MAPPING SOIL PROPERTIES IN THE WEST CASTLE WATERSHED, LETHBRIDGE, ALBERTA, CANADA

Trevor W. Deering
B.Sc., University of Lethbridge, 2017

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirement for the Degree

Master of Science

Department of Geography
University of Lethbridge
Lethbridge, Alberta, Canada

MAPPING SOIL PROPERTIES IN THE WEST CASTLE WATERSHED, LEHBRIDGE, ALBERTA, CANADA

TREVOR W. DEERING

Date of Defence: February 6, 2019

Dr. S. Kienzle Co-Supervisor	Professor	Ph.D.
Dr. R. Virk Co-Supervisor Algonquin College Ottawa, Ontario	Professor	Ph.D.
Dr. P. Bonnaventure Thesis Examination Committee Member	Assistant Professor	Ph.D.
Dr. T. Jensen Thesis Examination Committee Member	Instructor	Ph.D.
Dr. C. Coburn Chair, Thesis Examination Committee	Associate Professor	Ph.D.

Dedication

To my family: Thank you for your awesome love, support, and encouragement
to reach my full potential and to pursue my passions.
It is possible "one day at a time"!!

To my friends and colleagues: Thank you for all the fun times!

To my professors: Thank you for cultivating and growing my love
for agriculture and learning.

Abstract

Information on soil properties is limited in the West Castle watershed. Therefore, this research aimed to aid environmental research in the watershed. Data on soil restrictive depth, soil texture, soil organic matter (SOM), and solum depth were acquired from 131 sites in the watershed. Stepwise linear regression modelling was used to find relationships between soil and environmental variables. The results were used to create maps predicting spatial distribution of the soil properties in the study area. Modelling results indicated soils were highly sandy and contain more organic matter than previously thought. Vegetation type, height and density, and land cover type, were found to be important variables in predicting soil properties, and as such could be beneficial for mapping soils in the future. Additionally, the research showed deeper soil restrictive depth than previously thought in the watershed and solum depth met expectations.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACRU   Agricultural Catchments Research Unit
AGRASID   Agricultural Region of Alberta Soil Inventory database
ANOVA   Analysis of Variance
asl   Above Sea Level
CART   Classification and Regression Trees
DEM   Digital Elevation Model
DSS   Detailed Soil Survey
g   Gram
GLM   General Linear Model
HRZ   Horizon
LOI   Loss On Ignition
MAE   Mean Absolute Error
MLR   Multiple Linear Regression
MSE   Mean Squared Error
NDVI   Normalized Difference Vegetation Index
NN   Neural Network
OM   Organic Matter
PP-plot   Probability-Probability Plot
R2   Coefficient of determination
RMSE   Root Mean Squared Error
SHMP   Sodium Hexametaphosphate
SLC   Soil Landscapes of Alberta
SLR   Stepwise Linear Regression
SMU's   Soil Mapping Units
SOM   Soil Organic Matter
SPSS   Statistical Package for the Social Sciences
TWI   Topographic Wetness Index
VIF   Variance Inflation Factor

Chapter 1 Introduction

1.1 Research Statement

Without soils, plant growth in nature is rarely possible. Soils provide ecosystem services, the regulation of infiltration rates, the potential attenuation of floods, and influence groundwater recharge (Shukla, 2014). Miller and White (1998) explain that understanding relationships between vegetation, soil, and the atmosphere for climate and hydrology modelling requires the input of soil property data. Especially crucial are physical soil properties (e.g. solum depth and soil texture) that help to explain the hydrological processes within the soil such as water holding capacity, infiltration rates, and hydraulic conductivity (Miller & White, 1998).

Measuring the magnitude of soil properties is important for understanding hydrological and ecological behavior of watersheds (Oltean, 2016; Wösten, Pachepsky, & Rawls, 2001). One of the difficulties is that little to no soil property data has been gathered in mountainous watersheds (Hitziger & Lieb, 2014). This poses issues for accurate hydrological modelling or other modelling which rely on such data for accurate predictions. These headwaters provide large amounts of fresh water for human use and ecological health, especially from snowmelt. For example, 60% of streamflow in the Cline River watershed in Alberta is attributed to snow melt (Kienzle, Nemeth, Byrne, & MacDonald, 2012). The prairie provinces of Southern Alberta highly rely upon the headwater stream flow from the Rocky Mountains for agricultural, industrial, and residential use (Schindler & Donahue, 2006).

In the mountainous West Castle watershed study area, little soils information is available. While general soil classes have been mapped, more detailed information of soil properties in the watershed, such as soil horizons and their depth and texture, have not been mapped. This lack of soil information for the West Castle watershed creates gaps in our understanding of the environmental processes and functioning of the watershed. Detailed soil sampling, lab analysis, and spatial modelling of soil properties creates new soil information, increasing our understanding of the soils and improving future research in need of soil information.

1.2 Research Objectives

      The first objective was to collect data on the position and magnitude of soil restrictive depth, soil texture, soil organic matter, and solum depth properties present in the West Castle watershed. These properties are important to measure because they are viewed as fundamental soil properties that help to inform the functioning of the soils such as water holding and transport in the soils. To accomplish this objective, soil sampling and extensive lab analysis procedures were carried out. The objective answered questions such as: What are the minimum, mean, and maximum restrictive depths of soils at the sample sites? Are the soil textures predominantly sandy or silty? Is there any variation in the soil organic matter content in the soil, and if yes, how is it related with land cover, elevation, slope, or other terrain variables? The second objective was to model the spatial distribution of soil properties over the West Castle watershed and to construct the maps based on the sampled data. Modelling soil properties is important to understanding relationships that exist with soil properties data and the environmental variables at each sample site. Mapping the soil properties for the entire watershed is important for investigating and understanding large scale patterns between soil and environmental properties that could not be seen with point soil data. Stepwise linear regression was used to model the relationship of soil with many environmental variables. Slope, elevation, aspect, vegetation, geology and a range of other terrain variables were tested for their influence on the range of soil properties gathered through objective 1. The third and final objective was to compare the resulting soil property maps with the two best existing soils maps, SOILGRID's and the detailed soil survey available for the study area. This final objective allows for the investigation into the extent to which the maps produced through this study improve the existing knowledge of soil properties in the West Castle watershed. This research is also crucial for increasing awareness on how soil sampling and mapping techniques can affect the modelling of our environment.

