 60 km from the INCO (Copper Cliff) smelter to measure the deposition rates located at study sites that are 15, 25 and 30 km from the smelter. Their results showed that most of the contaminants are observed during periods when the wind is blowing directly to the smelter sites, which is 20% of the year, as compared to when the prevalent wind was not directly on the study sites, that is 80% of the year (figure 36). From their results, they determined that approximately 100% of the metals were deposited within 35 km from the INCO smelter. This observation can be seen from the current study's geochemical maps were much of the elements are distributed around the Copper Cliff smelter decreasing with distance (figure 27, 28). From figure 35 it can be concluded that soil metal concentration declines with an increase in distance from the source. However, for feasible

comparisons the fate of metals over long periods of time in these soils must be studied (Spektor, 2003).

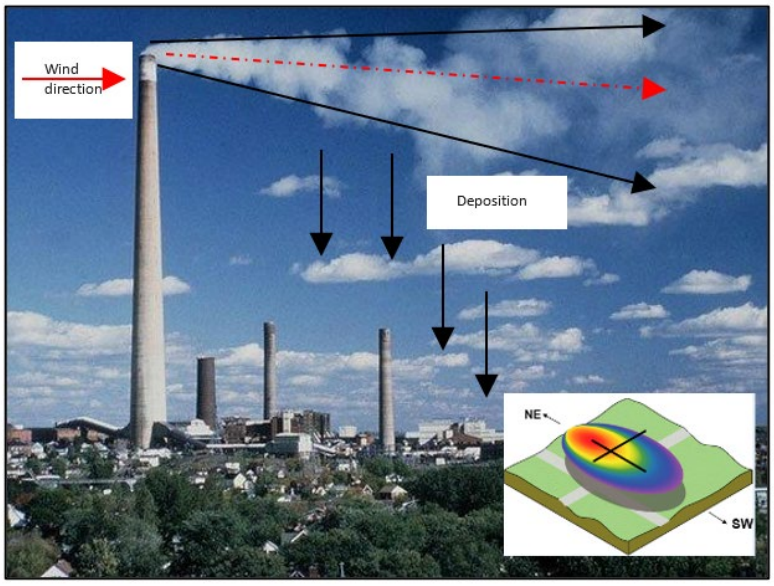


Figure 35: An image from the INCO smelter in Sudbury showing the influence of wind direction on metal deposition. The insert image (lower right corner) shows how metal concentrations are distributed with respect to the wind direction, the warm colours show high concentrations and cool colours low concentrations. https://www.myespanolanow.com/35204/the-sudbury-superstack-turns-another-chapter-in-its-history/

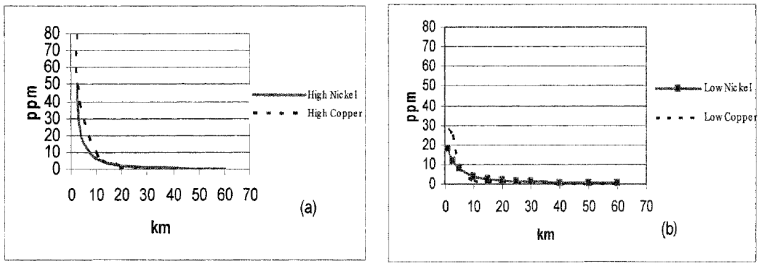


Figure 36: Deposition rates analysed by Hutchinson and Freeman (1980). Left image shows a period when the prevailing wind was blowing directly to the smelter sites and the right shows the opposite.

4.7.3 Trend variability based on metal mobility

In the Sudbury soils, studies (e.g., Dudka, Ponce-Hernandez and Hutchinson, 1995) have shown that copper is available in more mobile forms as compared to nickel. Copper is found bounded to

organic matter of which it can be easily removed. The geochemical maps from this study agreed with this. The maps show elongated trends for Cu as compared to Ni. This is due to their varying mobilities as copper is removed at a faster rate than nickel. Cu has a high affinity for soluble organic ligands, the formation of these complexes greatly increases its mobility. Ni on the other hand will adsorb to clays, iron and manganese oxides, and organic matter, hence gets removed from the soil solution (Mclean and Bledsoe, 1992), the formation of Ni with inorganic ligands limits its mobility. But however, acidic soils, as those found in Sudbury, increases the mobility of Ni (Wuana and Okieimen, 2011). The results observed from Ni and Cu are very similar to what was found in Noril'sk Russia, whereby ecosystem degradation was in a south-south-west direction at 120 kms from the smelter point.

The spatial distribution of lead appears to be patchy showing no predictable trend. This might be the result of smelting and atmospheric pollution not being the only sources of lead distribution, and that plants do not absorb lead in general (Mclean and Bledsoe, 1992).

The metalloid arsenic is also viewed to have lower values and distributed for shorter distances. Many compounds of arsenic can adsorb strongly to soils but are then transported for just over shorter distances in ground and surface water. This could explain why arsenic only has moderate values around the smelter centroid.

Concentrations of selenium are the lowest of all the elements. They are only high around the smelters, especially Copper Cliff (based on the OK maps), although the simulation maps refine the possible trends. Selenium is more mobile at higher pH values. Factors of the soil as alkaline pH, high selenium concentrations and oxidising conditions favour selenium mobility. The low mobility of selenium in these soils could be due to its low concentrations.

Cobalt is not widely distributed across the area and forms a rather unique trend. The high values of cobalt are in the vicinity of the three smelters. A detailed study on the mode of occurrence and minerals associated with the chemicals of concern in these soils would be essential in assessing further these observed behaviours.

4.6.3. Vegetation

The influence of local to regional scale biogeochemical processes, from carbon accumulation in individual trees and their underlying soils, to watershed cycling of elements is due to the reaction between soil and vegetation (Richardson and Friedland, 2016). Forest soils are responsible for the accumulation and cycling of micronutrients, macronutrients and toxic metals in terrestrial ecosystems, the cycling of plant essential metals is important for plant growth and any loss of these metals can pose a threat to the trees (Richardson and Friedland, 2016). Mining and smelting, and timber harvest are some of the main contributors to Sudbury's soil disturbance which in turn disrupted metal accumulation and cycling processes causing some land to be barren by decreasing plant growth. The type of trees in a region can be used as a valuable tool in evaluating the role of airborne pollution in plant deterioration. In the study area, the dominant tree species are deciduous and coniferous trees. Based on their leaves, deciduous trees have been found to accumulate more pollution from air than coniferous trees (Spektor, 2003). Deciduous trees shed their leaves during the fall season, hence dropping accumulates into the litter horizon. During the winter season, with falling snow the accumulates get into the snow and runs off during the melt season. This kind of process does not really explain the role of vegetation type as a driving force for the observed soil metal contamination trends. A study on the direction of the drainage systems might further elaborate such. There are a variety of roles that plant type might play in metal accumulation in the

soil environment. Further detailed study is proposed for the understanding of vegetation on metal variability trends.

