Spatial Patterns of Soil Organic Carbon Distribution in Canadian Forest Regions: An Eco-region Based Exploratory Analysis

by

Junzhu Li

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Geography

Waterloo, Ontario, Canada, 2013

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

As the largest carbon reservoir in ecosystems, soil accounts for more than twice as much carbon storage as that of vegetation biomass or the atmosphere. The goal of this study is to examine spatial patterns of soil organic carbon (SOC) in Canadian forest area at an eco-region scale and to explore its relationship with different ecological variables. In this study, the first Canadian forest soil database published in 1997 by the Canada Forest Service was analyzed along with other long-term eco-climatic data (1961 to 1991) including precipitation, air temperature, Normalized Difference Vegetation Index (NDVI), slope, aspect, and elevation. Additionally, an eco-region framework established by the Environment Canada was adopted in this study for SOC distribution assessment.

Exploratory spatial data analysis techniques, with an emphasis on spatial autocorrelation analysis, were employed to explore how forest SOC was spatially distributed in Canada. Correlation analysis and spatial regression analysis were applied to determine the most dominant ecological factors influencing SOC distribution in different eco-regions. At the national scale, a spatial error model was built up to adjust for spatial effects and to estimate SOC patterns based on ecological and ecosystem property factors. Using the significant variables derived in the spatial error model, a predictive SOC map in Canadian forest area was generated.

Findings from this study suggest that high SOC clusters tend to occur in coastal areas, while low SOC clusters occur in western boreal eco-region. In Canadian forest area, SOC patterns are strongly related to precipitation regimes. Although overall SOC distribution is influenced by both climatic and topographic variables, distribution patterns are shown to differ significantly among eco-regions, thus verifying the eco-region classification framework for SOC zonation mapping in Canada.

Acknowledgements

First of all, I want to give the greatest gratitude to my supervisor, Dr. Su-Yin Tan, for accepting me as one of her graduate students, providing me valuable suggestions and guidance during regular supervision meetings, and encouraging me to fight for my progress. Without her, I would not be able to successfully complete my master degree. I am so appreciated for her friendship, encouragement, and edification.

I would also like to give my sincerest gratitude to my committee members, Dr. Richard Kelly, Dr. Jonathan Li, and Dr. Maren Oelbermann, for reading my paper and providing critical suggestions and comments.

Next, I want to express my thanks to Mike Lackner and James McCarthy from Mapping, Analysis and Design (MAD) Help desk for their technical support and to Que Ren for providing me useful learning materials.

Finally, I want to thank my family and friends who give me endless love and faith to support me to overcome difficulties. I want to thank my parents for being understandable and supportive through my two academic years study. Particularly, I want to dedicate this paper as a present to my grandfather. May his soul rest in peace in heaven.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

Soil is an essential resource on our planet. It has three major ecological functions: (1) it provides a foundation layer with water and a variety of nutrients to support the growth of rooted plants; (2) it maintains an important function in transferring energy between land and the atmosphere; and, (3) soil plays an important role in global organic carbon fluxes by storing organic matter (Grunwald, 2006; Plaster, 1992). As the largest organic carbon reservoir in ecosystems, soil accounts for more than twice as much carbon storage as vegetation biomass or the atmosphere (Galbraith et al., 2003; Liu et al., 2011). Globally, about 30% of soil organic carbon (SOC) is estimated to be preserved in tundra and boreal ecosystems (Lee et al., 2010; Siltane, 1997).

In forest ecosystems, three major reservoirs contribute to the carbon exchange cycle, namely vegetation, soil, and the atmosphere. Soil carbon fluxes generally consist of two processes, growth and decay. In the carbon-growth sub-cycle, forest absorbs carbon through photosynthesis processes (Fonseca et al., 2011; Trofymow et al., 2008). Around 50 percent of the carbon is then transferred back to the atmosphere by vegetation respiration (Malhi, 2002). For the remaining vegetation-absorbed carbon, some stocks are stored in vegetation in the form of photosynthate (Fonseca et al., 2011), whereas some stocks are transferred into the soil through litterfall accumulation and root systems (Malhi, 2002). In the carbon-decay sub-cycle, soil releases carbon into the atmosphere via soil respiration and SOC decomposition (Janzen, 2004). In this integrated carbon cycle, the amount of SOC is calculated as the difference between organic carbon inputs and releases. Consequently, due to effective vegetation-soil interactions and decades of accumulation, considerable organic carbon has been stored in forest soils (Conen et al., 2004; Simmons et al., 1996).

The dynamics of such large quantities of organic carbon stored in forest soil not only influence soil fertility and forest productivity (Jobbágy & Jackson, 2000), but also partly account for changes in atmospheric carbon concentration (Fonseca et al., 2011; Ju & Chen, 2005; Mishra et al., 2010; Powers & Schlesinger, 2002). Many studies have pointed out that SOC distribution is temperature-sensitive and small fluctuations in SOC could greatly affect atmospheric carbon concentration (Jenkinson et al., 1991; Shakiba &

Matkan, 2005; Tewksbury & Van Miegroet, 2007). Previous evidence suggests that the average global surface temperature has increased about 0.6 ˚C in past three decades (Bhatti et al., 2006). Thus, global warming could release carbon stocks from forest soils by altering SOC decomposition rates (Jenkinson et al., 1991).

As a result, policy making and forest resource management necessitates having a solid understanding of SOC distribution and the variables that influence observed spatial patterns (Powers & Schlesinger, 2002). Moreover, the amount of SOC is influenced by historic land use changes, human disruptions, temporal accumulation or loss, environmental impacts, and many other factors (Conen et al., 2004; Grunwald, 2006; Yuan et al., 2013). Recent studies have used geostatistical techniques and spatial regression analysis to incorporate environmental information into SOC mapping, rather than simply replying on ground soil surveys and measurements (Mishra et al., 2010; Zhang et al., 2011). Numerous studies have been conducted to improve the modelling of SOC-environment relationships (Kurz & Apps, 1999; Liu et al., 2011; Mishra et al., 2010; Powers & Schlesinger, 2002). These studies are based on two main assumptions. First, specific soil properties (e.g., SOC distribution) vary in space and through time across ecosystems, because different ecological conditions have varying impacts on pedogenic processes (Chen, et al., 2003; Grunwald, 2006; Tsui et al., 2004). This indicates that a SOC model established in one ecosystem could be inadequate for capturing SOC dynamics in other ecosystems (Powers & Schlesinger, 2002). Second, ecological factors contribute to SOC-environment relationships unequally and to varying levels in different environments (Powers & Schlesinger, 2002).

In Canada, about 4,690,000 km^2 (47% of total area) are covered by intact forest (Lee et al., 2010). This forest-dominant landscape indicates that Canada is one of the vital carbon reservoirs in the world. Consequently, many efforts have been made to estimate Canadian SOC distribution and to model SOC-environment relationships. However, most Canadian SOC studies have been conducted at the local scale with limited SOC distribution and SOC-environment modelling conducted at a national or regional scale of analysis. This thesis adopts a Canadian eco-climatic region (eco-region for short) framework to compare spatial distribution patterns of SOC in different eco-regions and to examine relationships between SOC and environmental variables in Canadian forest areas. Exploratory spatial data analysis (ESDA) techniques, particularly spatial autocorrelation analysis, are employed to explore how forest SOC is spatially distributed in Canada. Correlation analysis and spatial regression analysis are applied to determine the most dominant ecological factors influencing SOC distribution. At the national scale, a spatial error model is used to adjust for spatial effects and to estimate SOC patterns based on ecological and ecosystem property factors. A predictive SOC map is produced based on the significant variables identified in the spatial error mode.

1.1. Research Goal & Objectives

This study was conducted to address four key research questions: (1) How is forest SOC spatially distributed in Canadian forests? (2) How can the relationships between SOC and ecological variables, including climatic conditions and terrain attributes be quantified? (3) In Canada, do these relationships vary across different eco-regions and is this a sufficient classification scheme for soil zonation mapping? (4) What are the dominant ecological factors influencing Canadian forest SOC distribution?

The main research goal of this study is to examine spatial patterns of SOC distribution in Canadian forest regions at the eco-region scale and to explore relationships between SOC and various ecological variables. More specifically, the objectives of this study include:

(1) To explore the spatial distribution of SOC levels in Canadian forests in seven eco-regions: the Subarctic, Boreal, Cool Temperate, Subarctic Cordilleran, Cordilleran, Interior Cordilleran, and Pacific Cordilleran,

(2) To assess the influence of ecological factors on forest SOC stock in growing seasons at regional scales of analysis, including precipitation, maximum/mean/minimum air temperatures, Normalized Difference Vegetation Index (NDVI), slope, aspect, and elevation, and

(3) To assess how SOC-environment relationships vary geographically with respect to ecosystem properties and ecological factors.

1.2. Thesis Structure

This thesis consists of seven chapters as follows:

Chapter 1 – Introduction: provides a brief introduction to the important role that forest soils have in global carbon fluxes.

Chapter 2 – Literature Review: reviews soil surveys of Canada and the theoretical background on modeling relationships between SOC and ecological factors.

Chapter 3 – Study Area: provides a brief description of ecological conditions within the study area at an eco-region scale of analysis.

Chapter 4 – Data: describes available datasets used in this study.

Chapter 5 – Methodology: details the study's exploratory analysis workflow, focusing on the Exploratory Spatial Data Analysis (ESDA) approach and spatial regression modelling.

Chapter 6 – Results: elaborates on empirical findings of this study, including results from the ESDA analysis and spatial regression models. A predictive SOC distribution map is presented.

Chapter 7 – Discussion and conclusion: analyzes and interprets the implications of key findings of this research. Recommendations and improvements on future studies are also discussed.

Chapter 2. Literature Review of Current SOC Studies

2.1. Soil Surveys of Canada

Field-soil-survey development in Canada generally consists of three phases: early government-driven soil surveys (from the 1920s to 1974), mature government-driven soil surveys (from 1975 to 1995), and increasing private-sector-driven soil surveys (from 1996 to present) (Anderson & Scott Smith, 2011; McKeague & Stobbe, 1978). Before 1995, field soil surveys in Canada were mainly undertaken by federal pedologists and associated university professors. In recent years, however, increasing demands for detailed soil information in a diversity of applications have promoted the growth of private sector soil surveys, which are usually site- and motivation-specific (Anderson & Scott Smith, 2011).

2.1.1. Soil Surveys of Canada: Phase I - 1920s to 1974

After the completion of the first field soil survey in Ontario in 1914, Canadian soil surveys made substantial progress in the following decades. In order to maintain consistency among national field soil surveys, standard soil classification schemes and survey-related principles and instructions were established in the first National Soil Survey Committee (NSSC) meeting in 1945 (Anderson & Scott Smith, 2011; Coen, 1987; McKeague & Stobbe, 1978). Afterwards, NSSC meetings were held every three years in the following three decades, and NSSC was renamed as the Canada Soil Survey Committee (CSSC) in the early 1970s. Also, the soil cartographer group had been enlarged during the 1950s to assist with national soil mapping (McKeague & Stobbe, 1978). According to Canadian pedologists, the 1970s should be considered as a pivotal period for Canadian soil surveys development (McKeague & Stobbe, 1978). Over the last 50 years, the evolution of techniques employed in field soil surveys from hand-drawn to aerial-photographs led to more efficient and accurate field soil surveys (McKeague & Stobbe, 1978). As a result, once unattainable areas such as mountainous and northern areas became accessible and assessable (McKeague & Stobbe, 1978). Advances in transportation also sped up field soil survey development.

McKeague and Stobbe (1978) traced and reported the approximate number of published Canadian soil maps in the period from 1920 to 1974. As shown in *Table 2.1*, about 200 soil maps were completed and published across Canadian provinces until the mid-1970s. In early years, only a small number of soil maps (approximate 28) were generated due to external limitations such as lack of accessibility and transportation. At the national scale, significant increases in published soil maps were observed since 1960.

Table 2.1 Quantities of published soil maps in Canadian provinces (1920-1974)

Province	***1920-1929***	***1930-1939***	***1940-1949***	***1950-1959***	***1960-1969***	***1970-1974***
British Columbia		1	2	3	4	7
Alberta	3	5	5	5	12	8
Saskatchewan	8	2	5	2	4	2
Manitoba		1	4	6	5	7
Ontario	3	3	6	14	15	2
Quebec		2	15	9	15	2
New Brunswick			2	2	1	
Nova Scotia			2	5	8	1
Prince Edward Island			1		1	1
Newfoundland						1
Yukon and Northwest Territories				1	3	2
Total	14	14	42	47	68	33

Source: McKeague & Stobbe (1978)

2.1.2. Soil Surveys of Canada: Phase II - 1975 to 1995

The introduction of geospatial soil-environment modelling in the mid-1970s and geo-databases in the late-1970s led to more efficient soil data storage and information delivery (Grunwald, 2006). According to McKeague and Stobbe (1978), around 35% of Canada's total area had been investigated by 1975. Moreover, researchers' perception of soil science also positively led to the maturing of field soil surveys in Canada. The use of soil information in a variety of studies such as land use regulation and environmental impacts evaluation gradually became the focus of the 8th CSSC meeting in 1970 and the following epoch of soil survey development (Anderson & Scott Smith, 2011). Field soil surveys continued and were led by governmental pedologists during this period, which became the peak time for such activities (Anderson & Scott Smith, 2011; Schut, et al., 2011). For example, notable progress was achieved in British Columbia soil surveys from the 1960s to mid-1980s, resulting in an almost complete provincial landmass coverage (Anderson & Scott Smith, 2011). Also, funds were provided to Saskatchewan to support soil investigation in provincial forest areas in 1968, leading to increased capabilities for soil mapping and forest soils assessment (Anderson & Scott Smith, 2011).

2.1.2.1. *Comparison of Existing Soil Carbon Databases*

Increasing demand for detailed soil information consequently led to the establishment of the Canada Soil Information System (CanSIS) to organize and deliver soil data to its users. Specifically, three versions of a Soil Landscapes of Canada (SLC) map series (at 1:100000 scale) were produced, updated, and released via CanSIS since the 1980s (Geng et al., 2010; Schut, et al., 2011). The system's success lay in: (1) assembling and publishing existing soil surveys and maps at various spatial scales; and (2) standardizing the map-related attribute structures (Anderson & Scott Smith, 2011). Although the generalized SLC maps enable national-wide soil evaluations, they may be inefficient for more intensive and in-depth soil studies (Siltane, 1997). Meanwhile, a global soil carbon database was established by Zinke et al. based on an eco-region classification scheme (Siltane, 1997; Zinke et al., 1984). However, only 117 soil profile records were contained in the Zinke database, and uneven distribution of soil samples limited the capability of providing a full SOC description for all eco-regions (Kurz & Apps, 1999; Siltane, 1997). Thus, in order to better estimate the terrestrial soil carbon distribution, the Canada Forest Service (CFS) was developed by assembling more extensive historical SOC data (1,462 records in total).

Figure 2.1 below shows the differences between Zinke's database (solid bar) and the CFS database (shaded bar) in describing Canadian forest SOC distribution (Kurz & Apps, 1999). In each eco-region, obvious differences between the two databases can be observed from *Figure 2.1*. Due to a lack of sufficient samples, Zinke's database may give rise to biases when representing SOC distribution in the real world. SOC stock may be under- or over-estimated in small eco-regions such as the Pacific Cordilleran and Subarctic Cordilleran eco-regions. Thus, the CFS database is considered to be an improved data source for large-scale Canadian SOC studies due to its extensive sampling coverage.

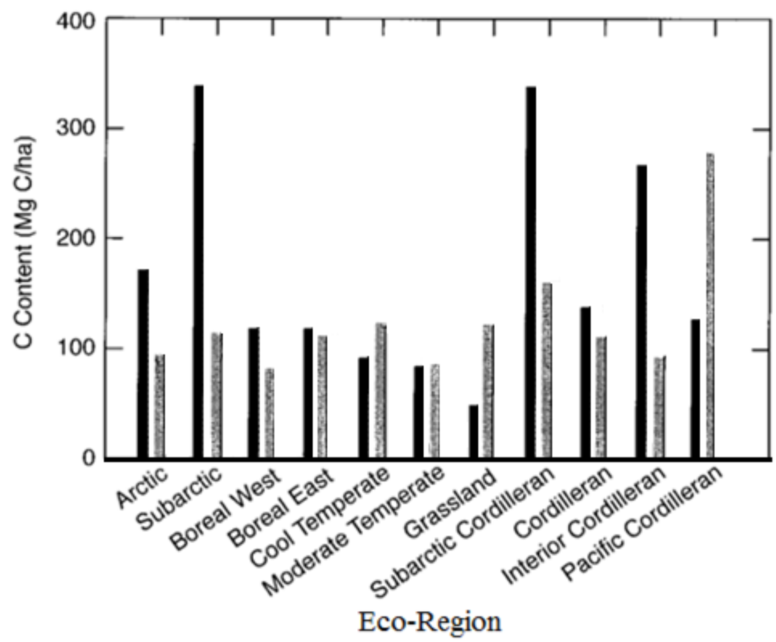


Figure 2.1 Comparison of Canadian SOC between Zinke's database and CFS database
Source: Zinke et al. (1984)

2.1.3. Soil Surveys of Canada: Phase III – 1996 to Present

Nevertheless, government-driven field soil surveys have decreased from 1995 to the present. Anderson and Scott Smith (2011) indicated that only Newfoundland and Manitoba still currently maintain their soil surveys. Administrative changes and retirement of experienced field surveyors are two major limiting factors that hinder acquisition of soil information (Anderson & Scott Smith, 2011). The key emphasis in work has currently switched from data collection to maintaining and refining the existing soil database. In contrast, private sector-driven field soil surveys have increased over the years. However, private surveys are usually undertaken at local scales or for specific projects, and not made available to public users. Insufficient availability of updated soil data is still a major challenge, especially when assessing nation-wide soil properties.

2.2. Methodology Evolutions in Estimating SOC Distribution

Existing SOC estimation approaches can be categorized into two groups, (a) the Measure and Multiply Approach (MMA) and (b) the Soil Landscape Modeling (SLM) approach (including geostatistical analysis and spatial regression analysis techniques used in this study) (Grunwald, 2006; Mishra et al., 2010; Schimel & Potter, 1995; Thompson & Kolka, 2005; Zhang et al., 2011). In early years, SOC estimations were mainly conducted at regional scales by calculating the average SOC stock within certain areas based on soil survey maps (Thompson & Kolka, 2005). MMA techniques, which have been widely employed, divide the study area into several strata. For each stratum, the SOC stock can be extrapolated by multiplying the average of observations by the stratum's total area (Mishra et al., 2010; Schimel & Potter, 1995). For example, Schlesinger (1977) selected eleven ecosystems to estimate the total SOC stock in the world. In his work, the average SOC stock was calculated within each ecosystem and multiplied by each ecosystem's total area. In doing so, Schlesinger (1977) concluded that the worldwide estimated SOC stock in the top one meter range was approximately 1,515 Pg. This number is in agreement with results obtained by Eswaran (1993) and Batjes (1996) who used similar MMA approach and obtained the results of 1,576 Pg and 1462-1548 Pg, respectively.

Although MMA is easy to implement, its accuracy is limited due to the lack of spatial variation considerations (Meersmans et al., 2008; Mishra et al., 2010). Uncertainty is substantial when using estimates at local scales of analysis (Meersmans et al., 2008). It is inaccurate to estimate an entire region's SOC stock based on a small quantity of samples and assuming that SOC distribution is homogenous at continental scales (Mishra et al., 2010; Thompson & Kolka, 2005). According to previous studies, SOC estimation is sensitive to sampling design and sampling density (Galbraith et al., 2003; Mishra et al., 2010; Thompson & Kolka, 2005). This suggests that estimation errors at un-sampled locations could likely be due to spatial variability of SOC (Eswaran 1995; Zhang et al., 2011). On the other hand, environment-induced variations, such as climatic conditions and terrain attributes, also influence SOC stock at national and regional scales (Hontoria et al., 1999; Zhang et al., 2011). Thus, more reliable approaches are required to improve the accuracy of SOC estimation at various spatial scales of analysis.

Compared to MMA, SLM is a more-effective approach for SOC estimation, since it models a statistical relationship between the SOC (dependent variable) and a set of environmental determinants (independent variables), based on samples collected across the study area (Grunwald, 2006; Mishra et al., 2010; Thompson & Kolka, 2005). Once the relationship is developed, it is able to estimate SOC stock at un-sampled locations, although at various levels of certainty (Mishra et al., 2010). According to Grunwald (2006), the first factorial-based model was developed by Jenny in the 1940s. Jenny's model is:

$$S = f\,(c, o, r, p, t) \tag{2.1}$$

where s is the specific soil properties, f is the quantitative model, c is the climatic conditions, o is the soil organisms, r is the terrain attributes, p is the parent materials, and t is the time. This factorial-based model has been widely applied by researchers in SOC distribution studies (Grunwald, 2006; Webster, 1994). It succeeds in describing the combined impacts of environmental determinants on pedogenic processes, as well as providing a baseline for assessing soil-environment interactions. Two principles should be taken into consideration when applying Jenny's model:

(1) In soil-environment modelling, it is very difficult to include all the independent variables mentioned in Jenny's model because not all the variables can be measured quantitatively, especially categorical variables such as parent materials (Grunwald, 2006). In addition, incorporating categorical variables would limit the performance of regression models due to the added complexity.

(2) The magnitude of each independent variable is restricted to specific study areas (Bergstrom et al., 2001). For example, Wang et al (2002) found that slope had no obvious impacts on forest SOC distribution in northeastern Puerto Rico. Conversely, slope is identified as one of the statistically significant environmental determinants that influence forest SOC distribution in southern Taiwan (Tsui et al., 2004). Results from the Pearson correlation analysis showed that slope is negatively related to forest

SOC collected from different depths (0-5 cm and 5-15 cm), with correlation coefficients of -0.4 and -0.36, respectively.

The selection of quantitative models (*f*) also influences SOC estimation results. Traditionally, non-spatial statistical approaches, such as Pearson correlation and multiple linear regression (MLR) analyses were widely accepted (Meersmans et al., 2008; Mishra et al., 2010; Thompson & Kolka, 2005; Zhang et al., 2011). MLR models are implemented based on three hypotheses: (1) SOC samples are independent of each other; (2) regression residuals are independent of each other; and (3) the SOC- environment relationship remains the same across the study area (Lichstein, 2002; Mishra et al., 2010). However, early soil scientists proposed that an autocorrelated variable would violate the aforementioned hypotheses (Grunwald, 2006; Webster, 1994). This viewpoint is consistent with Tobler's First Law of Geography stating that adjacent objects are more related to each other (Tobler, 1970). This has also been supported by previous research showing that SOC is strongly auto-correlated in space, because similar soil properties, as well as soil-environment relationships, are usually observed in proximal geographic areas (e.g., Legendre & Fortin, 1989; Trangmar et al, 1985; Wang et al., 2002). Leung (2000) argued that, in reality, the relationships between SOC and pertinent environmental determinants vary spatially; thus the regression coefficient for each determinant should not be assumed to be constant across the entire study area. Consequently, such a "global" statistical approach is not capable of capturing spatial variations and spatial dependence in SOC distribution at local scales (Wang et al., 2002; Zhang et al., 2011). For large-scale studies, MLR models are mainly used to represent the relationship between dependent and independent variables.

In recent years, quantitative models employed in the SLM approach have adopted advanced statistical algorithms to address the aforementioned limitations. Primary techniques include geostatistical analysis and spatial regression models, which are described in the following sections. Thus, SOC distribution can be estimated and predicted on the basis of spatial dependency and environmental correlations (Goovaerts, 1999).

2.2.1. Geostatistical Analysis

Geostatistical techniques were introduced in soil studies in the mid-1960s (Grunwald, 2006). SOC values at unsampled locations are estimated using interpolation methods, such as Inversed Distance Weighting (IDW) and Kriging, based on a set of collected SOC samples within a user-defined neighbourhood radius. Thus, continuous SOC distribution throughout the study area is estimated from discretely collected samples (Liu et al., 2011). In the Kriging method, spatial dependence is examined by the semi-variogram *γ*, which calculates the variance between each pair of samples (Bergstrom et al., 2001; Grunwald, 2006; Liu et al., 2011; Wang et al., 2002):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \tag{2.2}$$

where γ is the semi-variogram, h represents each lag distance, $Z(x_i)$ is the sample's value at location x_i, and N is the number of data pairs. The semi-variogram is illustrated in *Figure 2.2*: range is the smallest distance at which SOC samples are not spatially correlated with each other, C_0 is the nugget which is caused by measurement errors and micro-scale (distances that are shorter than sampling intervals) variations (Cressie, 1988), and *partial sill* is the variance caused by environmental factors (Qiu et al., 2011).

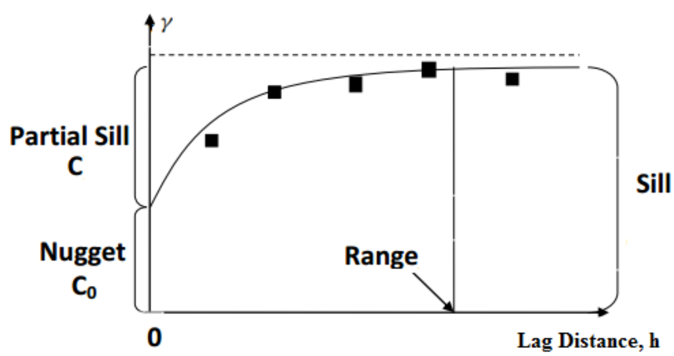


Figure 2.2 Illustration of semi-variogram parameters
Source: Qiu et al. (2011)

Ettema and Wardle (2002) went a step further to explore different semi-variograms and the corresponding surface patterns of soil properties (e.g., SOC and soil biota distribution). As shown in *Figure 2.3*: (a) in this situation, strong local variations are observed, and surface patterns are spotty, reflecting spatial clusters at local scales; (b) represents large-scale variations, showing smoother surface patterns; and (c) represents a combined situation, indicating that variations are observed at different spatial scales. Thus, Ettema and Wardle (2002) provide deeper insight into the interpretation of spatial patterns using geostatistical techniques.

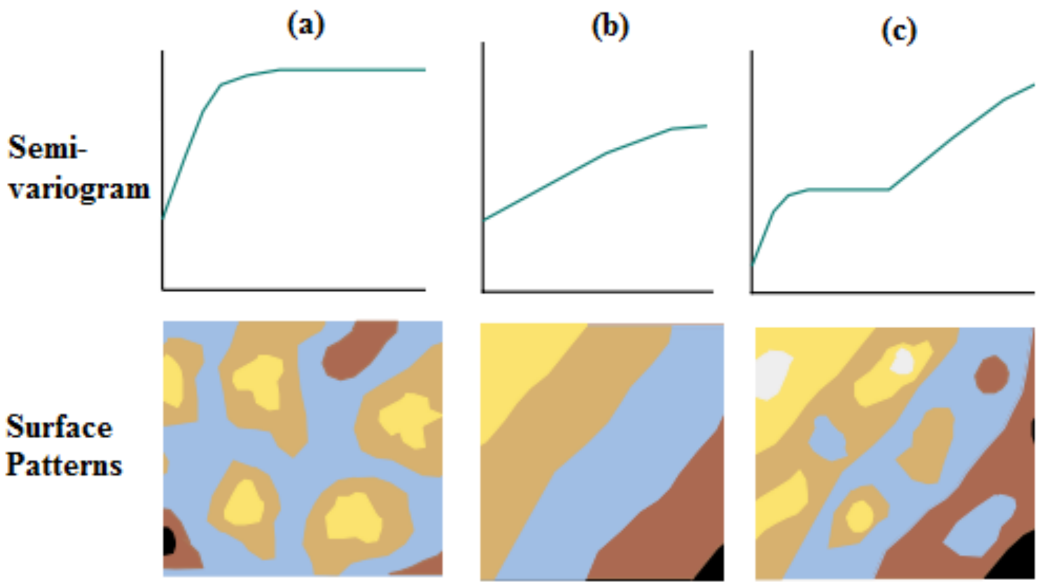


Figure 2.3 Semi-variogram and corresponding surface patterns of soil properties
Source: Ettema and Wardle (2002)

McGrath and Zhang (2003) employed a semi-variogram to examine the spatial pattern of SOC distribution in Ireland grassland. Results revealed that SOC distribution showed strong local variations, with a relatively high nugget and sill ratio (45%). Chuai et al. (2012) used a Kriging method (with 12 neighbours) to generate a SOC stock map in Jiangsu Province, China. Interpolation results showed that surface SOC stock varied from 3.25 kg/ m^3 to 32.43 km/m^3, which was a narrower range than that of the raw field collected samples. They explained that this was an expected result because spatial interpolation techniques remove outliers and make the spatial patterns of SOC distribution relatively smooth (Chuai et al., 2012).

A similar study has been conducted by Wang et al. (2002), who employed geostatistical techniques to investigate the spatial distribution of SOC in Puerto Rico's Luquillo Experimental Forest area (LEF, approximately 110 km^2) and to explore spatial relationships between SOC and a set of ecological variables, which include soil moisture (SM), elevation, slope, and aspect. They found that SOC stock shows a relatively large range of spatial dependence (approximately 3 km), indicating that SOC is not randomly distributed in LEF and that the spatial distribution of SOC is likely controlled by both climatic conditions and terrain attributes within proximal areas (Wang et al., 2002). In addition, Pearson correlation analysis was applied to provide a baseline for evaluating relationships between SOC and ecological variables. Results revealed that SOC is positively related to SM and elevation across the entire study area, with the correlation coefficients of 0.86 and 0.57, respectively. However, no detailed information on how these relationships vary in space was provided. Wang et al. (2002) went one step further to examine SOC-SM relationship, by using a cross-correlation approach. Results suggested that the strength of the positive correlation between SOC and SM decreased with increasing separation distances. As shown in *Figure 2.4*, when the separation distance exceeded a certain range (approximately 5,000 m for SM), a negative correlation was observed between SOC and SM.

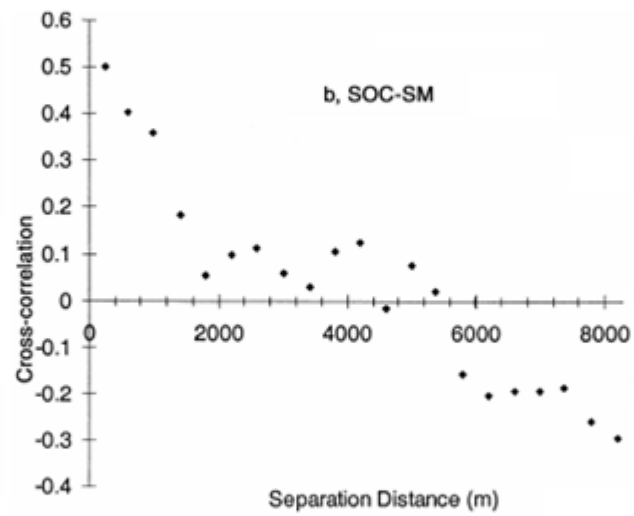


Figure 2.4 Spatial correlation between SOC and soil moisture
Source: Wang et al. (2002)

2.2.2. Spatial Regression Models

Statistically, a set of spatial autocorrelated dependent variables can lead to clustered regression residuals (Collins et al., 2006). Voss et al. (2006) explained that if independent variables fail to account for the spatial autocorrelation in the dependent variable, the autocorrelated component will remain in the regression model's error terms, with higher or lower errors tending to cluster together. Since the phenomenon of spatial dependence intrinsically breaks the assumptions of non-spatial regression models, spatial autocorrelation test statistics (e.g., Moran's *I* test) are usually applied prior to any statistical analysis to test for spatial dependence. Using the Moran's *I* test, spatial patterns (e.g., the magnitude of spatial autocorrelation and the distribution of spatial clusters) of the targeted objects are identified. If a high magnitude of spatial autocorrelation is detected, spatial regression models should be employed for further statistical analysis. For example, Huo et al. (2011, 2012) used the Moran's *I* test statistic to examine the spatial distribution of heavy metals in Beijing agricultural soils. They found that heavy metals show positive and relatively strong spatial autocorrelation in space.

In general, spatial dependence is incorporated into regression models in two ways: (1) a spatial lag effect on the dependent variable; and (2) an error term, which refers to spatially correlated residuals (Anselin, 2009; Ward & Gleditsch, 2007). Accordingly, spatial lag models and spatial error models are designed to adjust for the effects of spatial dependence. The former add a spatially lagged component to regression models, while the latter ones assume that a regression model's errors are spatially correlated (Dormann et al., 2007; Ward & Gleditsch, 2007). When the influences on a dependent variable at location i, y_i, are dominated by its neighbours, y_n, spatial lag models should be applied (Ward & Gleditsh, 2007). Otherwise, when the influences on the dependent variable are caused by omitted independent variables, spatial error models is the appropriate model specification (Dormann et al., 2007).

Consequently, both spatial regression models are superior to traditional regression models in that they allow researchers to take spatial dependence into account. However, although spatial regression models have been widely applied in social science studies (e.g., crime mapping and disease diffusion), these statistical methods, or models, are seldom employed in current ecological studies (Dormann et al., 2007). This leads to a gap in the literature that has been identified for this study for analyzing SOC-environment interactions using spatial statistics and modelling techniques.

An alternative method to account for spatially correlated residuals is the use of a Geographic Weighted Regression (GWR) model. It is a function based on a pre-defined spatial kernel and is used to capture local variations in SOC-environment relationships. Only the samples within the spatial kernel are involved in statistical calculations (Zhang et al., 2011). The further a sample observation is located away from the given center point, the less weight will be assigned to it (Foody, 2004). Thus, instead of assuming the SOC distribution follows the same spatial pattern within the entire study area, the calculated regression coefficients for different environmental determinants vary in space because of different weights (Mishra et al., 2010). For example, Mishra et al. (2010) applied a GWR model to estimate SOC distribution at the regional scale in the Midwestern United States. A spherical semivariogram model was used to determine the appropriate radius of the spatial kernel. Results revealed that the GWR model was capable of reducing smoothing issues caused by interpolation. Thus, Mishra et al. (2010) concluded that GWR could improve the overall estimation accuracy and generate more-reliable results. Similar conclusions have been found in other studies, such as by Zhang et al. (2011) and Scull (2010).

2.3. Useful Ecological Factors in Soil-Environment Relationship Modelling

In general, key environmental determinants in modelling relationships between SOC and the environment at regional scales can be classified into two groups: climatic regime variables and terrain attributes (Jobbágy & Jackson, 2000; Mishra et al., 2010; Scull, 2010; Zhang et al., 2011). In the following two sections, the impact of each environmental determinant on forest SOC distribution is discussed.

2.3.1. Climate Factors

Shakiba and Matkan (2005) stated that specific soil properties, such as SOC distribution, are relatively sensitive to different climatic conditions. Similar conclusions were drawn by Birkeland (1974), namely that most soil types and soil morphology are observed only in specific climatic regions. In his publication, Birkeland (1974) compared the American soil classification map with accumulated precipitation and averaged temperature maps in summer and winter months, and concluded that American soil zonation generally followed climatic gradients. Therefore, at national and regional scales, the most notable impacts on SOC distribution are considered to be climate variables

(Birkeland, 1974). Many researchers have provided the supportive viewpoint that abiotic factors (e.g., climatic conditions and terrain attributes) are more dominant than the amount and magnitude of organic carbon inputs (e.g., litterfall, fine root turnover, soil microbial residues, and animal residues) in terms of influencing SOC distribution in forest ecosystems and at synoptic scales (Kirschbaum, 1995; Simmons, 1996).