1.3 Thesis Organization

The thesis is organized into six chapters. Chapter 1 describes the motivations for this research and the objectives addressing needs in the study area. Chapter 2 is the literature review, which provides context in current research to understand soil information gaps, soil modeling methods and climate change concerns. Chapter 3, research methods, describes in detail field and laboratory methods undertaken to collect/measure and analyze soil data, as well as the statistical analyses and spatial modelling procedures. Chapter 4 presents the results for field, laboratory, and modelling. The discussion of the results, important conclusions, and recommendations for future research of the pedosphere in the study area are presented in Chapter 5.

Chapter 2: Literature Review

2.1 Soil as a Natural Body

2.1.1 A Definition of Soil

An agreed upon, direct, and encompassing definition of soil seemed elusive from the time

Vasilij Dokuchaev fathered pedology and gave rise to the acknowledgement of soil as a natural

body (Certini & Ugolini, 2013; Huang, Li, & Sumner, 2012). Dokuchaev, and Hans Jenny

following in his footsteps, described soil as a natural three-dimensional entity that is weathered

from rock by chemical, biological, and physical forces in turn creating unique chemical,

biological, and physical properties in the soil. The formation of these properties are affected by

the combination of climate, organisms, relief, parent material, and time which result in

horizonation and distribution of soils (Certini & Ugolini, 2013; Jenny, 1980). The ability of soil to

maintain plant growth seems not to have been important in early descriptions of soil; however,

this has changed. Presently, soil is most commonly explained in the context of its function or

influence. Whether soil characteristics are described for growing crops, building upon, or as a

layer of the lithosphere, the context shapes the definition. For example, soil may be defined

through the lens of plant growth as *"(1) A dynamic natural body composed of mineral and*

organic solids, gases, liquids, and living organisms which can serve as a medium for plant

growth. (2) The collection of natural bodies occupying parts of the Earth's surface that is capable

of supporting plant growth and that has properties resulting from the integrated effects of climate

and living organisms acting upon parent material, as conditioned by topography, over periods of

time"(Brady & Weil, 2008, p. 947). Tan (2005, p. 2) define soil as "*an independent body in*

nature with a unique morphology from the surface down to the parent material as expressed by

the soil profile. Soil is the product of biochemical weathering of parent material, and its

formation is influenced by the soil forming factors: climate, organisms, parent material, relief,

and time". The level of consolidation, depth, and level of reconfiguration through chemical

weathering from the parent material are all of high importance for a concise definition of soil.

4

In mountainous ecosystems, it can be hard to delineate the boundary between soils and rock. Upon closer inspection of a seemingly lifeless mountain slope, it can be observed that soil is present in a small amount, of approximately a centimeter or thicker. Life may or may not be sustained by a soil, however, elements accumulate and erode due to physical, chemical, and biological processes. Humans are reaching into the heavens, exploring far flung planets, asking the question are there other "Earths" out there and are there soils present on celestial bodies. By fitting the definition of soil, possibly contained on celestial bodies, we can encompass soils in seemingly soilless landscapes such as mountainous ecosystems. Certini and Ugolini (2013, p. 379) propose that "*soil is a centimetric or thicker unconsolidated layer of fine-grained mineral and/or organic material, with or without coarse elements and cemented portions, lying at or near the surface of planets, moons, and asteroids, which shows clear evidence of chemical weathering*". This definition encompasses all the important characteristics that make up soil, therefore, this definition has been adopted as the suppositional foundation of this research in mountainous ecosystems. Soil is defined in this manner in order to understand the interaction of soils with environmental forces.

2.1.2 Soil Formation and Classification

In order to arrive at the definition of soil, a fundamental understanding of soil was and is required. Physical, chemical, and biological forces (soil forming processes) act on geological material over time, creating a multitude of soil properties. Soil properties such as sand, silt, and clay proportions (soil texture), colour, depth, cation exchange capacity, structure, pore space, and pH. The combination and intensity of environmental factors of climate, topography, organisms, and parent material act on the soil through time to influence the formation of the soil properties (Steila & Pond, 1989). Pedology is the study of the processes of soil genesis due to the individual and combined influence of these forces (Duchaufour, 1998). Soil genesis leads to the formation of layers in the soil termed soil horizons, that are differentiated based on color, texture, consistence,

structure, and physical, biological, and chemical composition (Soil Classification Working Group, 1998).

Soil taxonomy is the practice of classifying soils, based on individual soils (soil pedons), that are differentiated on the basis of their horizonation (Brady & Weil, 2008). A pedon is 1 m by 1 m hole that extends down to the control section (25 cm into the C or IIC, to permafrost, or 2 m whichever is smallest) (Soil Classification Working Group, 1998). Soil taxonomy follows a hierarchical structure that identifies broad characteristics to more specific characteristics of a soil properties over space (Brady & Weil, 2008). Canada has established a soil classification system consisting of 10 soil orders which can be broken down into great group and subgroup levels of more detailed characteristics (University of Saskatchewan, n.d.). Soil orders are named in regards to soil genesis factors such as the climate degree of weathering. The Regosolic soil order lacks a developed B horizon (< 5 cm) and is referred to as weakly developed due to alluviation, colluviation, climate and other conditions that hinder soil formation (Soil Classification Working Group, 1998). Dissecting the Regosolic soil order name *Rego* refers to regolith and *solic* to solum which conveys the "young" development of the order. Common soils found in forested regions of Canada are Organic, Luvisolic, Brunisolic and Podzolic soil orders. In foothills near mountainous regions and in mountainous regions Regosolic and Chernozemic soil orders can also be found (University of Saskatchewan, n.d.). For more explanation of soil genesis and taxonomy refer to Brady and Weil (2008); Duchaufour (1998); Soil Classification Working Group (1998); and/or Steila and Pond (1989).