4.6.4. Bedrock geology

When compared with soils of other regions on the Canadian shield, the background concentrations of soils in the Sudbury region were found to have the same values. The high mineralization of the bedrock in the region is not reflected in the background concentrations. This observation has been suggested to be due to dilution with upstream material due to glaciation and that the base metal-rich mineral phases that are hosted within the sulphide rich units of the regional bedrocks are relatively soft, and hence might have been transported and then dissolved from the surficial materials due to glacial activity and weathering (SARA, 2008). In any highly mineralized region, it is normal to find that the soil is also highly mineralized as soils are derived from the bedrocks, but the Precambrian rocks of the Sudbury region resisted ice abrasion that is responsible for forming soil materials during glacial periods. This finding rules out bedrock mineralization as a possible driving force to the high metal levels around the smelter centroid. To further validate this, rock sampling in and around the soil sampled regions might prove to be useful.

Chapter 5: Conclusions

This study attempted to use geostatistics and remote sensing information to determine the regional and geospatial distribution of trace elements in the Sudbury region and to further correlate this to the vegetation distribution in the area. The study managed to meet all the objectives that were proposed.

The study achieved the following

- The existence of two key trends operating at different scales were observed for the elements, a secondary narrow northeast-southwest and a primary broader northwest-southeast. The northeast-southwest trend can be attributed to the prevailing wind direction which forms similar patterns, and the existence of the northwest-southeast trend could be due to the geographic locations of the smelter stacks.
- NDVI images further confirmed the existence of these trends through the ovoid distribution of vegetation around the vicinity of the smelters elongated in the same directions s the observed trends. The reduction of the barren soil/land could be attributed to improved dust removal technologies, mining methods and the regreening program.
- The elements were found to have a high uncertainty in the smelter centroid, especially Copper Cliff decreasing with increasing soil profile depth and moving distance away from the smelters.
- Implementation of Gaussian Conditional Simulation to the estimates proved to be robust in elucidating the metal patterns and overcoming some of the draw backs of kriging.

- The current study's results reproduced what was observed by the Sudbury Soil Study using different approaches, however a new discovery was made which is the existence of a northwest-southeast trend in the data.
- A brief study on wind patterns, vegetation, bedrock geology and metal mobility of the region assisted in the understanding and assessing the observed trends.

The main challenge encountered during the study was that most of the elements were not normally distributed even after data transformation was applied, which also posed issues when modeling the variogram. This was not much of a concern as it is nearly impossible to find normally distributed data when dealing with sensitive values such as those of soil contamination. However, this might have been the cause of most of the issues encountered such as that the estimated and simulated values were not close to the sample data on average. Detailed research on the role in which the occurrence and behaviour of the chemicals of concern in the region played in influencing contamination levels is proposed. The results of this study can be used in future soil sampling programs, risk assessments and remediation of soil contamination by trace elements in the region.

References

Adamo, P. *et al.* (1996) 'Chemical and mineralogical forms of Cu and Ni in contaminated soils from the Sudbury mining and smelting region, Canada', *Environmental Pollution*, 91(1), pp. 11–19. doi: 10.1016/0269-7491(95)00035-P.

Berrow, M. L. and Ure, A. M. (1986) 'Trace element distribution and mobilization in Scottish soils with particular reference to cobalt, copper and molybdenum', *Environmental Geochemistry and Health*, 8(1), pp. 19–24. doi: 10.1007/BF02280117.

Borkowski, A. S. and Kwiatkowska-Malina, J. (2017) 'Geostatistical modelling as an assessment tool of soil pollution based on deposition from atmospheric air', *Geosciences Journal*, 21(4), pp. 645–653. doi: 10.1007/s12303-017-0005-9.

Cambardella, C. A. *et al.* (1994) 'Field-Scale Variability of Soil Properties in Central Iowa Soils', *Soil Science Society of America Journal*, 58(5), pp. 1501–1511. doi: 10.2136/sssaj1994.03615995005800050033x.

Chander, G., Markham, B. L. and Helder, D. L. (2009) 'Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors', *Remote Sensing of Environment*. Elsevier Inc., 113(5), pp. 893–903. doi: 10.1016/j.rse.2009.01.007.

Chilès, J.-P. (2012) 'The Generalized Variogram', (October 2012). Available at: http://cg.ensmp.fr/bibliotheque/public/CHILES_Rapport_02272.pdf.

Compositing, S., Mory, D. F. M. and Deutsch, C. V (1989) 'Inference of the Nugget Effect and Variogram Range with Sample Compositing', pp. 1–10.

Cressie, N. and Hawkins, D. M. (1980) 'Robust Estimation of the variogram. Math. Geo. 11, 12 (2).

Department of the Interior U.S. Geological Survey (2016) 'Landsat 8 Data Users Handbook', *Nasa*, 8(June), p. 97. Available at: https://landsat.usgs.gov/documents/Landsat8DataUsersHandbook.pdf.

Deutsch, J. (2014) 'Experimental Variogram Tolerance Parameters', *Geostatistics Lessons*, pp. 1–5. Available at: http://www.geostatisticslessons.com/lessons/variogramparameters.

Division, O. *et al.* (2004) 'Metal Levels in the Soils of the Sudbury Smelter Footprint', pp. 1–61.

Dudka, S., Ponce-Hernandez, R. and Hutchinson, T. C. (1995) 'Current level of total element concentrations in the surface layer of Sudbury's soils', *Science of the Total Environment*, 162(2–3), pp. 161–171. doi: 10.1016/0048-9697(95)04447-9.

Gandhi, G. M. *et al.* (2015) 'Ndvi: Vegetation Change Detection Using Remote Sensing and Gis - A Case Study of Vellore District', *Procedia Computer Science*. Elsevier Masson SAS, 57, pp. 1199–1210. doi: 10.1016/j.procs.2015.07.415.

Goovaerts, P. (1999) 'Geostatistics in soil science: State-of-the-art and perspectives', *Geoderma*, 89(1–2), pp. 1–45. doi: 10.1016/S0016-7061(98)00078-0.

Gringarten, E. and Deutsch, C. V (2001) 'Variogram interpretation'. Math. Geol. Vol. 33. No. 4.

Guagliardi, I., Cicchella, D. and De Rosa, R. (2012) 'A geostatistical approach to assess concentration and spatial distribution of heavy metals in urban soils', *Water, Air, and Soil Pollution*, 223(9), pp. 5983–5998. doi: 10.1007/s11270-012-1333-z.

Gunn, J. *et al.* (1995) 'Ecosystem recovery after emission reductions: Sudbury, Canada', *Water, Air, & Soil Pollution*, 85(3), pp. 1783–1788. doi: 10.1007/BF00477238.