In general, the impact of temperature on northern forest SOC distribution is two-fold (Jenkinson et al., 1991). On one hand, air temperature changes alter forest SOC stock indirectly by influencing vegetation growth and regeneration (Bhatti et al., 2006; Jenkinson et al., 1991). Increasing temperatures extend the length of growing seasons, and thus enlarge the potential quantity of litterfall accumulation (Bhatti et al., 2006; Jenkinson et al., 1991; Kirschbaum, 1995).

On the other hand, air temperature changes directly alter northern forest SOC stock by influencing soil temperatures. Oechel and Vourlitis (1995) argued that high-latitude soils in northern forest ecosystems, including boreal forest soils and transitional forest-tundra soils, are more influenced by temperature than low-latitude forest soils. They explained that soils in northern forest ecosystems are usually characterized as cold and wet. The increases in global air temperature caused by climate change could potentially give rise to higher soil temperatures, and thus shift the balance between soil respiration and SOC accumulation, as well as increase the rate of SOC decomposition (Oechel & Vourlitis, 1995). A similar viewpoint was supported by Jobbágy and Jackson (2000), who suggested that temperature influences SOC distribution mainly through accelerating or decelerating the organisms' decomposition processes.

To date, the influence of temperature on SOC distribution remain controversial. Kirschbaum (1995) and Tewksbury and Van Miegroet (2007) suggest that increasing temperatures result in a higher SOC decomposition rate, and thus lead to reduced SOC stock. A divergent result was suggested by Gifford (1992) that temperatures generate very weak impacts on changes in SOC stock. Thornley and Cannell (2001) also pointed out that no obvious evidence has been found to support a negative correlation between temperature and SOC because of this contradictory mechanism in forest ecosystems: increasing temperatures tend to increase the amount of litterfall inputs and accelerate

SOC decomposition rates. Thus, conclusions have been made that the impacts of temperature on SOC distribution are difficult to simulate, and vary according to different study areas.

In comparison to temperature effects, the impact of precipitation on SOC distribution is usually observed to be dominant and positive. First, a large amount of precipitation results in high levels of soil moisture, which tend to reduce SOC decomposition rates by slowing and restricting oxygen diffusion processes (Deluca & Boisvenue, 2012). In addition, an adequate amount of precipitation promotes forest growth, and thus increases potential carbon inputs into the soil (Bhatti et al., 2006). Therefore, soils in humid eco-regions usually accumulate more organic carbon (Buringh, 1984; Davidson et al., 2000).

2.3.2. Terrain Factors

Previous studies have suggested that primary terrain attributes (e.g., elevation, slope, aspect, drainage capacity, and vegetation biomass) have notable influences on SOC distribution (Birkeland, 1974; Grunwald, 2006). For example, Terra et al. (2004) and Bou Kheir et al. (2010) pointed out that quantifying SOC-environment relationships is quite sensitive to study areas characteristics. They suggested that different topography features could result in unique spatial patterns of terrestrial carbon distribution. In SOC-landscape modelling, topography is considered to be the most dominant influence on pedogenic processes, due to the dependence of both micro-climatic regimes and other terrain attributes on elevation gradients (Birkeland, 1974; Tewksbury & Van Miegroet, 2007). For example, mountainous areas usually experience colder temperatures compared to flat areas, resulting in a shorter growing season (Colpitts et al., 1995).

Elevation gradients also create different slope and aspect conditions, which greatly contribute to spatial variations in SOC distribution by changing the movement of underground waterflow (Birkeland, 1974; Colpitts et al., 1995). In steep slope areas, waterflow velocities are relatively higher than those in flat areas, resulting in rapid nutrient loss from soils. Therefore, downslope soils tend to receive and hold more organic matter than upslope soils due to higher levels of water-saturation (Colpitts et al., 1995), indicating a negative impact of slope position on SOC distribution. A similar relationship between SOC distribution and slope was supported by Terra et al. (2004), who reported a negative correlation between SOC and slope in central Alabama. Rather than affecting waterflow velocity, the aspect of slopes influences SOC distribution by controlling the direction of waterflow, and thus determining the location of organic matter accumulation (Colpitts et al, 1995; Tsui et al., 2004). Another useful terrain attribute is soil drainage capacity, which has similar effects on SOC distribution as slope. In rapid- and well-drained areas, soils are usually observed as loose and penetrable, and thus organic matter can easily run off with underground waterflow (Clopitts et al., 1995). This phenomenon is widely observed in various ecosystems (e.g., Ju & Chen, 2005; Meersmans et al., 2008). For example, Meersman et al (2008) found a negative relationship between SOC stock and drainage capacity in Flanders, Belgium.

In addition, forest soil properties are influenced by vegetation cover and composition (Tsui et al., 2004). As previously discussed, vegetation cover influences SOC input and distribution through litterfall accumulation. A study by Chen et al. (2003) found that organic carbon distribution is positively related to forest age. This viewpoint was supported by Luyssaert et al. (2008), which found that old-growth forest is more capable of SOC sequestration than young-growth forest. *Figure 2.5* shows the distribution of forest age in Canada from 1973 to 1998. Young-growth forests (mainly due to forest fires) are commonly found in mid-western Canada, whereas old-growth forests are predominant in western and eastern Canada (Luyssaert et al., 2008). This distribution is consistent with the CFS database records, which show high SOC stock in coastal regions.

2.4. Chapter Summary

To date, Canadian SOC studies mainly focus on carbon estimations made at local scales of analysis (e.g., Bhatti et al., 2002; Chen et al., 2003; and Kurz & Apps, 1999). Little effort has been made to simulate and evaluate the SOC-environment relationships in Canadian forest areas regional scales of analysis. More specifically, there is a lack of accurate and updated national soil survey data available for current Canadian forest SOC estimations.

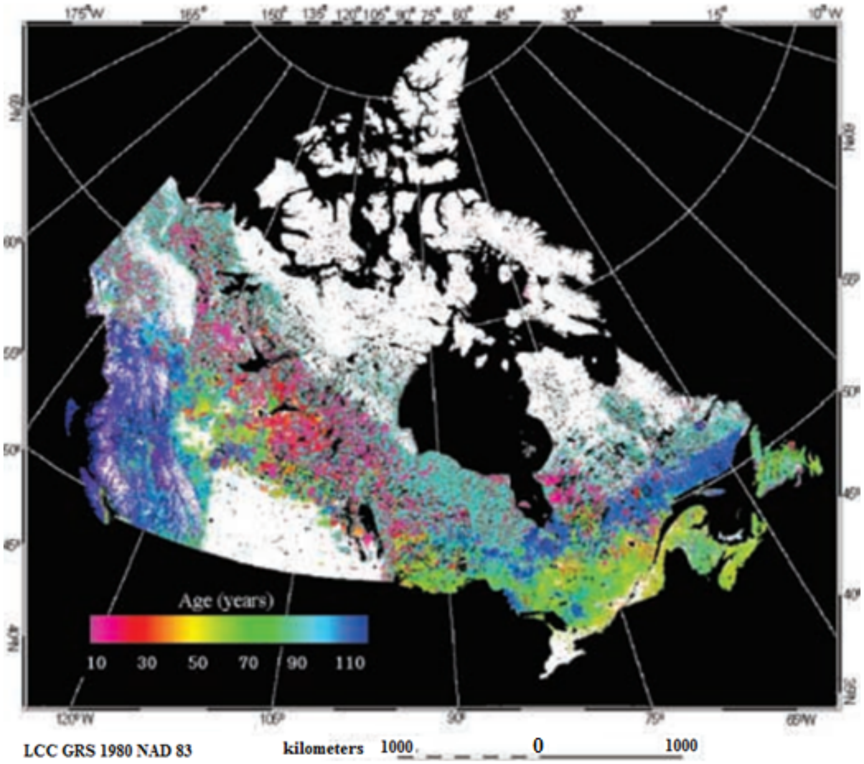


Figure 2.5 The distribution of forest age for the previous decades from 1973 to 1998
Source: Chen et al. (2003)

A geostatistical analysis and spatial regression modelling approach can be used to incorporate spatial effects for predicting patterns of SOC distribution and for modelling SOC-environment relationships. Although spatial regression models have been widely accepted by researchers, the use of such models still remains uncertain because it is difficult to determine whether the correlated residuals are caused by local variations or by the omitted environmental determinants (Jetz et al., 2005). Thus, an assumption is made that spatial regression models provide better performance when a set of independent variables that are capable of fully describing SOC-environment relationships is available for use. Moreover, the SOC-environment relationships are difficult to estimate, because the ecological influences on SOC distribution are manifold and complex. Organic carbon accumulation and decomposition processes occur simultaneously in ecosystems (Grunwald, 2006), which are difficult to quantify in statistical models. Based on the aforementioned discussion, useful attributes in soil-environment modelling are summarized in *Table 2.2*.

Table 2.2 Useful criteria in modelling the SOC-environment relationships

Criteria		***Description***	***Sample Research Papers***
Terrain Attributes	Parent Material	Foundational materials in pedogenic processes that provide a basis for soil generation.	Grunwald, 2006.
	Land Cover Land Use	The coverage of land surfaces. Historical changes of land cover and land use influence soil generation, resulting in different soil properties.	Conen et al., 2004; Grunwald, 2006; Yuan et al., 2013.
	Topography	Primary terrain attributes, including slope, aspect, elevation, and drainage capacity.	Birkeland, 1974; Tsui et al., 2004.
	Vegetation Biomass	Vegetation provides carbon inputs into soils.	Chen et al., 2003; Tsui et al., 2004.
Climate Conditions	Temperature	Major environmental determinants in SOC-environment relationships which influence SOC distribution by changing decomposition rates and litterfall accumulation.	Jenkinson et al., 1991; Jobbágy & Jackson, 2000.
	Precipitation		Birkeland, 1974; Deluca & Boisvenue, 2012.
Other Soil Properties		Internal determinants in SOC-environment relationships, including Potential of Hydrogen (PH) values, soil nitrogen, and different soil types.	Islam & Weil, 2000. Riha et al., 1986.
Human Disturbances		Human-induced impacts on SOC distribution, include forest clearance and forest fires.	Grunwald, 2006; Rashid, 1987.
Time		The impacts on SOC distribution caused by the changes of environmental determinants.	Grunwald, 2006.
Sampling Considerations		The model parameters that affect the performance of statistical analysis, including sampling design, density of observations, and the total number of observations.	Gallardo, 2003; Grunwald, 2006; Yuan et al., 2013.

Chapter 3. Study Area

The study area for this thesis is mapped in *Figure 3.1*, focusing on forest covered areas in Canada, while excluding tundra areas, grassland, and main waterbodies. Rowe (1972) provides a description of Canadian forest areas, focusing on a general identification of vegetation categories and their areal distribution. In his work, each forest region is delineated as one geographic zone that shares similarities in both ecological conditions and dominant vegetation species compositions. *Figure 3.2* below shows the distribution of Canadian forests based on statistics undertaken by the Canada Forest Service in the 1970s. In Canada, eight forest regions are identified on a macro-scale level, including: the Boreal region, Subalpine, Montane, Coast, Columbia, Deciduous, Great Lakes-St. Lawrence, and Acadian (Rowe, 1972). The Boreal region is notably the largest constituent of Canadian forest areas.

As shown in *Figure 3.1*, seven eco-climate regions are delineated: the Subarctic, Boreal, Cool Temperate, Subarctic Cordilleran, Cordilleran, Interior Cordilleran, and Pacific Cordilleran. The treeline (represented by a black dashed line) or altitude above which less trees grow is located in the subarctic eco-region is also visible in *Figure 3.1*. MacDonald and Gajewski (1992) stress that the tree line is not a specific curve that explicitly separates the forest and non-forest areas. Instead, it represents a transitional zone consisting of forest and other northern surface features (e.g., tundra). Due to lack of sufficient soil samples, the moderate temperate eco-region (containing only two soil samples and located in the most southern part of Canada), is excluded from the study. Since vegetation growth is restricted by climatic conditions, Rowe (1972) suggested that the distribution of forest regions follows the patterns of macro-scale eco-climatic gradients quite closely. Comparing the distribution of Canadian forest regions (*Figure 3.2*) to the eco-climatic zones (*Figure 3.1*), we can observe great a high degree of similarity between the spatial distribution of different forest types and eco-climatic zonation across Canada. According to the Ecoregions Working Groups (1989), each broad eco-region is distinguished and defined based on its ecological responses (e.g., vegetation types, soils types, hydrological conditions, and biota) to different climatic regimes. Detailed descriptions of ecological conditions for each eco-region are summarized in *Table 3.1*.

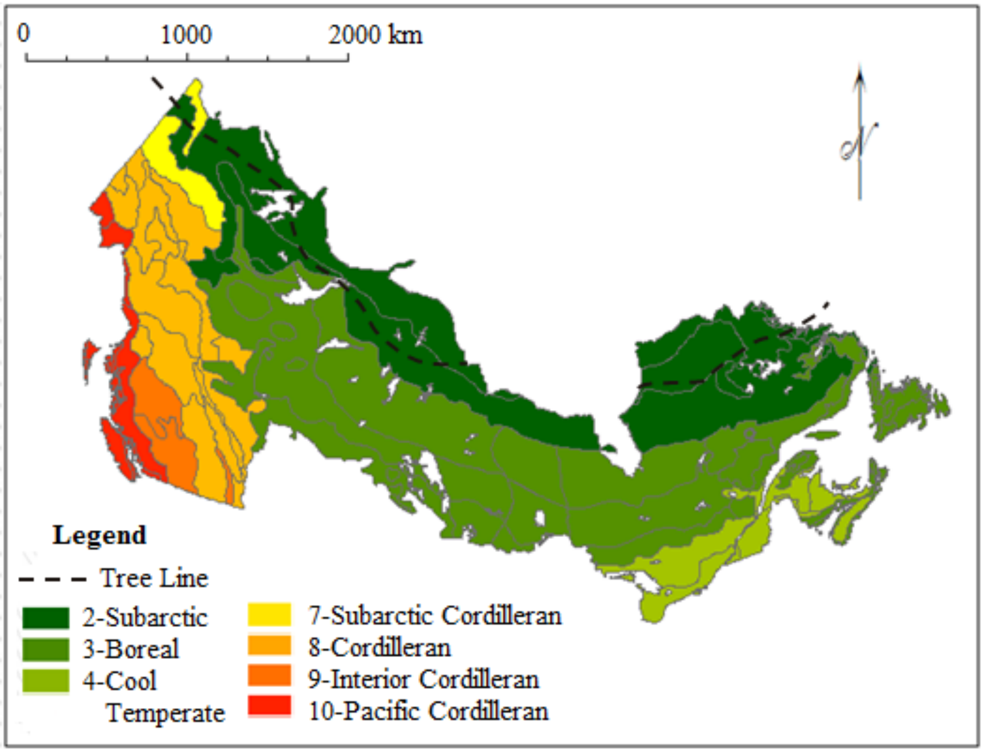


Figure 3.1 Study area

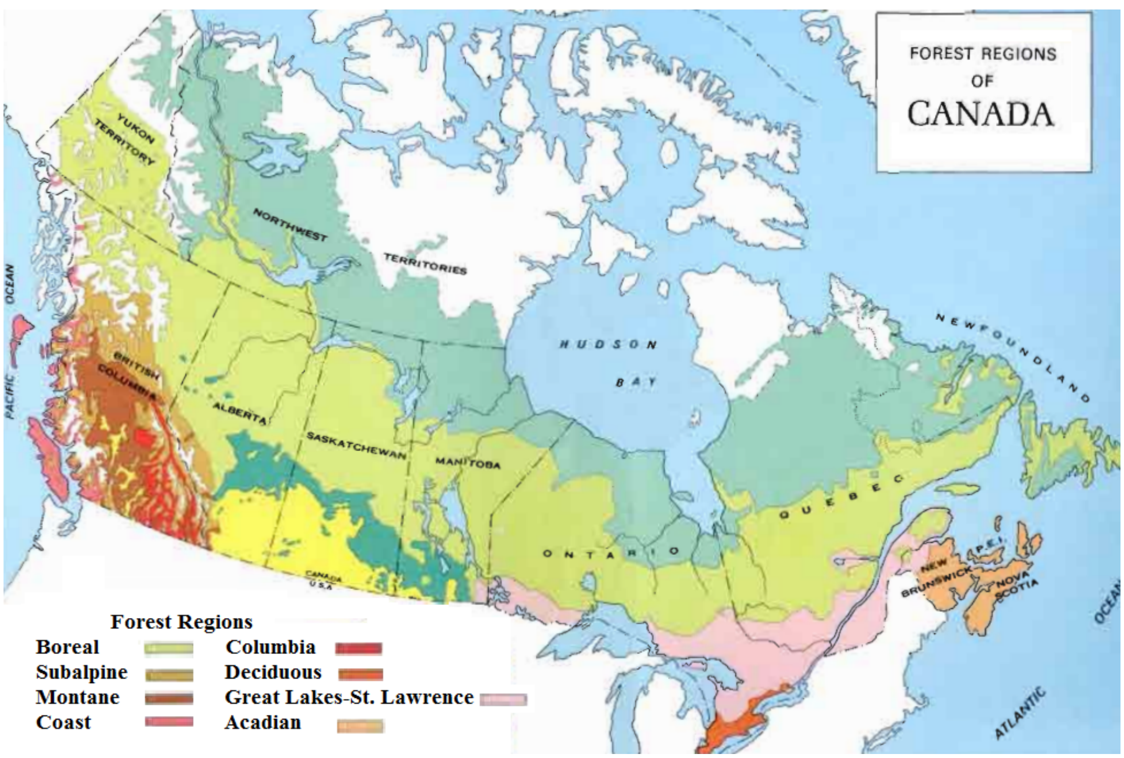


Figure 3.2 The distribution of Canadian forest regions
Source: Rowe (1972)

Table 3.1 Descriptions of the ecological conditions for each eco-region

Eco-region	***Ecological Description***
Eco-region 2 Subarctic 	The transitional forest-tundra belt area distributed in high-latitude continent, with a tree line traverse the central eco-region. Beyond the tree line, vegetation growth is limited mainly due to the cold and dry climatic conditions (Timoney et al., 1992). Generally, the terrain is relatively flat, and is covered by coniferous forest (e.g., spruce) The growing season is cooler and shorter, and usually lasts for five months. The average volume of precipitation in growing season is about 500 mm, with more rainfall distributed in the coastal areas (the Eastern Subarctic) and less in the inland areas.
Eco-region 3 Boreal 	The largest and continuous climatic belt stretching from Alberta to central Quebec and Nova Scotia, thus experiences more variability of ecological conditions. The western boundary of this eco-region is relatively sharped by the increasing elevation in British Columbia's (B.C.) mountainous areas. Thus, the Western Boreal experiences a cooler and drier growing season, contrary to the eastern parts whose growing season is longer, warmer, and rainy. The Boreal eco-region is also dominated by coniferous forests, and most SOC distribution is found in northern Ontario and eastern Quebec (Ju & Chen, 2005; Lee et al., 2010).
Eco-region 4 Cool Temperate 	This eco-region is characterized by mixed forests, including both coniferous and deciduous forests. Compared to those of high-latitude eco-regions, the Cool Temperate eco-region's growing season is much warmer and longer, with daily mean temperature above 0 °C generally extend from late March to November. Additionally, most rainfall is received from mid-summer to early-fall.
Eco-region 7 Subarctic Cordilleran 	The alpine and sub-alpine areas sparsely covered by coniferous forests in high-latitudes. The highest elevation exceeds 2,200 m. This eco-region undergoes a cold climate, with the mean annual temperatures around -5 °C to -10 °C. Inadequate precipitation is another main characteristic of this eco-region, with a range of 250 mm to 450 mm on an annual base.
Eco-region 8 Cordilleran 	The north-to-south climatic belt composed of mountainous topography. Thus, aspect along the elevation gradient is considered as an important factor controls soil moisture and forest growth. Coniferous forest is the dominant vegetation type. Growing season in the northern part is generally cold and short, while the southern parts are warmer and receives more rainfall.

Eco-region 9 Interior Cordilleran 	This eco-region is observed in southern B.C., with a mixed forest group. Mountains and valleys form the primary topography. The growing season is characterized as warm and semi-arid because it is located at the lee-side of the coastal mountains which prevents maritime moisture transferring into inland. The monthly precipitation ranges from 30mm to 50 mm.
Eco-region 10 Pacific Cordilleran 	A long and narrow climatic belt occupies the coastal areas. The growing season is cold and dry in the northern part of the eco-region; however, heavy rainfall is distributed throughout the southern parts, with monthly precipitation ranging from 150 mm to 350 mm. Soils in this eco-region are wet and relatively rich of nutrients and organic matters due to the accumulation of forest litterfall.

In summary, the boundaries of the Canadian forest, as well as the eco-region framework, have been selected to delineate the study area. This is based on two reasons. First, the Canadian forest is one of the largest carbon reservoirs in the global carbon cycle, and thus plays a significant role in global carbon regulations. Second, carbon regulations require deeper insight and understanding of SOC-environment interactions, yet few efforts have been made to explore the spatial relationships between Canadian forest SOC distribution and pertinent environmental determinants at regional scales. Therefore, the Canadian forest regions were adopted as the study area boundaries, while the seven eco-region boundaries were employed to provide a zonation framework for assessing SOC distribution.

Chapter 4. Data

4.1. SOC Data

As discussed in Section 2.1.2.1, after comparing the three nation-wide soil databases, the CFS database was employed in this study. As the first forest SOC database in Canada, CFS assembled and compiled data from field surveys undertaken before 1991, making it a useful mechanism for preserving historical SOC data. The development of the CFS database began in 1991. In addition to provincial- and federal-led field surveys, soil profiles obtained by private research groups and individual research papers were also included in the database. To ensure high-quality soil carbon data compilation, private research was strictly reviewed with two major criteria: (1) the study area was adequately described; and (2) a standard methodology framework for soil data analysis was established (Siltane, 1997). For example, in one of the referenced sources, Strong and La Roi (1985) collected soil samples from the Boreal forests in central Alberta. Detailed description of study area, field work processes, and laboratory procedures on soil nutrient content calculation were provided. As a result, 1,462 records from 170 referenced sources across all eco-regions and administration provinces in Canada were selected and added to the database (Siltane, 1997). After omitting records from tundra areas and only focusing on forest regions, the remaining 1,317 records are used in this study.

A map showing the distribution of original SOC sample points is presented in *Figure 4.1*. Most SOC samples were collected from the southern regions of Canadian forest ecosystems with the number of samples gradually deceasing in concentration towards northern regions. No samples were collected in mid-Quebec and northern Ontario that are highlighted by red ellipses in *Figure 4.1*. The absence of data collected in these two areas is due to the unavailability of field surveys in these areas during the study period from 1961 to 1991.

It is important to note that a number of constraints exist with the use of this dataset. Siltane (1997) pointed out in the database documentation that soil carbon information from peat land and agricultural areas were excluded from the database because of different pedogenic processes and rapid soil properties changes. Also, due to the lack of available field surveys, soil carbon data for northern Ontario and mid-Quebec were excluded from the database (Siltane, 1997). Consequently, soil carbon data were derived from relatively undisturbed forest- and tundra-environments. For each soil profile record, the SOC stock (measured in kg per square meters) was measured to a depth of one meter of mineral soil. The corresponding drainage capacity levels from 1 to 6 defined by Agricultural Canada for each record was also coded in the database, with 1 representing rapid drainage and 6 for very poor drainage (Siltane, 1997).

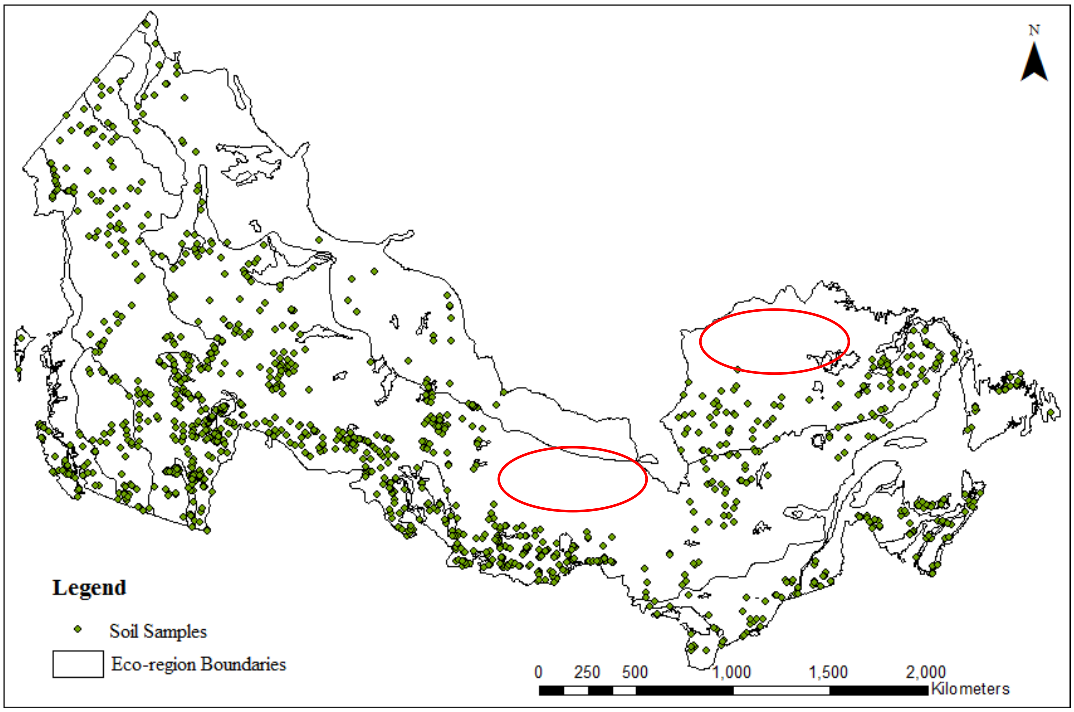


Figure 4.1 The spatial distribution of 1317 CFS soil samples collected in Canadian forest areas before 1991. Areas without soil samples are labelled in red.
Source: Canada Forest Service (CFS) soil database

4.2. Climate Data

Long term climate data in growing seasons (April to October) from 1961 to 1991 were collected in the study. This time period was selected based on two reasons: (1) the forest soil database consists of extensive historical data before 1991; and (2) the number of soil surveys in Canada considerably increased in the 1960s and reached a peak in the 1970s and 1980s, meaning that most soil data was collected within this time period. Daily 10 km Gridded Climate datasets for Canada provided by the National Land and Water Information Service, Agricultural Canada, were selected due to their large landmass coverage (up to 60 degrees north). Climate data from 7,514 weather stations were recorded and implemented in the ANUsplin model using a thin plate smoothing spline surface fitting method (Agricultural Canada, 2008). The climate dataset consists of daily maximum and minimum air temperatures (°C) and daily precipitation (mm), enabling the temperature-sensitivity of SOC distribution to be examined.

In order to obtain complete landmass coverage, seasonal climate records beyond the 60 degrees north line (in .xlsx format) were collected from Environment Canada. Due to data availability, climate data from 34 weather stations across three Northern provinces, the Yukon, the Northwest Territories, and Nunavut were selected and added to the dataset.

4.3. Terrain Data

The 16-day Normalized Difference Vegetation Index (NDVI) datasets employed in the study were acquired from the Global Inventory Modeling and Mapping Studies (GIMMS). NDVI, which is a vegetation index calculated from the red and near-infrared bands, has been widely used to assess vegetation status and biomass (Zhou et al., 2007). The geometric- and atmospheric-calibrated GIMMS NDVI products were obtained from the Advanced Very High Resolution Radiometer (AVHRR) satellite instrument at 8 km spatial resolution (Tucker et al., 2004). GIMMS NDVI products for a period of 25 years, from 1981 to 2006, are available for public use. To maintain consistency between climate and NDVI datasets, 16-days GIMMS NDVI datasets from1981 to 1991 were selected.

In addition, other primary topographic attributes – elevation, slope, and aspect – were derived from a digital elevation model (DEM) product (Canada3D) provided by the Canadian Forest Service. Ground elevation with a full landmass coverage from 41˚-83˚ N and from 52˚-148˚ W were obtained at 30 arc-seconds (approximately 926 m) spatial resolution (National Resource Canada, 2001). As discussed in Section 3, Rowe's forest boundaries provided by the Canada Forest Service and eco-region boundaries from Environment Canada were employed to delineate the boundary of the study area.

Since the main objective of the present study is to explore spatial relationships between SOC and ecological variables (climatic conditions and terrain attributes), soil properties such as parent material, soil pH values, bulk density, and other soil compositions (e.g., the percentage of clay and silt) are not considered. The soil properties, as well as other factors such as land cover and land use changes and forest harvest and clearance, were also not considered due to data availability. The original data sources for the present study are listed in *Table 4.1.* Specifically, the variables used in soil-environment modelling are shown in *Table 4.2* below.

Table 4.1 The original data sources

<table>
<tr><th colspan="2">Data</th><th>Date</th><th>Data Type</th><th>Temporal Scale</th><th>Spatial Scale</th><th>Source</th></tr>
<tr><td colspan="2">Carbon Density (kg/m^2)</td><td>< 1991</td><td>Point</td><td>N/A</td><td>N/A</td><td>Canada Forest Service</td></tr>
<tr><td colspan="2">AVHRR NDVI</td><td>1981 to 1991 (Apr. to Oct.)</td><td>Grid Raster</td><td>15 Days</td><td>8 km * 8 km</td><td>Global Land Cover Facility</td></tr>
<tr><td rowspan="2">Climate Data below 60 ˚N</td><td>Max. and Min. Temperature (˚C)</td><td rowspan="2">1961 to 1991 (Apr. to Oct.)</td><td rowspan="2">Grid Raster</td><td rowspan="2">Daily</td><td rowspan="2">10 km * 10 km</td><td rowspan="2">Agricultural Canada</td></tr>
<tr><td>Precipitation (mm)</td></tr>
<tr><td rowspan="2">Climate Data beyond 60 ˚N</td><td>Max. and Min. Temperature (˚C)</td><td rowspan="2">1961 to 1991 (Apr. to Oct.)</td><td rowspan="2">Point</td><td rowspan="2">Monthly</td><td rowspan="2">N/A</td><td rowspan="2">Environment Canada</td></tr>
<tr><td>Precipitation (mm)</td></tr>
<tr><td colspan="2">Elevation (m)</td><td>2001</td><td>Grid Raster</td><td>N/A</td><td>926 m * 926 m</td><td>GeoGratis</td></tr>
<tr><td colspan="2">Eco-Climate Region</td><td>1989</td><td>Polygon</td><td>N/A</td><td>N/A</td><td>Environment Canada</td></tr>
</table>

Table 4.2 Dependent and independent variables for statistical modelling of the SOC-environmennt relationships

Dependent Variable	***Description***
SOC (kg/m^2)	SOC stock measured to a depth of 1 meter mineral soils
Independent Variables	***Description***
Precipitation (mm)	Average total precipitation in growing seasons (April to October) from 1961 to 1991
Maximum Temperature (°C)	Average maximum temperature in growing seasons from 1961 to 1991
Average Temperature (°C)	Average mean temperature in growing seasons from 1961 to 1991
Minimum Temperature (°C)	Average minimum temperature in growing seasons from 1961 to 1991
Elevation (m)	A DEM model containing ground elevation information
Slope	Terrain relief derived from the DEM model
Aspect	Aspect information derived from the DEM model
NDVI	16-day average NDVI data from 1981 to 1991 representing the mean vegetation biomass

Chapter 5. Methodology

With the aim of examining spatial patterns of SOC distribution in Canada's forest areas and modelling relationships between SOC distribution and environmental determinants, an exploratory-based methodology that includes geostatistical analysis, Exploratory Spatial Data Analysis (ESDA), and spatial regression analysis was conducted for this study. The exploratory-based methodology scheme is illustrated in *Figure 5.1*. Specifically, descriptive statistics were first applied to describe and summarize the SOC samples collected from the Canada Forest Service (CFS) database. Pearson correlation analysis was applied to provide a first insight into the strength of associations between SOC and ecological variables. In addition, Ordinary Kriging was employed for spatial visualization by generating a continuous SOC distribution map for Canadian forest areas.

Then, global and local spatial autocorrelation statistics were applied to explore spatial dependency of Canadian forest soil distribution. In this study, the Moran's *I* test statistic was selected because it is the most popular statistic of spatial autocorrelation (Cliff & Ord, 1981; Huo et al., 2012) and is easily accessible in GIS software (e.g., ArcGIS and Geoda). The spatial cluster patterns derived from Moran's *I* tests verified the significant spatial effects occurring among SOC samples and verified the use of a spatial regression modelling approach. In this study, the Lagrange Multiplier diagnostics test was applied for model specification due to its ability of testing for omitted significant environmental determinants or a missing spatially lagged dependent variable. Ultimately, a predictive SOC map was produced to assist with evaluating the model's performance.

5.1. Data Preprocessing

5.1.1. Climate Datasets

Based on the previous discussion in Section 2.3.1, the ecological responses of SOC distribution to the various temperatures still remain controversial. This study chose the average maximum, mean, and minimum temperatures over growing seasons (April to October) from 1961 to 1991 to examine temperature sensitivity of SOC distribution. In addition, the average precipitation accumulation in the same period was selected as one of the environmental determinants. Since the climate datasets were obtained from different sources in different formats, a workflow (shown in *Figure 5.2*) showing the development of the final climate datasets for the study area.

The daily climate data obtained from Agriculture Canada covers the majority of Canada's forest area, with a full coverage below the 60° N latitudinal line. A python script (See *Appendix I*) was written to automatically calculate the average climate data. First, the average maximum/mean/minimum air temperatures for each year's growing seasons were calculated. Then, the daily precipitation data was summed for growing seasons. In summary, the mean seasonal temperatures and precipitation values were calculated over the period of 1961 to 1991.

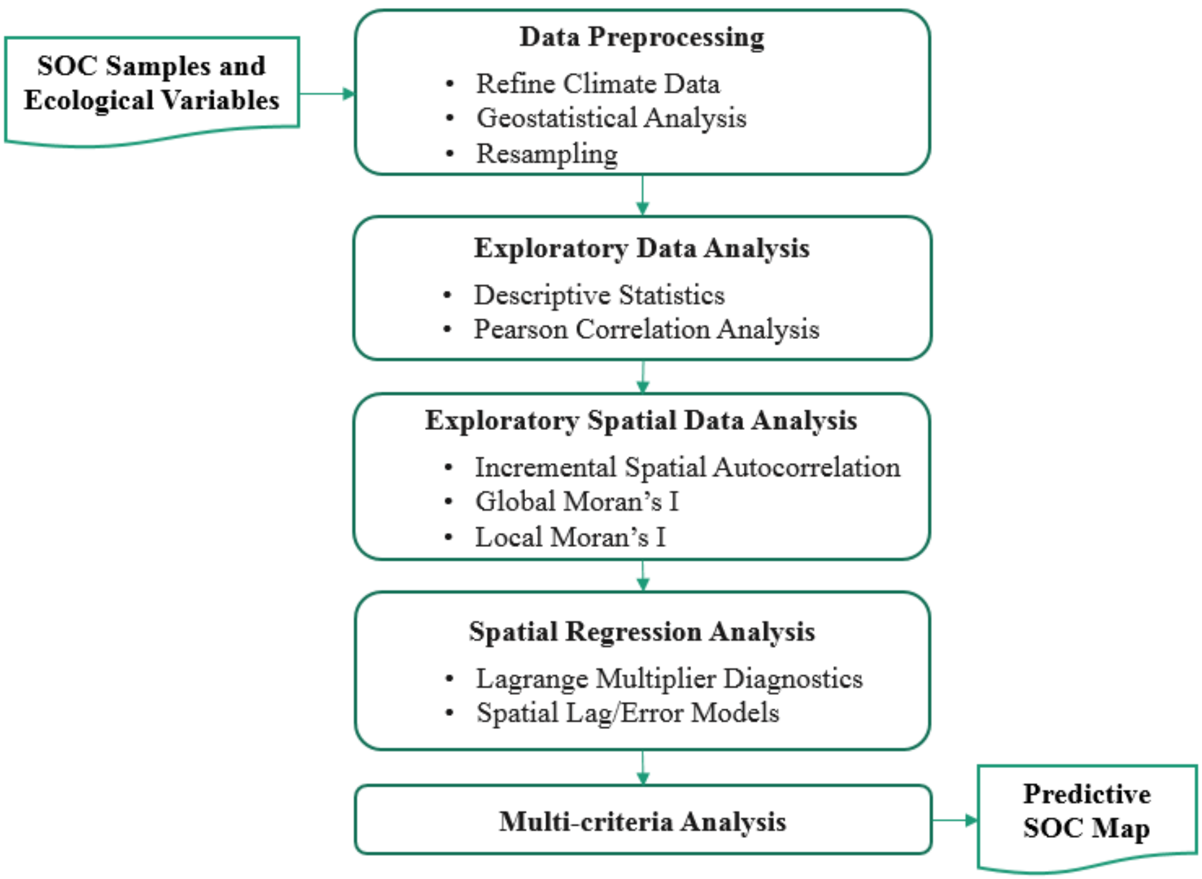


Figure 5.1 The exploratory-based workflow proposed by this study, mainly consisting of data preprocessing, descriptive statistics of SOC samples in Canadian forest areas, spatial autocorrelation analysis, and spatial regression modelling.