To identify connections between environmental characteristics, much investigation has been conducted through soil surveying efforts. This is observed in the soil survey methods, where environmental relationships are recognized to map same and similar soil types across a landscape with similar geology, slope position, and vegetation patterns (Nikiforuk, 1998; Soil Classification Working Group, 1998). Broad environmental relationships of the Chernozemic order are the accumulation of organic matter in the A horizon due to grass, forb, and shrub plant communities

that are capable of growing in well to imperfectly drained soils. They occur in cool, subarid to subhumid environments that include prairie and mountain valley regions. More specific properties of the soil determine the great group and subgroup classifications, such as eluviation of the A horizon in Eluviated Brown Chernozem due to wetter climates moving soil constituents down the profile, leading to an illuviated B horizon (Soil Classification Working Group, 1998). Therefore, classifying soils into soil orders and more detailed groups help soil scientists and all others interested in soil (e.g. farmers and environmental managers) understand how soil types may function in broad terms. For example, Chernozemic soil orders are commonly used for crop production, as they have well developed soil profiles, contain much organic matter, and have good soil structure along with other desirable properties that often allow for favorable crop growth conditions. Deeper understanding of soil property quantification and distribution would need to be undertaken for more specific uses. For example, in precision agriculture initial agriculture activity was established in regards to soil classification followed by more intensive soil sampling for more in depth understanding and management of soil (Flores, Butaslac, Gonzales, Dumlao, & Reyes, 2016). Concepts of soil classification have implications in this research that are discussed throughout.

2.2 Soil Surveying

2.2.1 Conventional Soil Surveying

Taking stock of important soil resources, knowingly or unknowingly, has been an important activity for humans, especially for agricultural practices. The earliest known record of soil management through tillage comes from Iraq, approximately 11,000 years BP (Brevik & Hartemink, 2010). A deeper understanding of soil through scientific methods has only recently grown in the early 19th century (Churchman, 2010). Soil surveying can be described as "*the systematic examination, description, classification, and mapping of soils in an area*" (Brady & Weil, 2008, p. 948). Soil surveying evolved into a deeper understanding of the formation of soils due to environmental forcing, which is used to classify soils such as the Canadian soil surveying

projects (Dent & Young, 1981). Soil surveying started in Ontario, Canada, in 1914, which led to increased soil survey work in the 1920s and beyond. The formation of national committees and working groups, soil classification systems, and extensive soil information gathering resulted from the foundational work done in the 1920s (McKeague & Stobbe, 1978).

The early focus of this foundational work was on assessing the soil resources of Canada for agricultural and environmental management (Anderson & Smith, 2011). Canada has been focused on generalized mapping projects, such as the soil landscapes of Alberta (1:1 million scale) and the detailed soil survey map (1:50,000 scale), due to the cost and effort needed to produce maps of greater detail (Anderson & Smith, 2011). There has been a noticeable decline of government supported conventional soil surveying projects in Canada and around the world since the early 1980's (Anderson & Smith, 2011; Brevik et al., 2016; Geng et al., 2010). Description of the soil survey as "conventional" refers to the use of hard copy maps and information to hand draw soil type boundaries in relation to landscape associations observed in the field, supplemented with aerial photograph interpretation (Geng et al., 2010; Roecker, Howell, Haydu-Houdeshell, & Blinn, 2010). Canada has undergone a dismantling of multiple working groups, committees, and research stations, leading to a shortage of resources such as labour, equipment, knowledge, and money available for conventional surveying efforts (Anderson & Smith, 2011).

Conventional soil surveying has laid the foundation for understanding the spatial distribution of soils, which has allowed for effective management and utilization of soil functions. There has been an increased interest in soil understanding to meet the need for clear communication of soil information for effective soil and environmental management due to dynamic soil conditions under a changing climate and anthropogenic pressure (Hartemink, 2008; Hartemink & McBratney, 2008). Colleges and universities are experiencing an upsurge of enrolment and interest in soil courses, especially in the departments of geography and environmental sciences (Diochon et al., 2016). This seems to be a shift from emphasis placed on soil science for agronomic importance (Diochon et al., 2016). Digital soil mapping (DMS) has

emerged as the new paradigm for soil mapping practices utilizing computer power and other technologies (satellite data, radiometrics etc.) (Boettinger, Howell, Moore, Hartemink, & Kienast-Brown, 2010; Brevik et al., 2016).

In the following sections, Alberta's soil surveying efforts are described, and new opportunities in soil science will be described, enabled by state of the art technology and statistical practices.

2.2.2 Alberta Soil Survey

Two crucial soil survey projects have been conducted in Alberta, Canada: the detailed soil survey (DSS) and the soil landscapes of Canada (SLC). One other survey, less crucial to this research, is the Land Suitability Rating survey (LSRS), updating and improving the Canadian Land Inventory (CLI), which are available for the study area for forestry, agriculture, and other (Agriculture and Agri-Food Canada, 1998). Version 3 of the DSS and version 3.2 of the SLC (and earlier versions) have been created by Agriculture and Agri-Food Canada and compiled into CanSIS (Canadian Soil Information Service). The DSS covers most of the agricultural land of significance in Canada and some non-agricultural lands (Government of Canada, 2016). Alberta agricultural activities have benefited enormously from the DSS by allowing farmers to prioritize land-use according to soils of high potential crop production, which has been one of the major goals of soil survey from the beginning (McKeague & Stobbe, 1978). Chernozemic soils, for example, are greatly utilized in Alberta due to their high organic matter content in the deep Ah horizon, overall deep profile development with one or more B horizons, fine texture development for high moisture and nutrient holding capacity, strong aggregate stability, and usually high-quality tilth (Brady & Weil, 2008; Soil Classification Working Group, 1998). A large amount of land is unsurveyed because the land is not significant for agriculture. This presents a gap in the soil knowledge of Alberta. The detailed soil survey, which encompasses the Agricultural Region of Alberta Soil Inventory database (AGRASID) and soil information viewer data sets, covers approximately 40% of Alberta (Canada - Alberta Environmentally Sustainable Agriculture