Habashi, F. (2012) 'Norilsk, Russia's capital of nickel', *Metall*, 66(7–8), pp. 325–327.

Hengl, T. (2007) *7.* A Practical Guide to Geostatistical Mapping of Environmental Variables. Office for Official Publication of the European Communities, Luxembourg.

Isaaks, E.H.; Srivastava, R.M (1989). Applied Geostatistics; Oxford University Press: New York, NY, USA, 1989.

Johnston, K. et al. (2003) 'The principles of geostatistical analysis', Using ArcGIS geostatistical analyst, pp. 49–80.

Journel, A. G. (2003) 'Multiple-point Geostatistics : A State of the Art', Stanford Center for Reservoir Forcasting.

Journel, A. G. (1989) Fundamentals of Geostatistics in Five Lessons, Short course in Geology, 8, American Geophy., Union Press, Washington DC.

Klauberg, C. *et al.* (2018) 'Use of ordinary kriging and Gaussian conditional simulation to interpolate airborne fire radiative energy density estimates', *International Journal of Wildland Fire*, 27(4), pp. 228–240. doi: 10.1071/WF17113.

Lal, R. (2017) 'Trace Metal Contamination', *Encyclopedia of Soil Science, Third Edition*, (January 2016), pp. 2364–2368. doi: 10.1081/e-ess3-120053724.

Leuangthong, O., McLennan, J. A. and Deutsch, C. V. (2004) 'Minimum acceptance criteria for geostatistical realizations', *Natural Resources Research*, 13(3), pp. 131–141. doi: 10.1023/B:NARR.0000046916.91703.bb.

Lightfoot, P. C. and Zotov, I. A. (2005) 'Geology and geochemistry of the Sudbury igneous complex, Ontario, Canada: Origin of nickel sulfide mineralization associated with an impact-generated melt sheet', *Geology of Ore Deposits*, 47(5), pp. 349–381.

López-Granados, F. *et al.* (2005) 'Using geostatistical and remote sensing approaches for mapping soil properties', *European Journal of Agronomy*, 23(3), pp. 279–289. doi: 10.1016/j.eja.2004.12.003.

Mantha, N. M., Schindler, M. and Kyser, T. K. (2012) 'Silica- and sulfate-bearing rock coatings in smelter areas: Part II. Forensic tools for atmospheric metal(loid)- and sulfur-isotope compositions', *Geochimica et Cosmochimica Acta*. Elsevier Ltd, 90, pp. 221–241. doi: 10.1016/j.gca.2012.05.013.

Marcotte, D. (1995) 'Conditional simulation with data subject to measurement error: post-simulation filtering with modified factorial kriging', *Mathematical Geology*, 27(6), pp. 749–762. doi: 10.1007/BF02273536.

Matheron, G. Kleingeld, W. J. (1987) 'The Evolution of Geostatistics', *APCOM 87 - Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries*, 3(x), pp. 9–12. Available at: https://inis.iaea.org/search/search.aspx?orig_q=RN:19095057.

Mclean, J. E. and Bledsoe, B. E. (1992) 'Ground Water Issue Behavior of Metals in Soils', *Director*, pp. 1–25. doi: 10.1056/NEJMoa030660.

Mousavifard, S. M. *et al.* (2013) 'Determining and mapping some soil physico-chemical properties using geostatistical and GIS techniques in the Naqade region, Iran', *Archives of Agronomy and Soil Science*, 59(11), pp. 1573–1589. doi: 10.1080/03650340.2012.740556.

MuhaimeedAmal, A. S. (2013) 'Spatial Variability Analysis of Selected Soil Properties at Musayab, Babil, Iraq', *IOSR Journal of Agriculture and Veterinary Science*, 6(3), pp. 72–79. doi: 10.9790/2380-0637279.

Myers, D. E. and Journel, A. (1990) 'Variograms with zonal anisotropies and noninvertible kriging systems', *Mathematical Geology*, 22(7), pp. 779–785. doi: 10.1007/BF00890662.

Oliver, M., Webster, R. and Gerrard, J. (1989) 'Geostatistics in physical geograpgy. Part I theory'. 14(3). Trans. Inst. Br. Geogr. N.S. 14: 259-269 (1989) ISSN: 0020-2754 pp. 259–269.

Rajendran, S., Al-Sayigh, A. R. and Al-Awadhi, T. (2016) 'Vegetation analysis study in and around Sultan Qaboos University, Oman, using Geoeye-1 satellite data', *Egyptian Journal of Remote Sensing and Space Science*. National Authority for Remote Sensing and Space Sciences, 19(2), pp. 297–311. doi: 10.1016/j.ejrs.2016.03.005.

Rennen, G. (2009) 'Subset selection from large datasets for Kriging modeling', *Structural and Multidisciplinary Optimization*, 38(6), pp. 545–569. doi: 10.1007/s00158-008-0306-8.

Richardson, J. B. and Friedland, A. J. (2016) 'Influence of coniferous and deciduous vegetation on major and trace metals in forests of northern New England, USA', *Plant and Soil*, 402(1–2), pp. 363–378. doi: 10.1007/s11104-016-2805-5.

Saavedra, J. F., Spiers, G. and Dunn, P. G. (2007) Comparative Assessment of Geostatistical Tools : A Sudbury case study. Mining and the Environment IV Conference, Sudbury, Ontario, Canada, October 10-27, 2007.

SARA, Sudbury Area Risk Assessment Group (2008) Volume 1: Executive summary: 2001 Sudbury Soils data. SARA Sudbury Soils Study combined soils report. SARA Group January 2008.

Sparks, D. L. (2005) Toxic metals in environment: The role of surfaces. Department of Plant and Soil Sciences, University of Delaware, Newark DE 19716-2170, USA.

Spektor, V. (2003) Laurier Variability of metals in soils in the Sudbury area (Ontario). Theses and Dissertations (Comprehensive).446. Wilfrid Laurier University. Scholars Commons @Laurier.

Sui, D. Z. (2004) 'Tobler ' s First Law of Geography : A big Idea for a small world?. Annals of the Association of American Geographers, UK, 94 (2), pp 269-277.

Sundukov, Y. N. (2006) ‘First record of the ground beetle Trechoblemus postilenatus (Coleoptera, Carabidae) in Primorskii krai’, *Far Eastern Entomologist*, 165(April), p. 16. doi: 10.1002/tox.