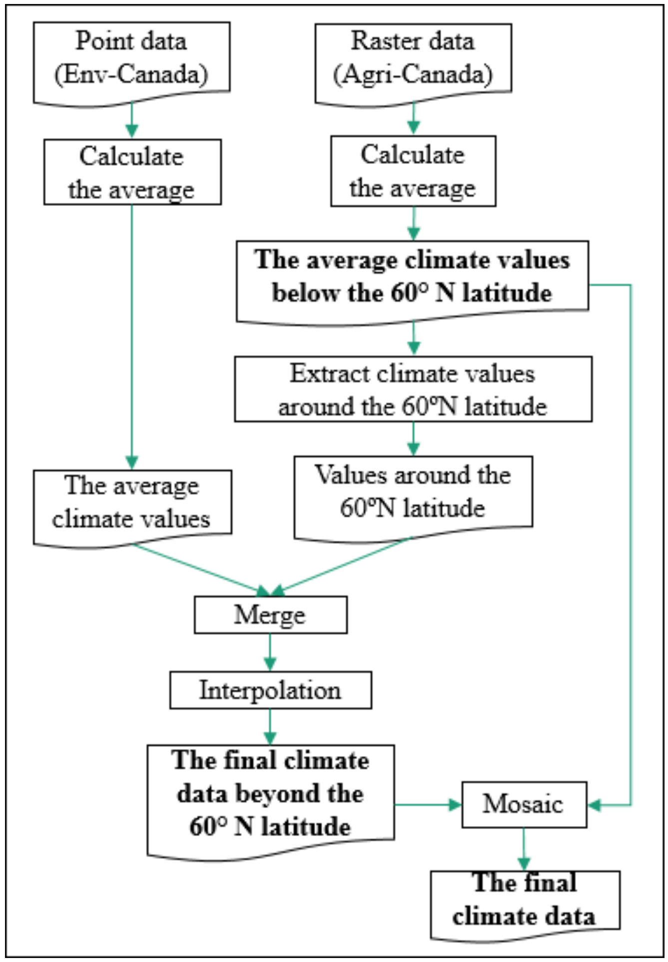


Figure 5.2 The process of finalizing climate data: merging the interpolated climate values beyond the 60° N latitude with the ones below the 60° N latitude

The initial climate datasets obtained from Environment Canada account for the landmass beyond 60° N latitude. However, these datasets are stored in point format, with a total of 34 available points. Thus, an interpolation approach was applied to estimate spatially continuous climate-distribution maps in northern Canada. To increase the accuracy of the interpolation results, the climate values around the 60° N latitudinal line was extracted from Agriculture Canada's datasets, and then merged with the point data extracted from Environment Canada's datasets. *Figure 5.3* helps to illustrate this process (the "Merge" step in *Figure 5.2*), using the average maximum air temperature for growing seasons as an example. In doing so, the final point data on which interpolation analysis could be performed were obtained and used to generate climate maps beyond the 60° N latitudinal line.

The next step was to estimate unknown climate values for the areas beyond the 60° N latitudinal line. In this study, the performance of IDW and Ordinary Kriging approaches was compared, because they are the most widely used approaches (Azpurua & Ramos, 2010). For IDW approach, the neighbourhood effects on any unknown value decrease with increasing distance (Childs, 2004). In addition, theoretical background of the Ordinary Kriging approach is explained in Section 5.1.2. To assess the performance of IDW and Ordinary Kriging, the criterion of the minimum Root-Mean-Square Error (RMSE) was adopted, namely that the lower the RMSE is, the more improved interpolated results are obtained (Johnston et al., 2001). In this study, Ordinary Kriging was selected due to the lower estimation errors (see *Appendix II*).The final climate dataset for the entire study area was generated by using a Mosaic tool to merge the climate-distribution maps from above and below the 60° N latitudinal line.

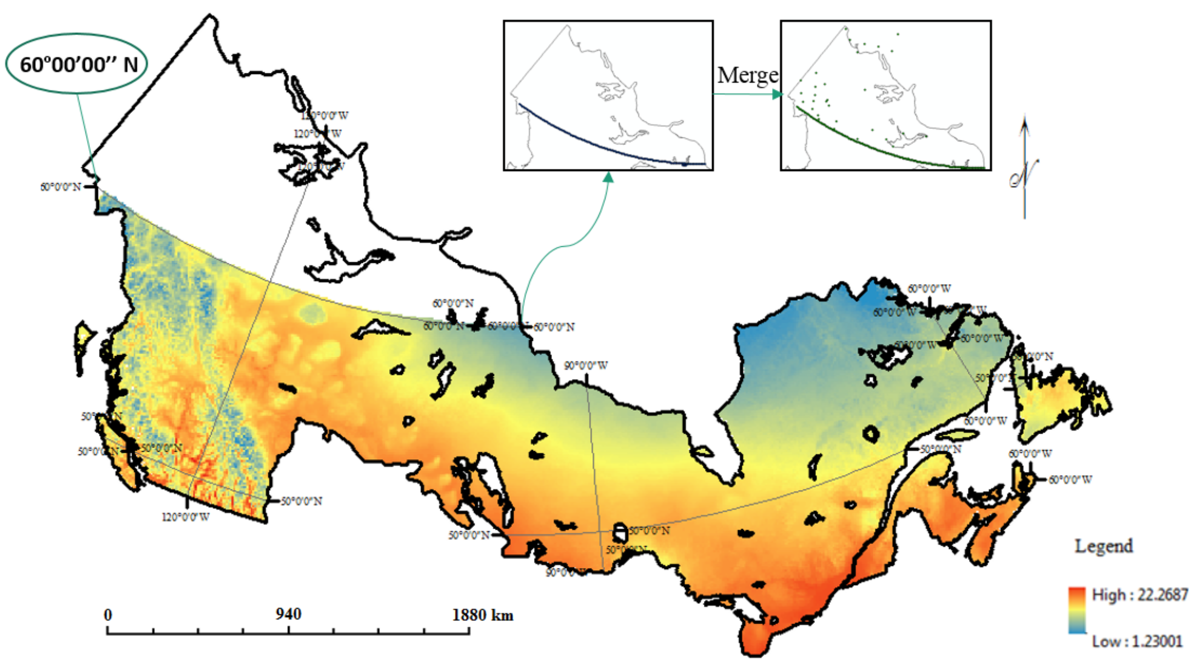


Figure 5.3 Merging the average maximum temperature in the growing season (April to October) around the 60° N latitude with the one beyond the 60° N latitude

5.1.2. Geostatistical Analysis of Canadian Forest SOC

According to Goovaerts (1997), all kriging approaches are derived from the basic equation:

$$Z^*(\mathbf{u}) = m(\mathbf{u}) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})[Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)] \tag{5.1}$$

where u_α is the location, $\lambda_\alpha\ (u)$ is the weights assigned to the observations within a user-defined neighbourhood radius, $m(u_\alpha)$ is the mean value within the radius, $n(u)$ is the number of observations within the radius, $Z(u_\alpha)$ is the sample's value at location u_α, and $Z^*(u)$ is the estimated sample's value. The present study adopted the Ordinary Kriging approach because it minimizes biases by standardizing the sum of weights to equal to one (Goovaerts, 1997). In addition, Ordinary Kriging assumes the mean value is constant within the radius of each estimated sample, indicated as $m(u)$ equals to $m(u_\alpha)$ (Goovaerts, 1997). Thus, the equation for Ordinary Kriging is transformed into:

$$Z_{OK}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u}) Z(\mathbf{u}_\alpha) \tag{5.2}$$

In practice, many researchers have suggested that Ordinary Kriging is the most appropriate interpolation method when no strongly global trend exists in the samples (Childs, 2004; Lefohn et al., 2005; Negreiros et al., 2010). The global trend refers to any directional distribution, such as the impact of prevailing wind on air pollution (Johnston et al., 2001). Johnston et al. (2001) stressed the necessity of detrending data in a geostatistical analysis:

(1) If the global trend exists, the assumption of Ordinary Kriging that the mean value is constant within the neighbourhood radius of each estimation will likely be violated.
(2) After removing the global trend, the local-scale variance in the original data can be more effectively examined.

In an ArcGIS software environment, the global trend is detected through Trend Analysis, by creating a three-dimension plot. The *X* and *Y* axes represent longitude and latitude respectively, while the *Z* axis records the sample's value at each location (example shown in *Figure 5.4 (b)*). Since samples are projected onto each plane, the global trend can be examined with a user-defined direction (Johnston et al., 2001). In addition, Ordinary Kriging requires a normalized distribution of samples (Johnston et al., 2001). Thus, a log-transformation is usually applied to skewed datasets. Finally, before the interpolated distribution-map is created, the data is back-transformed to its original scale; also, the trend will be added back to the data (Johnston et al., 2001). *Figure 5.4 (a) s*hows the sub-workflow of performing Ordinary Kriging, with solid lines that represent detailed steps of this study. *Figure 5.4 (b)* shows an example of the trend in the north-south and east-west directions. Since the trend lines are relatively flat, it was concluded that no significant trend was observed in the SOC samples of this study. Finally, the accuracy of Ordinary Kriging was assessed by the leave-one-out cross validation[1] approach.

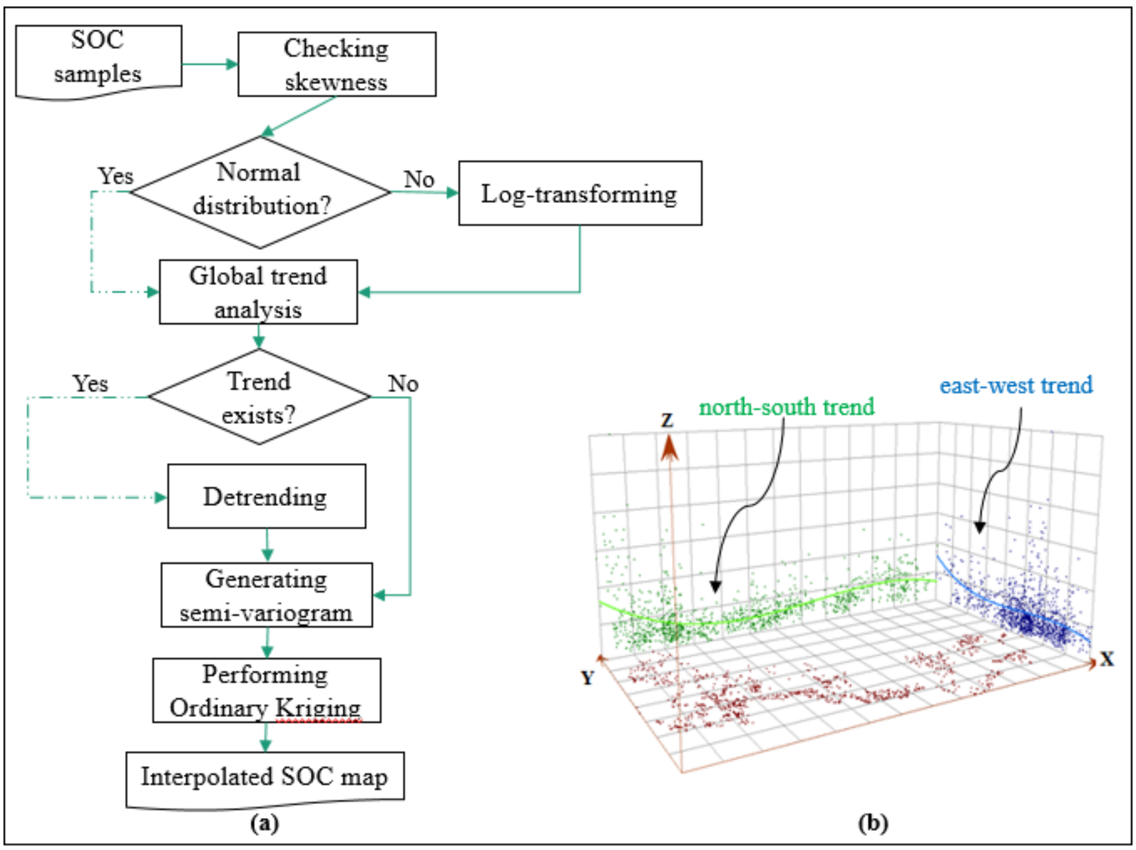


Figure 5.4 (a) The sub-workflow of performing Ordinary Kriging, and (b) trend analysis of the SOC samples in north-south and east-west directions.

[1] Leave-one-out cross validation is a three-step procedure: (1) excluding one sample and estimating its value by the remaining samples, (2) repeating the first step for all the samples, and (3) comparing the measured and estimated values (Johnston et al., 2001).

In summary, the average NDVI over growing seasons from 1961 to 1991 was calculated. Two primary terrain attributes, slope and aspect, were calculated from the DEM. All datasets, including climate, NDVI, and elevation, were re-sampled to 10 km resolution in order to have comparable cell size. In addition, slope and aspect were calculated based on the re-sampled elevation data. Finally, corresponding environmental data for each soil sample was extracted.

5.2. Exploratory Spatial Data Analysis of SOC Distribution

The exploratory analysis on Canadian forest SOC distribution started with a two-step Exploratory Data Analysis (EDA) approach, descriptive statistics and the Pearson correlation analysis. However, one major limitation of EDA is that no spatial information is taken into account. For spatially referenced data, spatial arrangement is of essential importance because geographic proximity tends to results in spatial dependency (Chakraborty, 2011). This phenomenon can be explained by Tobler's First Law of Geography which states that adjacent objects are more related to each other than distant objects (Chakraborty, 2011). To this end, Exploratory Spatial Data Analysis (ESDA) is designed as an extension of conventional EDA to examine the intrinsic characteristics of spatially referenced data. Main functions include the visualization of data properties, detection of spatial dependency, and identification of spatial clusters (Anselin, 1999; Haining et al., 1998; Oliveau & Guilmoto, 2005). In this study, a two-step ESDA scheme was subsequently applied to fully explore any spatial information underlying in the SOC distribution at national and eco-region scales, including (1) optimal modelling scale calculation, (2) global and local autocorrelation analysis.

5.2.1. Global Spatial Autocorrelation Analysis

The basis of exploring spatial patterns is the detection of spatial autocorrelation. Rooted in Tobler's First Law of Geography, global spatial autocorrelation is defined as the similarity among one sample and its neighbours (Valcu & Kempenaers, 2010). In this study, the global Moran's *I* test was applied to examine whether the null hypothesis of spatially independent SOC samples should be accepted or not and to measure the degree of spatial dependency among SOC samples (Ord & Getis, 1995). Based on the samples' locations and values (Wulder et al., 2007), the global Moran's *I* test measures the strength of correlation between a target object and its neighbours based on the equation below (Cliff & Ord, 1981; Huo et al., 2011):

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \quad i \neq j \tag{5.3}$$

where x_i is the observation at location i, $\bar{x}$ is the mean of x, w_{ij} is the distance-based weight between x_i and x_j, and n is the number of observations. If global spatial autocorrelation among the target objects does not equal to zero, the null hypothesis of spatial randomness is rejected and spatial statistical approaches can be adopted. A positive global Moran's *I* value suggests the comparability among proximal observations, meaning that similar observations are clustered together, while a negative value indicates a spatially dispersed pattern (e.g., a checkerboard pattern) (Oliveau & Guilmoto, 2005). In addition, a Moran's *I* scatterplot is usually employed to assist in illustrating global spatial autocorrelation: the original observations' values are recorded on the *X* axis, the *Y* axis records the weighted average of neighbouring values of each corresponding observation, and the slope of the regression line corresponds to the global Moran's *I* value (Anselin, 2003). Typically, the Moran's *I* scatterplot contains four quadrants to visually assess the different types of spatial association between each sample and its neighbours (Anselin, 1996). The samples in the right upper and left lower quadrants represent potential spatial clusters, while the samples in the right lower and left upper quadrants represent potential spatial outliers (Anselin, 1996).

5.2.1.1. *Scale Effects of Spatial Autocorrelation*

Since the Moran's *I* test statistic is measured based on different user-defined neighbourhoods to assess the similarity between one sample and its neighbours (Knegt et al., 2010), one critical question relates to how to select "neighbours". For spatially referenced point data, "neighbours" are usually identified by a distance-based approach. For each point, surrounding samples located within a certain distance threshold will be considered as "neighbours". However, this user-defined distance threshold selection may be subjective. Thus, in order to minimize human subjectivity, this study adopted the Incremental Spatial Autocorrelation approach to calculate the optimal distance threshold.

Incremental Spatial Autocorrelation is a new tool included in the ArcGIS software, which automatically calculates global Moran's *I* at a series of incremental distances[2] (ESRI, 2013[b]). This entails calculating a Z-score[3] associated with each global Moran's *I* value at each distance increment to quantify the strength of spatial dependency. According to ESRI (2013[c]) and Mitchell (2005), a high positive Z-score (i.e. larger than 1.96, p = 0.05) indicates a significant cluster pattern, while a high negative Z-score (i.e. less than -1.96, p = 0.05) indicates a dispersed pattern. Thus, statistically significant and positive peak Z-scores signify the spatial scales at which the ecological responses of targeted objects to cluster-patterns are most notable (ESRI, 2013[b]). In particular, the peak Z-scores associated with larger distances indicate general distribution trends (e.g., decreasing SOC stock from coasts to interior continental regions), while the peaks associated with smaller distances could preserve local variations. Thus, this study used the distance where the first peak Z-score occurs as the optimal distance of the proposed spatial analysis.

In addition, when data contains outliers (i.e. samples that are spatially isolated), the optimal distance may be over-estimated, because ensuring each outlier to have at least one neighbour likely causes some samples to have excessive neighbours (ESRI, 2013[a]). In this study, the Nearest Neighbour Distance (NND) of each soil sample was calculated. Any sample with a NND three times larger than its standard deviation (SD) was considered as outliers and should be removed (ESRI, 2013[a]). This process is illustrated in *Figure 5.5*. In this study, the optimal model scales were calculated at both national and eco-region scales of analysis.

---

[2] For a set of samples, the average nearest neighbor distance is usually used as the distance increment. The ArcGIS software allows a maximum of 30 iterations of distance increment. Since the distance where the first peak Z-score occurs was selected as the optimal distance, this study chose 30 iteration of distance ncrement to ensure the occurrence of the first peak Z-score.

[3] A Z-score is defined as: $Z(I) = \frac{I-E[I]}{\sqrt{Var[I]}}$
where *I* is the observed Moran's *I* value, E[*I*] is the expected value which assumes a random distribution of samples and is equals to $\frac{-1}{N-1}$, N is the number of samples, and Var[*I*] is the variance (ESRI, 2013[c]).

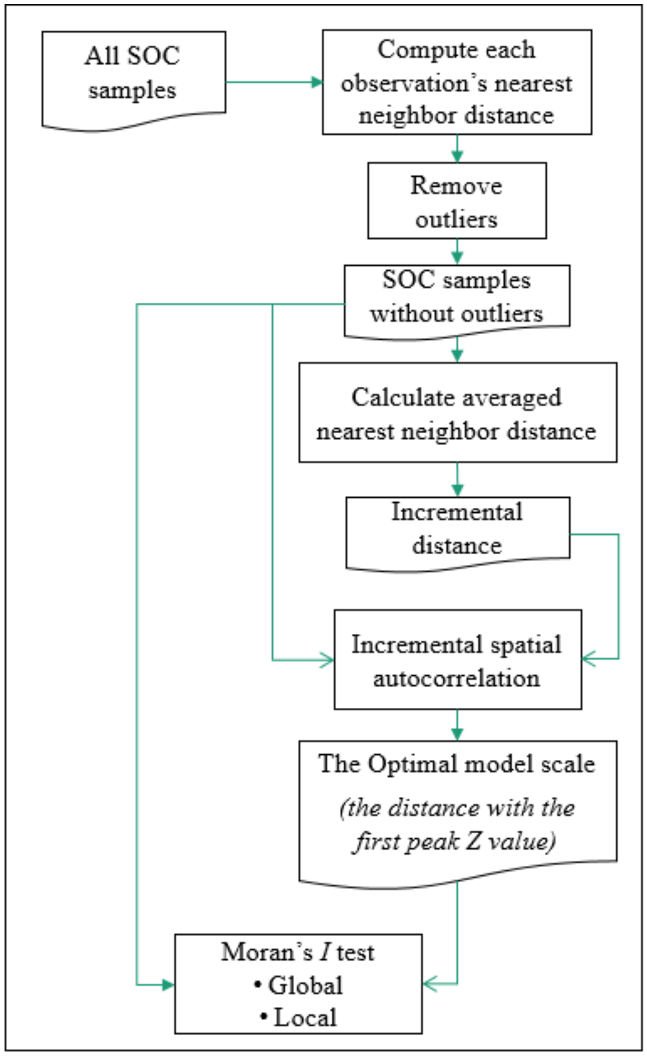


Figure 5.5 Determining the optimal modelling scale

5.2.2. Local Spatial Autocorrelation Analysis

The global Moran's *I* test measures the overall degree of spatial autocorrelation and returns a single value applied to the entire study area to indicate an overall spatial clustered or dispersed pattern (Anselin, 1995; Wulder et al., 2007). However, local variations remain unacknowledged from the results of a global Moran's *I* test (Anselin, 1995; Oliveau & Guilmoto, 2005; Wulder et al, 2007). Miller (2004) suggested that attention should be paid when analyzing the spatial patterns of proximal objects, since localized interactions among target objects could result in a complex ecological process. Thus, Local Indicators of Spatial Association (LISA) test statistics are often used to make up for this shortcoming of the global Moran's *I* test statistic. LISA assesses the null hypothesis of spatial randomness by measuring the spatial association of each observation to its neighbours[4] (Anselin, 1995; Unwin & Unwin, 1998). If spatial patterns (cluster or disperse) are found around particular samples, the null hypothesis of spatial randomness should be rejected. Similar with global Moran's *I* test, the optimal distance threshold derived from Section 5.2.1.1 was used to define "neighbours" when applying LISA test statistics.

The local spatial autocorrelation analysis highlights four types of local spatial patterns: the High-High (HH), Low-Low (LL), High-Low (HL), and Low-High (LF) patterns. The HH and LL patterns are known as spatial clusters within which the observations have positive Moran's *I* values and share similar spatial information, with high values located around high values and low values around low values respectively (Zhang et al., 2008). The HL and LH patterns are considered as spatial outliers, whose values are significantly different from their neighbours (Zhang et al., 2008). All the four types of spatial patterns provide unique insights into the spatial distribution of objects of interest, thus making spatial autocorrelation a fundamental step in spatial regression analysis.

In this study, the global and local spatial autocorrelation is measured at both the national and eco-region scales, with the Subarctic Cordilleran eco-region excluded. According to ESRI (2013[c]) and Mitchell (2005), the optimal number of samples used to calculate the Moran's *I* index should not be less than 30; otherwise the results may not be reliable. This is because any potential outliers would likely affect the identification of spatial patterns. Thus, this eco-region was excluded, since it contained a small sample of only 14 observations.

5.3. Spatial Regression Analysis of Modelling SOC-environment Relationships

The next step of this study involves investigating whether the detected spatial patterns of targeted objects are generated by specific processes (Goodchild et al., 1992). To achieve this goal, the relationships between SOC and ecological variables at the national and eco-region scales were tested based on traditional Ordinary Least Squares

4 The sum of local Moran's *I* is proportional to the global Moran's *I* (Anselin, 1995).

(OLS) regression models. As a non-spatial approach, OLS models are calculated based on the variables' absolute values, without considering the effects of spatial autocorrelation. Haining (1993) highlight that the spatial arrangement of locations where these values are recorded is also an important aspect of spatial association detection because it measures to what degree alike (or unalike) values of two variables are clustered together. Statistically, this combined effect of locations and values is known as spatial autocorrelation. Disregarding such effects can lead to biased regression coefficients and inaccurate estimation errors (Baller et al., 2001; Collins et al., 2006; Dormann et al., 2007). In addition, Chakraborty (2011) and Knegt et al. (2010) point out that if the spatial autocorrelation in the dependent variable cannot be completely explained by the estimated regression model, spatial autocorrelation will be detected among the regression residuals. Thus, the assumption of independent residuals is violated. Consequently, it is necessary to measure the strength of residual spatial autocorrelation to assess the performance of OLS models. For this purpose, the global Moran's I test statistic was applied on OLS residuals. Significant Moran's I values indicate model misspecification and suggest spatial regression models as alternatives.

Spatial autocorrelation is added to regression models as an independent variable to account for potential spatial effects (Chakraborty, 2011). As a result, a location-based component can be introduced into traditional OLS regression models in the form of a spatially lagged term or a spatial error term, thus developing the spatial lag model and spatial error model respectively (Anselin, 2001; Baller et al., 2001; Haining, 1993; Kelejian & Robinson, 1993). Although both models are capable of adjusting for the influences of spatial dependence, they are designed to simulate different situations in nature (Ward & Gleditsch, 2007). The spatial lag model is considered to be more appropriate when: (1) the existence of interactions among dependent variable, y, is pronounced (Anselin, 2001), (2) the impacts of nearby dependent variables on y_i are greater than those of independent variables, $x_{n,i}$, at location i (Ward & Gleditsch, 2007). In contrast, the spatial error model takes unobservable factors into consideration, suggesting that omitted variables showing certain spatial patterns should account for most of the spatial dependence of estimation errors (Ward & Gleditsch, 2007). Thus, the spatial error model is frequently used to adjust for immeasurable spatial dependence and to generate unbiased regression coefficients (Anselin, 2001; Ismail, 2006). However, due to the lack of solid understandings of ecological activities in forest soils (e.g., the debate about temperature effects are previously in Section 2.3.1.), it is usually difficult to assert which model is the most appropriate. Thus, the Lagrange Multiplier (LM) test illustrated in *Figure 5.6 (a)* should be first applied to assist with model selection (Anselin, 1988b; Velandia et al., 2008).

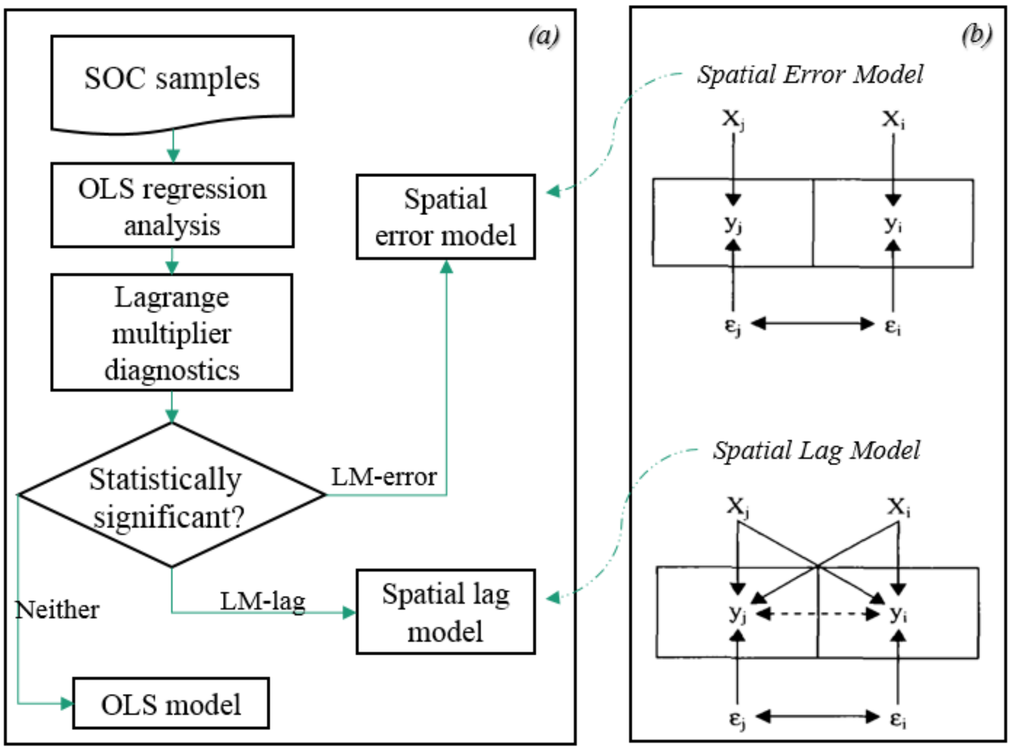


Figure 5.6 (a) Sub-workflow of model selection, and (b) Illustration of spatial processes described by the two spatial regression models
Source: Anselin (2005); Baller et al. (2001)

5.3.1. Lagrange Multiplier Diagnostics

The theoretical background of Lagrange Multiplier (LM) diagnostics for spatial dependence has been provided by many (e.g., Anselin, 1988[a]; Born & Breitung, 2011; Engle, 1984; Fazekas & Lauridsen, 1999; Haining, 1993; Velandia et al., 2008). According to Anselin (2009), spatial dependence is incorporated into regression models in two ways: (1) a spatially lagged dependent variable, and (2) an error term. Thus, the traditional OLS function:

$$y = \alpha + \beta X + u \tag{5.4}$$

could be re-written as:

$$\begin{cases} y = \alpha + \rho W y + \beta X + \varepsilon \\ \varepsilon = \lambda W \varepsilon + u \end{cases} \tag{5.5}$$

where y is the dependent variable, X is a set of independent variables, α is the intercept, ρ, β, and λ are the parameters, W is the spatial weights, Wy is the spatially lagged component of y, $W\varepsilon$ is the spatial autocorrelated error terms, and u is the independent and identically distributed (i.i.d.) errors.

LM diagnostics[5,6] are empirical tests with the null hypothesis of $\lambda = 0$ or $\rho = 0$, where λ is the parameter for an autocorrelated spatial error term and ρ is the parameter for an autocorrelated spatially lagged term (Anselin, 1988[a]):

(1) When the null hypothesis, $\lambda = 0$ and $\rho = 0$, is accepted, a traditional OLS regression model is the appropriate model specification.
(2) If $\lambda = 0$ and $\rho \neq 0$ is true, a spatial error model should be applied.
(3) If $\lambda \neq 0$ and $\rho = 0$ is true, a spatial lag model should be applied.

A robust version of LM diagnostics is also available to test the statistically significance of a spatial error, or, lag model. The robust LM diagnostics test against the interactions of the dependent variable under the condition that a spatially autocorrelated error term is inclusive (Velandia et al., 2008). Similarly, the robust LM diagnostics test the significance of a spatial error model by containing a spatially lagged component of the dependent variable (Velandia et al., 2008). Thus, spatial misspecifications will be minimized. In particular, five test statistics are provided in results of LM diagnostics,

---

[5] The equation for test against the spatial error is given by Burridge: $LM_{\lambda} = \frac{(u'Wu)^2}{\sigma^4 tr(W^2 + W'W)}$
where u is the OLS residuals, W is the spatial weights matrix, σ^2 is the residual covariance matrix, and *tr()* is the function that sums up the elements on the main diagonal (Anselin, 1988[a]; Born & Breitung, 2011).

[6] The equation for test against the spatial lag by Anselin is: $LM_{\rho} = \frac{y'MWy}{(\sqrt{\sigma^4 tr(W^2 + W'W) + \sigma^2 \beta' X' W' M W X \beta}}$
where M equals *to* $I - X(X'X)^{-1}X'$, β is the OLS estimator, and the others are as above (Anselin, 1988[a]; Born & Breitung, 2011).

namely the Moran's I test on OLS residuals, LM tests for a missing autocorrelated error term (LM-error) or spatially lagged dependent variable (LM-lag), and two robust versions (Robust LM-error and Robust LM-lag). The rule of deciding the model specification was provided by Anselin (2001) and Gallo et al. (2003). If LM-error test is more significant than LM-lag test and Robust LM-error is also statistically significant, while Robust LM-lag is not, the spatial error model should be used, and vice versa.

In this study, LM diagnostics were selected due to its capability of testing how spatial effects are associated with the targeted relationship. In general, spatial lag model specification simulates a "diffusion" process, indicating that nearby ecological activities would influence the target object under examination. While a spatial error model partly accounts for the unobservable factors and is effective when important determinants are omitted due to data availability. The two regression functions are explained in the following two sections.

5.3.2. Spatial Lag Regression Analysis

Spatial lag models simulate the situation when the dependent variable at one location promotes similar values in its surroundings (Anselin, 2001). The spatially lagged component of the dependent variable at each location is defined as a weighted average value within a neighbourhood radius which is distance-based and is calculated from Section 5.2.1.1. The equation of a spatial lag model is given as:

$$y = \alpha + \rho W y + \beta X + u \tag{5.6}$$

where y and X are dependent and independent variables respectively, α is the intercept, Wy is the spatially lagged component of y, W is the spatial weights, u is the error term which is i.i.d., and ρ, β are parameters. The equation can be reduced as:

$$y = (\mathrm{I} - \rho W)^{-1}\, \alpha + (I - \rho W)^{-1}\, \beta X + (I - \rho W)^{-1}\, u \tag{5.7}$$

where I is the unit matrix. In this case, we consider that the spatially lagged components of y are caused by the "spill over" effects of a set of independent variable x (Baller et al., 2001). Haining (1993) explain that rather than limited at location i, the impact of an independent variable, x_i, could propagate to its nearby locations. For example, along an elevation gradient, soils at lower positions tend to accumulate more organic matter due to soil erosions and waterflow movements (Birkeland, 1974). In addition, the spatial lagged component of y shows correlation with the independent error term (Anselin, 2001). The diffusion processes are delineated in *Figure 5.6 (b)*.

5.3.3. Spatial Error Regression Analysis

The spatial error model assumes that the spatial dependence occurs in the error terms, because given a regression model, the error term is the only part that produces uncertainties into an estimated y_i. Since Pace and LeSage (2010) suggested that the latent variables that are unobservable and unmeasurable would give rise to uncertainties in modelling the targeted relationships, and thus an autocorrelated error component is added to the spatial error model to account for the effects of unobservable factors on the dependent variable. The equation of spatial a spatial error model is given below:

$$\begin{cases} y = \alpha + \beta X + \varepsilon \\ \varepsilon = \lambda W\varepsilon + u \end{cases} \tag{5.8}$$

where y and X are dependent and independent variables respectively, α is the intercept, β and λ are parameters, $W\varepsilon$ is the spatial autocorrelated error terms, and u is the i.i.d. errors.