Agreement (CAESA), 2018). The SLC soil delineations were first published in 1977 to provide the major soil attributes of importance for growing plants, managing land, and evaluating soil degradation on a large scale (1:5 million) (Clayton, Ehrlich, Cann, Day, & Marshall, 1977). The SLC covers the entire Canadian landscape. Version 3.2 of the SLC provides information at 1:1 million scale (Government of Canada, 2013). Individual soil components of the SLC are included in each polygon, however, the location of the soil samples taken to define the components is not always near the soil component or defined explicitly. This can lead to generalization of component information and leaving detail unknown to the user of the information. Currently, in the Southern Rocky Mountains of Alberta, soil survey data consists mainly of the SLC, which creates much uncertainty for those interested in the location of the individual soil components and soil sample locations. Soil property information is missing over vast stretches of mountainous land, with soil group and great group, and other generalized soil information for soil depth, water holding capacity and other broad soil properties (Government of Canada, 2013).

Upon downloading both the DSS and the SLC data, I found a wealth of information supplied by both of the data sets. For example, querying any polygon in the DSS information (with ArcMap 10.5), it provides detailed information for Component 1 (Dystric Brunisol soil group) only, whereas a similar query for the same polygon using the online Alberta Soil Information Viewer, resulted in less information for component 1 and 2 (Dystric Brunisol soil group and Orthic Dark Brown Chernozems). The DSS provides greater delineation of soil components than the SLC, however, they both encounter similar limitations of providing limited and generalized soil property information for mountainous regions of Alberta. For research and other activities that require detailed soil data in the West Castle watershed the detailed soil surveys cover only 13% of the area, they are generalized polygons containing limited information on soil depth, soil texture, organic matter, bulk density, and other properties that may be of interest to researchers. The points described here become an issue when users require high-resolution soil data, for example, for hydrological modelling research.

I believe that studying the soils is incomplete without the application of the data. Therefore, finding that there is a gap in soil knowledge in the West Castle watershed, for hydrological research specifically, establishes a practical need for soil research in that location. This research sets out to create a digital soil database of soil properties, for use in hydrological research being conducted in the Westcastle watershed. A deeper understanding of soil will increase certainty of hydraulic soil properties used in modeling hydrologic cycling, and thus, potentially decrease the uncertainty of hydrological models. Kienzle et al. (2012) use the ACRU (Agricultural Catchments Research Unit) model in the Cline River Basin of Alberta and suggested that detailed soil information in their study could possibly lead to greater accuracy of the ACRU model predictions. Cornelissen, Diekkruger, and Bogena (2016) found that lower resolution soil data of a mesoscale catchment tended to increase overestimation of discharge rates compared to the higher resolution soil data of the sub-catchment. These studies (Cornelissen et al., 2016; Kienzle et al., 2012) support the need for more detailed soil data for physically-based hydrological models. Increasing the resolution of soils data in the West Castle watershed and distributing the data to hydrologists can lead to higher confidence in the hydrological or ecological modelling.

2.3 Changing State of Soil Mapping

2.3.1 Soil Mapping for Data Acquisition

Soil mapping is moving from conventional surveying techniques to that of digital soil mapping (DSM) techniques. DSM is the input of soil data into a computer model for conducting spatial and non-spatial mathematical and statistical computations with regard to soil variables and their connections to environmental variables (Grunwald, 2010; Minasny & McBratney, 2016). The computations result in predictions of soil property information that can be easily distributed in map and database formats, and can then be recalculated and updated when new information is available (Minasny and McBratney, 2016). Accessible environmental data that can be measured with minimal effort are rapidly growing in their application. Large remote sensing databases have

been and are being created by government, researchers, and private organizations (e.g. AltaLIS, GeoDiscover, ABMI, Farmers Edge, and AGRI-Trend). Remote sensing data can be used to create digital elevation models (DEM), which can be used to calculate slope, aspect, wetness index, solar radiation, and other environmental properties that can be used to predict soil property occurrence without having to measure those properties individually. Using accessible environmental data from a DEM has been implemented by many soil studies (Chai, Shen, Yuan, & Huang, 2008; Gessler, Moore, McKenzie, & Ryan, 1995; Hitziger & Lieb, 2014; McKenzie & Ryan, 1999). Much research into the nature of soils and how they form has been conducted by many DSM researchers, utilizing Jenny's soil forming function, adapted from Dokuchaev's research (Florinsky & Florinsky, 2012; McBratney, Mendonça Santos, & Minasny, 2003; Wiesmeier, Barthold, Blank, & Kögel-Knabner, 2011). Jenny's function, recognized by many as the CLORPT model after the elements in his equation, focuses heavily on quantitative reasoning (McBratney et al., 2003).

$$S = f (cl, o, r, p, t \dots) \hspace{3cm} \text{Equation 1.}$$

where cl is for climate, o is for biotic factor, r is for relief or topography, p is for parent material, and t is for time or age of the soil (Jenny, 1980). It can be observed that DSM is built heavily upon the theory of Jenny's model, because the discipline uses variables such as terrain, climate, vegetation, land use, geomorphology, and solar radiation, termed environmental variables, to describe soil properties of interest (Afzali, Varvani, & Jafarinia, 2015; Debella-Gilo & Etzelmüller, 2009; Moore, Gessler, Nielsen, & Peterson, 1993; Zhang, Huang, Shen, Ye, & Du, 2012) . Variables influencing soil formation contribute to specific soil property formation such as soil depth, soil water holding capacity, soil temperature, soil organic carbon, soil texture, and soil organic matter content (Jin et al., 2015; Kuriakose, Devkota, Rossiter, & Jetten, 2009; Piedallu, Gégout, Bruand, & Seynave, 2011; Seibert, Stendahl, & Sørensen, 2007).