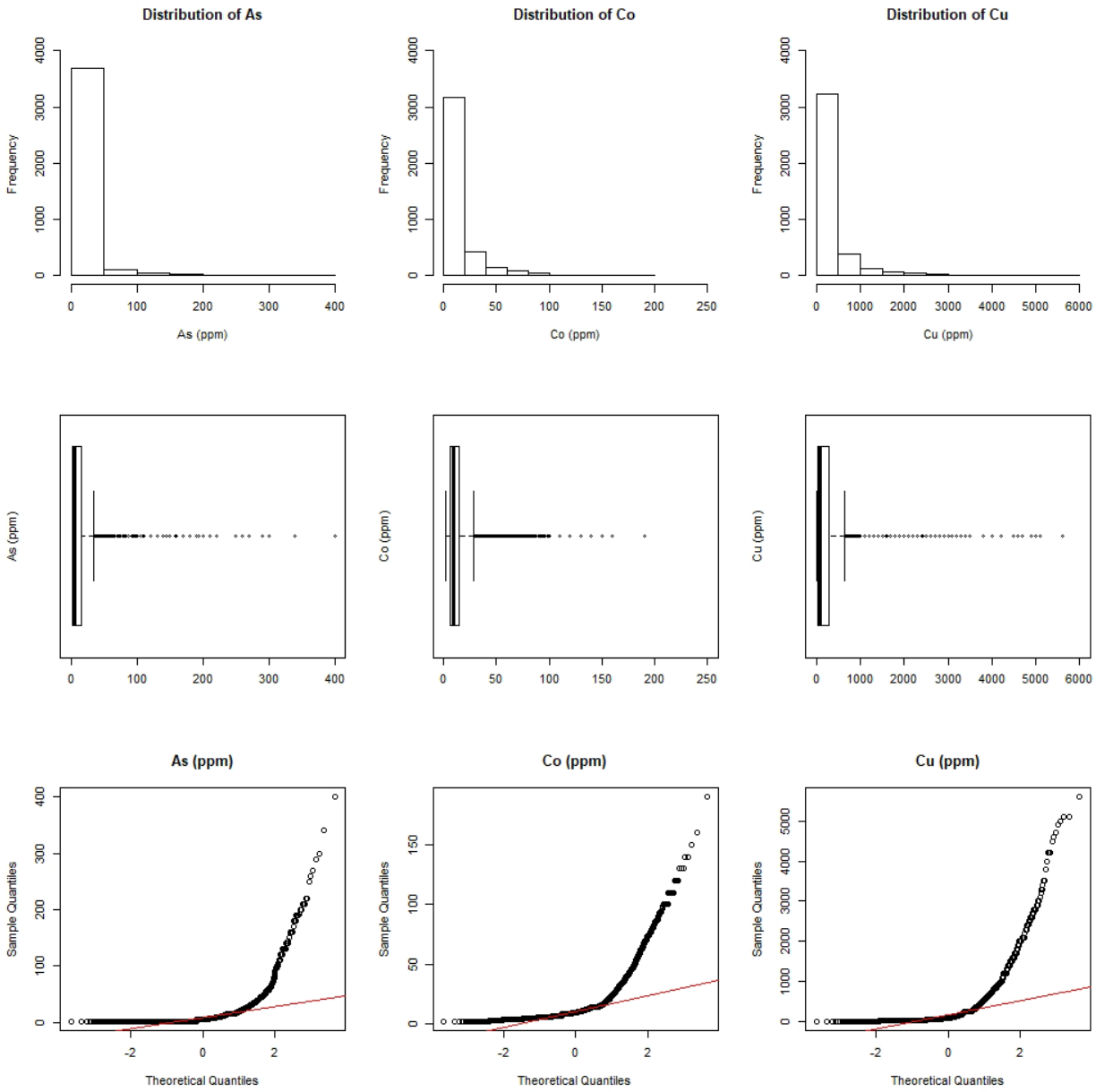
Torcal, F. *et al.* (1999) ‘Application of conditional geostatistical simulation to calculate the probability of occurrence of earthquakes belonging to a seismic series’, *Geophysical Journal International*, 139(3), pp. 703–725. doi: 10.1046/j.1365-246X.1999.00972.x.

Wuana, R. A. and Okieimen, F. E. (2011) ‘Heavy Metals in Contaminated Soils: A Review of Sources, Chemistry, Risks and Best Available Strategies for Remediation’, *ISRN Ecology*, 2011, pp. 1–20. doi: 10.5402/2011/402647.

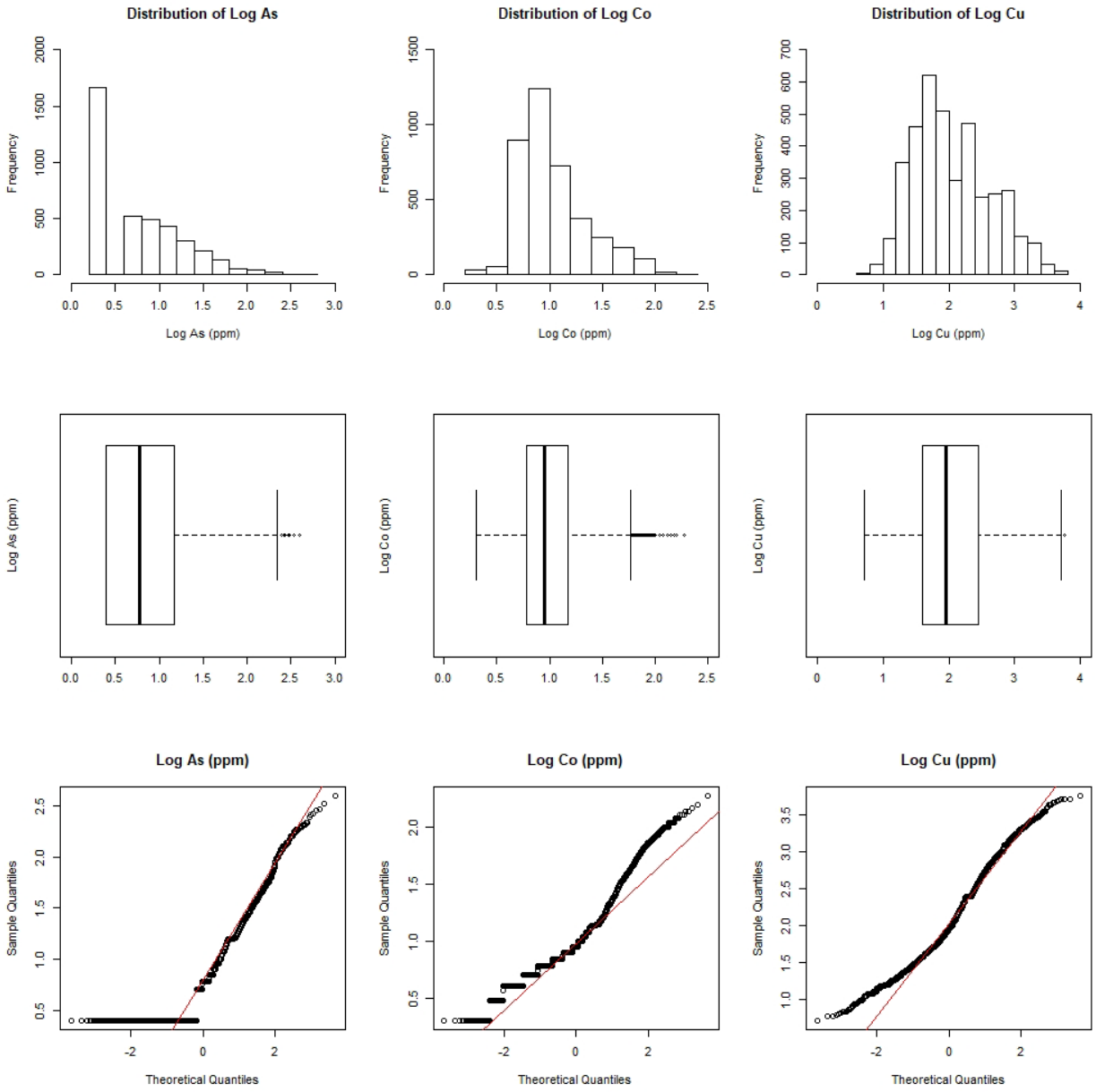
Yamamoto, J. K. (2005) ‘Correcting the smoothing effect of ordinary kriging estimates’, *Mathematical Geology*, 37(1), pp. 69–94. doi: 10.1007/s11004-005-8748-7.