In summary, ecological phenomena always show certain spatial structures (e.g., local-scale spatial clusters, large-scale gradients) due to interdependent activities between organisms and similar generating processes (Ettema & Wardle, 2002; Legendre & Fortin, 1989), thus providing a justification for the use of spatial regression models in this thesis. By employing spatial regression models, certain questions can be answered: (1) in what ways is the targeted spatial relationship promoted, (2) how spatial dependence is associated with the targeted relationship (e.g., spatial lags or spatial errors), and (3) whether the unobservable factors notably influence the targeted relationship or not.

Moreover, to compare the performance of spatial regression models to that of an OLS model, the present study adopted the Akaike Information Criterion (AIC)[7] instead of the coefficient of determination (R^2). The R^2 indicates the degree to which the variances

---

[7] $AIC = -2L + 2K$, where L is the maximized logarithmic likelihood, and K is the number of variables (Haining, 1993; Anselin, 1988[b]).

in SOC distribution can be explained by a set of selected environmental determinants. In an OLS model, the sum of the predicted errors equals to zero; however, a spatial regression model does not (Wang, 2006). Thus, for the spatial regression models, the statistics of coefficient of determination is defined as a “pseudo R^2”. Although it also represents to what degree a spatial model is fitted, it is not comparable with the R^2 statistic of an OLS model (Anselin, 2005; Wang, 2006). In this case, the AIC is a more effective method to assess the model’s performance by taking both the model’s maximized logarithmic likelihood and the number of independent variables into consideration (Burnham & Anderson, 2004). Gagne and Dayton (2002) interpret the AIC as the loss of information when true values are replaced by the estimated values derived from maximum likelihood mechanism. Thus, the model with a minimum AIC value is considered to be optimal when the same datasets are used.

5.4. Predictive SOC Map

The last part of this study will employ a multi-criteria analysis to produce a predictive map of SOC in Canadian forest areas. At a national scale, a spatial error model informs four significant environment determinants ($p < 0.1$) that influence pedogenic processes in Canadian forest areas: seasonal accumulated precipitation, average minimum temperature, elevation, and slope. Thus, they were selected as useful criteria, and were weighted by corresponding regression coefficients derived from the spatial error model.

In this study, the Analytic Hierarchy Process (AHP) scheme was employed to calculate the weights for each environmental determinant. The AHP divides criteria into several hierarchical levels, with the lower level as one criterion in the higher level (Chen et al., 2009). Within each level, the importance of each criterion is rated on a pair-wise comparison basis (Saaty, 1977). Since the present study only has one hierarchical level with four criteria, the weights calculation is based on the pair-wise comparison between the variables’ regression coefficients. In doing so, the final predictive SOC map can be created from a three-step procedure: (1) multiplying each environmental determinant (the resampled raster data generated in the data preprocessing step) with its corresponding weight, (2) summing up the weighted environmental determinants on a pixel by pixel basis, and (3) standardizing the pseudo SOC-stock range as zero to one.

In summary, the final predictive map is not used to rigorously represent the real amount of SOC stock across the Canadian forest area. The main intention is to map the forest SOC distribution gradient under certain climatic conditions and terrain attributes on a national scale. By comparing the predictive SOC distribution map with the interpolated one, how the spatial patterns of SOC distribution differ between the two SOC maps could be visualized.

Chapter 6. Results

In order to explore how forest SOC is distributed in Canada under different ecological regimes, this study applied the methods described in Chapter 5 based on a set of Canada Forest Service (CFS) soil samples. Other long-term ecological datasets collected during growing seasons (April to October) from 1961 to 1991 were analyzed to investigate ecological influences on the spatial distribution of SOC in Canadian forest areas based on an eco-region zonation framework. This chapter describes the primary results from this study, beginning with a description of ecological conditions characteristic of Canadian forest ecosystems, followed by the results of the SOC spatial analysis. The results of EDA, including descriptive and graphical statistics of SOC samples and environmental determinants, are provided in Section 6.2. The ESDA results are then described in Section 6.3. Section 6.4 discusses the SOC-environment relationships on the national and eco-region scales. Finally, a predictive map is presented in Section 6.5 that shows the SOC distribution in response to statistically significant environmental determinants.

6.1. Ecological Background of Canada's Forest Area

Forest is a major ecosystem in Canada with a distribution ranging from British Columbia to Canada's east coast (Rowe, 1972). Prominent climatic gradients are observed within this large geographic coverage. *Figure 6.1* below shows the overall climatic conditions of Canada's forest ecosystems in the growing season. As shown in *Figure 6.1 (a), (b)*, and *(c)*, all three measures of temperature, namely the seasonal average maximum, mean, and minimum, generally show a north-to-south distribution in the entire study area, suggesting marked eco-climate regions with temperature gradually increasing with decreasing latitude. In Canadian forest areas, the average maximum temperature in the growing season (April to October) from 1961 to 1991 approximately range from 0.85 to 22.26 ˚C, and the average minimum temperature range from -5.77 ˚C to 10.21 ˚C. Higher maximum and minimum temperatures of approximate 22 ˚C and 10 ˚C (represented by red) are observed in Southern Ontario, indicating a longer and warmer growing season. Comparing the three temperature values, interesting spatial patterns are observed in Canada's west coast, namely smaller temperature variations due to buffering influences of the Pacific Ocean. In the growing season, relatively low maximum temperatures and higher minimum temperatures of about 13.5 ˚C and 9 ˚C are observed in this area. Moreover, since temperatures usually decrease with increasing elevation, B.C. mountainous areas in western Canada have complex temperature spatial patterns due to high local relief, as labelled in *Figure 6.1 (a), (b)*, and *(c)*.

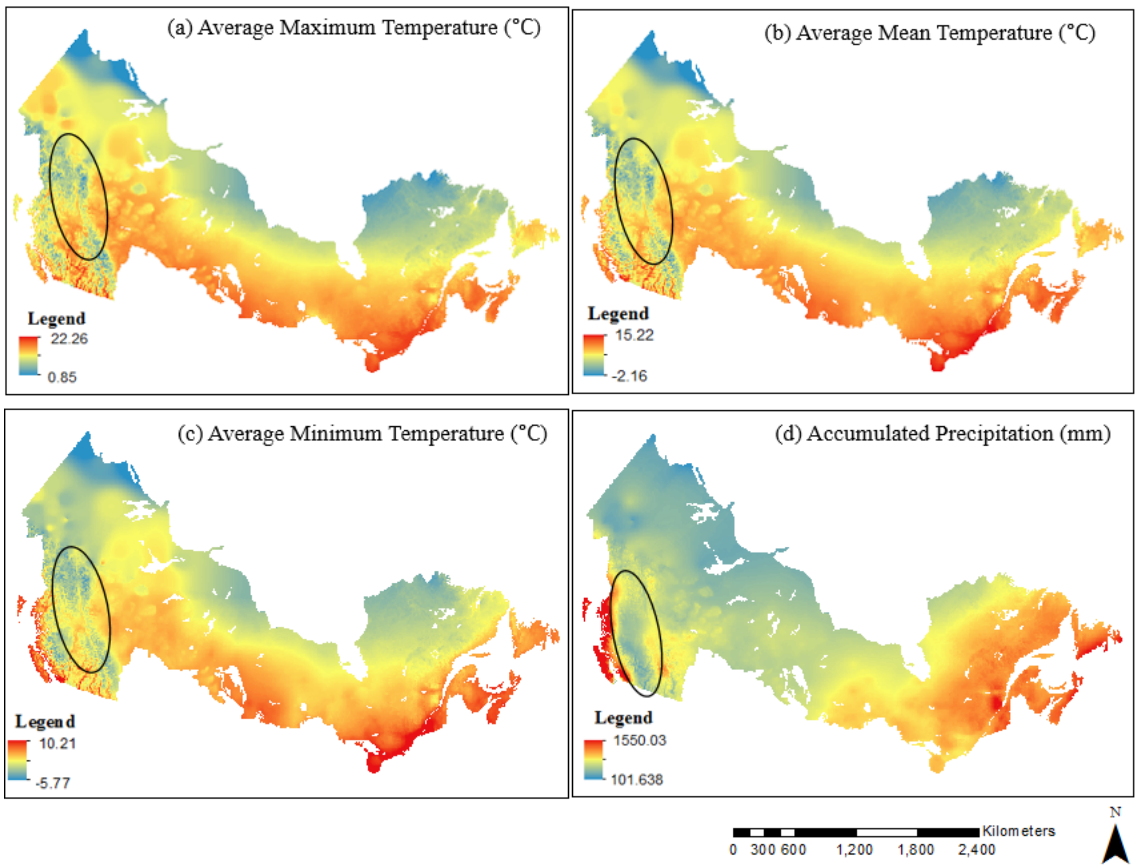


Figure 6.1 Average climatic conditions of Canadian forest areas in the growing season (April to October) from 1961 to 1991: (a), (b), and (c) represent the seasonal average maximum, mean, and minimum temperature, respectively; (d) represents the seasonal accumulated precipitation. B.C. mountainous areas with complex temperature patterns and insufficient amount of rainfall are labelled in (a), (b), (c), and (d), respectively. Source: Monthly average climate data from Environment Canada; Daily 10 km gridded climate data from Agricultural Canada

In contrast, the spatial pattern of precipitation is highly variable across the entire study area. The amount of precipitation is high along the west and east coasts, but gradually decreases towards to the central Great Plains areas. Northern regions experience a much drier or arid climate. In particular, the moist airflow from the Pacific

Ocean is forced to rise due to physical barriers imposed by coastal mountains, thus resulting in more precipitation falling on west-facing slopes as elevation increases. The maximum amount of seasonal accumulated precipitation (approximately 1,550 mm) is observed along Canada's west coast. The central regions of British Columbia are located between two main mountain ranges, the Coast Mountains in the west and Rocky Mountains in the east. Not surprisingly, an inadequate moisture supply, which is highlighted by an ellipse in *Figure 6.1 (d)*, is evident in central regions of British Columbia due to the interference of mountain ranges.

Figure 6.2 below maps the overall terrain attributes of Canadian forest areas. As previously mentioned, western Canada exhibits a complex topography due to the combination of mountain ranges and plateaus. Higher elevation areas are shown in red in *Figure 6.2 (a)*, while lower elevation is coloured blue. Rapid changes in elevation result in relatively steeper slopes, which are highlighted in red in *Figure 6.2 (c)*. Obvious east- and west-facing slopes, which are highlighted by black dashed lines, are visible in *Figure 6.2 (b)*. Terrain fluctuations are also shown in other areas, with small local relief in central Canada (e.g., Saskatchewan and Ontario) and uplands in eastern Quebec.

In addition, it is observed that the distribution of vegetation biomass closely follows temperature spatial patterns. The highest vegetation biomass (the largest NDVI values are approximately around 0.7) is measured in Southern Ontario and the east coast due to the longer growing season and adequate precipitation. As mentioned in Section 3, a treeline can be observed in the northern forest ecosystem. Around the tree line, the growth of vegetation is quite limited by adverse climatic conditions, such as low moisture and temperature (MacDonald & Gajewski, 1992). Thus, as shown in *Figure 6.2 (d)*, low vegetation growth and biomass (represented by the purple color) is observed and mapped in northern forest ecosystem areas even during the growing season. Little vegetation biomass (e.g., NDVI values around or below 0.4) is measured in arid mountainous areas due to decreasing temperatures and the presence of snow-cover on mountainous peaks.

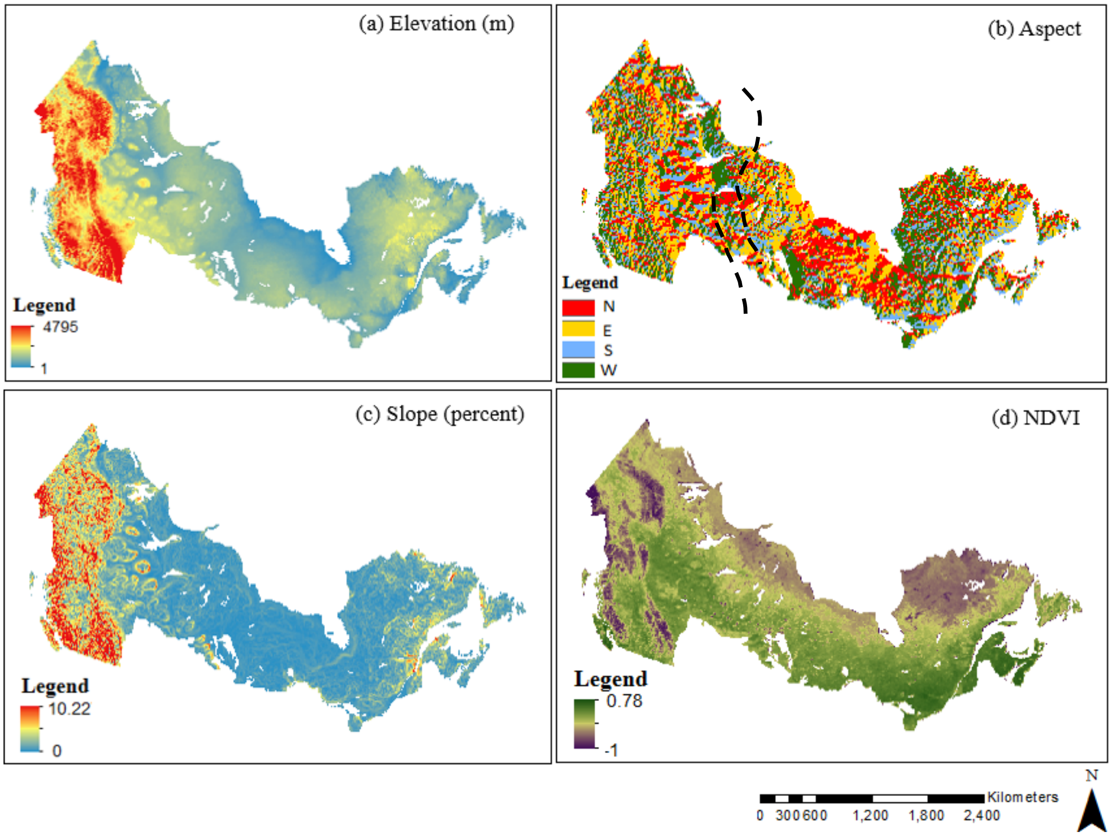


Figure 6.2 Terrain attributes of Canadian forest areas: (a), (b), and (c) show elevation, aspect, and slope, respectively; (d) shows the seasonal (April to October, 1961-1991) average NDVI. The east- and west-facing slopes in western mountainous areas are labelled in (b).
Source: Elevation data from GeoGratis; AVHRR NDVI data from Global Land Cover Facility (GLCF).

6.2. Non-Spatial Analysis of Canadian Forest SOC

In this subsection, soil samples are analyzed based on a non-spatial EDA approach in order to develop a basic understanding of forest SOC relationships. Results are divided into two parts. First, the graphical statistics are provided: (1) to visualise the mean SOC stock in each eco-region; and (2) to assess the influence of drainage capacity on the mean SOC stock. In addition, SOC information and pertinent ecological conditions within the entire study area and each eco-region are assessed. Second, the results of Pearson correlation analysis are summarized to quantify the association between SOC and each environmental variable identified in this study.

6.2.1. Statistical Description of Canadian Forest SOC and Ecological Variables

The mean organic carbon stock (kg/m^2) in the top one meter forest soils of each Canadian eco-region is mapped in *Figure 6.3*. Calculated from the historical soil profiles (the CFS soil data collected before 1991), B.C. coastal areas (the Pacific Cordilleran eco-region) holds the maximum SOC stock of about 28 kg/m^2. In the northern woodlands (the Subarctic Cordilleran and the Subarctic eco-region), SOC stock is also relatively high ranging from 12 kg/m^2 to 17 kg/m^2. In mountainous areas, the Cordilleran and Interior Cordilleran eco-regions also have lower SOC stock of approximately 10 kg/m^2. Moderate SOC stock is observed in Southern Ontario (the Cool Temperate eco-region), which has a warmer and moister climate located at southern latitudes. Moreover, the lowest mean SOC stock of about 9.7 kg/m^2 was surprisingly observed in the largest eco-region area: the Boreal eco-region. *Figure 6.4* below plots the mean SOC stock against six drainage capacity levels. According to Siltane (1997), the six levels coded from 1 to 6 are defined as: rapidly, well, moderately well, imperfectly, poorly, and very poorly drained soils. Naturally, little organic carbon is held by forest soils with good or high drainage conditions, because organic matter easily runs off with fast-moving water flow.

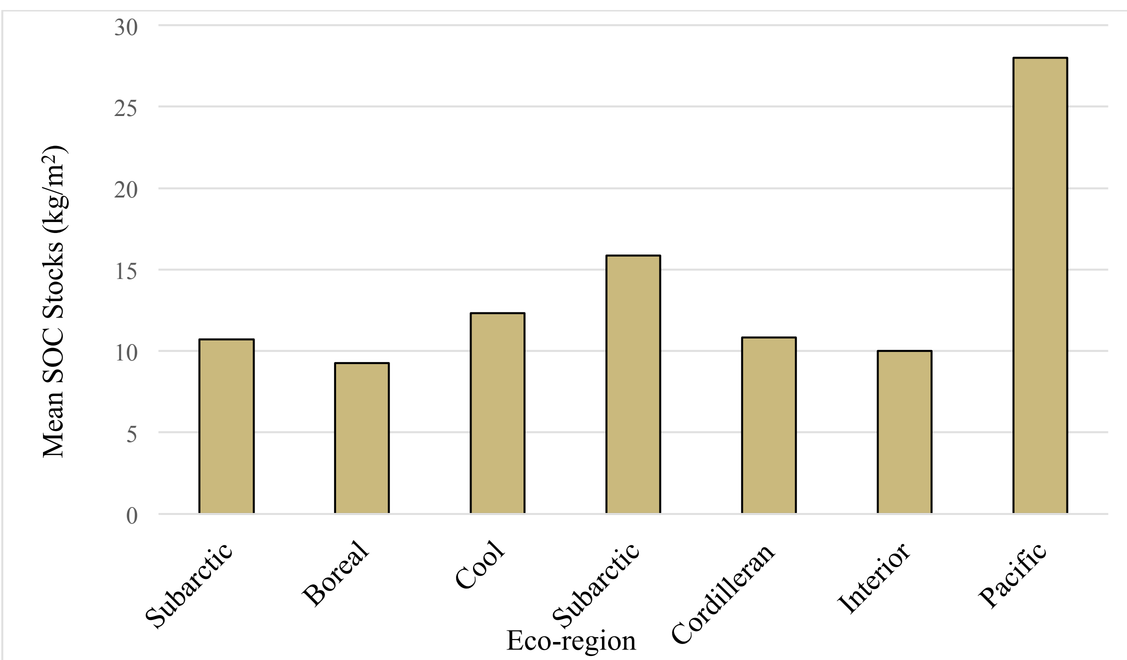


Figure 6.3 Mean SOC stocks (1961-1991) of each eco-region in Canadian forest areas
Source: Canada Forest Service (CFS) soil database

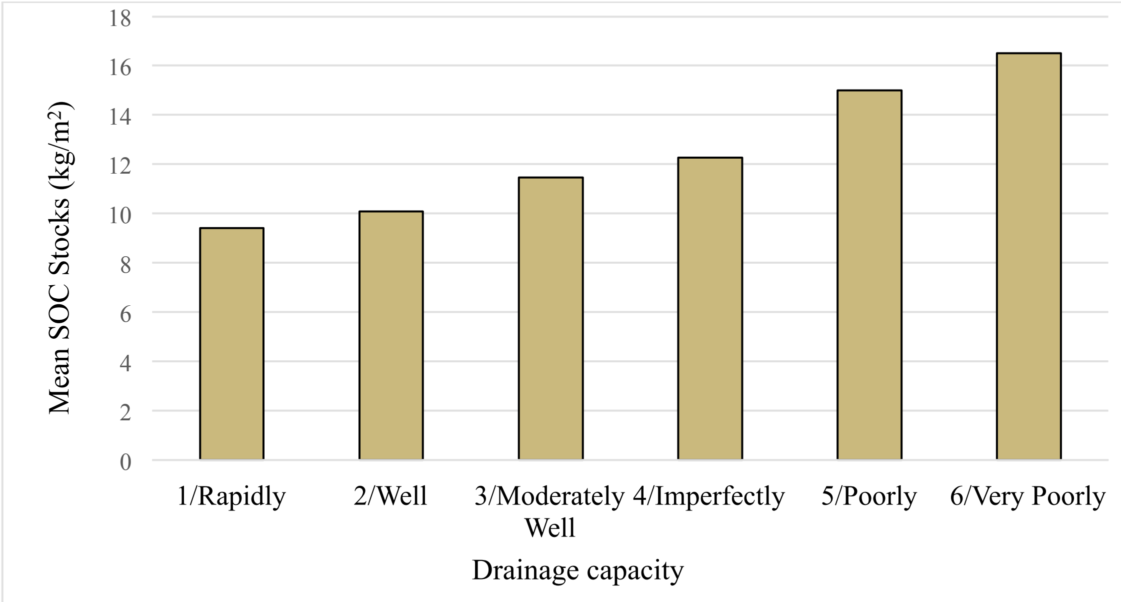


Figure 6.4 Mean SOC stocks (1961-1991) of six drainage capacity levels in Canadian forest areas
Source: Canada Forest Service (CFS) soil database

The statistical description of SOC and environmental determinants in the study area is summarized in *Table 6.1*. In Canadian forest areas, SOC stock ranges from 0.8 kg/m^2 to 57.8 kg/m^2. Compared to the maximum SOC stock (57.8 kg/m^2), the mean (11.12 kg/m^2) and median (9 kg/m^2) values are relatively small. The standard deviation of SOC stock was 7.73 kg/m^2, suggesting that only a small number of samples have very high SOC stock. In the growing season, the amount of precipitation greatly varies across the entire study area, ranging from 138.66 mm to 1216.80 mm with a standard deviation of 160.33 mm. Some soil samples were collected from the areas with very low vegetation biomass (NDVI = 0.15), such as mountainous areas and the forest-tundra zone. Moreover, all soil samples were collected from low-relief areas, with the maximum percent of slope[8] as 5.64 (equals to 3.23 degree). Summary statistics of the SOC stock and ecological conditions in each eco-region are shown in *Table 6.2* to *Table 6.8*, and are further described in the remainder of this section.

[8] The percent of slope is also known as percent rise. It equals to the rise (vertical distance) divided by the run (horizontal distance).

As shown in *Table 6.2*, the maximum SOC level in Canadian forest regions is observed in the Pacific Cordilleran eco-region. The overall SOC stock is quite high with a range of 8.5 kg/m^2 to 57.8 kg/m^2. Both maximum precipitation (1,216.80 mm) and vegetation biomass (0.73) are also observed in this eco-region. This result is consistent with previous literature, which suggests that humid soils tend to hold more organic carbon because the decomposition rate of organic matter is limited under good water-saturation conditions (Bhatti et al., 2006; Buringh, 1984; Davidson et al., 2000; Deluca, & Boisvenue, 2012).

Table 6.1 Descriptive statistics of SOC stock and pertinent ecological variables within the entire study area (n=1317)

Parameter	***Minimum***	***Maximum***	***Mean***	***Median***	***Standard deviation***
SOC (kg/m^2)	0.80	57.80	11.12	9.00	7.73
Max. Temp. (°C)	1.49	19.69	14.38	15.06	2.85
Mean. Temp. (°C)	-2.12	14.54	8.75	9.33	2.53
Min. Temp. (°C)	-5.73	9.55	3.20	3.26	2.41
Precipitation (mm)	138.66	1216.80	443.34	431.11	160.33
Elevation (m)	6.00	2690.00	627.53	439.00	507.10
Slope (percent)	0.01	5.64	0.68	0.33	0.83
Aspect	0	359.21	169.44	158.91	105.42
NDVI	0.15	0.73	0.52	0.55	0.12

Table 6.2 Descriptive statistics of SOC stock and pertinent ecological variables in the Pacific Cordilleran eco-region (n=62)

Parameter	***Minimum***	***Maximum***	***Mean***	***Median***	***Standard deviation***
SOC (kg/m^2)	8.50	57.80	27.98	24.50	15.02
Max. Temp. (°C)	8.44	17.98	14.37	14.56	2.34
Mean. Temp. (°C)	3.54	13.11	9.91	10.49	2.50
Min. Temp. (°C)	-1.36	8.25	5.46	6.71	2.87
Precipitation (mm)	257.77	1216.80	795.43	794.34	314.32
Elevation (m)	16.00	2311.00	601.35	479.00	443.67
Slope (percent)	0.12	5.27	1.69	1.51	1.11
Aspect	0	341.47	162.93	172.67	99.55
NDVI	0.20	0.73	0.54	0.59	0.15

The statistics of SOC samples from the two northern eco-regions, the Subarctic and Subarctic Cordilleran, are shown in *Table 6.3* and *Table 6.4*, respectively. Since the two eco-regions are located at high latitudes, the climate is characterized as being cold and dry. Precipitation amounts in the Subarctic and Subarctic Cordilleran eco-regions are 450.65 mm and 216.19 mm; and the maximum temperatures are 10.05 °C and 9.35 °C, respectively. The overall vegetation biomass is consequently limited by the dry and cold climatic conditions. Compared to other eco-regions, much lower NDVI values are observed, with a mean of 0.38 for the Subarctic eco-region and 0.33 for the Subarctic Cordilleran. In terms of SOC stock, both of the two eco-regions tend to hold relatively larger amounts of SOC than other eco-regions, with mean values of 12.18 kg/m^2 for the Subarctic eco-region and 15.86 kg/m^2 for Subarctic Cordilleran.

Table 6.3 Descriptive statistics of SOC stock and pertinent ecological variables in the Subarctic eco-region (n=129)

Parameter	***Minimum***	***Maximum***	***Mean***	***Median***	***Standard deviation***
SOC (kg/m^2)	1.9	43.5	12.18	10.7	7.72
Max. Temp. (°C)	6.81	14.27	10.05	10.18	1.42
Mean. Temp. (°C)	1.98	8.21	5.27	5.49	1.21
Min. Temp. (°C)	-2.90	2.27	0.50	0.78	1.16
Precipitation (mm)	181.44	654.72	450.65	464.35	150.49
Elevation (m)	10.00	763.00	355.67	367.00	171.76
Slope (percent)	0.01	3.59	0.32	0.23	0.39
Aspect	0	356.87	182.66	180.55	103.74
NDVI	0.21	0.58	0.38	0.38	0.07

Table 6.4 Descriptive statistics of SOC stock and pertinent ecological variables in the Subarctic Cordilleran eco-region (n=14)

Parameter	***Minimum***	***Maximum***	***Mean***	***Median***	***Standard deviation***
SOC (kg/m^2)	5.5	37.3	15.86	14.2	8.27
Max. Temp. (°C)	1.48	13.99	9.35	12.45	5.13
Mean. Temp. (°C)	-2.12	7.09	3.64	5.86	3.74
Min. Temp. (°C)	-5.73	0.18	-2.07	-0.66	2.36
Precipitation (mm)	138.66	288.09	216.19	227.3	49.21
Elevation (m)	176	1626	902.57	824.5	430.58
Slope (percent)	0.18	3.10	1.13	0.97	0.74
Aspect	10.45	354.95	136.27	97.55	103.87
NDVI	0.24	0.41	0.33	0.33	0.04

As shown in *Table 6.5* and *Table 6.6*, although the Boreal and Cool Temperate eco-regions have similar topographic features (e.g., both are characterized by relatively flat land surfaces and moderate vegetation biomass), different SOC levels are observed. The lowest SOC stock is measured in soil samples from the Boreal eco-region, while a moderate amount of SOC is measured from the Cool Temperate eco-region. Such differences are likely due to different climatic regimes. Namely, the Boreal eco-region tends to experience a relatively drier climate than the Cool Temperate eco-region. For example, the minimum and mean precipitation of the Boreal eco-region are 192.77 mm and 425.4 mm, while the values of the Cool Temperate eco-region are 480.18 and 593.88, respectively.

Table 6.5 Descriptive statistics of SOC stock and pertinent ecological variables in the Boreal eco-region (n=649)

Parameter	***Minimum***	***Maximum***	***Mean***	***Median***	***Standard deviation***
SOC (kg/m^2)	0.8	31.5	9.28	8.5	4.90
Max. Temp. (°C)	9.29	18.69	15.48	15.92	1.76
Mean. Temp. (°C)	4.62	12.54	9.71	9.92	1.48
Min. Temp. (°C)	-0.59	7.49	3.94	3.87	1.38
Precipitation (mm)	192.77	787.79	425.4	378.04	117.04
Elevation (m)	6.00	926.00	372.86	346.00	173.15
Slope (percent)	0.01	1.78	0.29	0.17	0.32
Aspect	0	359.21	171.9	158.91	108.19
NDVI	0.25	0.72	0.56	0.57	0.07

Table 6.6 Descriptive statistics of SOC stock and pertinent ecological variables in the Cool Temperate eco-region (n=86)

Parameter	***Minimum***	***Maximum***	***Mean***	***Median***	***Standard deviation***
SOC (kg/m^2)	2.90	38.40	12.32	10.60	6.98
Max. Temp. (°C)	16.41	19.69	18.13	18.09	1.11
Mean. Temp. (°C)	10.60	14.54	12.70	12.46	1.04
Min. Temp. (°C)	4.80	9.55	7.27	7.41	1.14
Precipitation (mm)	480.18	725.5	593.88	597.74	71.67
Elevation (m)	15.00	466.00	176.35	181.5	112.35
Slope (percent)	0.01	1.15	0.27	0.24	0.21
Aspect	0	343.66	192.46	199.7	95.85
NDVI	0.45	0.72	0.63	0.66	0.07

Last, the descriptive statistics of SOC samples collected from mountainous areas are provided in *Table 6.7* and *Table 6.8*. Compared to other eco-regions, the SOC stock in the Cordilleran and Interior Cordilleran eco-regions are comparatively lower, but still within an average range (e.g., an overall mean value of 11.12 kg/m^2 in *Table 6.1*). In general, the amount of precipitation is lower (e.g., the mean values of Cordilleran and Interior Cordilleran eco-regions are 406.22 mm and 306.88 mm, respectively), but is evenly distributed across the two eco-regions (indicated by the lower standard deviations, 99.09 and 60.67, respectively). Low vegetation biomass in soil sample sites is measured in the Cordilleran eco-region, perhaps due to less litterfall input.

Table 6.7 Descriptive statistics of SOC stock and pertinent ecological variables in the Cordilleran eco-region (n=306)

Parameter	***Minimum***	***Maximum***	***Mean***	***Median***	***Standard deviation***
SOC (kg/m^2)	1.00	47.00	10.84	9.00	6.65
Max. Temp. (°C)	7.54	17.07	12.89	13.38	2.50
Mean. Temp. (°C)	3.05	10.49	7.09	7.20	1.99
Min. Temp. (°C)	-2.01	3.91	1.28	1.30	1.66
Precipitation (mm)	195.04	618.73	406.72	433.59	99.09
Elevation (m)	591.00	2690.00	1273.00	1117.5	0.49
Slope (percent)	0.05	5.64	1.36	1.08	0.99
Aspect	1.71	358.42	149.29	128.2	100.25
NDVI	0.14	0.63	0.46	0.50	0.14

Table 6.8 Descriptive statistics of SOC stock and pertinent ecological variables in the Interior Cordilleran eco-region (n=71)

Parameter	***Minimum***	***Maximum***	***Mean***	***Median***	***Standard deviation***
SOC (kg/m^2)	2.00	30.7	10.02	8.10	6.28
Max. Temp. (°C)	8.45	18.39	15.07	15.48	2.05
Mean. Temp. (°C)	4.03	11.47	8.68	8.96	1.52
Min. Temp. (°C)	-1.22	4.84	2.29	2.23	1.13
Precipitation (mm)	226.29	573.92	306.88	289.20	60.67
Elevation (m)	696.00	1989.00	1182.5	1044.00	334.55
Slope (percent)	0.16	4.63	1.47	1.23	0.96
Aspect	0	352.75	194.05	232.26	107.91
NDVI	0.19	0.63	0.55	0.57	0.07

6.2.2. Correlation between SOC and Ecological Variables

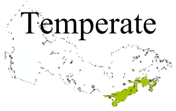
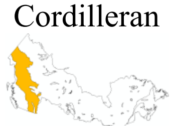
The results of Pearson correlation analysis between the SOC and environmental determinants at three significance levels of 1%, 5%, and 10%, are shown in *Table 6.9*, with significant correlations highlighted in green. In general, without considering spatial effects, a statistically significant association between SOC and precipitation was found at both the national and eco-region levels. The relationship between SOC and temperature in some eco-regions was significant, but generally quite weak. More detailed interpretation of Pearson correlation results are summarized in *Table 6.10*.

Table 6.9 Results of Pearson correlation analysis between SOC and select environmental and climatic factors

Eco-region	***Max. Temp***	***Mean Temp***	***Min. Temp***	***Prep.***	***Elev.***	***Slope (Percept)***	***Aspect***	***NDVI***
The Entire Study Area *Sig.*	-0.102*** *0.000*	-0.015 *0.585*	0.089*** *0.001*	0.436*** *0.000*	0.042 *0.130*	0.222*** *0.000*	0.018 *0.519*	-0.026 *0.340*
Boreal *Sig.*	-0.115** *0.003*	-0.054 *0.173*	0.031 *0.423*	0.292*** *0.000*	-0.098** *0.012*	0.087** *0.026*	-0.036 *0.363*	0.000 *0.997*
Cool Temperate *Sig.*	0.136 *0.211*	0.061 *0.578*	-0.022 *0.842*	0.265** *0.014*	0.146 *0.179*	0.310*** *0.004*	0.150 *0.168*	0.255** *0.018*
Subarctic *Sig.*	0.024 *0.786*	0.001 *0.992*	-0.028 *0.757*	0.014 *0.877*	0.115 *0.196*	0.170* *0.055*	-0.073 *0.411*	0.231*** *0.008*
Subarctic Cordilleran *Sig.*	0.050 *0.865*	0.053 *0.857*	0.059 *0.840*	-0.235 *0.418*	-0.264 *0.361*	0.179 *0.541*	-0.175 *0.551*	0.253 *0.383*
Cordilleran *Sig.*	-0.172*** *0.002*	-0.15*** *0.009*	-0.097* *0.091*	0.102 *0.075*	0.164*** *0.004*	0.150*** *0.008*	0.066 *0.252*	-0.168*** *0.003*
Interior Cordilleran *Sig.*	-0.166 *0.166*	-0.148 *0.218*	-0.098 *0.416*	0.318*** *0.007*	0.337*** *0.004*	0.149 *0.213*	0.135 *0.261*	-0.172 *0.151*
Pacific Cordilleran *Sig.*	0.158 *0.220*	0.300** *0.018*	0.393*** *0.002*	0.545*** *0.000*	-0.177 *0.168*	-0.029 *0.824*	0.225* *0.079*	0.288** *0.023*

$p < 0.1$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

Table 6.10 Interpretation of the Pearson correlation coefficients between Canadian forest SOC and ecological variables in each eco-region

Eco-region	***Interpreting the Pearson Correlation Results***
The Entire Study Area	At the national scale, SOC stock and precipitation resulted in the strongest relationship. Other significant determinants are: slope, maximum temperature, and minimum temperature. The negative influence of maximum temperature on SOC can be explained by a temperature-induced rapid decomposition rate. Increasing air (soil) temperatures would potentially stimulate soil biota activates, thereby accelerating SOC decomposition rate. Thus, a negative association is observed. However, a positive, but weak, correlation was found between SOC and minimum temperature (0.089). This is likely due to the possibility that increasing minimum temperature potentially lengthens the growing season, and thus plants produce more biomass and increase the amount of organic-carbon inputs into forest soils
Boreal 	Referring to *Figure 6.2*, the vegetation biomass is evenly distributed across the Boreal eco-region. Precipitation dominates the SOC distribution in this eco-region. In addition, quite weak correlations were found between terrain attributes (i.e. slope) and SOC distribution.
Cool Temperate 	Precipitation and vegetation biomass were significantly related to SOC distribution in this eco-region. In addition, a relatively strong correlation was detected between SOC and slope.
Subarctic 	This eco-region is the transitional forest-tundra zone. From the Pearson correlation results, vegetation biomass was highly related to SOC distribution. This is a reasonable result, since SOC sequestration benefits from root exudates and litterfall accumulation
Subarctic Cordilleran 	No significant relationships between SOC and environmental factors was found. This is probably due to the small sample size collected within this eco-region (n=14), thus typical trend or relationships may be neglected.
Cordilleran 	Terrain attributes tend to have more significant relationships with SOC distribution. A negative correlation was found between the maximum temperature and SOC distribution, likely due to carbon loss caused by temperature-induced decomposition.
Interior Cordilleran 	Similar with the Cordilleran eco-region, elevation was strongly related to SOC in the Interior Cordilleran eco-region, with a correlation coefficient of 0.337. Precipitation also showed a strong association with SOC distribution in this semi-arid eco-region.
Pacific Cordilleran 	Factors significantly associated with SOC include mean and minimum temperatures, precipitation, and NDVI. Precipitation was strongly associated with SOC, likely due to the ability of high soil-moisture content soils to preserve SOC accumulation.