The fast-paced improvement and integration of technology, such as increasing computer power, into the field of soil mapping is greatly increasing our ability to understand and map soil

properties. There appears to be a common objective inherent in mapping soils through both conventional and cutting-edge soil mapping techniques (e.g. random forest modelling, Latin Hypercube sampling). Soils are mapped based on the most important need. The most important need has risen from the agricultural industry, where it is crucial to know and understand the soil to be able to produce sustainable, healthy, and productive crops (Grunwald, 2010). Land has been investigated intensively on the basis of its agricultural suitability and has been utilized accordingly (Hartemink & McBratney, 2008). Land is also investigated for building infrastructure, taxation, and environmental management, but from a soil perspective these purposes have only become more popular than agriculture in approximately the last 30 years (Brevik & Hartemink, 2010; McKeague & Stobbe, 1978). Land that is important for agriculture and infrastructure use has been mapped extensively by conventional surveys and digital techniques (Government of Canada, 2016; Hartemink, 2008). There is a clear demand for soil mapping in complex terrain regions for environmental management, especially in the face of climate change (Hartemink, 2008; Wang et al., 2018; Were, Bui, Dick, & Singh, 2015). Soil is a dynamic geosphere that interacts with the lithosphere and atmosphere. It is inextricably linked to the hydrosphere and biosphere creating a complex life supporting web (Grunwald, 2005). Understanding changes that occur due to anthropogenic and environmental activities and their influence on soil is important for future soil and environmental management. Soil scientists, hydrologist, mathematicians, and professionals from many other disciplines have and must continue to work together to manage the important soil resource (Shukla, 2014).

2.4 Soil Modelling

Relating the soil variability to the environmental aspects was greatly implemented in the conventional soil surveying efforts, as explained above. Also, DSM research has focused on refining modelling methods that are capable of predicting soil properties that explain high levels of soil variability. Many statistical modelling techniques have been researched by the DSM community in order to effectively achieve this goal. Based on the literature, multiple linear

regression (MLR), generalized linear models (GLM), many forms of kriging, neural networks (NN), decision trees, and fuzzy logic techniques are most commonly implemented to study soil variability (Brevik et al., 2016; Heung, Bulmer, & Schmidt, 2014; Menezes, Silva, Mello, Owens, & Curi, 2014; Mosleh, Salehi, Jafari, Borujeni, & Mehnatkesh, 2016; Were et al., 2015; Zhu & Lin, 2010). Grunwald (2009) reviewed 90 journal articles and found that the majority of studies used forms of regression statistics (41.1%), followed by classification and discrimination methods (32.2%), univariate kriging (18.9%), and lastly tree-based models (13.3%). Also, many of the articles compare two or more methods (60%) as opposed to just one (40%) as reported by Grunwald (2009). While reviewing the literature it became indeed evident that many of the studies compared two or more statistical methods (Forkuor, Hounkpatin, Welp, & Thiel, 2017; Hitziger & Lieb, 2014; Ryan et al., 2000; Were et al., 2015). Machine learning techniques, such as tree-based models and NN, are receiving much attention for their ability to model non-linear and complex relationships of soil properties in relation to environmental variables, as well as find hierarchical importance of environmental variables (Wang et al., 2018). Decision tree models are able to "handle" categorical variables and continuous variables in the same data set, which is useful when considering vegetation type and/or geological features that may influence soil property formation (De'ath & Fabricius, 2000). One-hot encoding is the creation of binary variables (zeros and ones) to create numerical representations of categorical variables (Chang & Lipson, 2018). For example, gender as an independent variable has two categories, male and female, two separate categories would be created. A male category where every sample that was male would be given a one and every sample that was female would receive 0, same for the female category. These types of variables are also called dummy variables (Suits, 1957). There are those who argue that categorical variables should be one-hot coded (Breiman, 2001), and there are others who argue that one-hot coding is not necessary and is actually detrimental to tree modelling (e.g. Random Forest Model) (Dingwall & Potts, 2016; Estrada, Ahneman, Sheridan, Dreher, & Doyle, 2018).

Stepwise linear regression (SLR) was chosen as it has been tried and tested many times to model soil property spatial distribution (Grunwald, 2009). It was also observed that SLR performed similarly to RF (Random Forest) and other methods (Forkuor et al., 2017; Hitziger & Lieb, 2014; Zhang et al., 2012). SLR was viewed as highly interpretable as to the equations, outputs, and connection to relationships between the dependent and independent variables. Forkuor et al. (2017) explain that linear regression model performance can suffer greatly from multicollinearity of independent variables. Therefore, testing for multicollinearity was regarded important. Geostatistical methods were considered through spatial autocorrelation analysis with Moran's I and semi-variograms. It was assumed that the chosen sampling strategy would not be sufficient for geostatistical modelling across the entire watershed, however, spatial autocorrelation could be useful in considering spatial dependence of dependent variables to independent variables.

Sampling techniques have also been researched and applied to the modelling methods to best sample for the modelling method being used. For example, in geostatistical modelling, sampling strategies that follow standard statistical rules that require uniform and independent samples over a study area are implemented (Chiles & Delfiner, 2012). A sample strategy can be planned on the basis of covering the geographic space, the covariate range, or both (Kidd, Malone, McBratney, Minasny, & Webb, 2015). Random sampling, stratified sampling and variations of such (Tan, 2005), Conditioned Latin Hyper Cube sampling (stratified random sampling from multivariate distributions of covariates, to produce a Latin Square of probabilities) (Minasny & McBratney, 2006; Roudier, Beaudette, & Hewitt, 2012), fuzzy k-Means clustering of covariates (clustering of multivariate centroids based on minimizing the mean squared distance between centroid values, where observations are given a degree of belonging to a cluster) (Kidd et al., 2015), and other variations of sampling (e.g. Integrative Hierarchical Stepwise sampling as suggested by (Yang et al., 2013) have been tested and used to map soils properties. When choosing a sampling method to characterize the study area, existing soil data or exploratory

sample data, and modelling type are important to take into consideration to plan for time, effort, expense, and suitability of the sample regime. Backcountry areas in mountainous regions may be too hard to access, like in the Rocky Mountains. Choosing a method that takes this into consideration is important for conducting the sampling of soils. Rationale for selection of a sampling method is provided in the methods section.