Yang, P., Drohan, P. J. and Yang, M. (2020) ‘Patterns in soil contamination across an abandoned steel and iron plant: Proximity to source and seasonal wind direction as drivers’, *Catena*. Elsevier, 190(March), p. 104537. doi: 10.1016/j.catena.2020.104537.

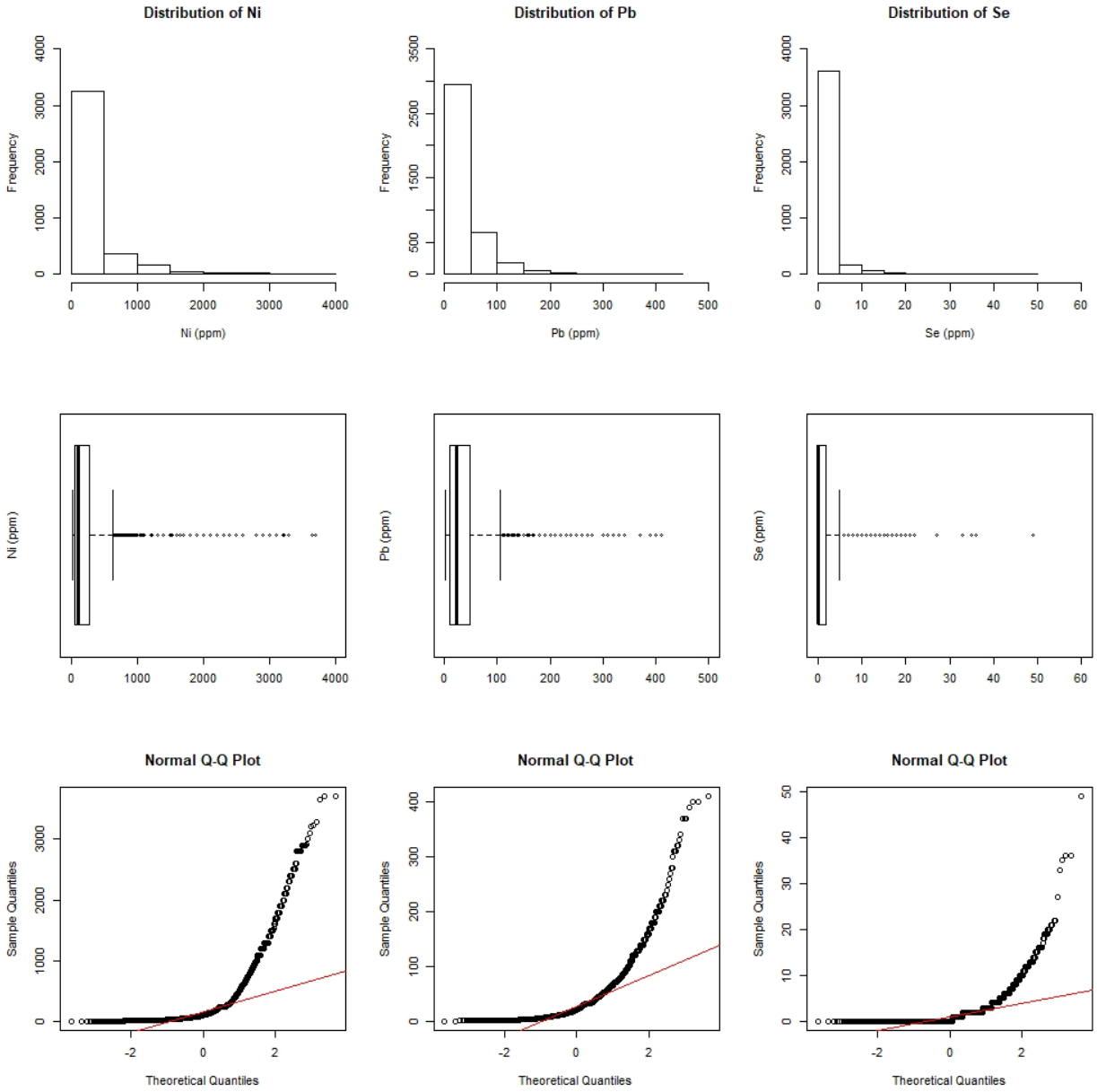
Appendix



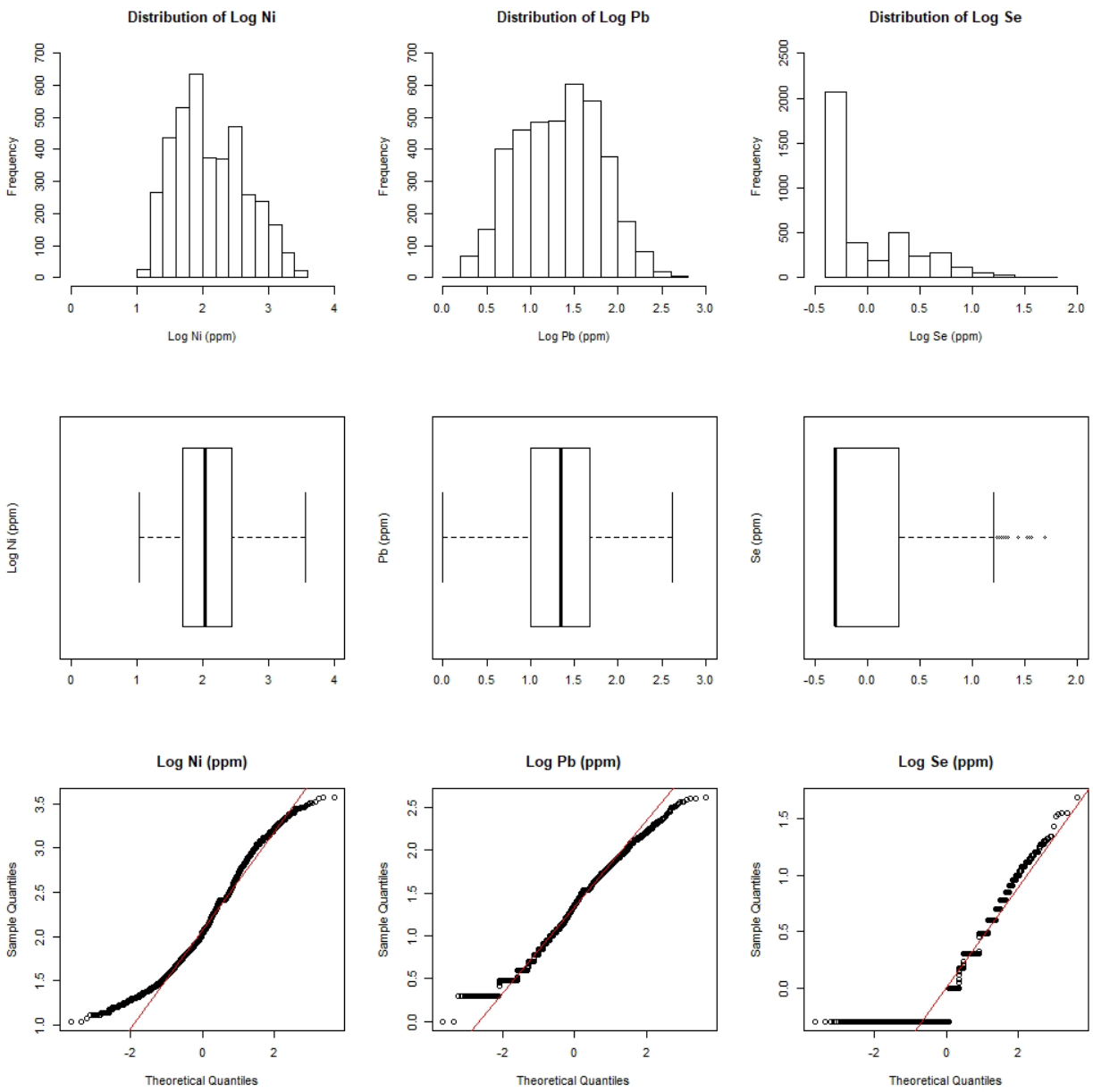