6.3. Spatial Analysis of Canadian Forest SOC

This section presents the results of a spatial analysis of SOC distribution in Canadian forest areas. First, a continuous SOC distribution map is derived from original soil samples (n=1,317) for data visualization. Then, spatial patterns of SOC levels are presented based on a spatial autocorrelation analysis.

6.3.1. Geostatistical Estimation of Canadian Forest SOC

A fitted semi-variogram of SOC levels in the entire study area is shown in *Figure 6.5*. Recall the different types of semi-variogram presented earlier in *Figure 2.3*, our model indicates the presence of local variations and a relatively uneven SOC distribution. In Canadian forest areas, SOC was not randomly distributed, and hot spots of SOC should be observed. From *Figure 6.5*, it can be observed that the nugget effect is 0.211, total sill is 0.391, and range is about 1,000 km. Thus, the high nugget-to-sill ratio of 54% indicates that strong local-scale variations exist in the SOC distribution. In addition, the large range shown in the semi-variogram suggests that SOC is spatially autocorrelated within a neighbourhood radius of 1,000 km. Similar results were found in other large-scale SOC studies. For example, Mishra et al. (2010) found that SOC in the mid-western United States was spatially autocorrelated in a range of about 657 km. While McGrath and Zhang (2003) found that the spatial autocorrelation of SOC in grassland, Ireland, existed in a range of about 120 km.

By applying Ordinary Kriging, a spatially continuous SOC distribution map was estimated from the original SOC samples and shown in *Figure 6.6*. As expected, a west-to-central gradient of decreasing SOC stock and an increasing trend from central to eastern regions are observed. SOC and precipitation spatial patterns were quite similar, which is consistent with the previous Pearson correlation results (r = 0.436). Consistent with the strong local-scale variations of SOC distribution detected from the semi-variogram in *Figure 6.5*, some spotty or disperse patterns were observed in central and western forest ecosystems. Moreover, from the interpolated map below (*Figure 6.6*), forest SOC stock ranges from 3.66 kg/m^2 to 35.89 kg/m^2, which is a shrinking range than previously shown in *Table 6.1*. In order to assess the accuracy of Ordinary Kriging interpolation results, the leave-one-out cross validation approach was applied. *Figure 6.7* shows the plot of the measured SOC stocks (x-axis) against the interpolated SOC values (y-axis). The Pearson correlation coefficient was 0.76, indicating that the interpolated SOC values are in relatively good agreement with the measured SOC stocks.

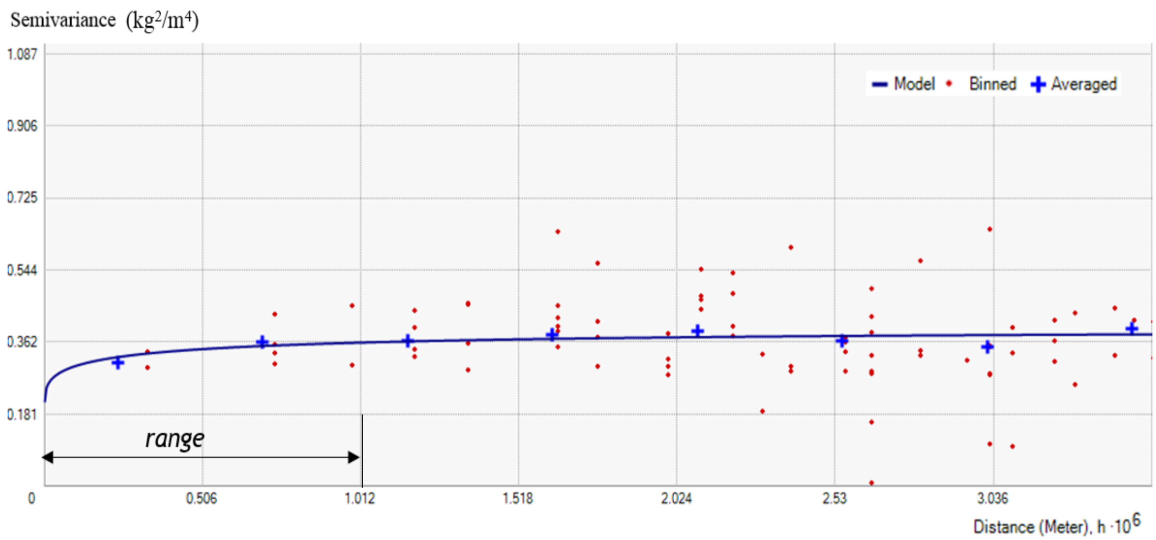


Figure 6.5 The Semi-variogram model of Canadian forest SOC distribution using Ordinary Kriging based on 1317 samples collected from Canada Forest Service

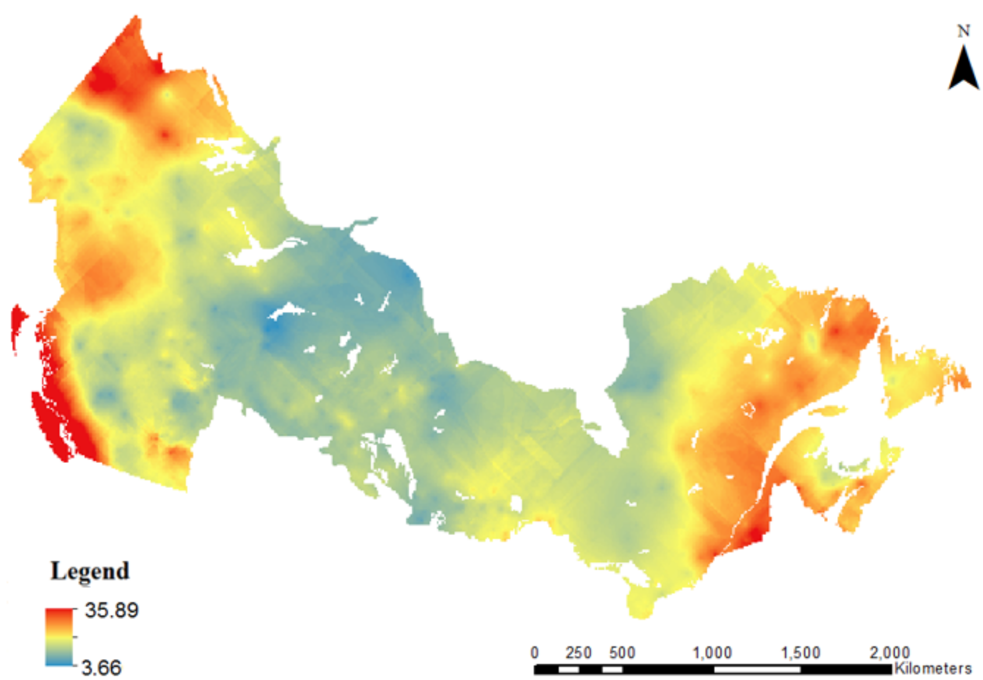


Figure 6.6 10 km gridded SOC data (before 1991) for Canadian forest areas using Ordinary Kriging based on 1317 soil samples collected from Canada Forest Service.

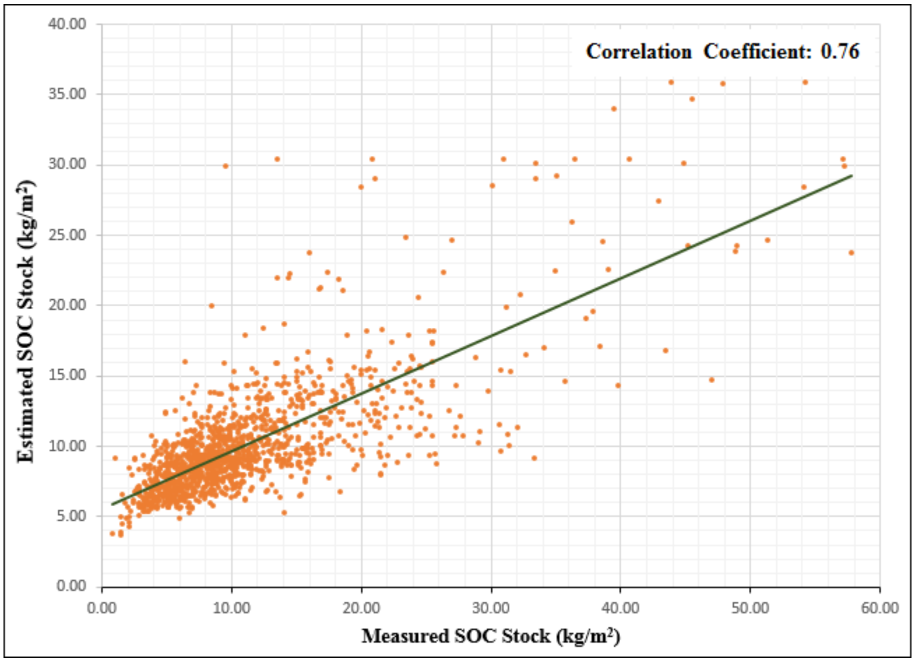


Figure 6.7 Leave-one-out cross validation of the Ordinary Kriging interpolation

6.3.2. Spatial patterns of Canada's Forest SOC Distribution

Based on the methodology for measuring spatial patterns outlined in Section 5.2.1.1, the optimal neighbourhood-sizes for testing for spatial dependence and for performing the spatial regression analysis at the national and eco-region scales were calculated and presented in *Table 6.11*. Since soil samples were not evenly distributed, potential outliers may exist in areas with low sampling densities. Ensuring each outlier to have at least one neighbour causes some samples to have excessive neighbours, thus over-estimating the optimal-neighbourhood-sizes. This study applied a three-step "standard deviation (SD)" assessing scheme: (1) calculating each sample's nearest neighbour distance, (2) calculating the SD of this set of distances, and (3) excluding the samples with a distance three times larger than the SD. Thus, potential outliers that may bias the optimal distance measurement were identified and removed before the calculation (refer to *Table 6.11*).

Then, the optimal neighbourhood-sizes were measured based on Incremental Spatial Autocorrelation analysis, which calculates global Moran's *I* and associated Z-score at a series of distance increments. Since the distances associated with the peak Z-scores indicate the spatial scales at which ecological responses of targeted objects to cluster-patterns are most notable (ESRI, 2013[b]), the first peak Z-score distance was selected as the optimal distance in order to preserve more local information. At the national scale, the first peak Z-score occurred at a distance of about 313,805 m. Thus, this distance was considered as the optimal-neighbourhood-size at which ecological activities were believed to promote a most intensive cluster-pattern. In addition, the optimal distances determined for each eco-region are summarized in *Table 6.11*. As mentioned in Section 5.2.2., in this study, the Subarctic Cordilleran eco-region was excluded due to the small sample size.

Table 6.11 Optimal distance for spatial analysis

Eco-region	Number of Samples	Standard Deviation	Outliers	Average Nearest Neighbour (m)	First Peak Z-Score Distance (m)
The entire study area	1,317	22,678.25	53	16,392.04	313,805.45
Subarctic	129	38,757.55	5	33,741.42	148,467.05
Boreal	649	22,504.70	21	18,371.85	260,811.32
Cool Temperate	86	21,150.86	6	17,706.42	114,605.65
Cordilleran	306	21,076.25	15	13,633.76	72,983.24
Interior Cordilleran	71	14,473.06	4	10,810.77	57,580.51
Pacific Cordilleran	62	23,451.71	1	11,746.60	94,840.38

6.3.2.1. *Global Spatial Autocorrelation*

In this study, the Moran's I test statistic was applied to measure the strength of spatial dependence of SOC distribution at the national and eco-region scales. The corresponding Moran's I scatterplots are shown in *Figure 6.8* and *Figure 6.9* respectively. The standard SOC stock is recorded on the x-axis and the spatially lagged SOC stock is recorded on the y-axis. As shown in *Figure 6.8*, the global Moran's I index of SOC at the national scale is about 0.289, which is relatively low yet statistically significant ($p < 0.01$). The positive global Moran's I index indicates that forest SOC is not randomly distributed.

This suggests that spatial clusters of SOC may be found across the study area, with similar values clustering close together.

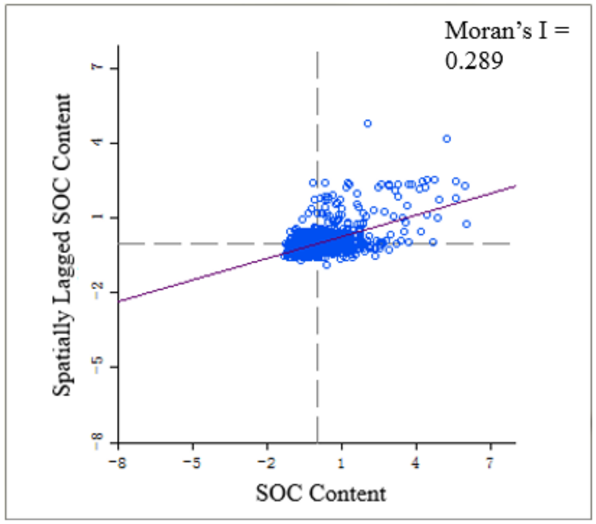


Figure 6.8 Global Moran's *I* scatterplot for SOC stock in Canadian forest areas at the national scale

To explore how SOC is distributed in different climatic zones, the global Moran's *I* scatterplots of each eco-region are shown in *Figure 6.9*. All Moran's *I* values are positive, suggesting that similar SOC values are spatially clustered together at the eco-region scale. The maximum global Moran's *I* (0.391) is observed in the Subarctic eco-region, while the minimum global Moran's *I* (0.069) is found in the Cordilleran eco-region. This suggests that although the global Moran's *I* index in the Cordilleran eco-region is significant, very weak spatial dependence, or spatial pattern, is detected.

6.3.2.2. *Local Spatial Autocorrelation*

Although global Moran's *I* test statistic quantifies the strength of SOC spatial autocorrelation, the types of spatial arrangement of SOC distribution (e.g., locations of significant hot spots) in Canadian forest areas still remains unknown. To further explore spatial patterns of SOC distribution on both national and eco-region scales, corresponding Local Spatial Autocorrelation (LISA) maps were produced based on the local Moran's *I* test statistics. Specific spatial clusters, as well as a set of outliers, are identified from this analysis. In the LISA maps below, potential High-High clusters wherein high SOC values are surrounded by high values are represented in red, and Low-Low clusters wherein low values are surrounded by low values are labelled in dark blue. In addition, potential outliers (High-Low and Low-High) are colored in pink and light blue, respectively.

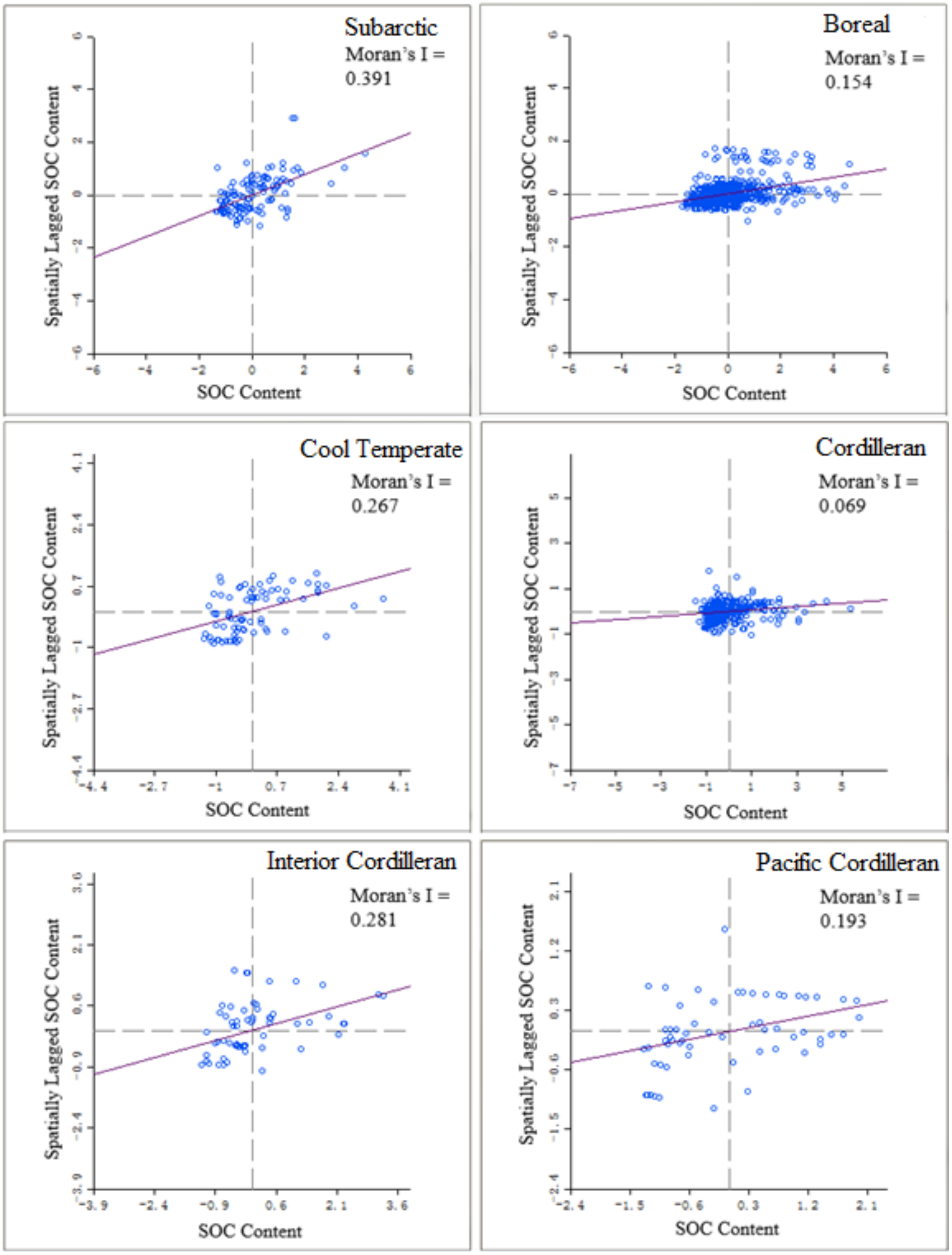


Figure 6.9 Global Moran's *I* scatterplots for Canadian forest SOC in six eco-regions: the Subarctic, Boreal, Cool Temperate, Cordilleran, Interior Cordilleran, and Pacific Cordilleran

Beginning with the SOC distribution at the national scale, from *Figure 6.10*, we found that the SOC outliers are distributed throughout HH and LL clusters across the entire study. This results in the strong nugget effect that was previously observed. For example, due to local variations, any sample of relatively higher SOC stock (e.g., caused by poor drainage system) is considered to be an outlier within a LL cluster. In addition, multiple HH clusters are identified at the national scale, including along the southwest coast. This finding is consistent with the descriptive statistics, confirming that the B.C. forest coastal areas are rich in terms of organic carbon storage. In addition, a small HH cluster is observed in the south-east of Canadian forest ecosystems in Quebec, where the growing season is warmer and longer. Another significant HH cluster is found at the border of Yukon and Northwest Territories, namely the Peel Watershed. The cold temperatures in the growing season promote low decomposition rates by limiting microbial activities in the soils. In addition, relatively higher soil moisture preserves a considerable amount of SOC accumulation. When comparing the location of HH clusters to a map of forest age distribution (*Figure 2.5*), all three HH clusters are located in old-growth forest areas (represented by the green-blue color gradient in *Figure 2.5*). This result supports previous literature that suggests old growth temperate forests accumulate high carbon stock in soils (e.g., Chen et al., 2003; Luyssaert et al., 2008).

Most SOC LL clusters were situated in central forests ecosystems, suggesting that soils in these regions have comparatively low carbon stock compared to the rest of the study area. Compared to the eco-region classification map and climatic conditions map (*Figure 3.1* and *Figure 6.1*, respectively), this LL cluster encompasses the Western Boreal eco-region, which has a drier and warmer climate compared to other eco-region classifications, which may not encourage SOC sequestration. This eco-region also consists of forest groups of various growth stages with forest age ranging from 10 years to 70 years, with a majority within the range of 10 years to 30 years (*Figure 2.5*). All of these attributes potentially explain why lower SOC levels are observed in this region.

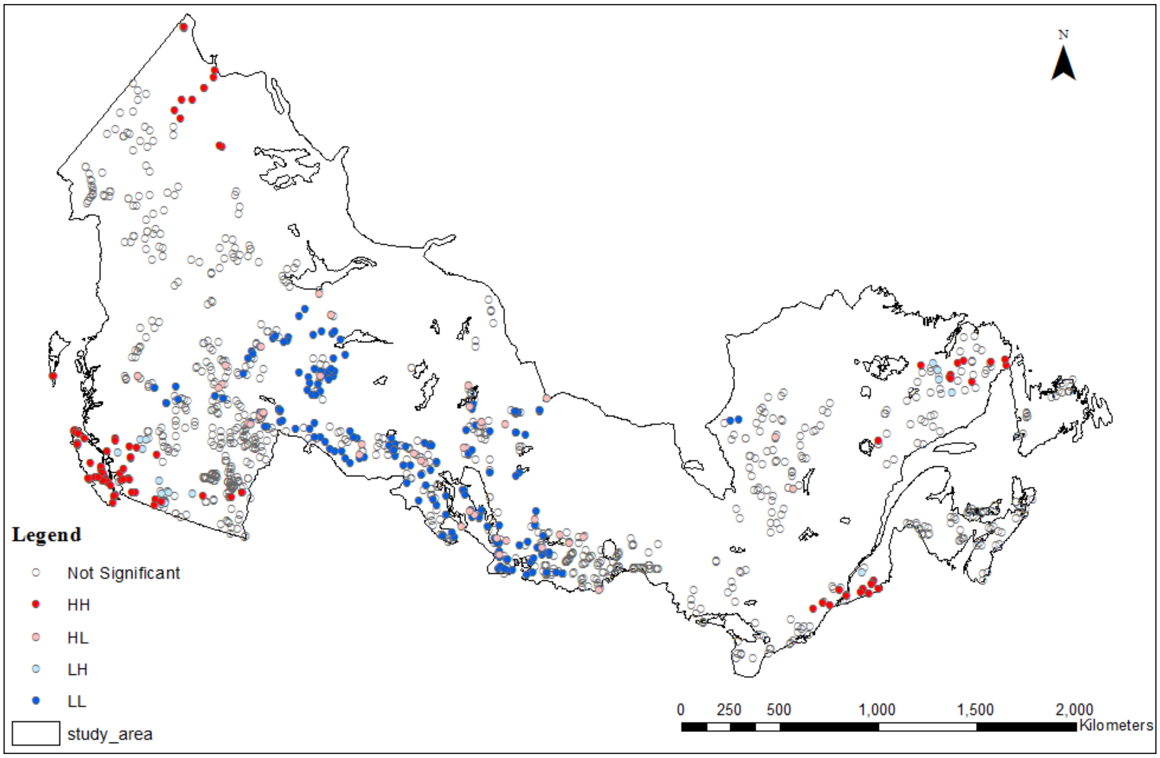


Figure 6.10 LISA cluster map of SOC distribution at the national scale

The LISA maps of SOC samples in each eco-region are shown in *Figure 6.11* to *Figure 6.15*. The Subarctic Cordilleran eco-region was excluded due to a sample size (n=14) that was too small for spatial autocorrelation calculation, as well as the Cordilleran eco-region due to the weak global Moran's *I* (0.069).

The LISA map SOC distribution in the Subarctic eco-region is shown in *Figure 6.11*. Few outliers exist in this eco-region as shown in *Figure 6.11*. Soil samples with higher organic carbon stock (highlighted by a red ellipse) are clustered in the northwest part of the Subarctic eco-region. Specifically, it is located in Peel Plateau and Peel Plain region. Soils in this area are usually cold and wet (likely affected by snowmelt during the growing season), and thus accumulate a considerable amount of organic matter, as evident in a HH cluster (Meikle & Waterreu, 2008). Another small HH cluster occurs along the east coast, mainly located in Melville Lake estuary (Newfoundland and Labrador), where the climate is humid and moist. This area is also one of the hot spots where ongoing soil organic matter research is being undertaken (e.g., the Earth Science Laboratory of Memorial University). Comparing this map to the terrain attributes (*Figure*

6.2), we find that the detected LL cluster is also located in central Quebec with little vegetation cover and biomass, which may also play a role in influencing SOC distribution in the Subarctic eco-region.

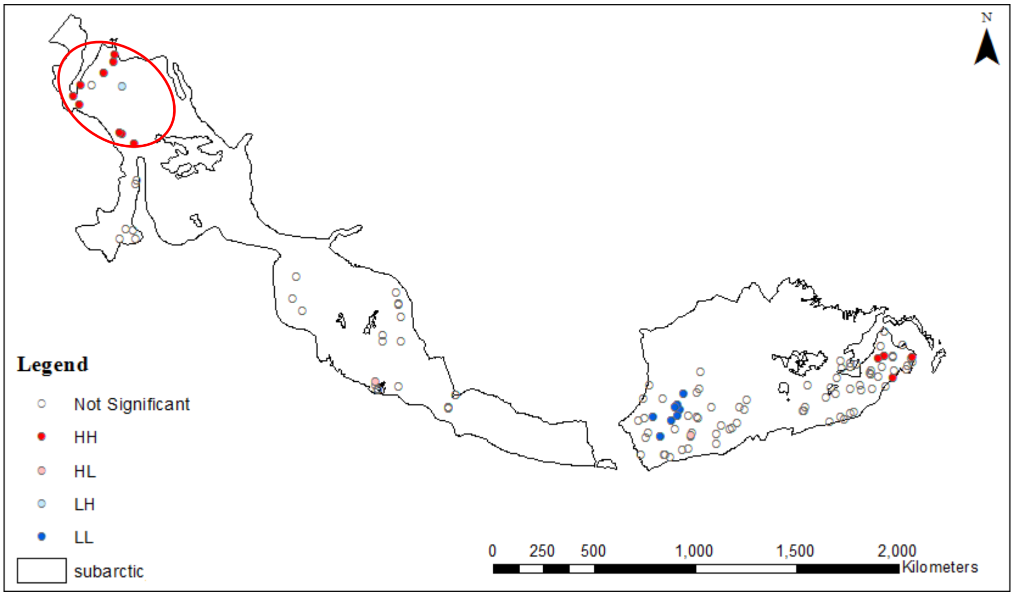


Figure 6.11 LISA cluster map of SOC distribution in the Subarctic eco-region, with Peel Watershed area labelled.

As shown in *Figure 6.12*, LL patterns mainly observed in the west Boreal and HH patterns in the east Boreal. This typical spatial pattern is quite consistent with the precipitation regime and forest age distribution characteristic of this region. In general, multiple hot spot clusters are observed in this eco-region. From *Figure 6.12*, the distribution of LL clusters in western Boreal (e.g. Alberta, Saskatchewan, and Manitoba) was not homogeneous with many HL outliers also prevalent. As previously discussed, this may be partly due to the strong local variations in SOC stock. On one hand, western Boreal areas experience a relatively mixed forest age distribution. This is likely due to the disturbance such as forest fire caused by seasonal high-temperature and human interference. Thus, different amounts of litterfall inputs potentially result in variations in SOC levels. On the other hand, differences in other terrain attributes, such as drainage capacity, soil types, and soil nutrients, also partly account for the variations observed in SOC distribution in western boreal forest ecosystems.

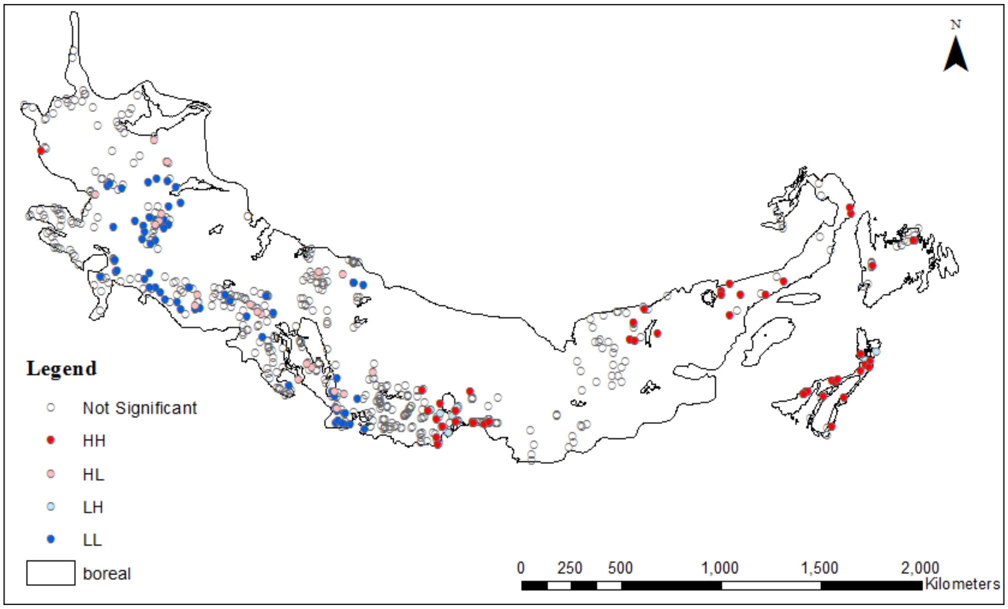


Figure 6.12 LISA cluster map of SOC distribution in the Boreal eco-region

Unlike the Boreal and Subarctic eco-regions previously discussed, the SOC distribution pattern in the Cool Temperate eco-region is mainly characterized by a single HH cluster and LL cluster. Comparing *Figure 6.13* to *Figure 6.1* (the climatic condition map), soil samples with high organic-carbon stock are clustered in the St. Lawrence watershed that receives the highest level of precipitation. The LL cluster is observed around Prince Edward Island (PEI). As shown in *Figure 6.14* and *Figure 6.15*, the cluster patterns of the Interior Cordilleran and Pacific Cordilleran eco-regions are not apparent. A small LL cluster is observed in the lower part (elevation is about 800 to 900 m) of the Interior Plateau (*Figure 6.14*), where the climate is dry and the vegetation does not effectively contribute to carbon accumulation in soils. For the Pacific Cordilleran eco-region, although the overall SOC stock is high, a LL cluster is observed in the northern region (*Figure 6.15*). This is likely caused by insufficient rainfall input and less vegetation biomass. Another LL cluster is identified at the Lower Fraser Basin in southern B.C. Province.

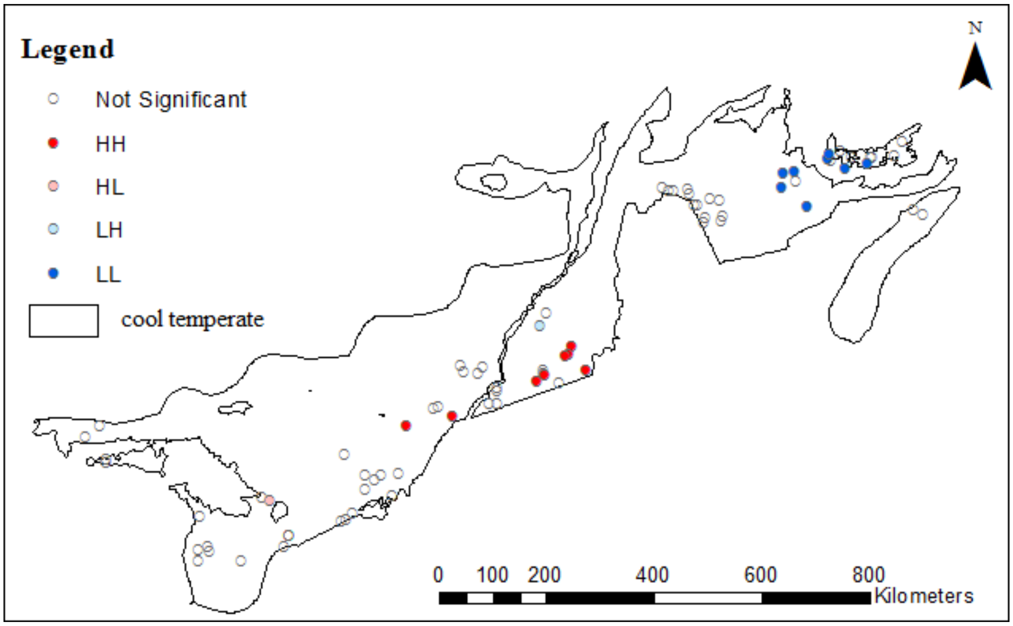


Figure 6.13 LISA cluster map of SOC distribution in the Cool Temperate eco-region

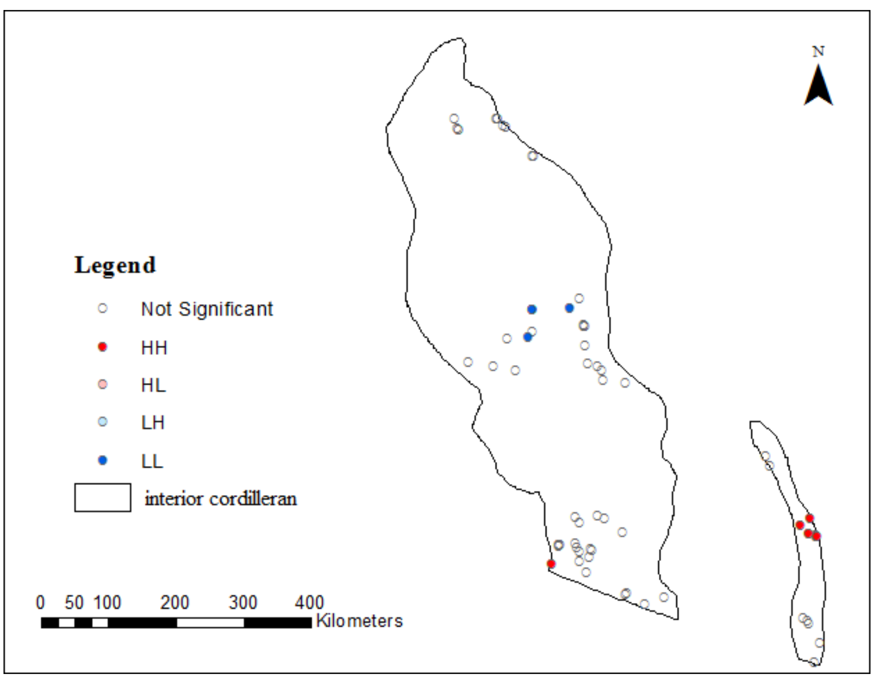


Figure 6.14 LISA cluster map of SOC distribution in the Interior Cordilleran eco-region

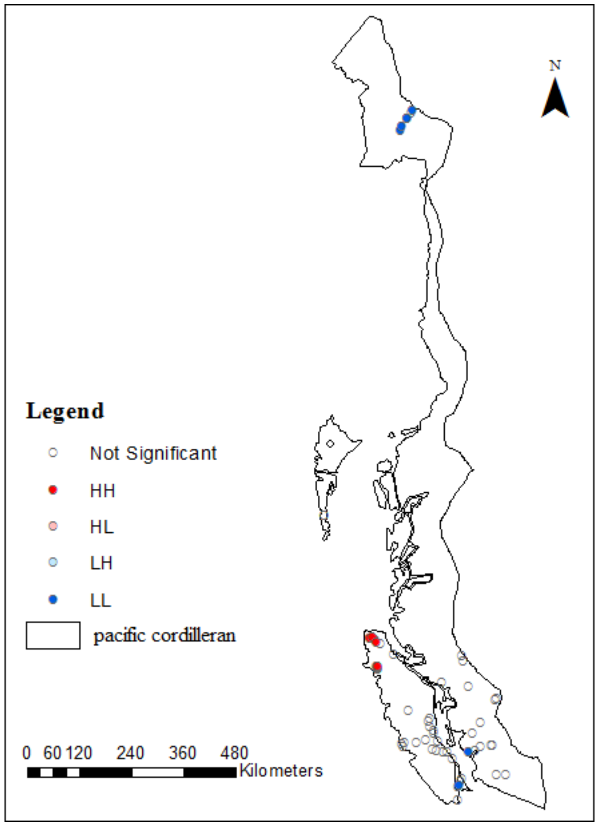


Figure 6.15 LISA cluster map of SOC distribution in the Pacific Cordilleran eco-region

6.4. Spatial Soil-Environment Modelling

Results from the exploratory analysis of Canadian forest SOC presented in Section 6.2 provides evidence of significant associations between SOC and various ecological variables; however, no causality is implied. Spatial regression models were thus developed to explore the relationships between SOC and environmental determinants. Specifically, two questions are addressed. First, what are the dominant environmental determinants that influence Canadian forest SOC distribution? Second, how do these SOC-environment relationships vary spatially across Canadian forest areas?