2.5 Soil Mapping in a Changing Climate

2.5.1 Soil Properties and Change Dynamics

Foundational knowledge of soil properties that tend not to change over short periods of time is critical for studying a changing climate. Soil texture is one property that does not change rapidly over time because it takes thousands of years of physical and chemical weathering of rock to break down rock particles into smaller soil particles (Brady & Weil, 2008). Changing the texture of a soil to be less sandy for agricultural production, for example, would take a prohibitive amount of clay material and effort to increase the clay content in the soil to induce better texture and thereby structural change in the soil. Another example, is that of soil depth, which also forms over many years, sometimes in short periods if the soil is being rapidly eroded, but for the majority of soils the depth does not change substantially over hundreds of years (Shukla, 2014). Changing the depth of a soil takes massive effort, due to the amount of soil or sediment required. One acre of medium textured soil with a 15 cm cultivated zone and a bulk density of 1.3 g/cm^3 would weigh approximately 870 tons (Brady & Weil, 2008). This example illustrates that changing the depth over a field (>= 1 acre) would be an unattainable feat. Soil organic matter content can change drastically over short periods of time where anthropogenic drivers are present, however, it tends to be stable or increasing in forest ecosystems depending on the age of the forest and many other drivers (Brady & Weil, 2008; Fisher, Binkley, & Pritchett, 2000; Smith, Janzen, Scherloski, Larney, & Ellert, 2016). Gathering soil property information for basic and stable soil properties such as these can help researchers investigating environmental change.

2.5.2 Climate Change Induced Issues

Climate change is said to be the most important concern of the 21st century (Tanzeeba & Gan, 2012). It has the potential to disrupt the operations of agricultural production, industrial manufacturing, functioning of ecological systems, and combined possibly leading to degraded human health. Many studies have reported that the climate of the prairies is changing to one with less snowfall in the winter, earlier occurrence of the spring freshet, and lengthening growing degree days (Bonifacio, 2016; Cutforth et al., 1999; Kienzle et al., 2012; Lemmen, Warren, & Lacroix, 2007; Schindler & Donahue, 2006). Hydrological and environmental research concerning climate change is important to understand which changes will occur and the magnitude of the changes. Integral to many of these studies is knowledge of the soil.

Rood et al. (2008), for example, studied the changing flow regimes of Rocky Mountain Rivers of Alberta and the possible challenges facing forests in the flood plain ecosystem. They found increased winter flow, early onset of spring freshet, and decreased late summer flows resulting from the changing climate. Of great concern is the decreased late summer flow which could highly stress cottonwoods species because they require water during the hottest period of the year, potentially leading to increased mortality of forests in the flood plain environment (Rood et al., 2008). If the health of our riparian zones decline this could lead to declined water quality for downstream users due to vegetation change (Lemmen et al., 2007) and the possibility of weed invasion (Rood et al., 2008). Die off of vegetation like this in the riparian zone could lead to a great amount of soil erosion, thus decreasing water quality, due to decreased root systems available to hold the soil in place (Naiman & Décamps, 1997). Mapping soil and/or sediment types in the riparian zones could help understand how distribution of tree species such as the cottonwoods will change and the magnitude of erosion that could occur due to water scarcity events (Botero-Acosta, Chu, Guzman, Starks, & Moriasi, 2017; Naiman & Décamps, 1997). This is especially important when considering the possibility of increasing water scarcity due to climate changes.

Another example of land cover change is the potential for tree species to move into new habitat, where once the temperature and moisture regime was not suitable. Of concern is the loss of habitat, for cold tolerant species in particular, that will lose habitat to warmer temperatures (Zolkos et al., 2015). Agricultural expansion may occur in Northern areas of Alberta due to increasing growing season length and increasing CO_2, whereas southern parts of Canada and the US may see decreased agriculture production due to increasing drought occurrence and water shortages (Motha & Baier, 2005). The soil is one variable that has a large impact on suitable habitat for plant growth and is used to investigate where plants may be redistributed (Zolkos et al., 2015).

Another concern is that of water availability and management for dryland and irrigated crop production, animal production, and food processing (Fischer, Tubiello, van Velthuizen, & Wiberg, 2007; Hanjra & Qureshi, 2010). Hanjra and Qureshi (2010) explain that approximately 450 million people in 29 countries are facing water shortage problems and that two-thirds of the world population could face water stress by 2025, straining agricultural production. Starvation and many other issues arise from water shortages (Hanjra & Qureshi, 2010). If the wasting away of glaciers continues, snowpack decreases, earlier freshet occurs, and temperatures increase many issues can arise. One such issue is the decline of soil water recharge. Cutforth et al. (1999) found decreasing snowfall patterns during their research in Saskatchewan that will lead to decreased soil water recharge. This could increase crop stress, especially for dryland crops in the semi-arid prairie region where it is expected that the prairies face greater aridity in the future (Jiang, Gan, Xie, Wang, & Kuo, 2017; Sauchyn, Barrow, Hopkinson, & Leavitt, 2002). Also, the availability of water for irrigation and residential use could decline when it is needed most in the mid-summer, which may lead to a water crisis in Alberta and elsewhere (Schindler & Donahue, 2006). Droughts are not a new phenomenon in the Canadian prairies, however, if trends continue the severity and length of droughts could increase leading to water scarcity and degradation of water quality (Sauchyn et al., 2002).

To adapt to the drought threats facing southern Alberta, and to protect head water sources, the hydrological regimes must be better understood to map groundwater recharge regions and watersheds with high water yields. As soils play an essential part in the hydrological cycle, a better understanding of the soils in the head water regions will certainly improve the understanding of the respective hydrological regimes.