Histogram, boxplot and QQ plot for the original data set of As, Co and Cu.



Histogram, boxplot and QQ plot for the log transformed data set of As, Co and Cu.

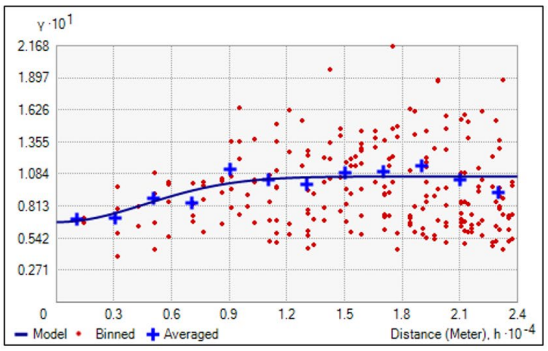


Histogram, boxplot and QQ plot for the original data set of Ni, Pb and Se.

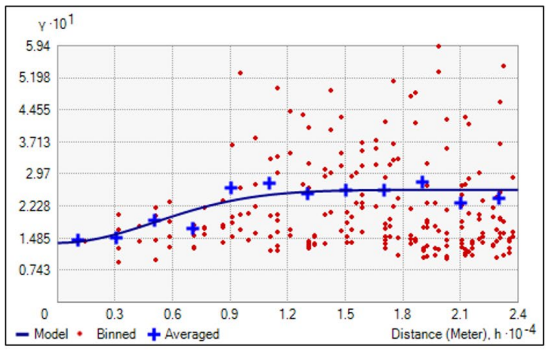


Histogram, boxplot and QQ plot for the transformed data set of Ni, Pb and Se.