In order to answer these questions, traditional OLS models are tested at two spatial scales of analysis, namely national and eco-region scales. The models are based on six independent variables: precipitation, temperature, NDVI, elevation, slope, and aspect. As previously discussed, environmental determinants that are important in one ecosystem may not be equally important in another area (Powers & Schlesinger, 2002). For example, Chuai et al. (2012) found a positive SOC-elevation relationship in woody areas in Jiangsu Province (China); however, no association between SOC (measured to a depth of 50 cm of mineral soil) and elevation in the Great Smoky Mountains National Park (U.S.) was found by Tewksbury and Miegroet (2007). Thus, an eco-region classification framework was adopted in this study to assess the SOC-environment relationships at a local level and to explore dominant ecological variables within each eco-region.

In order to avoid statistical problems such as unstable parameters and unreliable significance tests, spatial dependency in OLS regression models were tested. Significant autocorrelation in OLS residuals violates the assumption of independence and indicates potentially biased regression coefficients and estimation errors. In this study, from the spatial autocorrelation results presented in Section 6.3, significant spatial dependency in SOC distribution at the national and eco-region scales was noted. If the spatial autocorrelation in the dependent variable (e.g., SOC stock) cannot be adequately explained by independent variables (e.g., environmental determinants), autocorrelation will be detected among the regression residuals, and thus lead to potential misspecification in the modelling of SOC-environment relationships (Chakraborty, 2011; Collins et al., 2006).

In this study, the strength of spatial autocorrelation of OLS models' residuals was measured based on the Moran's I test statistic. Any statistically significant Moran's I value indicates the inappropriate use of using a non-spatial OLS model, suggesting that it is necessary to take spatial effects into consideration. In this study, the selection of spatial models is determined based on the Lagrange Multiplier diagnostic (refer to Section 5.3 for details): a statistically significant LM-lag test suggests the inclusion of a spatial lagged dependent variable, while a statistically significant LM-error test suggests adding an autocorrelated error term. When both tests are statistically significant, robust versions are tested for model specification. In the following subsections, the results of model specification and testing at the national and eco-region scales are presented.

6.4.1. Soil-Environmental Modelling at a National Scale

OLS regression models were applied at the national scale based on six independent variables: precipitation, temperature, elevation, slope, aspect, and NDVI. The SOC data (the dependent variable) compiled by Canada Forest Service (CFS) is a set of historical soil records collected before 1991. *Table 6.12* below shows the results of the initial estimation of three OLS models. Regression model *(1)*, *(2)*, and *(3)* differ in terms of the type of temperature readings included in the model specification, based on maximum, mean, and minimum temperature. For the three estimated OLS models, the R^2 values were approximately 0.24, indicating that about 24% of variation in SOC distribution was explained by the initial OLS models. With the exception of NDVI (e.g., p = 0.708 from OLS model *(3)*) and aspect (e.g., p = 0.253 from OLS model *(3)*), all other four independent variables were statistically significant at a 10% level. Compared to other independent variables, precipitation (p = 0.000) was shown to have the most significant influence on SOC distribution at the national scale.

To evaluate the performance of three OLS models, the strength of residual spatial autocorrelation was measured based on the Moran's *I* test statistic. As shown in *Table 6.13*, all three Moran's *I* tests were positive and highly significant ($p < 0.01$), indicating that spatial clustering existed in the regression residuals. Thus, spatial regression models were tested to explore whether taking spatial information into consideration would improve the estimation of relationships between SOC and ecological variables.

As discussed in Section 5.3, Lagrange Multiplier diagnostics were applied to assist the optimal model specification. The results of Lagrange Multiplier tests are shown in *Table 6.13*. Since both standard LM-lag and LM-error tests were highly significant (p = 0.001), robust versions were tested for model specification. As shown in *Table 6.13*, the robust LM-error tests (p = 0.000) were slightly more significant than the robust LM-lag tests (p = 0.001). Therefore, a spatial error model specification was selected.

Table 6.12 OLS and spatial error models of SOC-environmental relationships in Canadian forest areas at the national scale (n = 1317). Model (1) includes maximum temperature as one of the independent variables, while model (2) and (3) includes mean and minimum temperature, respectively.

Independent Variables	***OLS Model (1)***	***Spatial Error Model (1)***	***OLS Model (2)***	***Spatial Error Model (2)***	***OLS Model (3)***	***Spatial Error Model (3)***
Intercept *Sig.*	4.794*** *0.000*	-1.675 *0.557*	3.375*** *0.006*	-1.951 *0.438*	2.312* *0.092*	-0.001 *0.999*
Precipitation (cm) *Sig.*	0.203*** *0.000*	0.240*** *0.000*	0.212*** *0.000*	0.236*** *0.000*	0.217*** *0.000*	0.219*** *0.000*
Max. Temp. (˚C) *Sig.*	-0.304*** *0.006*	0.213 *0.192*	--	--	--	--
Mean Temp. (˚C) *Sig.*	--	--	-0.226* *0.099*	0.413** *0.046*	--	--
Min. Temp. (˚C) *Sig.*	--	--	--	--	-0.300** *0.019*	0.581** *0.011*
Elevation (km) *Sig.*	-0.853* *0.072*	0.821 *0.319*	-0.990* *0.054*	1.254 *0.151*	-1.094* *0.041*	1.681* *0.067*
Slope (%) *Sig.*	1.972*** *0.000*	0.574** *0.063*	1.998*** *0.000*	0.575* *0.062*	2.033*** *0.000*	0.570* *0.065*
Aspect (˚) *Sig.*	-0.002 *0.278*	0.0001 *0.973*	-0.001 *0.265*	0.0001 *0.925*	-0.002 *0.253*	0.0002 *0.885*
NDVI *Sig.*	2.353 *0.402*	-2.621 *0.362*	1.110 *0.680*	-3.267 *0.237*	-0.907 *0.708*	-2.611 *0.303*
Spatial Error Term (λ) *Sig.*	--	0.793*** *0.000*	--	0.805*** *0.000*	--	0.810*** *0.000*
R^2 (Pseudo R^2)	0.238	(0.344)	0.237	(0.346)	0.235	(0.347)
Log Likelihood	-4383.48	-4298.36	-4384.49	-4297.35	-4385.89	-4296.241
AIC	8780.96	8610.72	8782.99	8608.71	8785.78	8606.48

$p < 0.1$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

Table 6.13 Lagrange Multiplier diagnostic tests for the SOC-environment relationships at the national scale based on six ecological determinants. Model (1), (2), and (3) represent the inclusion of maximum, mean, and minimum temperature, respectively.

Dependence Test	***Value***		
	Model (1)	***Model (2)***	***Model (3)***
Moran's *I* (Residual) *Sig.*	0.129 *0.000*	0.133 *0.000*	0.134 *0.000*
LM-lag *Sig.*	358.725 *0.000*	370.806 *0.000*	382.745 *0.000*
Robust LM-lag *Sig.*	12.832 *0.000*	11.338 *0.001*	11.244 *0.001*
LM-error *Sig.*	576.000 *0.000*	605.507 *0.000*	618.851 *0.000*
Robust LM-error *Sig.*	230.107 *0.000*	246.038 *0.000*	247.350 *0.000*

A spatial error model specification suggests that cluster patterns in Canadian forest SOC distribution are likely due to the omission of other spatially autocorrelated variables that potentially influence SOC distribution, such as soil pH and nitrogen content, which could not be accounted in this study due to data availability. Thus, the correlated error term (λ) is added to the regression model as an independent variable to partly account for the effects of unobservable and unmeasurable factors. The results of initial spatial error models are presented in *Table 6.12*.

Compared to initial OLS models, general improvements in model fit (e.g., lower AIC values and higher log likelihood values) are observed. Similar to the results of the traditional OLS models, precipitation (p = 0.000) was the most significant environmental determinant influencing SOC. The spatial error term (λ) was also statistically significant (p = 0.000), further supporting the notion that important unobservable or unmeasurable variables are missing from the model specification. As shown in *Table 6.12*, initial estimation of spatial error models suggested that aspect of slope (e.g., p = 0.885 from spatial error model *(3)*) and NDVI (e.g., p = 0.303 from spatial error model *(3)*) had a weaker relationship with SOC distribution. In addition, compared to maximum and mean temperature (p = 0.192, p = 0.046), the minimum temperature regime (p = 0.011) had a more significant relationship with SOC distribution in Canadian forest areas. In order to improve the performance of spatial error models, independent variables that were less significant ($p > 0.1$) were removed from the model specification. *Table 6.14* shows the results of the modified spatial error models.

The re-estimated spatial error model (a) includes two independent variables, namely precipitation and slope. While spatial error model (b) also takes elevation and minimum temperature into account. Comparing the two models, the AIC values remained almost the same, suggesting minimal improvement in the model goodness of fit. In addition, compared to the initial spatial error model *(3)* (in *Table 6.12*), a general increase in significance of independent variables were observed in the re-estimated spatial error model (b) ($p < 0.05$). This indicated that the most optimal model fit was achieved by a spatial error model (b). As shown in spatial error model (b) (*Table 6.14*), all environmental determinants and the error term were positively related to the SOC distribution, and precipitation was proven to be the most important variable ($p = 0.000$). Thus, it can be concluded that four dominant ecological variables influence Canadian forest SOC distribution at the national scale, namely precipitation, minimum temperature, elevation, and slope.

Table 6.14 Re-estimated spatial error models of SOC-environmental relationships in Canadian forest areas at the national scale (n = 1317). Spatial error model (a) includes two independent variables: precipitation and slope. While spatial error model (b) takes minimum temperature and elevation into account.

Independent Variables	***Spatial Error Model (a)***	***Spatial Error model (b)***
Intercept *Sig.*	0.942 *0.441*	-1.299 *0.425*
Precipitation (cm) *Sig.*	0.229*** *0.000*	0.221*** *0.000*
Min. Temperature (˚C) *Sig*	--	0.500** *0.021*
Elevation (km) *Sig.*	--	1.877** *0.036*
Slope (%) *Sig.*	0.626** *0.032*	0.570* *0.065*
Spatial Error Term (λ) *Sig.*	0.779*** *0.000*	0.810*** *0.000*
Pseudo R^2	0.342	0.347
Log Likelihood	-4299.498	-4296.780
AIC	8605.000	8603.560

$p < 0.1$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

6.4.2. Soil-Environmental Modelling at an Eco-Region Scale

After identifying dominant ecological variables influencing SOC levels at the national scale, the next stage of this study is to explore relationships at the local scale. In this section, the SOC-environment relationships and dominant ecological variables are examined within each eco-region in Canada, namely the Boreal, Subarctic, Cool Temperate, Cordilleran, Interior Cordilleran, and Pacific Cordilleran. Details about model specifications at the eco-region scale are further discussed.

6.4.2.1. *Local-scale Regression Analysis of the Boreal Eco-region*

The Boreal eco-region is the largest eco-climatic zone in Canada, stretching from Alberta to central Quebec and Nova Scotia. Different climate regimes are observed within this east-to-west running belt. Alberta, Saskatchewan, and Manitoba mainly experience a cooler and drier growing season, while the growing seasons of Ontario and Quebec are usually longer, warmer, and rainy. To start estimating the relationships between SOC and ecological variables, three initial OLS models were fitted based on all six independent variables (see *Appendix III*). The global Moran' *I* test statistic on OLS residuals were shown to be highly autocorrelated (p = 0.000), thus suggesting the necessity of considering spatial effects in the model.

Results from Lagrange Multiplier diagnostic tests suggested that a spatial lag model should be used to estimate the SOC-environment relationship in the Boreal eco-region. As shown in *Table 6.15*, both standard LM-lag and LM-error tests were highly significant at the p = 0.000 level. Thus, the robust versions should be considered to guide the model specification process. Since the robust LM-error tests (e.g., p = 0.157 from Model *(1)*) were less significant at the p = 0.1 level, a spatial lag model was shown to be the appropriate approach. This indicates that neighbourhood effects of SOC stock were more important than the ones caused by unobserved variables. This implies that the ecological influences on SOC stock in one area are more likely to "spill-over" to its neighbours. Thus, spatial lag models were tested based on a full set of six independent variables to explore the SOC-environment relationships in the Boreal eco-region (see *Appendix III*). However, initial estimation of spatial lag models suggested that NDVI, aspect, elevation, and temperature were not significantly related to SOC distribution at the p = 0.1 level. Thus, the spatial lag model was re-estimated to exclude these insignificant variables to improve model specification. Results of this spatial lag model are shown in *Table 6.16* below.

In the Boreal eco-region, a positive association between precipitation and SOC stock was observed (coefficient β = 0.058). This is an expected result because this eco-region experiences an obvious increase of rainfall supply from west to east. In humid areas such as eastern Boreal region, SOC decomposition rates are potentially lowered because sufficient precipitation ensures high levels of water-saturation in soils and thus limits oxygen diffusion processes. In addition, the positive relationship between SOC and slope (coefficient β = 1.055) indicates that higher SOC stock tend to be observed in areas with relatively steeper slopes. In general, soils in lower slope positions along a slope gradient tend to hold higher level of soil moisture and SOC stock because of sediment movement. However, due to data availability, slope position information for each sample is missing from the CFS database, making it difficult to further assess the influence of slope attributes on SOC distribution in this study.

Table 6.15 Lagrange Multiplier diagnostics for SOC-environment relationships in the Boreal eco-region based on six ecological variables as determinants. Model (1), (2), and (3) represent the inclusion of maximum, mean, and minimum temperature, respectively.

Dependence Test	***Value***		
	Model (1)	***Model (2)***	***Model (3)***
Moran's *I* (Residual) *Sig.*	0.053 *0.000*	0.054 *0.000*	0.054 *0.000*
LM-lag *Sig.*	35.793 *0.000*	36.684 *0.000*	38.179 *0.000*
Robust LM-lag *Sig.*	8.942 *0.003*	9.424 *0.002*	10.765 *0.001*
LM-error *Sig.*	28.851 *0.000*	29.615 *0.000*	30.425 *0.000*
Robust LM-error *Sig.*	2.000 *0.157*	2.355 *0.124*	3.011 *0.083*

Table 6.16 Spatial lag model of SOC-environmental relationships in the Boreal eco-region (n = 628).

Independent Variables	***Spatial Lag Model***
Intercept *Sig.*	1.892** *0.022*
Precipitation (cm) *Sig.*	0. 058*** *0.001*
Slope (%) *Sig.*	1.055* *0.061*
Spatial lag Term (ρ) *Sig.*	0.496*** *0.000*
Pseudo R^2	0.132
Log Likelihood	-1890.940
AIC	3789.880

$p < 0.1$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

6.4.2.2. *Regression Analysis of the Cordilleran and Pacific Cordilleran Eco-region*

Next, the relationships between SOC and ecological variables in the Cordilleran and Pacific Cordilleran eco-regions were examined. First, initial OLS models were estimated based on the full set of six independent variables (see *Appendix IV* and *V,* respectively). Then, in order to evaluate the performance of initial OLS models, the strength of residual spatial autocorrelation was measured based on the Moran's *I* test statistic, and results were shown in *Table 6.17*. As shown in *Table 6.17*, all six Moran's *I* values were not statistically significant ($p > 0.1$), suggesting that spatial dependence among OLS residuals is insignificant. Thus, non-spatial OLS regression models were applied to estimate the SOC-environment relationships in these two eco-regions. After excluding insignificant independent variables ($p > 0.1$), important environmental determinants used to estimate the SOC-environment relationships in these two eco-regions are presented in *Table 6.18*.

In the Pacific Cordilleran eco-region, the significant independent variable was shown to be precipitation ($p < 0.01$), indicating that SOC stock was not significantly sensitive to temperature regimes and terrain attributes (e.g., elevation, slope, and NDVI). A positive association between precipitation and SOC stock was observed (coefficient β = 0.26). The R^2 was about 0.297, indicating that approximately 29.7% of the variation in SOC stock was explained by the precipitation regime. The Cordilleran eco-region is a north-to-south climatic belt which occupies most areas of B.C. Province and a small part of the southern Yukon Territory. Unlike the climatic regime in the Pacific Cordilleran eco-region, growing season in the Cordilleran eco-region is usually cold and short in this mountainous area. As shown in *Table 6.18*, precipitation is no longer statistically significant at the 10% level. Instead, SOC stock is negatively influenced by the maximum temperature (coefficient $\beta = -0.458$, $p = 0.002$). The R^2 of 0.03 was quite low, indicating that the maximum temperature regime could only explain 3% of the variation in SOC stock in the Cordilleran eco-region. Nevertheless, maximum temperature is the most significant ecological variable ($p = 0.002$) among the six independent variables.

Table 6.17 The Moran's *I* test statistic of the initial OLS residuals in the Cordilleran and Pacific Cordilleran eco-regions. Model (1) includes maximum temperature as an independent variable, while model (2) and (3) includes mean and minimum temperature, respectively.

Dependence Test	***Moran's I Values***		
	Model (1)	***Model (2)***	***Model (3)***
Pacific Cordilleran (n=61) *Sig.*	-0.079 *0.698*	-0.077 *0.728*	-0.075 *0.752*
Cordilleran (n=291) *Sig.*	0.029 *0.168*	0.031 *0.147*	0.030 *0.155*

Table 6.18 OLS models of SOC-environmental relationships in the Cordilleran and Pacific Cordilleran eco-regions.

Eco-region	***Independent Variable***	***Regression Coefficient***	R^2	***AIC***	***Log Likelihood***
Pacific Cordilleran (n=61)	Precipitation (cm) *Sig.*	0.260 *0.000*	0.297	493.005	-244.502
Cordilleran (n=291)	Max. Temperature (˚C) *Sig.*	-0.458 *0.002*	0.030	2021.35	-1008.68

6.4.2.3. *Local-scale Regression Analysis of the Subarctic, Cool Temperate, and Interior Cordilleran Eco-regions*

Finally, initial OLS models of the remaining three eco-regions, the Subarctic, Cool Temperate, and Interior Cordilleran, were estimated based on the full set of six independent variables. For each of the three eco-regions, the strength of OLS model's residuals were measured based on the Moran's *I* test statistics. Results from spatial autocorrelation tests on OLS residuals in the three eco-regions suggested that OLS residuals were spatially autocorrelated. Thus, it is necessary to take spatial effects into consideration in order to improve the model fit. However, for each of the three eco-regions, none of the Lagrange Multiplier tests was statistically significant, making it difficult to identify an appropriate model specification (e.g., spatial lag model or spatial error model). *Table 6.19* shows the results of Lagrange Multiplier diagnostics of the Subarctic eco-region. Although the standard LM-lag and LM-error tests were proven to be significant at the $p = 0.000$ level, the less significant robust tests (e.g., robust LM-lag $p = 0.089$ from Model *(1)*) indicated that applying either a spatial lag or error models specification would be inappropriate in the Subarctic eco-region. As shown in *Table 6.20* and *Table 6.21*, similar results were observed in the Cool Temperate and Interior Cordilleran eco-regions, respectively. Thus, OLS regression models were tested for the three eco-regions.

Table 6.19 Lagrange Multiplier diagnostics for SOC-environment relationships in the Subarctic eco-region based on six ecological variables as determinants. Model (1), (2), and (3) represent the inclusion of maximum, mean, and minimum temperature, respectively.

Dependence Test	***Value***		
	Model (1)	***Model (2)***	***Model (3)***
Moran's *I* (Residual) *Sig.*	0.336 *0.000*	0.332 *0.000*	0.331 *0.000*
LM-lag *Sig.*	41.319 *0.000*	40.605 *0.000*	40.125 *0.000*
Robust LM-lag *Sig.*	2.897 *0.089*	2.949 *0.086*	2.634 *0.105*
LM-error *Sig.*	38.460 *0.000*	37.682 *0.000*	37.491 *0.000*
Robust LM-error *Sig.*	0.038 *0.845*	0.027 *0.870*	0.001 *0.974*

Table 6.20 Lagrange Multiplier diagnostics for SOC-environment relationships in the Cool Temperate eco-region based on six ecological variables as determinants. Model (1), (2), and (3) represent the inclusion of maximum, mean, and minimum temperature, respectively.

Dependence Test	***Value***		
	Model (1)	***Model (2)***	***Model (3)***
Moran's *I* (Residual) *Sig.*	0.051 *0.052*	0.077 *0.012*	0.101 *0.002*
LM-lag *Sig.*	2.201 *0.138*	3.200 *0.074*	4.821 *0.028*
Robust LM-lag *Sig.*	3.617 *0.057*	2.345 *0.126*	2.432 *0.119*
LM-error *Sig.*	0.757 *0.384*	1.771 *0.183*	3.001 *0.083*
Robust LM-error *Sig.*	2.173 *0.140*	0.916 *0.338*	0.619 *0.431*

Table 6.21 Lagrange Multiplier diagnostics for SOC-environment relationships in the Interior Cordilleran eco-region based on six ecological variables as determinants. Model (1), (2), and (3) represent the inclusion of maximum, mean, and minimum temperature, respectively.

Dependence Test	***Value***		
	Model (1)	***Model (2)***	***Model (3)***
Moran's *I* (Residual) *Sig.*	0.125 *0.015*	0.122 *0.019*	0.120 *0.020*
LM-lag *Sig.*	2.810 *0.094*	2.597 *0.107*	2.520 *0.112*
Robust LM-lag *Sig.*	1.231 *0.267*	1.021 *0.312*	0.969 *0.325*
LM-error *Sig.*	2.051 *0.152*	1.95 *0.164*	1.884 *0.170*
Robust LM-error *Sig.*	0.472 *0.492*	0.358 *0.549*	0.333 *0.564*

For each of the three eco-regions, the Subarctic, Cool Temperate, and Interior Cordilleran, less significant independent variables ($p > 0.1$) were excluded in order to improve model fit. Results of the re-estimated OLS models are shown in *Table 6.22* and *Table 6.23*, respectively. Compared to other independent variables, NDVI was shown to be the dominant ecological factor in northern woodlands ($p = 0.008$). In this area, vegetation density is relatively sparser than other eco-climate zones, mainly due to the cold and dry climate regime limiting vegetation growth. Thus, the importance of vegetation biomass in SOC-environment relationship estimation is evident in this area.

The positive association (β = 24.945) between NDVI and SOC indicates that litterfall inputs and root exudates potentially play a role in increasing SOC accumulation. In the Interior Cordilleran eco-region, SOC stock was significantly related to elevation (p = 0.004), but elevation only accounts for about 11.4% variation of SOC distribution in the Interior Plateau. The regression coefficient (β = 6.323) associated with elevation indicated a positive influence of elevation on SOC stock.

The Cool Temperate eco-region is mainly located in southern Ontario and southern Quebec. Comparing the three re-estimated OLS models in the Cool Temperate eco-region, about 20% of SOC stock variation was explained by the three environmental determinants, namely precipitation, temperature, and slope. All three environmental determinants were shown to positively influence SOC stock. Considering the temperature sensitivity of SOC distribution, all types of temperature measures were proven to be significant at the p = 0.1 level, especially maximum temperature (p = 0.000). One possible explanation is that increasing temperature potentially lengthens the growing season in the Cool Temperate eco-region. Forest growth adds carbon input (e.g., litterfall) into soils and is more responsive to temperature rise than SOC decomposition in the Cool Temperate eco-region, thus encouraging SOC accumulation.

Table 6.22 OLS models of SOC-environmental relationships in the Subarctic and Interior Cordilleran eco-regions.

Eco-region	***Independent Variable***	***Regression Coefficient***	R^2	***AIC***	***Log Likelihood***
Subarctic (n=124)	NDVI *Sig.*	24.945 *0.008*	0.054	889.266	-442.633
Interior Cordilleran (n = 67)	Elevation *Sig.*	6.323 *0.004*	0.114	456.816	-226.408

Table 6.23 OLS models of SOC-environmental relationships in the Cool Temperate eco-region (n = 80). Model (1), (2), and (3) represent the inclusion of maximum, mean, and minimum temperature, respectively.

Independent Variables	***Model (1)***	***Model (2)***	***Model (3)***
Intercept *Sig.*	-49.940*** *0.006*	-37.798** *0.017*	-17.124 *0.124*
Precipitation (cm) *Sig.*	0.229*** *0.000*	0.368*** *0.003*	0.300** *0.015*
Max. Temperature (˚C) *Sig*	2.103** *0.005*	--	--
Mean. Temperature (˚C) *Sig*	--	2.049** *0.014*	--
Min. Temperature (˚C) *Sig*	--	--	1.262* *0.096*
Slope (%) *Sig.*	6.837** *0.047*	8.057** *0.019*	8.954** *0.011*
R^2	0.212	0.195	0.162
Log Likelihood	-278.334	-279.268	-281.013
AIC	564.669	566.535	570.026

$p < 0.1$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

6.5. Predictive SOC Distribution Map

At the national scale, spatial regression results informed four dominant ecological variables that influence Canadian forest SOC distribution, namely precipitation, minimum temperature, elevation, and slope. Using a pair-wise comparisons scheme, the weights of four dominant ecological variables were calculated using the AHP approach based on corresponding regression coefficients derived from the spatial error model estimated at the national scale (see *Appendix VI*). Accordingly, a predictive SOC distribution map in the period 1961-1991 was created from a three-step procedure: (1) assigning weights to each criteria, (2) summing up the weighted criteria on a pixel by pixel basis, and (3) standardizing the pseudo SOC stock range as zero to one. Since it is difficult to obtain absolute SOC stock from the modelled SOC distribution map based on four ecological variables, the range of modelled SOC stock was standardized as zero to one to map the forest SOC distribution gradient on a national scale. In order to make the modelled SOC distribution map comparable with the interpolated SOC map, the latter was also standardized to a range from zero to one. In this way, it was possible to examine the differences in spatial patterns between the modelled and interpolated SOC distribution maps. *Figure 6.16* shows the final predictive SOC distribution map estimated by spatial error model parameters.

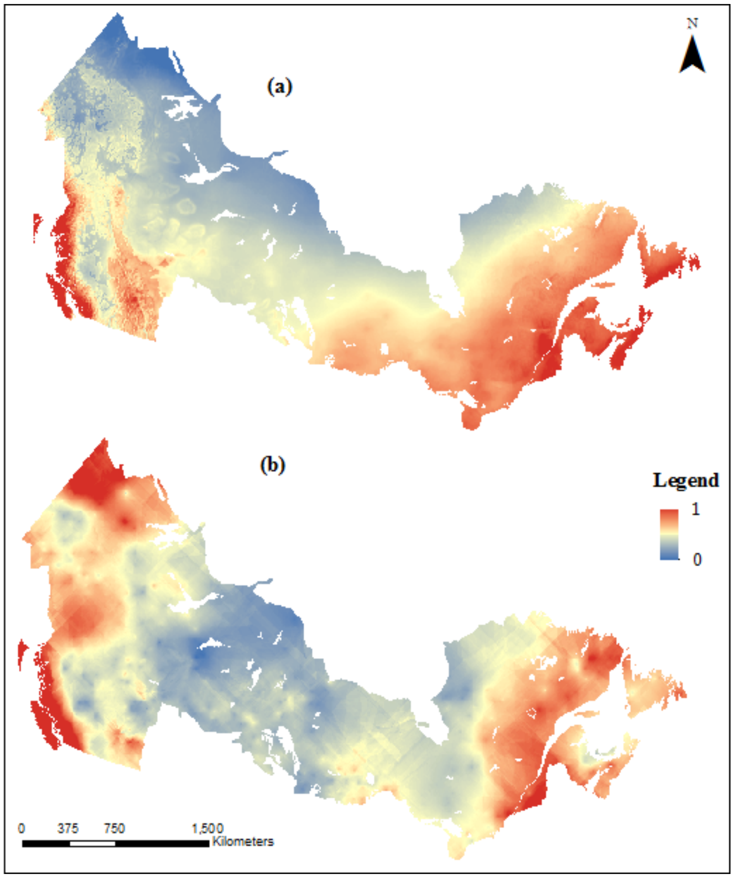


Figure 6.16 (a) Predictive SOC distribution (1961-1991) estimating from spatial error model parameters at the national scale: precipitation, minimum temperature, elevation, and slope. (b) Interpolated SOC distribution (1961-1991) using Ordinary Kriging based on the CFS SOC dataset. For each map, the SOC stock is standardized as zero to one: dividing the differnece between the raw SOC stock and the minimum value by the range of raw SOC stock

To assess the SOC-environment relationship estimation at the national scale, differences in the pseudo SOC stock between the predictive SOC map estimated from spatial error model parameters, precipitation, minimum temperature, elevation, and slope, and the interpolated SOC map using Ordinary Kriging are shown in *Figure 6.17*. An image differencing method was used by subtracting the interpolated SOC stock map from the previous predictive SOC map. It should be noted that the values shown in the legend do not have absolute meaning, since they are calculated from pseudo SOC stock.

Compared to the standardized interpolated SOC distribution map, it was found that the general trend of SOC distribution was captured by the four dominant ecological variables. In general, SOC stock in B.C. coastal areas was shown to be relatively accurate, since differences were close to zero. However, big differences in SOC stock were identified in the Subarctic Cordilleran eco-region (Yukon Territory), southern Ontario, and Prince Edward Island. Missing important ecological variables are considered to be a likely source of model estimation error. For example, SOC sequestration ability in Prince Edward Island may be limited due to forest clearance or land use practices that were not factored into the model.

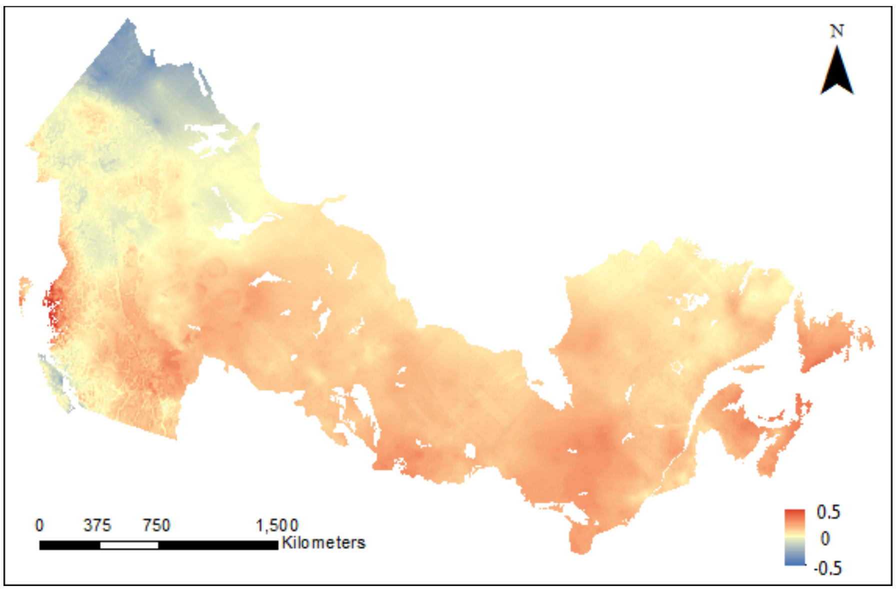


Figure 6.17 Spatial distribution of differences between the predictive SOC stock estimated from spatial error model parameters at the national scale and the interpolated SOC stock using Ordinary Kriging

6.6. Chapter Summary

In summary, this chapter presents the results of descriptive statistics of SOC stock in Canadian forest areas, describes the spatial patterns of SOC distribution at the national and eco-region scales, and examines the spatial relationships between SOC and six ecological variables. Results of descriptive statistics showed that most SOC stock was around 9 kg/m^2. Specifically, the highest SOC stock was observed in B.C. forest coastal areas. Soils in the Subarctic and Subarctic Cordilleran eco-regions, which mainly cover the woodland around the 60° N latitude mark, also contained a relatively large amount of organic-carbon stock. Although the Boreal eco-region is the largest eco-climatic zone covered with intensive forests, the lowest SOC stock was observed in this region. This is likely because the drier climatic conditions and younger growth forests in the western Boreal region (e.g., Alberta and Saskatchewan) limit SOC sequestration ability.

In addition, a spatially continuous SOC distribution map was estimated by applying Ordinary Kriging. Results showed that, at the national scale, Canadian forest SOC stock generally diminished from the east and west coasts towards the interior continental regions. In particular, results from the fitted semi-variogram indicated that, although SOC was spatially autocorrelated within a large range of 1,000 km, strong local-scale variations likely caused by terrain attributes (e.g., drainage capacity and soil texture) existed in Canadian forest areas.

To further explore how SOC was distributed, the spatial autocorrelation of SOC samples was measured at both national and eco-region scales based on the Moran's *I* test statistics. Results showed that, SOC was not randomly distributed at the national scale, with similar values tending to spatially cluster together (Moran's $I = 0.289$). Specifically, within a range of 313,805 m determined from Incremental Spatial Autocorrelation analysis, the ecological activities were considered to promote the most notable cluster-pattern of SOC distribution. Thus, this distance was used as the optimal scale for further spatial analysis at the national scale. The global Moran's *I* test statistic on SOC samples within the six eco-regions (the Subarctic, Boreal, Cool Temperate, Cordilleran, Interior Cordilleran, and Pacific Cordilleran) showed that all six Moran's *I* values were significantly positive, thus verifying the presence of spatial dependency of SOC distribution at the eco-region scale. Moreover, the strength of SOC spatial dependence varies across different eco-climatic zones. For example, SOC in the Subarctic eco-region (woodland around 60° N latitude) tends to be highly spatially autocorrelated (Moran's $I = 0.391$). However, in the Cordilleran eco-region (B.C. mountainous areas), the strength of SOC spatial autocorrelation was quite weak (Moran's $I = 0.069$).