2.6 Research Considerations: Field and Laboratory

Review of the topics up to this point all lead to the selection of soil properties of importance to sample and analyze. Soil nutrients, mainly nitrogen, phosphorus, and potassium, are sampled at multiple depth increments are analyzed for improving agriculture and forest production (Brady & Weil, 2008; Fisher et al., 2000).

This research is aimed at informing all hydrological and other environmental research that may benefit from soil information in the West Castle watershed. Also, there is little to no soil property information for the study area. Therefore, base soil properties are seen as important properties to gather. Base properties are those that are used to understand the fundamental nature of a soil, and they inform higher level soil properties, such as water holding capacity and hydraulic conductivity. Soil color, texture, depth, OM content, soil depth, bulk density, structure, strength, and consistency can all be considered base soil properties. For example, soil water holding capacity is affected by the texture of the soil, where it can be expected that sandier soils will hold less water at saturation and when nearly dry (Shukla, 2014). The soil texture is a base soil property that helps understand soil porosity which in turn helps understand soil water holding capacity better. Organic matter content is similarly important as soil texture for understanding soil water holding capacity and for assessing crop and other vegetation productivity potential because OM releases nutrients and adds to other soil properties (e.g. aggregate stability) (Brady & Weil, 2008; Hudson, 1994). Solum depth, which refers to the depth of the A and B horizons to the top of the C horizon, is regularly required for physically based hydrological modeling and must be assessed where data is not available (Herbst, Diekkrüger, & Vereecken, 2006). Sampling and

measuring soil horizons is required to gather data for solum depth analysis. A restrictive layer in a soil profile can impede percolation of water to lower soil depth thereby increasing horizontal water movement and decreasing water holding capacity of the soil. Measurement of restrictive depth can be important for understanding water movement and plant interactions in soil (Ezeaku & Anikwe, 2006; Laboski, Dowdy, Allmaras, & Lamb, 1998). Soil depth can be assessed for agriculture to understand rooting depth potential, water holding capability or other criteria (Brady & Weil, 2008). Electrical conductivity has been used extensively in agriculture to assess salinity and properties relating to crop growth (Corwin & Lesch, 2003). Soil hydraulic conductivity is also important for hydrological research because it can give insight into the rate of infiltration and subsequent water movement through a watershed (Davie, 2008).

Next, sampling methods must be considered to best sample for the purpose and the soil property. A soil probe, shovel, and/or an auger are commonly used soil sampling tools (Brady & Weil, 2008; Tan, 2005). For agriculture purposes, it is common to use long metal probes to sample at multiple depth increments. In shallow soils a hand held probe can be sufficient. In rocky and or dense rooted soils a shovel and/or pick axe may be the only tools that work. Another consideration is that of horizon sampling for soil classification and solum depth analysis. To analyze the horizons, a pedon or small hole can be used to closely inspect the horizons, or if a soil probe is able to easily penetrate the soil and hold the soil in the tube with little disturbance to the core in the tube, then a probe can be used. In mountainous regions with many rocky and sandy soils a tube may get damaged, may not penetrate into the soil well, and may not hold the core therefore a shovel may be the best soil sampling tool for digging holes and gathering samples.

In general, data collection is followed by laboratory analysis of samples. Selection of the lab methods depends on cost, time, equipment availability, and knowledge of the method. Some methods can be highly technical and others more easy. For example, soil texture can be analyzed using sieves or by using laser diffraction. The laser diffraction equipment and understanding can be more sophisticated and require training and/or education of the concepts to use, whereas

sieving methods use sieves, a shaker, and commonly water to separate the soil particles thereby being easier to conduct. Another method of consideration, for the determination of soil organic carbon, is that of the Walkley-Black wet combustion method (Tan, 2005). This method requires the use of sulfuric acid, which is highly volatile, and other elements that may be difficult to obtain. This method is accurate and precise, but is not used mostly due to the dangers of using sulfuric acid (Stuart, personal communication, October 2018). Assessing equipment and multiple methods were integral to conducting lab methods effectively.

2.7 Summary

Knowledge of soils is important to the functioning of many human activities. Taking stock of the distribution of soils through soil surveys and advanced soil mapping techniques is important for managing soil, food, and water resources, now more than ever in the middle of large climactic changes. Soils of the Rocky Mountains have not been mapped sufficiently for the needs of environmental research. Therefore, mapping soils in the Rockies by implementing statistical methods that have been used by other soil mapping studies will help to fill the soils knowledge gap. In the next chapter, methods are chosen based on my ability to implement the methods, equipment availability, time, and comparative studies of statistical methods found through the literature review process.