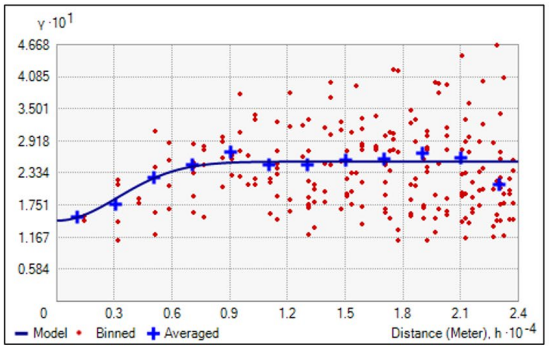
As



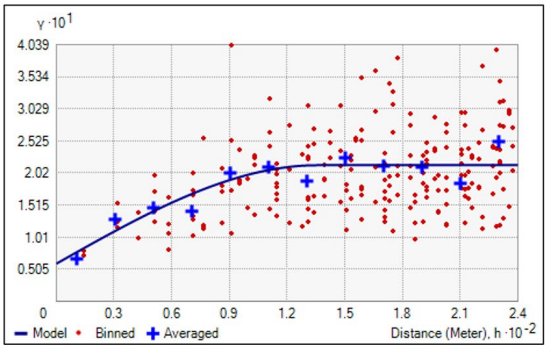
Co



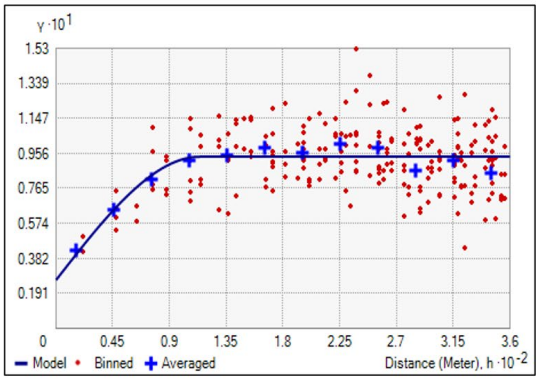
Ni



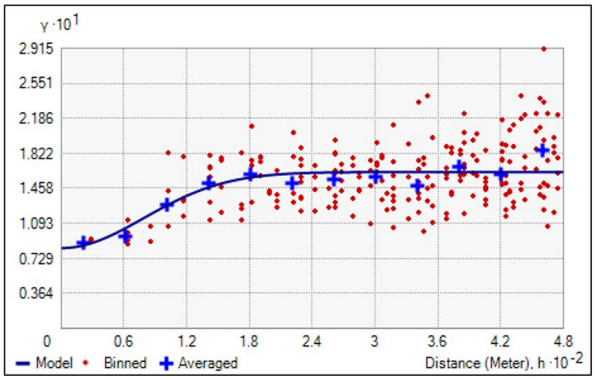
Cu



Se

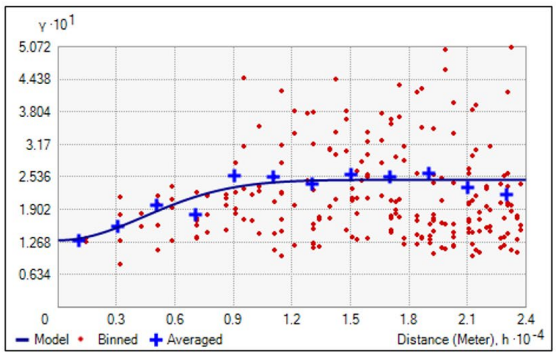


Pb

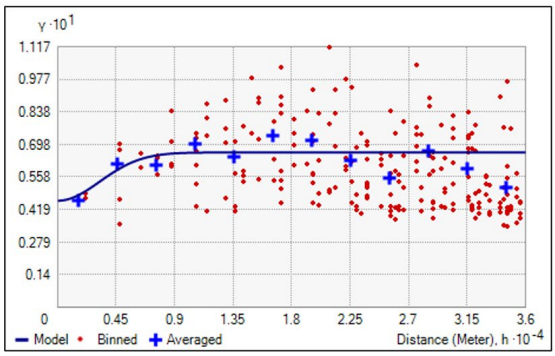


Theoretical variograms for the elements at 5-10 cm soil profile depth

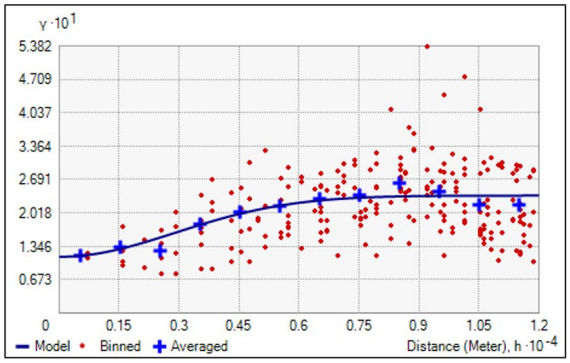
As



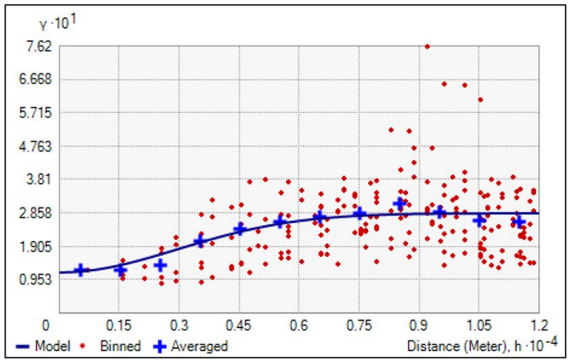
Co



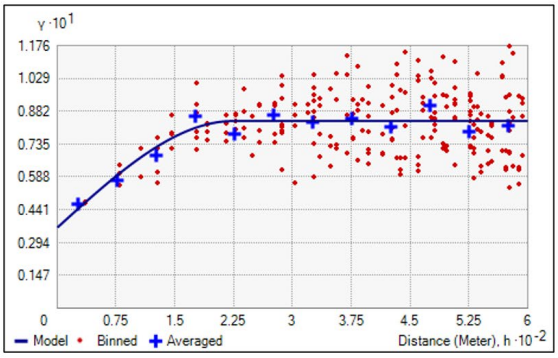