Results from local spatial autocorrelation analysis allowed us to identify the locations of potential hot spots and outliers in SOC distribution. At the national scale, multiple high-high (HH) and low-low (LL) clusters were identified. Results suggest that these cluster patterns of SOC distribution were quite similar with that of precipitation and forest age. The HH clusters were mainly distributed in areas covered by old-growth forests and with sufficient rainfall supply. In contrast, LL clusters were observed in areas covered by younger growth forests. At the eco-region scale, unique patterns of SOC distribution were detected, indicating that SOC stock was significantly influenced by micro-scale level climatic conditions and terrain attributes, such as rainfall supply and forest age.

Thus, based on the results of the spatial autocorrelation analysis, we confirmed that: (1) the SOC distribution in Canadian forests was not randomly distributed at the national and eco-region scales, (2) the spatial patterns of SOC distribution were quite eco-region unique, supporting the significant contribution of micro-scale ecological effects on pedogenic processes, and (3) the spatial patterns, or spatial information, were important factors in exploring and interpreting SOC-environment relationships. However, most previous SOC studies were conducted based on traditional OLS models. Thus, potential model misspecification in SOC-environment relationships estimation may be caused due to neglecting spatial effects.

With this in mind, Lagrange Multiplier diagnostics were applied to test the suitability of using a spatial regression model (a spatial lag model or spatial error model) as an alternative to estimate SOC-environment relationships. This study suggests that at the national scale, a spatial error model was tested to be the most appropriate model specification. Results showed that the correlated error term was highly significant (p = 0.000), indicating the importance of the unobservable and unmeasurable ecological effects on SOC distribution. In addition, SOC distribution was determined by both climatic regimes and terrain attributes, with precipitation as the most dominant ecological factor. Compared to maximum and mean temperature, SOC distribution was shown to be more sensitive to the minimum temperature regime at the national scale than compared to maximum or mean temperature measurements.

When examining the ecological effects on SOC distribution in each eco-region, there were some noteworthy findings. This study suggests that in the B.C. coastal areas and mountainous areas, traditional OLS models were suitable for estimating SOC based on the selected six environmental variables. As expected, in B.C. coastal areas, precipitation was the most dominant variable, while terrain attributes had less influence on the variability of SOC distribution. In B.C. mountainous areas, a negative relationship found between SOC and maximum temperature indicated that increasing temperature may release carbon from forest soils in B.C. mountainous areas.

Moreover, weak associations between SOC and temperature regimes were found in the Boreal eco-region (stretching from central Alberta to Nova Scotia). Compared to other eco-climatic zones, neighbourhood effects were shown to be highly significant in the Boreal eco-region. Other important environmental determinants were precipitation and slope. In particular, this study suggested that neither the OLS models nor the spatial regression models were the most appropriate model specification when estimating the SOC-environment relationships of the Subarctic, Cool Temperate, and Interior Cordilleran eco-regions. Thus, further discussion is required to test alternative model specifications.

In summary, precipitation was considered to be the most significant ecological variable at both the national and eco-region scales, with the only exception of B.C. mountainous area and northern woodlands. Thus, it was suggested that precipitation has the strongest influence on SOC distribution in Canadian forest areas. However, the relationships between SOC distribution and ecological variables were not constant across different eco-regions. In some areas, SOC stock was influenced by both climatic regimes and terrain attributes, such as the Boreal eco-region and south Ontario. In other areas, SOC distribution is more sensitive to a specific ecological variable. For example, SOC distribution in B.C. coastal areas was more influenced by precipitation. Finally, a predictive SOC distribution map was estimated from spatial error model parameters at the national scale. Although spatial patterns of modelled SOC did not necessarily coincide with the interpolated results in some areas (e.g., Yukon Territory and Prince Edward Island), the general distribution of SOC at the national scale was captured by the four dominant ecological variables: precipitation, minimum temperature, elevation, and slope.

Chapter 7. Discussion and Conclusions

In this study, an exploratory-based method was proposed to examine the spatial patterns of Canadian forest SOC distribution and to explore the spatial relationships between SOC and ecological variables at the national and eco-region scales. In this chapter, key empirical findings are discussed by answering the four research questions defined at the beginning of this study. First, how is forest SOC spatially distributed in Canadian forests? Second, how can the relationships between SOC and ecological variables, including climatic conditions and terrain attributes be quantified? Third, in Canada, do these relationships vary across different eco-regions and is this a suitable regionalization scheme for studying SOC patterns? Finally, what are the dominant factors in Canadian forest SOC distribution modelling? This chapter will also discuss limitations of the research and future work.

7.1. Summary of Key Findings

7.1.1. The Spatial Patterns of SOC Distribution in Canadian Forests

In order to explore the spatial patterns of SOC distribution in Canadian forest areas, descriptive statistics and spatial analysis techniques, including Ordinary Kriging and spatial pattern analysis, were applied on the Canada Forest Service (CFS) soil samples at the national and eco-region scales. First, results suggested that Canadian forest SOC stock varied across different eco-regions. The coastal areas and northern woodlands contain higher SOC stock, while lower SOC stock was found in Central Canada. From the estimated SOC distribution map, the pattern of SOC distribution can be summarized as – diminishing from the east and west coasts to inland, with strong local variations existing across the entire study area. In addition, this SOC distribution pattern was in good agreement with the precipitation regime in Canadian forests. Similar findings were described by Brady and Weil (2010) that in forest ecosystems, climatic regime contributes to pedogenic processes over broad geographic areas and promotes a general trend of SOC distribution, whereas terrain attributes (e.g., drainage capacity) should be responsible for local variations in SOC distribution.

More specifically, the range of estimated SOC stock was smaller than that obtained from the original CFS soil samples. This issue has been discussed by many researchers. For example, Raty and Gilbert (1998) and Rezaee et al. (2011) point out that when using Ordinary Kriging to estimate unknown values, the maximum value usually tends to be underestimated, while the minimum value is often overestimated, because Ordinary Kriging calculates the unknown value based on the mean value within a user-defined neighbourhood around each estimation. Nevertheless, Ordinary Kriging is still considered to be an optimal interpolation approach when no external factors are taken into account (e.g., ecological influences) since it maintains much of the local variance (Goovaerts, 1997).

Second, Canadian forest SOC was not randomly distributed. Premo (2004) indicated that local spatial autocorrelation statistics are capable of examining spatial patterns of ecological phenomena by identifying potential hot spots. In this study, several High-High SOC clusters were found in areas with sufficient rainfall supply and litterfall inputs. In particular, the Incremental Spatial Autocorrelation approach (refer to Section 5.2.1.1. for details) was applied to calculate the optimal neighbourhood size, within which the most intensive cluster-patterns of SOC distribution was determined. Results show that, compared to other eco-regions, large neighbourhood sizes were identified in the Subarctic and Boreal eco-regions (148,467.05 and 260,811.32 m, respectively). Since the CFS soil database is a collection of historical field surveys and individual research, no specific sampling method is considered and applied. As a result, sampling densities vary across different eco-regions. In particular, less soil samples were collected in northern woodlands, and no soil samples were collected from mid-Quebec and northern Ontario. Thus, although potential outliers were removed in the optimal distance calculation, the optimal neighbourhood sizes of the Subarctic and Boreal eco-region are probably biased due to the relatively sparse sampling density.

Third, according to spatial autocorrelation analytical results, the eco-region framework was proved to be a suitable classification scheme for SOC distribution patterns in Canada. For example, at the national scale, multiple HH clusters of high SOC levels were found in east and west coasts, which typically have a humid climate, while

LL clusters of low SOC levels were mainly concentrated in central continental regions, such as the Great Plains which is characterized as warm and arid, or semi-arid. This finding is consist with the conclusion made by Brady and Weil (2010) that low SOC stock is usually found in warm and dry environments. This type of spatial pattern is similar with the precipitation regime, suggesting a positive relationship between SOC levels and precipitation in Canadian forests, which is consistent with previous research. For example, Deluca and Boisvenue (2012) find that a sufficient rainfall supply maintains a good water-saturation in soils, which tends to limit SOC decomposition rates. Similarly, Buringh (1984) concludes that soils in humid eco-regions usually accumulate more organic-carbon.

In B.C. mountainous areas, the lowest global Moran's *I* (0.069) indicates that local variations in SOC distribution exist. This is partly due to complex geographic conditions and topography in mountainous areas. Ehrlich et al. (1977) found that high variability in soil clay content in B.C. mountainous areas were due to sediment movement. This movement is often caused by human interference (e.g., transportation) and soil erosion processes. Thus, the complex soil composition leads to variation in SOC distribution, causing the weak spatial autocorrelation (Moran' *I* = 0.069) observed in this study.

Observed SOC spatial patterns also closely correspond to forest age distribution. Results from Section 6.3.2.2 show that high SOC levels concentrate in areas covered by old-growth forests (e.g., B.C. coast areas). As previously mentioned, western Boreal eco-region forests (e.g., central Alberta, northern Saskatchewan and Manitoba) tend to be much younger than forests in Eastern regions. According to Chen et al. (2003), forests in the B.C. coast and east Quebec are generally around more than 100-years old with few disturbances detected, whereas young forests with average ages between 10- to 40-year old are identified in western Boreal areas. Such regrowth forests are largely from fire disturbance and human inference, providing less carbon inputs into soils (Chen et al., 2003), leading to poor regional SOC sequestration ability, and leading to observed low-low (LL) SOC clusters in this analysis. For example, clearance of original forests in PEI was widely observed due to agricultural practices (Meikle & Waterreu, 2008). Compared to other areas in the Cool Temperate eco-region, soils in PEI retain less SOC stock due to insufficient litterfall inputs. A similar situation is found in the Lower Fraser Basin area. Goldin and Lavkulich (1990) pointed out that the loss of old-growth forests caused by agricultural activities should account for an approximate 20% decrease in SOC stock from 1943 to 1976 in the southern Fraser Basin.

In contrast, Boyle et al. (1997) and Harmon et al. (1990) suggest that old-growth forests ensure a large quantity of litterfall inputs to forest soils, and removing old-growth forests reduces the amount of above-ground woody debris and organic detritus, leading to lower forest SOC levels. However, it should be noted that some research studies have argued that a forest's carbon sequestration ability weakens with increasing age (e.g., Chen et a., 2003; Yuan et al., 2013), indicating that young forests have a better ability to absorb atmospheric carbon than that of old-growth forests due to higher photosynthetic efficiency (Murty et al., 1996). This argument could mislead forest SOC managers that substituting old-growth forests with young forests could reduce atmospheric carbon concentration (Harmon et al., 1990). Harmon et al. (1990) point out that the amount of carbon stored in forest eco-systems is of greater importance than that absorbed from the atmosphere. For younger forests, Net Primary Productivity (NPP)[9] is mainly stored as standing biomass, while for old-growth forests, a similar or larger amount of carbon (compared to that of NPP) is transformed to soils as organic detritus (Boyle et al., 1997; Long, 1982). Thus, the difference in carbon allocation makes soils in old-growth forests a better terrestrial carbon reservoir. A similar viewpoint was expressed by Luyssaert et al. (2008) that old-growth forests continue to accumulate SOC reserves. Black et al., (2008) also point out that original old-growth forests can accumulate large amounts of SOC that cannot be achieved by young-growth forests. This literature lends support to the results derived from the spatial autocorrelation analysis, which observed a positive association between SOC stock and forest age.

These findings consequently support the importance of protecting old-growth forests for SOC management. In high SOC-concentration areas, potential practices

[9] Net Primary Productivity (NPP) refers to the difference between the amount of carbon absorbed through photosynthesis processes and that of consumed by vegetation respiration (Chen et al., 2003).

include protecting old-growth forests from pest disasters, reducing excessive cutting, and avoiding other human disturbances. In other areas, especially in Low-Low SOC clusters, sustainable forest management practices should be considered. Batjes (1996) indicated that land use changes, such as the transition from intact forests to agricultural land, greatly influence oxygen diffusion processes in surface soil layers. Thus, reducing forest-cutting intensity and leaving residues on-site are widely considered as efficient ways to preserve SOC stocks after forest harvest (Apps et al., 2006; Gershenson & Barsimantov, 2011).

7.1.2. Assessing Relationships between SOC and Ecological Variables

To date, few studies have been conducted to assess the effects of ecological variables on SOC distribution at the regional scale (Yuan et al., 2013). Moreover, only traditional linear regression models (e.g., OLS models) have been considered due to their simplicity (Mishra et al., 2010). However, an OLS model may not be an appropriate modelling approach when variables are spatially dependent or when high spatial autocorrelation is measured in OLS residuals. Although spatial regression models have been applied widely in the social sciences, this approach has seldom been applied in ecological literature or environmental studies (Dormann et al., 2007). Thus, in this study a spatial regression approach was adopted for modelling the SOC-environment relationship. In particular, the relationship between Canadian forest SOC distribution and six ecological predictor variables was modelled at the national and eco-region scales.

Results suggest that a spatial error model is considered to be more appropriate than the spatial lag model specification when estimating the relationship between SOC distribution and ecological variables at the national scale. Although ecological effects on SOC stock at one sample site tend to diffuse outward to its neighbours and promote the similarity in surrounding SOC distribution, the magnitude of this "spill over" effects (LM-lag value = 11.244 and p = 0.001) is less significant than that of the "omitted variables" effects (LM-error value = 247.350 and p = 0.000). Thus, it is concluded that apart from the six selected independent variables, other ecological factors may also influence the SOC distribution patterns at the national scale, but may have been missing from the model specification.

Regression analysis at the local scale suggested that in B.C. coastal and mountainous areas, OLS models are able to effectively describe the SOC-environment relationships based on the six independent variables. In B.C. mountainous areas, maximum temperature is tested to be negatively related to SOC stock. One potential explanation is that increasing maximum air temperature results in an increase in the soil temperatures, which tend to accelerate the SOC decomposition rate. According to Birkeland (1974) and Tewksbury and Van Miegroet (2007), micro-climatic regimes are quite dependent on elevation gradients. In B.C. mountainous areas, the temperature regime is influenced by the complex local topography. Any change in maximum air temperature will influence soil temperatures and eventually alter SOC decomposition rates. In B.C. coastal areas, precipitation is the primary ecological factor that promotes the overall high SOC stock. Due to a sufficient rainfall supply, moist soils in B.C. coastal areas limit SOC decomposition rates, and thus ensure a large amount of organic carbon accumulation. In addition, a spatial lag model was determined to be the most appropriate regression specification in the Boreal eco-region. The ecological activities occur at one sample site significantly influence SOC stock in the surrounding neighbourhood.

Moreover, in the Subarctic, Cool Temperate, and Interior Cordilleran eco-regions, the autocorrelated residuals indicate that SOC distribution patterns were not fully accounted for in the OLS model estimation. However, results from the Lagrange Multiplier diagnostic suggested that neither a spatial lag model, nor a spatial error model were suitable. To explain this phenomenon, the causes of autocorrelated residuals should be discussed first. Cliff and Ord (1970) summarize three possible reasons that would lead to the interdependence among an OLS model's residuals: (1) significant independent variables are excluded, (2) an autoregressive component is omitted from the regression model, and (3) the targeted relationship cannot be simulated by linear regression models.

In this study, it is suggested that the autocorrelated residuals in the three eco-regions are caused by omitted independent variables or the misuse of linear regression models. An autoregressive component can be introduced into the spatial regression models in two ways: a spatially lagged dependent variable and a spatially correlated error term. However, results from Lagrange Multiplier diagnostic tests reject including either type of autoregressive component. Although a spatially correlated error term can partly compensate for the impacts of omitted variables, it cannot substitute potentially significant environmental determinants. Since the ecological influences on SOC distribution are quite complicated to simulate in reality, the six selected independent variables in this study may not be able to explain the spatial patterns of SOC distribution in the three eco-regions.

7.1.2.1. *Dominant Variables at the National Scale*

In this study, a combined influential effect of climatic regimes and terrain attributes on SOC distribution is observed in Canadian forest areas. From the regression results, it is suggested that the influences of climatic regimes are more significant, relative to terrain attributes, such as elevation. As expected, precipitation was the most dominant ecological variable (p = 0.000) and positively related to SOC distribution. Minimum temperature proved to be another dominant variable (p = 0.021) influencing SOC distribution at the national scale. In the literature, temperature effects on SOC stock are controversial with no universal agreement. Some researchers have suggested that the association between temperature and SOC stock is quite weak (i.e. Gifford, 1992), whereas others have argued that the SOC stock decreases consistently with increasing temperatures (Kirschbaum, 1995; Tewksbury & Van Miegroet, 2007). In this study, compared to the maximum and mean temperature, minimum temperature was proved to have a significant and positive effect on SOC distribution in Canadian forest areas. Recalling that SOC stock is the difference between organic-carbon inputs and carbon decomposition, any changes in the two processes will alter SOC stock. At the national scale, the increasing minimum temperature tends to result in a longer growing season in the mid- and high-latitude forest ecosystems. Thus, it is concluded that, although increasing temperatures tend to accelerate the SOC decomposition rate in Canadian forest ecosystems, the amount of carbon loss is offset by the increasing vegetation biomass and litterfall accumulation.

Results from the spatial error model showed that the most significant terrain variable was elevation (p = 0.036), while slope (p = 0.065) was tested to be less important. A positive correlation was found between SOC distribution and elevation. One potential explanation would be that soil temperature is usually low in high elevation areas, thus maintaining a good environment for SOC accumulation. Tewksbury and Van Miegroet (2007) point out that soil temperature decreases with increasing elevation. Thus, it is assumed that organic carbon decomposition rates tend to be slower at high elevation areas due to lower soil temperatures. Consequently, higher SOC stock is observed in high altitudes.

Finally, the highly significant error term, λ (sig. level = 0.000), suggested that other ecological factors that were not considered in this analysis, including unobservable and immeasurable factors, should likely partly account for the variation in SOC distribution. In reality, the regional-scale interactions between SOC and its surroundings are quite complex to model. In particular, many researchers have suggested that there is a lack of research into quantifying SOC distribution and modelling SOC-environment relationships (Parry & Charman, 2013; Wang et al., 2013). In this study, many topographic variables were unavailable (e.g., forest ages, soil ph, and nitrogen content), which could have improved the model specification. Thus, the effects of ecological activities on the SOC distribution in Canada's forest ecosystems are potentially underrepresented. Although using a spatial error model partly moderates the influences of omitted variables, the low *pseudo* R^2 of 0.347 indicated that 65.3% of variation in SOC distribution pattern remains unexplained.

7.1.2.2. *Dominant Variables at the Eco-Region Scale*

The dominant ecological variables that influence SOC distribution vary across different eco-regions. We suggest that the SOC distribution in the boreal eco-region is influenced by the combination of precipitation regime and slope. While the pedogenic processes in the Pacific Cordilleran and Cordilleran eco-regions are more dominated by climatic conditions. For example, precipitation was tested to be a significant ecological variable in the Pacific Cordilleran eco-region. This was an expected result because this eco-region is located in coastal areas in B.C. province. This effect results with from high humidity and rainfall conditions due to local climate and the influence of the Pacific Ocean. Thus, compared to other ecological factors, precipitation was the most dominant factor that influences pedogenic processes in this region.

In the Subarctic, Cool Temperate, and Interior Cordilleran eco-regions, results from the re-estimated OLS models were discussed (refer to Section 6.4.2.3.). In the Subarctic and Interior Cordilleran eco-regions, SOC distribution was more related to terrain and land cover attributes: NDVI and elevation, respectively. The Subarctic eco-region is a typical area with a tree line traversing the central region. Due to the adverse climate conditions, vegetation density is very low in this region. Thus, the carbon dynamics between soils and vegetation become increasingly important. The Interior Cordilleran eco-region is located between the Coast Mountains and the Columbia and Rocky Mountains. Along the altitudinal gradient, temperature decreases with increasing elevation and potentially lowers SOC decomposition rates. In addition, Sprout et al. (1978) suggested that compared to higher elevation areas, many lower parts of the Interior Plateau were covered by coarse materials, which contain less SOC stock (e.g., gravelly and sandy soils) deposited by melt-water. Thus, a positive association between elevation and SOC stock was identified in this study.

In general, more organic carbon would be accumulated in the areas covered by vegetation through carbon-sequestration processes. In southern Ontario and along the east coast, SOC distribution was influenced by both climate regimes and terrain attributes. Positive associations were observed between SOC stock and temperature. One potential reason is that temperature increases potentially lengthen the growing season of the forest ecosystems. Thus, the loss of SOC through carbon decomposition and soil respiration is offset by the increasing amount of litterfall inputs and root exudates.

From the regression analysis at the local scale, different responses of SOC to various climatic regimes were observed. This indicates that ecological factors influence pedogenic processes differently among ecosystems at the regional scale, and thus verifies the eco-region classification framework for SOC zonation mapping in Canada. Since the eco-regions are integrated homogeneous geographic areas which share similar climatic conditions, vegetation types, wildlife groups, and pedogenic processes (Ecoregions Working Group, 1989), an assumption is made that the eco-region classification framework is capable to greatly distinguish the major differences in ecological effects on SOC distribution. Thus, it was selected as the regionalization scheme to test SOC-environment relationships at the local scale. Regression results showed typical spatial variability in dominant ecological variables that influence SOC distribution across different geographic areas. For example, vegetation was identified to promote SOC accumulation in northern woodlands, whereas precipitation was tested to positively influence SOC stock in B.C. coast areas. Therefore, it is concluded that the eco-region classification framework is a sufficient alternative in SOC studies in Canadian forest areas.

Finally, three reasons should account for the differences between the modelled SOC map and the interpolated result. First, the selection of a spatial error model indicates that unobservable and immeasurable ecological variables should explain part of the variation in SOC distribution in Canadian forest areas. Given the complexity in the real world, four variables are not sufficient to fully describe the ecological effects on SOC distribution. In addition, lacking important environmental determinants is another reason that leads to inappropriate model specification in the Cool Temperate eco-region. Thus, the effect of “omitted variables” should be considered to be a potential reason for the mis-estimation of SOC stock in southern Ontario. Second, based on the regression analysis results, it was obvious that a dominant environmental/climatic factor at the regional scale may not be as important at the local scale. Thus, the role of microclimate variation may potentially contribute to model mis-estimation in B.C. mountainous areas. Third, in the Subarctic Cordilleran eco-region, an insufficient number of soil samples and observations distributed throughout the study area is considered to be a primary source of error in the analysis. Since the Subarctic Cordilleran eco-region only contained 14 samples, it was excluded from the spatial pattern analysis and no significant environmental determinants were identified.

7.2. Limitations and Recommendations

7.2.1. Data Quality Issues

The CFS soil database is a compilation of historic data that was used for this study. Although it includes intensive SOC observations across Canada, it nevertheless lacks comprehensive coverage of the continent and the data collection is based on a predetermined sampling design, since the secondary dataset was contributed by other organizations. As a result, there is a higher certainty in spatial analysis results in areas with more available SOC samples, while certainty in modelling estimates is lower in areas with fewer observation points. Thus, the calculated optimal-neighbourhood-sizes of each eco-region are likely biased in areas with a low sampling density (refer to Section 6.3.2.)

The impact of sampling design on SOC distribution has been emphasized by previous research. For example, Yuan et al. (2013) emphasized the importance of sampling design from the perspective of preserving more local-scale SOC variations. They argue that regional-scale SOC studies would greatly benefit from an accurate representation of local information, thus it is better to include extensive and intensive SOC samples to ensure more reliable results. Similarly, Gallardo (2003) point out that an insufficient number of soil samples increases the difficulty of characterizing spatial variations in soil properties, including SOC and other soil nutrients. However, although increasing sampling density is recommended by many researchers to ensure reliable spatial analysis, no consensus is reached among soil scientists with regard to the minimum, or optimal, sampling density at regional scales of analysis (Yuan et al., 2013). Thus, more efforts are required to assess the influence of sampling design on regional-scale SOC distribution analysis.

7.2.2. Issues of Low Spatial Resolution

Another limitation of this study is the low spatial resolution datasets on which the analysis of SOC levels and climate variables was based. Since the climate and terrain datasets that were acquired had different spatial resolutions, the datasets were resampled to 10 km resolution in order to have comparable cell size. However, when data of higher spatial resolution is resampled to a lower spatial resolution, loss of unique local information inevitably occurs. This was especially true for datasets of terrain variables, such as elevation and slope. For a given remote sensing image, the extracted value (e.g., slope and aspect) at a sample site is actually a product of averaging effects of corresponding and neighbouring pixels. Resampling may introduce bias into correlation analyses of SOC and ecological variable relationships.

In this study, the low spatial resolution may partly account for the weak association observed between SOC distribution and terrain attributes (e.g., slope and aspect). The extracted slope measure at one sample site may not be reliable after being resampled to a coarse resolution. Ju and Chen (2005) also find similar issues related to spatial resolution and resampling of datasets in their research, which explores the Canadian forest and wetland SOC distribution based on a set of topography variables derived from remote sensing images. Ju and Chen (2005) conclude that the remote sensing images with a cell size of 1 km actually average the effects of local topography features, and thus potential errors and biases are introduced into the statistical modelling and analysis of SOC-environment relationships. In addition, Ju and Chen (2005) recommend that it is better to use remote sensing data with higher spatial resolution for regional-scale SOC studies in order to explore local variability and underlying processes.

7.2.3. Issues of Data Availability

In this study, an additional factor that potentially limits the performance of spatial analysis approaches is the lack of updated soil data and data related to other ecological variables that influence SOC. The CFS soil database has three primary limitations in this study. First, the CFS soil database consists of soil profiles collected before the year 1991 and lacks recent data. Thus, the database does not accurately reflect current SOC conditions in Canadian forest ecosystems. Second, the acquisition time of soil samples varies across different sample sites, generally ranging from 1961 to 1991, but is not consistent in terms of temporal sampling, availability, and collection methods. However, this long range of 31 years of soil data enables comparison with long-term climate datasets over the same time period and provides a general impression of climate and SOC relationships at regional scales. In general, a 31-year range for SOC distribution analysis at the national scale is considered to be acceptable, because the estimated average-turnover-time[10] for SOC to a depth of one meter is approximately 32 years (Manabe, 1983; Raich & Schlesinger, 1992; Trumbore, 1997). However, it should also be noted

[10] Turnover time of SOC refers to the period required for organic-carbon inputs to completely decompose and transferr back to the atmosphere (Six & Jastrow, 2002). Generally, the turnover periods expand with increasing depth of soil layers (Trumbore, 1997).

that SOC from different soil layers have different turnover times and this regional analysis is not able to explore local-scale variability and processes.

Trumbore (1997) states that organic-carbon at the soil surface can easily decompose with an average turnover time of less than one year. This indicates a rapid dynamic-cycle between SOC and vegetation biomass (e.g., litterfall and root exudates). With increasing depth of soil layers, SOC turnover periods could potentially expand to years, decades, or even centuries. Thus, by setting the study period as 31-years in duration, potential dynamics within a shorter period between SOC and vegetation in the soil surface are neglected, or offset, during this long period. This could partly attribute to the weak association between SOC distribution and NDVI observed in this study. A similar finding was reported by Batjes (1996): over a longer period of time, SOC distribution was more influenced by climatic conditions; however, within in a shorter period, effects from variations in vegetation biomass were more notable. Therefore, although a 31-year period allows general relationships between Canadian forest SOC distribution and ecological variables to be assessed over the long run, future research could involve data collected over shorter time periods to explore shorter-term variability in order to better understand SOC-ecological relationships.

This study is constrained by the availability of soil data, especially since the datasets analysed were secondary in nature and not from primary data collection. Other SOC studies have also discussed problems associated with data availability for SOC distribution analysis. For example, four major concerns of SOC studies at the global and regional scales are summarized by Batjes (1996), including: (1) lack of updated, completed, and reliable soil databases, (2) local-scale spatial variations in SOC distribution are difficult to be qualitatively and quantitatively measured and calculated, (3) the interactions between SOC and ecological variables are too complicated to simulate, and (4) the absence of data related to other local-scale soil properties (e.g., carbon to nitrogen ratio, clay content, and bulk density[11]). Another factor discussed by Eswaran

[11] Bulk density (g/cm^3) is the ratio between the dry weight of soil and its volume. Batjes (1996) stated that bulk density is an important indicator of the structural condition of soils, and is influenced by many factors, such as soil moisture, soil texture, and organic matter particles.

(1993) suggests that the frequent changes of vegetation types and land uses introduce uncertainties into large-scale SOC studies. As discussed previously in Section 2.1.3, in this study, the emphasis of Canadian field soil surveys in recent years has been switched to a private-driven mechanism. Consequently, problems include: (1) no standard sampling design applied to the private-driven field surveys, and (2) data are usually not freely available for public uses. Thus, a lack of updated and comprehensive soil data is considered to be a primary limitation of this study.

It is also important to note that CFS soil profiles are often referred to as aggregated data. During the period from 1961 to 1991, SOC stock for each sample site was recorded once, rather than measured repeatedly. Thus, temporal changes in SOC distribution in response to climate fluctuations are potentially underrepresented. With multi-temporal soil datasets, it is possible to examine how SOC distribution varies with temperature changes. Thus, potential responses of SOC stock to global warming in different climatic zones can be studied. In addition, multi-temporal data (1) allow us to evaluate the performance of estimated models in regard to future SOC distribution prediction, and (2) provide actual field observations for model validation. Previous research has found that time lags[12] generally exist between the process of SOC sequestration and ecological activities (e.g., vegetation growth and air temperature changes) (Meyer et al., 2012; Trumbore, 1997; Zheng et al., 1993). For example, Gaudinski et al. (2000) find an approximate 7 year of lag between root carbon inputs and carbon decomposition. Thus, with multi-temporal soil data, it is possible for us to identify the time lags between SOC sequestration and ecological activities, thus enabling SOC distribution and ecological conditions to be more accurately estimated. In this study, by using the aggregated CFS SOC data, only a general relationship between the SOC distribution and temperature regime is estimated. In Canadian forests, temporal responses

[12] Specifically, the time lag between SOC and vegetation refers to the period during which carbon is absorbed by vegetation through photosynthesis and transferred to soil layers through root exudates (Meyer et al., 2012; Trumbore, 1997).

The time lag between SOC and air temperature refers to the lag between air temperature and soil temperature. For instance, Zheng et al. (1993) adopted an approximate two-week time lag to estimate the averaged soil temperature from the given air temperature data.

of SOC stock to ecological activities still remain unknown. This study could be improved if multi-temporal soil data were made available.

As previously discussed, topographic factors such as forest age, soil pH values, soil temperature, soil moisture, nitrogen content, and other soil properties were not considered in this study due to data availability. For example, soil texture is a key factor that influences SOC accumulation. Brady and Weil (2010) found that fine-textured soils usually hold more organic-carbon stocks due to relatively poor aeration conditions. Therefore, decomposition is limited in fine-textured soils. In addition, fine-textured soils have a better ability to hold moisture and soil nutrients, which promote vegetation-growth (Brady & Weil, 2010; Plaster, 1992). More vegetation biomass is consequently produced and added into soils, thus encouraging SOC accumulation. Another important factor that is missing in this study is bulk density[13], which differs with different soil types and soil textures (Birkeland, 1974). In general, the amount of SOC increases with decreasing bulk densities. Plaster (1992) explained that although the smallest space (poor aeration conditions) is found between fine-textured soil particles, fine-textured soil has the largest total pore-space. Therefore, a lower bulk density characteristic of fine-textured soils tend to contain a larger amount of organic material, and it is often used as a key variable when estimating SOC stocks.

Potentially missing independent variables may limit the performance of regression models and the predictive-map that was generated. Results could be improved by taking these factors into consideration. For example, data quality may have resulted in descriptive statistics showing a negative correlation between SOC distribution and drainage capacity, likely due to the fact that soil nutrients may be depleted with waterflow in well-drained areas. However, drainage capacity data at each sample site was categorical in nature, ranked from “poor condition” to “rapid condition”. Although the regression model analysis was executed based on “dummy coding”, allowing categorical variables to be incorporated into a regression model by creating a set of new binary variables to represent the different categorical levels (e.g. coded as 1 or 0) (Walter et al.,

[13] Bulk density is a fundamental soil property that is expressed as the dry-weight of soil per unit volume (Plaster, 1992).

1987), the complexity of regression models is greatly increased when multiple categorical levels are added into the estimated models. Thus, with more available ecological variables included, the interactions between SOC and ecological activities could be better represented based on quantitative data, rather than categorical indices.

In summary, the performance of the exploratory-based methods and spatial regression analyses would be improved with more availability of ecological data and multi-temporal soil data. Future in-depth studies on modeling the SOC-environment relationships focus on exploring local-scale dynamics between SOC distribution and ecological factors in the Subarctic, Cool Temperate, and Interior Cordilleran eco-regions.

7.3. Conclusion and Significance

In general, this study is valuable in four aspects. First, it contributes to the current literature by exploring the SOC distribution patterns in Canadian forest ecosystems and examining dominant ecological variables among different eco-climate zones. Since previous research has mainly focused on estimating the amount of available SOC stock, few studies have explored how SOC is spatially distributed across Canada and what ecological factors have dominant roles in shaping regional SOC distribution patterns. By developing an exploratory-based method, conclusions can be drawn that in Canadian forest ecosystems, the SOC distribution reveals distinct spatial patterns, with coastal areas containing higher SOC stock, while western boreal eco-region forests has lower SOC stock. Such spatial patterns are closely tied with both climatic regimes and terrain attributes, such as elevation and slope.

Second, this study verifies the effectiveness of using the climatic eco-region framework for SOC distribution classification and analysis. Empirical findings from this study suggest that in B.C. coastal and mountainous areas, SOC distribution is mainly influenced by micro-scale climate, namely temperature and precipitation. While in the northern forest ecosystems, a more notable dynamic between SOC stock and vegetation biomass is identified. In addition, a combined influence of climate and topography on the SOC distribution was observed in the Boreal eco-region and southern Ontario. In particular, the positive association between SOC and precipitation is identified on both the national and eco-region scales, indicating the significance of precipitation regimes in influencing Canadian forest SOC distribution.

Third, this study provides insight into temperature sensitivity of SOC levels in Canadian forests. In mid- and high-latitude forest ecosystems, SOC stock tends to be more susceptible to minimum temperature variations. It is suggested that increasing minimum temperature actually helps promote SOC accumulation in Canadian forest regions. When minimum temperatures increase, the loss of organic-carbon from decomposition is offset by an increasing amount of litterfall accumulation and root exudates. However, the SOC stock responds differently to temperature changes at local scales. For example, the SOC stock decreases with increasing maximum temperature in B.C. mountainous areas.

Finally, the findings of this study are relevant to forest SOC management and practice. Inappropriate land use changes could potentially accelerate the SOC decomposition rate, which results in carbon release to the atmosphere. By modelling and identifying areas of high SOC stock, better land management practices for reducing soil disturbances can be identified. Studying regional SOC distribution patterns and SOC-environment dynamics consequently assists soil resource management in two ways: (1) determining influencing factors that are important for modelling SOC distribution, and (2) making effective forest management policies, which may include the protection of old-growth forest, logging management practices, and land use regulation.

In conclusion, soil is an important and fundamental element in the terrestrial carbon cycle. Small changes in SOC stock has the potential to greatly affect atmospheric carbon, which could moderate or intensify global warming. However, such SOC-ecological relationships are too complicated to simulate. This study builds upon previous research and attempts to incorporate a spatial dimension to exploring SOC levels and their relationships to ecological and terrain variables in Canadian forests. This includes identifying clusters of SOC distribution and estimating spatial effects in the SOC-environment relationship. With more available data and improved data quality, the relationships between SOC and pertinent ecological variables can be better described, providing valuable information to better inform SOC management practices.

References

Agriculture Canada (2008). Daily maximum temperatures (Celsius) for Canada. *National Land and Water Information Service*, *Agriculture and Ari-Food Canada.*

Anderson, D. W., & Scott Smith, C. A. (2011). A history of soil classification and soil survey in Canada: Personal perspectives. *Canadian Journal of Soil Science*, *91*(5), 675-694.