Ni



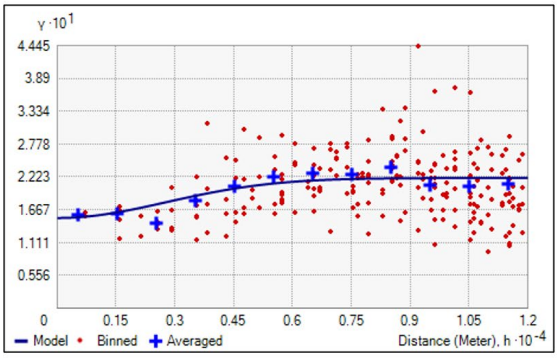
Cu



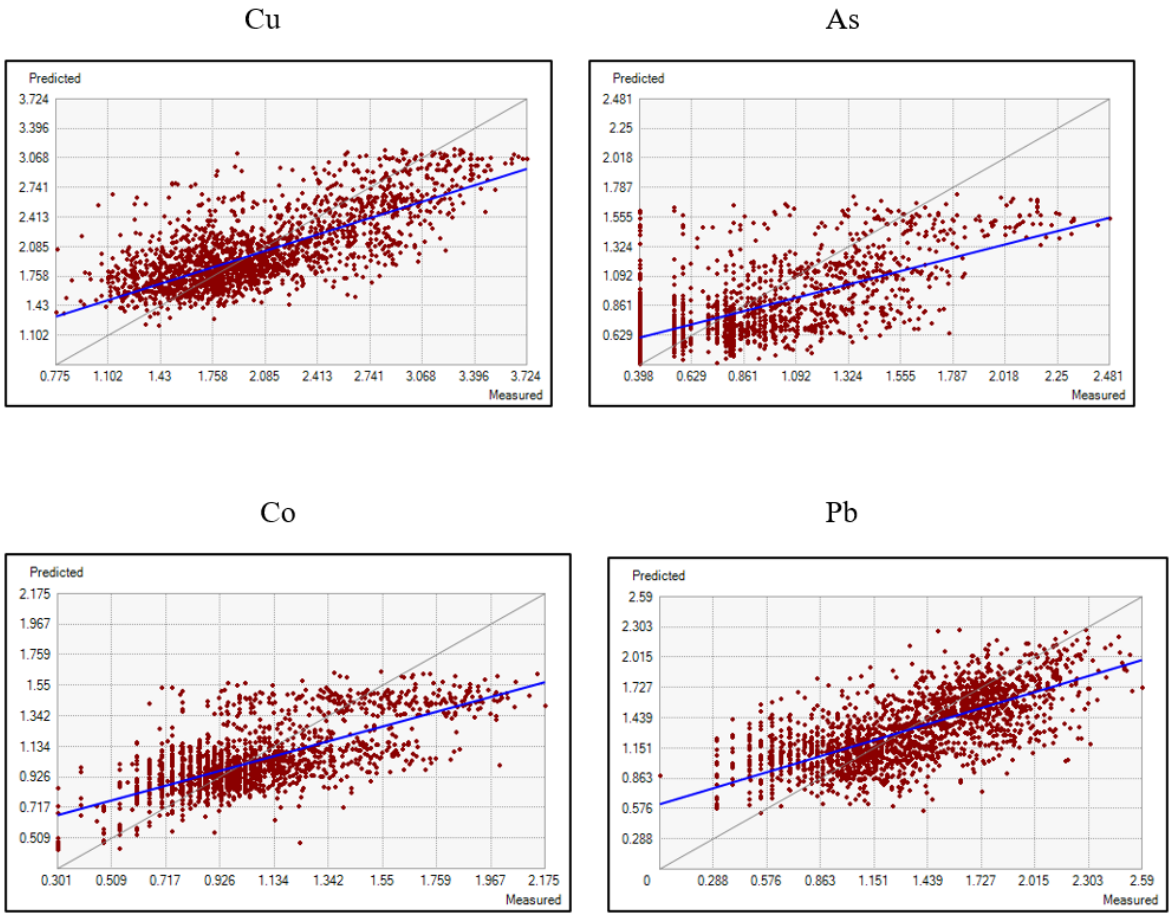
Se



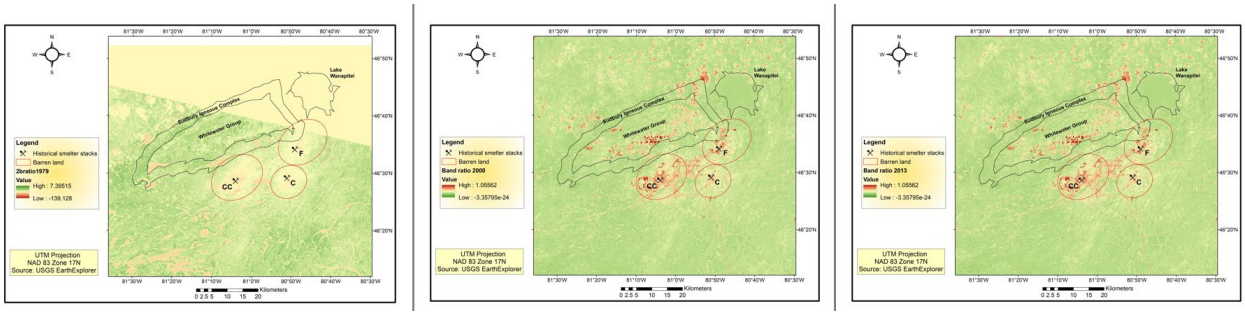
Pb



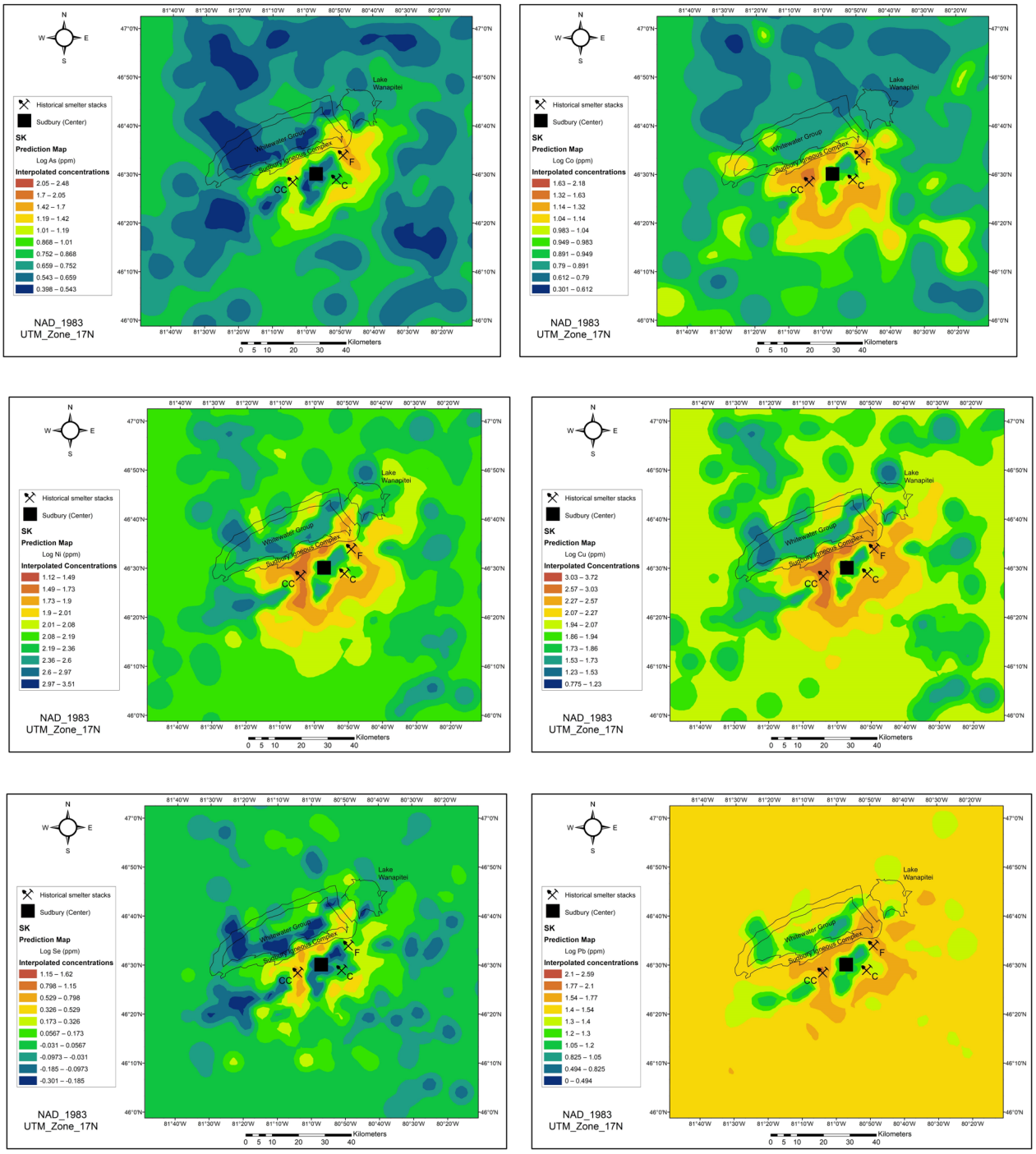
Theoretical variogram models for the elements at 10-20 cm soil profile depth.



Predicted vs. measured for the ordinary kriging estimates.



Band ratio images for the years 1979, 2000 and 2013.



Simple kriging prediction models that were used in the Gaussian Conditional Simulation.