Anselin, L. (1988[a]). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical analysis*, *20*(1), 1-17.

Anselin, L. (1988[b]). Spatial Econometrics: Methods and Models. Dordrecht, Netherlands: Kluwer Academic.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis, 27*(2), 93-115.

Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In *Spatial Analytical Perspectives on GIS in Environmental and Socio-Economic Sciences*. London: Taylor and Francis, 111-125.

Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. *Geographical Information Systems: principles, techniques, management and applications*, *1*, 251-264.

Anselin, L. (2001). Spatial econometrics. *A companion to theoretical econometrics*, 310-330.

Anselin, L. (2003). An introduction to spatial autocorrelation analysis with GeoDa. *Spatial Analysis Laboratory, University of Illinois, Champagne-Urbana, Illinois*.

Anselin, L. 2005. Exploring Spatial Data with GeoDa™: A Workbook. Spatial Analysis Laboratory (SAL). *Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign, IL.*

Anselin, L. (2009). Spatial regression. *The Sage handbook of spatial analysis*, 255-275.

Apps, M. J., Bernier, P., & Bhatti, J. S. (2006). Forests in the global carbon cycle: implications of climate change. *Climate Change and Managed Ecosystems*, 175-200.

Azpurua, M. A., & Ramos, K. D. (2010). A comparison of spatial interpolation methods for estimation of average electromagnetic field magnitude. *Progress in Electromagnetics Research M*, *14*, 135-145.

Baller, R. D., Anselin, L., Messner, S. F., Deane, G., & Hawkins, D. F. (2001). Structural covariates of U.S. county homicide rates: incorporating spatial effects. *Criminology, 39*(3), 561-588.

Batjes, N. H. (1996). Total carbon and nitrogen in the soils of the world. *European journal of soil science*, *47*(2), 151-163.

Bergstrom, D. W., Monreal, C. M., & St. Jacques, E. (2001). Spatial dependence of soil organic carbon mass and its relationship to soil series and topography. *Canadian Journal of Soil Science*, *81*(1), 53-62.

Bhatti, J. S., Apps, M. J., & Lal, R. (2006). Interaction between climate change and greenhouse gas emissions from managed ecosystems in Canada. *Climate change and managed ecosystems. Edited by JS Bhatti, R. Lal, MJ Apps, and MA Price. CRC Taylor & Francis, Boca Raton, Fla*, 3-15.

Bhatti, J. S., Apps, M. J., & Tarnocai, C. (2002). Estimates of soil organic carbon stocks in central Canada using three different approaches. *Canadian Journal of Forest Research*, *32*(5), 805-812.

Birkeland, P. W. (1974). Climate-soil relationship. In *Pedology, Weathering, and Geomorphological Research* (pp. 211-246). New York: Oxford University Press.

Black, T. A., Jassal, R. S., & Fredeen, A. L. (2008). Carbon sequestration in British Columbia's forests and management options. *University of Victoria (B.C.). Pacific Institute for Climate Solutions.*

Boyle, C. A., Lavkulich, L., Schreier, H., & Kiss, E. (1997). Changes in land cover and subsequent effects on Lower Fraser Basin ecosystems from 1827 to 1990. *Environmental Management*, *21*(2), 185-196.

Born, B. & Breitung, J. (2011). Simple regression-based tests for spatial dependence. *The Econometrics Journal*, *14*(2), 330-342.

Bou Kheir, R., Greve, M. H., Bøcher, P. K., Greve, M. B., Larsen, R., & McCloy, K. (2010). Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. *Journal of Environmental Management, 91(5),* 1150-1160.

Brady, N. C., & Weil, R. R. (2010). *Elements of the nature and properties of soils*. Pearson Educational International.

Buringh, P. (1984). Organic carbon in soils of the world. *The Role of Terrestrial Vegetation in the Global Carbon Cycle. Measurement by Remote Sensing, Vol. SCOPE*, *23*.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.

Chakraborty, J. (2011). Revisiting Tobler's first law of geography: Spatial regression models for assessing environmental justice and health risk disparities. In *Geospatial analysis of environmental health* (pp. 337-356). Springer Netherlands.

Chen, J. M., Ju, W., Cihlar, J., Price, D., Liu, J., Chen, W., Pan, J., Black, A., & Barr, A. (2003). Spatial distribution of carbon sources and sinks in Canada's forests. *Tellus B*, *55*(2), 622-641.

Chen, Y., Yu, J., Shahbaz, K., & Xevi, E. (2009). A GIS-based sensitivity analysis of multi-criteria weights. In *18th World IMACS/MODISM Congress, Cairns, Australia*.

Childs, C. (2004). Interpolating surfaces in ArcGIS spatial analyst. *ArcUser, July-September*, 32-35.

Chuai, X. W., Huang, X. J., Wang, W. J., Zhang, M., Lai, L., & Liao, Q. L. (2012). Spatial variability of soil organic carbon and related factors in Jiangsu Province, China. *Pedosphere*, *22*(3), 404-414.

Cliff, A. D., & Ord, K. (1970). Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography*, *46*, 269-292.

Cliff, A. D. & Ord, J. K. (1981). Spatial Processes, Models and Applications. London: Pion.

Coen, G. M. Eds. (1987). Soil survey handbook. Vol. 1. Technical Bulletin 1987-9E. Land Resource Research Centre Contribution Number 85-30. Research Branch. Agriculture Canada Ottawa, ON.

Collins, K., Babyak, C., & Molone, J. (2006). Treatment of Spatial Autocorrelation in Geocoded Crime Data. *Proceedings of the American Statistical Association Section on Survey Research Methods*, 2864-2871.

Colpitts, M. C., Fahmy, S. H., MacDougall, J. E., Ng, T. T. M., Mcilnnis, B. G., Zelazny, V. F. (1995). Forest soils of New Brunswick. *Department of Natural Resources & Energy, Timber Management Branch; Agriculture & Agri-Food Canada, Centre for Land & Biological Resources Research.*

Conen, F., Zerva, A., Arrouays, D., Jolivet, C., Jarvis, P. G., Grace, J., & Mencuccini, M. (2004). The carbon balance of forest soils: detectability of changes in soil carbon stocks in temperate and boreal forests. *Symposia-Society for Experimental Biology, Vol. 57,* 235.

Cressie, N. (1988). Spatial prediction and ordinary kriging. *Mathematical Geology*, *20*(4), 405-421.

Davidson, E. A., Trumbore, S. E., & Amundson, R. (2000). Biogeochemistry: soil warming and organic carbon content. *Nature*, *408*(6814), 789-790.

Deluca, T. H., & Boisvenue, C. (2012). Boreal forest soil carbon: distribution, function and modelling. *Forestry*, *85*(2), 161-184.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., ... & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, *30*(5), 609-628.

Ecoregions Working Group. (1989). Ecoclimatic regions of Canada, first approximation. *Ecological land classification series*, *23*.

Ehrlich, W. A., Cann, D. B., Day, J. H., & Marshall, I. B. (1977). *Soils of Canada* (Vol. 1, pp. 73-77). Research Branch, Canada Department of Agriculture.

Engle, R. F. (1984). Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. *Handbook of econometrics*, *2*, 775-826.

ESRI. (2013[a]). Modelling spatial relationships. *Retrieved from:* http://resources.arcgis.com/en/help/main/10.1/index.html#/Modeling_spatial_relationships/005p00000005000000/

ESRI. (2013[b]). Incremental Spatial Autocorrelation. *Derived from:* http://resources.arcgis.com/en/help/main/10.1/index.html#//005p0000004z000000

ESRI. (2013[c]). How spatial autocorrelation works. *Derived from:* http://resources.arcgis.com/en/help/main/10.1/index.html#//005p0000000t000000

Eswaran, H., Van Den Berg, E., & Reich, P. (1993). Organic carbon in soils of the world. *Soil science society of America journal*, *57*(1), 192-194.

Eswaran, H., Van den Berg, E., Reich, P., & Kimble, J. (1995). Global soil carbon resources. *Soils and global change*, 27-43.

Ettema, C. H., & Wardle, D. A. (2002). Spatial soil ecology. *Trends in ecology & evolution*, *17*(4), 177-183.

Fazekas, I., & Lauridsen, J. (1999). On the Lagrange multiplier test for spatial correlation in econometric models. *Journal of Mathematical Sciences*, *93*(4), 515-520.

Fonseca, W., Rey Benayas, J. M., & Alice, F. E. (2011). Carbon accumulation in the biomass and soil of different aged secondary forests in the humid tropics of Costa Rica. *Forest Ecology and Management*, *262*(8), 1400-1408.

Foody, G. M. (2004). Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-saharan endemic avifauna. *Global Ecology and Biogeography, 13(4),* 315-320.

Gagne, P. & Dayton, C. M. (2002). Best regression model using information criteria. *Journal of Modern Applied Statistical Methods, 1(2)*, 479-488.

Galbraith, J. M., Kleinman, P. J. A., & Bryant, R. B. (2003). Sources of uncertainty affecting soil organic carbon estimates in northern New York. *Soil Science Society of America Journal, 67(4),* 1206-1212.

Gallardo, A. (2003). Spatial variability of soil properties in a floodplain forest in northwest Spain. *Ecosystems*, *6*(6), 564-576.

Gallo, J. L., Ertur, C. & Baumont, C. (2003). A spatial econometric analysis of convergence across European regions, 1980-1995. In B. Fingleton (Ed.), *European regional growth* (pp. 99-129). Springer.

Gaudinski, J. B., Trumbore, S. E., Davidson, E. A., & Zheng, S. (2000). Soil carbon cycling in a temperate forest: radiocarbon-based estimates of residence times, sequestration rates and partitioning of fluxes. *Biogeochemistry*, *51*(1), 33-69.

Geng, X., Fraser, W., VandenBygaart, B., Smith, S., Waddell, A., Jiao, Y., & Patterson, G. (2010). Toward digital soil mapping in canada: Existing soil survey data and related expert knowledge. In *Digital Soil Mapping* (pp. 325-335). Springer Netherlands.

Gershenson, A. & Barsimantov, J. (2011). Accounting for Carbon in Soils. *Climate Action Reserve White Paper. EcoShift Consulting, LLC*.

Gifford, R. M. (1992). Implications of the globally increasing atmospheric CO2 concentration and temperature for the Australian terrestrial carbon budget: Integration using a simple-model. *Australian Journal of Botany*, *40*(5), 527-543.

Goldin, A. & Lavkulich, L. M. (1990). Effects of Historical land clearing on Organic matter and nitrogen levels in Soils of the Fraser lowland of British Columbia, Canada and Washington, USA. *Canadian Journal of Soil Science*, *70*(4), 583-592.

Goodchild, M., Haining, R., & Wise, S. (1992). Integrating GIS and spatial data analysis: problems and possibilities. *International Journal of Geographical Information Systems*, *6*(5), 407-423.

Goovaerts, P. (1997). Geostatistics for natural resources evaluation. Oxford university press.

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

Grunwald, S. (2006). What do we really know about the space-time continuum of soil-landscapes? *Environmental soil-landscape modeling: Geographic information technologies and pedometrics. Taylor & Francis, Boca Raton, FL*, 3-36.

Haining, R. (1993). Spatial data analysis in the social and environmental sciences. Cambridge University Press.

Haining, R., Wise, S., & Ma, J. S. (1998). Exploratory spatial data analysis in a geographic information system environment. *The Statistician, 47(3),* 457-469.

Harmon, M. E., Ferrell, W. K., & Franklin, J. F. (1990). Effects on carbon storage of conversion of old-growth forests to young forests. *Science*, *247*(4943), 699-702.

Hontoria, C., Saa, A., & Rodríguez-Murillo, J. C. (1999). Relationships between soil organic carbon and site characteristics in peninsular Spain. *Soil Science Society of America Journal*, *63*(3), 614-621.

Huo, X. N., Li, H., Sun, D. F., Zhou, L. D., & Li, B. G. (2012). Combining geostatistics with Moran's *I* analysis for mapping soil heavy metals in Beijing, China. *International journal of environmental research and public health*, *9*(3), 995-1017.

Huo, X. N., Zhang, W. W., Sun, D. F., Li, H., Zhou, L. D., & Li, B. G. (2011). Spatial pattern analysis of heavy metals in Beijing agricultural soils based on spatial autocorrelation statistics. *International journal of environmental research and public health*, *8*(6), 2074-2089.

Islam, K. R., & Weil, R. R. (2000). Land use effects on soil quality in a tropical forest ecosystem of Bangladesh. *Agriculture, Ecosystems & Environment*, *79*(1), 9-16.

Ismail, S. (2006). Spatial autocorrelation and real estate studies: A literature review. *Malaysian Journal of Real Estate*, *1*(1), 1-13.

Janzen, H. H. (2004). Carbon cycling in earth systems—a soil science perspective. *Agriculture, Ecosystems & Environment*, *104*(3), 399-417.

Jenkinson, D. S., Adams, D. E., & Wild, A. (1991). Model estimates of CO2 emissions from soil in response to global warming. *Nature*, *351*(6324), 304-306.

Jetz, W., Rahbek, C., & Lichstein, J. W. (2005). Local and global approaches to spatial data analysis in ecology. *Global Ecology and Biogeography*, *14*(1), 97-98.

Jobbágy, E. G., & Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications, 10(2),* 423-436.

Johnston, K., Ver Hoef, J. M., Krivoruchko, K., & Lucas, N. (2001). Using ArcGIS™ Geostatistical Analyst. *Redlands: ESRI.*

Ju, W., & Chen, J. M. (2005). Distribution of soil carbon stocks in Canada's forests and wetlands simulated based on drainage class, topography and remotely sensed vegetation parameters. *Hydrological Processes*, *19*(1), 77-94.

Kelejian, H. H., & Robinson, D. P. (1993). A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. *Papers in regional science*, *72*(3), 297-312.

Kirschbaum, M. U. (1995). The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. *Soil Biology and biochemistry*, *27*(6), 753-760.

Knegt, H. J., Van Langevelde, F., Coughenour, M. B., Skidmore, A. K., De Boer, W. F., Heitkönig, I. M. A., ... & Prins, H. H. T. (2010). Spatial autocorrelation and the scaling of species-environment relationships. *Ecology*, *91*(8), 2455-2465.

Kurz, W. A., & Apps, M. J. (1999). A 70-year retrospective analysis of carbon fluxes in the Canadian forest sector. *Ecological Applications*, *9*(2), 526-547.

Lee, P., Hanneman, M., Gysbers, J., & Cheng, R. (2010). Atlas of Key Ecological Areas within Canada's Intact Forest Landscapes. *Edmonton, Alberta: Global Forest Watch Canada. 10th Anniversary Publication, No. 4.*

Lefohn, A. S., Knudsen, H. P., & Shadwick, D. S. (2005). Using ordinary Kriging to estimate the seasonal W126, and N100 24-h concentrations for the year 2000 and 2003. A.S.L & Associates, 111 North Last Chance Gulch Suite 4A Helena. *Montana*, *59601*.

Legendre, P., & Fortin, M. J. (1989). Spatial pattern and ecological analysis. *Vegetatio, 80*(2), 107-138.

Leung, Y., Mei, C. L., & Zhang, W. X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, *32(1),* 9-32.

Lichstein, J. W., Simons, T. R., Shriner, S. A., & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological monographs*, *72*(3), 445-463.

Liu, J., Meng, J., & Huang, S. (2011). Spatial distribution and estimating of soil organic carbon on Jingyu county, Jilin province. 2011 International Conference on Electrical and Control Engineering, ICECE 2011 - Proceedings, pp. 3638-3641.

Liu, Z. P., Shao, M. A., & Wang, Y. Q. (2011). Effect of environmental factors on regional soil organic carbon stocks across the Loess Plateau region, China. *Agriculture, Ecosystems & Environment*, *142*(3), 184-194.

Long, J. N. (1982). Productivity of western coniferous forests. *Analysis of Coniferous Ecosystems in the Western United States (Edmonds RL ed). Academic Press, New York, USA*, 89-125.

Luyssaert, S., Schulze, E. D., Börner, A., Knohl, A., Hessenmöller, D., Law, B. E., ... & Grace, J. (2008). Old-growth forests as global carbon sinks. *Nature*, *455*(7210), 213-215.

MacDonald, G. M., & Gajewski, K. (1992). The northern treeline of Canada. In *Geographical Snapshots of North America: Commerating the 27th Congress of the International Geographical Union and Assembly*, 34-37.

Malhi, Y. (2002). Carbon in the atmosphere and terrestrial biosphere in the 21st century. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *360*(1801), 2925-2945.

Manabe, S. (1983). Carbon dioxide and climatic change. *Advances in Geophysics*, *25*, 39-82.

McGrath, D., & Zhang, C. (2003). Spatial distribution of soil organic carbon concentrations in grassland of Ireland. *Applied Geochemistry*, *18*(10), 1629-1639.

McKeague, J. A., & Stobbe, P. C. (1978). History of soil survey in Canada, 1914-1975. *Canada Department of Agriculture.*

Meersmans, J., De Ridder, F., Canters, F., De Baets, S., & Van Molle, M. (2008). A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium). *Geoderma*, *143*(1), 1-13.

Meikle, J. C. & Waterreus, M. B. (2008). Ecosystems of the Peel Watershed: A predictive approach to regional ecosystem mapping. Yukon Fish and Wildlife Branch Report. TR-08-01. 66 pp.

Meyer, S., Leifeld, J., Bahn, M., & Fuhrer, J. (2012). Free and protected soil organic carbon dynamics respond differently to abandonment of mountain grassland. *Biogeosciences*, *9*(2), 853-865.

Miller, H. J. (2004). Tobler's first law and spatial analysis. *Annals of the Association of American Geographers*, *94*(2), 284-289.

Mishra, U., Lai, R., Liu, D., & Van Meirvenne, M. (2010). Predicting the spatial variation of the soil organic carbon pool at a regional scale. *Soil Science Society of America Journal, 74(3),* 906-914.

Mitchell, A. (2005). The ESRI guide to GIS analysis, Volume 2: Spatial Measurements and Statistics. Redlands.

Murty, D., McMurtrie, R. E., & Ryan, M. G. (1996). Declining forest productivity in aging forest stands: a modeling analysis of alternative hypotheses. *Tree Physiology*, *16*(1-2), 187-200.

National Resource Canada. (2001). Canada 3D product standard. *Center for Topographic Information, National Resource Canada.*

Negreiros, J., Painho, M., Aguilar F., & Aguilar, M. (2010). Geographical Information Systems Principles of Ordinary Kriging Interpolator. Journal of Applied Sciences, 10: 852-867.

Oechel, W. C., & Vourlitis, G. L. (1995). Effects of global change on carbon storage in cold soils. Advances in Soil Science: Soils and Global Change, 117-129.

Oliveau, S., & Guilmoto, C. Z. (2005, July). Spatial correlation and demography. In *XXVe Congrès International de la Population*.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis*, *27*(4), 286-306.

Pace, R. K., & LeSage, J. P. (2010). Omitted variable biases of OLS and spatial lag models. *Progress in Spatial Analysis*. 17-28.

Parry, L. E., & Charman, D. J. (2013). Modelling soil organic carbon distribution in blanket peatlands at a landscape scale. *Geoderma*, *211*, 75-84.

Plaster, E. J. (1992). In Reiley H. E. (Ed.), *Soil science and management*. Albany, N.Y.: Delmar Publishers.

Powers, J. S., & Schlesinger, W. H. (2002). Relationships among soil carbon distributions and biophysical factors at nested spatial scales in rain forests of northeastern Costa Rica. *Geoderma, 109*(3), 165-190.

Premo, L. S. (2004). Local spatial autocorrelation statistics quantify multi-scale patterns in distributional data: an example from the Maya Lowlands. *Journal of Archaeological Science*, *31*(7), 855-866.

Qiu, W., Curtin, D., & Beare, M. (2011). Spatial variability of available nutrients and soil carbon under arable cropping in Canterbury. *Retrieved from:* http://www.massey.ac.nz/~flrc/workshops/11/Manuscripts/Qiu_2011.pdf

Raich, J. W., & Schlesinger, W. H. (1992). The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus B*, *44*(2), 81-99.

Rashid, G. H. (1987). Effects of fire on soil carbon and nitrogen in a Mediterranean oak forest of Algeria. *Plant and soil*, *103*(1), 89-93.

Raty, L & Gilbert, M. (1998). Large-scale versus small-scale variation decomposition, followed by Kriging based on a relative variogram, in presence of a non-stationary residual variance. *Journal of Geographic Information and Decision Analysis, 2(2)*, 91-115.

Rezaee, H., Asghari, O., & Yamamoto, J. K. (2011). On the reduction of the ordinary kriging smoothing effect. *Journal of Mining & Environment*, *2*(2), 25-40.

Riha, S. J., James, B. R., Senesac, G. P., & Pallant, E. (1986). Spatial variability of soil pH and organic matter in forest plantations. *Soil Science Society of America Journal*, *50*(5), 1347-1352.

Rowe, J. S. (1972). Forest regions of Canada (No. 1300). Ottawa: Information Canada.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of mathematical psychology*, *15*(3), 234-281.

Schimel, D. S., & Potter, C. S. (1995). Process modelling and spatial extrapolation. *Biogenic Trace Gases: Measuring Emissions from Soil and Water*, 358-384.

Schlesinger, W. H. (1977). Carbon balance in terrestrial detritus. *Annual Review of Ecology and Systematics*. *8,* 51-81.

Schut, P., Smith, S., Fraser, W., Geng, X., & Kroetsch, D. (2011). Soil landscapes of Canada: Building a national framework for environmental information. *Geomatica*, *65*(3), 293-309.

Scull, P. (2010). A top-down approach to the state factor paradigm for use in macro-scale soil analysis. *Annals of the Association of American Geographers, 100(1),* 1-12.

Shakiba, A., & Matkan, A. (2005). Sensitivity of Global Soil Carbon to different Climate Change Scenarios. *Environmental Sciences, 9,* 13-24.

Siltanen, R. M. (1997). *A soil profile and organic carbon data base for Canadian forest and tundra mineral soils*. Canadian Forest Service, Northern Forestry Centre.

Simmons, J. A., Fernandez, I. J., Briggs, R. D., & Delaney, M. T. (1996). Forest floor carbon pools and fluxes along a regional climate gradient in Maine, USA. *Forest Ecology and Management*, *84*(1), 81-95.

Six, J., & Jastrow, J. D. (2002). Organic matter turnover. *Encyclopedia of soil science. Marcel Dekker, New York*, 936-942.

Sprout, P. N., Baker, T. E., & Lavkulich, L. M. (1978). *The soil landscapes of British Columbia*. Resource Analysis Branch, Ministry of the Environment.

Strong, W. L. & La Roi, G. H. (1985). Root density-soil relationships in selected boreal forests of central Alberta, Canada. *Forest ecology and management*, *12*(3), 233-251.

Terra, J. A., Shaw, J. N., Reeves, D. W., Raper, R. L., Van Santen, E., & Mask, P. L. (2004). Soil carbon relationships with terrain attributes, electrical conductivity, and a soil survey in a coastal plain landscape. *Soil science*, *169*(12), 819-831.

Tewksbury, C. E., & Van Miegroet, H. (2007). Soil organic carbon dynamics along a climatic gradient in a southern Appalachian spruce-fir forest. *Canadian Journal of Forest Research, 37(7),* 1161-1172.

Thompson, J. A., & Kolka, R. K. (2005). Soil carbon storage estimation in a forested watershed using quantitative soil-landscape modeling. *Soil Science Society of America Journal*, *69*(4), 1086-1093.

Thornley, J. H. M., & Cannell, M. G. R. (2001). Soil carbon storage response to temperature: a hypothesis. *Annals of Botany*, *87*(5), 591-598.

Timoney, K. P., La Roi, G. H., Zoltai, S. C., & Robinson, A. L. (1992). The high subarctic forest-tundra of northwestern Canada: position, width, and vegetation gradients in relation to climate. *Arctic*, 1-9.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, *46*, 234-240.

Trangmar, B. B., Yost, R. S., & Uehara, G. (1985). Application of geostatistics to spatial studies of soil properties. *Advances in agronomy, 38(1)*, 45-94.

Trofymow, J. A., Stinson, G., & Kurz, W. A. (2008). Derivation of a spatially explicit 86-year retrospective carbon budget for a landscape undergoing conversion from old-growth to managed forests on Vancouver Island, BC. *Forest ecology and management*, *256(10),* 1677-1691.

Trumbore, S. E. (1997). Potential responses of soil organic carbon to global environmental change. *Proceedings of the National Academy of Sciences*, *94*(16), 8284-8291.

Tsui, C. C., Chen, Z. S., & Hsieh, C. F. (2004). Relationships between soil properties and slope position in a lowland rain forest of southern Taiwan. *Geoderma*, *123*(1), 131-142.

Tucker, C.J., Pinzon, J. E., & Brown, M. E. (2004), Global Inventory Modeling and Mapping Studies. *Global Land Cover Facility, University of Maryland, College Park, Maryland.*

Unwin, A., & Unwin, D. (1998). Spatial data analysis with local statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(3), 415-421.

Valcu, M., & Kempenaers, B. (2010). Spatial autocorrelation: an overlooked concept in behavioral ecology. *Behavioral Ecology*, *21*(5), 902-905.

Velandia, M., Rejesus, R. M., Bronson, K., & Segarra, E. (2008). Economics and marketing: Economics of management zone delineation in cotton precision agriculture. *The Journal of Cotton Science, 12,* 210-227.

Voss, P. R., Long, D. D., Hammer, R. B., & Friedman, S. (2006). County child poverty rates in the US: a spatial regression approach. *Population Research and Policy Review*, *25*(4), 369-391.

Walter, S. D., Feinstein, A. R., & Wells, C. K. (1987). Coding ordinal independent variables in multiple regression analyses. *American Journal of Epidemiology, 125*(2), 319-323.

Wang, H., Hall, C. A., Cornell, J. D., & Hall, M. H. (2002). Spatial dependence and the relationship of soil organic carbon and soil moisture in the Luquillo Experimental Forest, Puerto Rico. *Landscape Ecology*, *17*(8), 671-684.

Wang, G., Zhou, Y., Xu, X., Ruan, H., & Wang, J. (2013) Temperature Sensitivity of Soil Organic Carbon Mineralization along an Elevation Gradient in the Wuyi Mountains, China. PLoS ONE 8(1): e53914. Derived from: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0053914

Wang, L. (2006). Spatial econometric issues in Hedonic property value models: model choice and endogenous land use. *ProQuest*.

Ward, M. D. & Gleditsch, K. S. (2007). An introduction to spatial regression models in the social sciences. *Department of Political Science, University of Washington. Department of Goverment, University of Essex.*

Webster, R. (1994). The development of pedometrics. *Geoderma*, *62*(1), 1-15.

Wulder, M. A., White, J. C., Coops, N. C., Nelson, T., & Boots, B. (2007). Using local spatial autocorrelation to compare outputs from a forest growth model. E*cological modelling*, *209*(2), 264-276.

Yuan, Z., Gazol, A., Lin, F., Ye, J., Shi, S., Wang, X., ... & Hao, Z. (2013). Soil organic carbon in an old-growth temperate forest: Spatial pattern, determinants and bias in its quantification. *Geoderma, 195*, 48-55.

Zhang, C., Luo, L., Xu, W., & Ledwith, V. (2008). Use of local Moran's *I* and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Science of the Total Environment*, *398*(1), 212-221.

Zhang, C., Tang, Y., Xu, X., & Kiely, G. (2011). Towards spatial geochemical modelling: Use of geographically weighted regression for mapping soil organic carbon contents in Ireland. *Applied Geochemistry, 26(7),* 1239-1248.

Zheng, D., Hunt Jr, E. R., & Running, S. W. (1993). A daily soil temperature model based on air temperature and precipitation for continental applications. *Clim Res*, *2*, 183-191.

Zinke, P. J., Stangenberger, A. G., Post, W. M., Emanuel, W. R., & Olson, J. S. (1984). *Worldwide organic soil carbon and nitrogen data* (No. ORNL/TM-8857). Oak Ridge National Lab., TN (USA).

Zhou, M., Wang, G. Z., Guo, G., Heping, L., & Wang, L. (2007). Estimate soil organic carbon and nitrogen distribution in Huolin wetland with MODIS data. *In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Vol. 6679*, p. 23.

Appendix I. The Python Script for the Average Climate Data Calculation

```
### Importing ArcGIS Function
import os, sys, arcpy, string, arcgisscripting
from arcpy import env
from arcpy.sa import *

gp=arcgisscripting.create(10.0)
gp.CheckOutExtension("spatial")
gp.OverwriteOutput = True

### Setting the Workspace
gp.Workspace="F:/Data_Temp1/Document/Climate_Metadata/growing/Prep/output"

### Reading the Raster Data
tifs=gp.ListRasters("","tif")
n = len(tifs)

### Creating the FOR Loop to calculate the average climate values
a0 = Raster(tifs[0])
i = 1
loop = 1
while (i < n):
        a1 = Raster(tifs[i])
        output = a0 + a1
        loop = loop + 1
        a0 = output
        i = i+1

### Exporting the Calculated Raster data
output = a0/loop
output.save("F:/Data_Temp1/Document/Climate_Metadata/growing/Prep/prep_gr")
```

Appendix II. Accuracy Assessment of Ordinary Kriging and IDW Approaches Using Different Neighbourhood bands (Interpolated climate maps in the areas beyond the 60 ˚N Latitude).

Max. Temperature	***Max. Neighbours***	***Min. Neighbours***	***RMSE***
Ordinary Kriging	5	2	0.9078
	8	4	0.8580
	10	8	0.9810
	12	8	0.9731
IDW	5	2	0.9306
	8	4	0.9397
	10	8	0.9938
	12	8	0.9946
Min. Temperature	***Max. Neighbours***	***Min. Neighbours***	***RMSE***
Ordinary Kriging	5	2	0.6150
	8	4	0.5893
	10	8	0.5926
	12	8	0.6201
IDW	5	2	0.6562
	8	4	0.6680
	10	8	0.6948
	12	8	0.6874
Precipitation	***Max. Neighbours***	***Min. Neighbours***	***RMSE***
Ordinary Kriging	5	2	19.8572
	8	4	17.7153
	10	8	18.4524
	12	8	18.1708
IDW	5	2	19.5303
	8	4	18.9616
	10	8	19.3719
	12	8	19.3339

Appendix III. Initial estimation of full OLS and spatial error models of SOC-environmental relationships in the Boreal eco-region (n = 628). Model (1) includes maximum temperature as one of the independent variables, while model (2) and (3) includes mean and minimum temperature, respectively.

Independent Variables	***OLS Model (1)***	***Spatial Lag Model (1)***	***OLS Model (2)***	***Spatial Lag Model (2)***	***OLS Model (3)***	***Spatial Lag Model (3)***
Intercept *Sig.*	7.843*** *0.000*	4.634** *0.022*	6.714*** *0.000*	3.928** *0.027*	5.471*** *0.001*	3.092* *0.068*
Precipitation (cm) *Sig.*	0.116*** *0.000*	0. 061*** *0.001*	0.122*** *0.000*	0.065*** *0.001*	0.127*** *0.000*	0.068*** *0.001*
Max. Temp. (˚C) *Sig.*	-0.257** *0.057*	-0.161 *0.226*	--	--	--	--
Mean Temp. (˚C) *Sig.*	--	--	-0.279* *0.069*	-0.180 *0.236*	--	--
Min. Temp. (˚C) *Sig.*	--	--	--	--	-0.251 *0.122*	-0.168 *0.294*
Elevation (km) *Sig.*	-0.833 *0.478*	-0.111 *0.923*	-1.124 *0.334*	-0.290 *0.799*	-1.414 *0.228*	-0.474 *0.679*
Slope *Sig.*	1.260** *0.045*	0.917 *0.134*	1.326** *0.033*	0.951 *0.116*	1.444** *0.018*	1.018* *0.087*
Aspect *Sig.*	-0.002 *0.158*	-0.002 *0.212*	-0.002 *0.161*	-0.002 *0.214*	-0.002 *0.164*	-0.002 *0.217*
NDVI *Sig.*	1.452 *0.678*	0.601 *0.860*	0.903 *0.790*	0.314 *0.924*	-0.231 *0.164*	-0.357 *0.908*
Spatial lag Term (ρ) *Sig.*	--	0.464*** *0.000*	--	0.466*** *0.000*	--	0.471*** *0.000*
R^2 *(Pseudo R^2)*	0.101	0.136	0.101	0.136	0.100	0.136
Log Likelihood	-1898.850	-1888.980	-1899.020	-1889.010	-1899.480	-1889.160
AIC	3811.700	3793.970	3843.310	3794.020	3812.950	3794.330

$p < 0.1$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

Appendix IV. Initial estimation of OLS models of SOC-environmental relationships in the Pacific Cordilleran eco-region (n = 61). Model (1) includes maximum temperature as one of the independent variables, while model (2) and (3) includes mean and minimum temperature, respectively.

Independent Variables	***OLS Model (1)***	***OLS Model (2)***	***OLS Model (3)***
Intercept *Sig.*	-15.821 *0.375*	-13.141 *0.346*	-6.132 *0.595*
Precipitation (cm) *Sig.*	0.268*** *0.001*	0.239*** *0.002*	0.195** *0.016*
Max. Temp. (˚C) *Sig.*	1.193 *0.325*	--	--
Mean Temp. (˚C) *Sig.*	--	1.626 *0.180*	--
Min. Temp. (˚C) *Sig.*	--	--	1.896 *0.103*
Elevation (km) *Sig.*	5.247 *0.444*	6.765 *0.334*	8.209 *0.249*
Slope *Sig.*	0.184 *0.926*	0.242 *0.901*	0.390 *0.839*
Aspect *Sig.*	0.015 *0.487*	0.016 *0.447*	0.018 *0.403*
NDVI *Sig.*	-0.837 *0.969*	-1.948 *0.923*	-0.187 *0.992*
R^2	0.325	0.336	0.347
Log Likelihood	-239.712	-239.239	-238.744
AIC	493.425	492.478	491.489

$p < 0.1$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

Appendix V. Initial estimation of OLS models of SOC-environmental relationships in the Cordilleran eco-region (n = 291). Model (1) includes maximum temperature as one of the independent variables, while model (2) and (3) includes mean and minimum temperature, respectively.

Independent Variables	***OLS Model (1)***	***OLS Model (2)***	***OLS Model (3)***
Intercept *Sig.*	10.015** *0.038*	9.131** *0.023*	11.594** *0.014*
Precipitation (cm) *Sig.*	0.333 *0.474*	0.362 *0.416*	0.112 *0.831*
Max. Temp. (˚C) *Sig.*	-0.095 *0.786*	--	--
Mean Temp. (˚C) *Sig.*	--	0.106 *0.816*	--
Min. Temp. (˚C) *Sig.*	--	--	0.459 *0.362*
Elevation (km) *Sig.*	0.727 *0.609*	0.741 *0.603*	0.842 *0.555*
Slope *Sig.*	0.486 *0.279*	0.513 *0.256*	0.546 *0.226*
Aspect *Sig.*	0.001 *0.738*	-0.001 *0.727*	-0.001 *0.746*
NDVI *Sig.*	-2.421 *0.742*	-5.189 *0.487*	-8.323 *0.239*
R^2	0.040	0.041	0.043
Log Likelihood	-1006.950	-1006.960	-1006.56
AIC	2027.890	2053.980	2027.120

$p < 0.1$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

Appendix VI. Weights for Climate and Terrain Dominant Variables at the National Scale

Environmental Determinant	***Regression Coefficient***	***Weight***
Precipitation (cm)	0.221	0.070
Min. Temperature (°C)	0.500	0.158
Elevation (km)	1.877	0.592
Slope (percent)	0.570	0.180