Chapter 3: Research Methods

3.1 Study Area

The research was conducted in the West Castle watershed of the Southern Rocky Mountains of Alberta, Canada, centered at 114.37 º W, 49.37 º N (Figure 3.1). The study area covers approximately 825 km^2, with a minimum elevation of 1,109 m and a maximum elevation of 2,811 m above sea level (asl). Mean annual temperature reaches approximately 4.9ºC and annual precipitation (snow and rain) accumulates to approximately 679.8 mm (Government of Canada, 2017). These climatic conditions support a wide variety of vegetation, including, White pine (*Pinus strobus*), Lodgepole pine (*Pinus contorta*), Bearberry (*Arctostaphylos uva-ursi*), Poplar (*Populus*), Common Juniper (*Juniperus communis*), Solomon's Seal (*Maianthemum racemosum*), Bear Grass (*Xerophyllum tenax*), Oregon Grape (*Mahonia aquifolium*), and many other shrubs, flowers, trees, and grasses (Alberta Biodiversity Monitoring Institute, 2010).
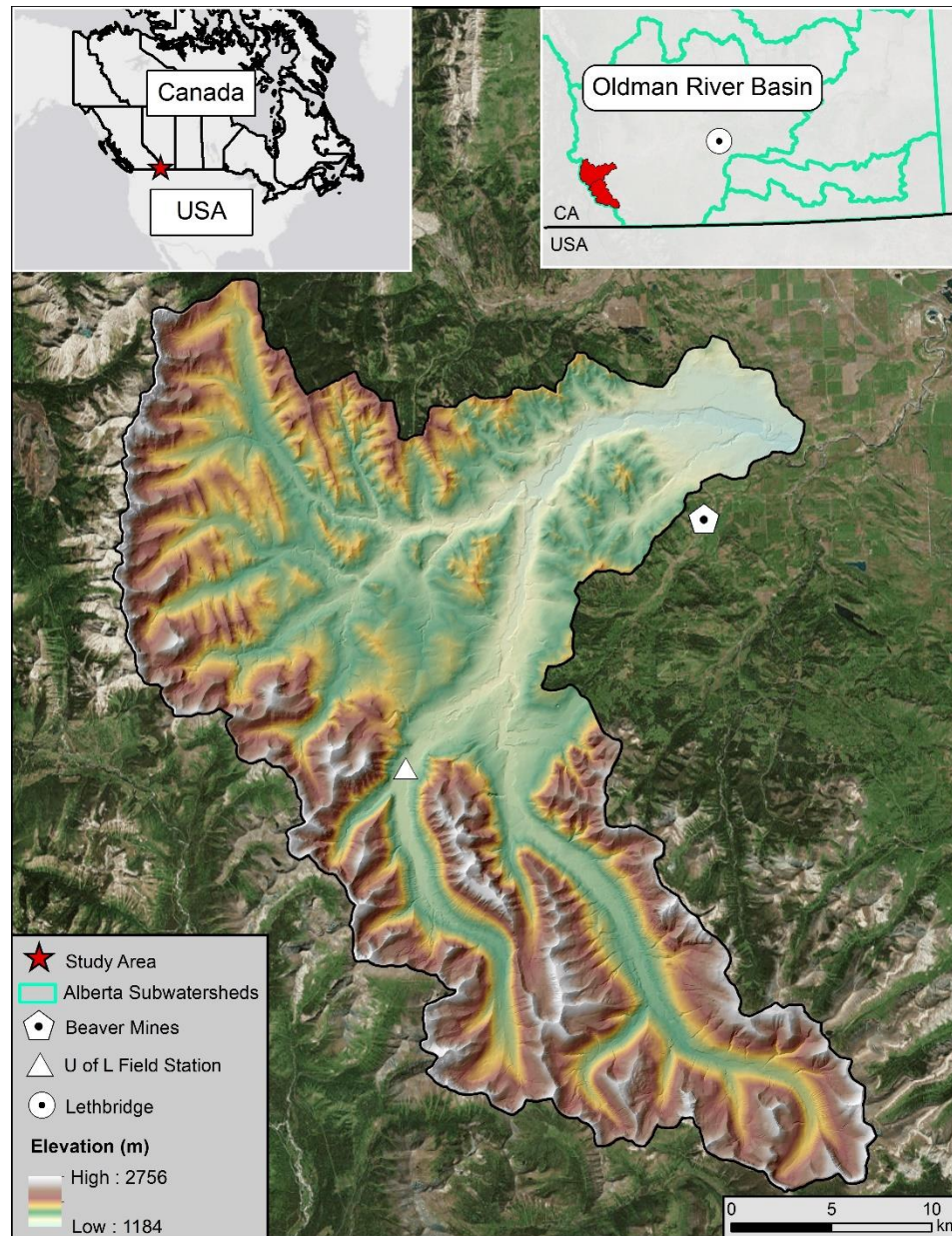
Figure 3.1: Map showing the extent of the West Castle watershed study area with smaller inset maps showing the location of the study area in relation to North America and Canada. *(Sources: Base map from Esri (2018), Alberta watersheds from Government of Alberta (2014), DEM from Natural Resources Canada (2016), and Canada boundaries from Statistics Canada (2016)).*

The study area is dominated by Regosolic, Brunisolic, Luvisolic, and Chernozemic soil orders (Government of Canada, 2013). The Regosolic, Brunisolic, and Luvisolic orders are young soils, which are weakly developed with either no B horizon, or the early formation of a B horizon, or slight illuviation of clays in the B horizon from the A horizon. These are typical soils of mountainous areas that are young, eroded, and under forest vegetation.

The Luvisolic order is characterized by high base saturation (leading to dominantly neutral to alkaline conditions) occurring under forest vegetation. Luvisols can be moderately to strongly acidic and, based on visual inspection of vegetation, e.g. pine species, in the Castle region, this is likely to be the case in the study area. Furthermore, presence of Dystric Brunisol great group suggests that the soil is acidic with a pH < 5.5 in the B horizon, which is at least 5 cm thick.

The black Chernozemic soils in the area are more developed, with a dark A horizon having a chroma of at least 3.5 moist, at least 10 cm thick A horizon, 5 to 8.5 % organic matter content, and less acidity than the Regosolic, Brunisolic and Luvisolic orders (Soil Classification Working Group, 1998).

The parent material of the area is dominated by colluvial and alluvial material transported and deposited by gravitational forces and water movement. The lithology of the bedrock consists of sandstone, shale, mudstone, limestone, dolomite, argillite, dolostone, and basalt (Government of Alberta, 2013). The depth to bedrock ranges from ~1.5 m, in the low lying grassland areas, to 0 m, nearest the mountain outcrops (Alberta Agriculture and Forestry, 2018).

Conducting soil mapping in the complex terrain of this watershed will help in filling gaps in soil property knowledge through DSM. The knowledge will be useful to the larger soil mapping research community due to the lack of research on complex terrain soils (Hitziger & Lieb, 2014).

3.2 Field Data, Experimental Design, and Sampling Methods

3.2.1 Selection of Sampling Sites

Soil genesis depends on the complex interaction of soil forming factors such as climate, vegetation, topography, time, and anthropogenic influence (Jenny, 1980). It was recognized that vegetative growth, aspect and slope are all variables that influence climate, biological functioning,

and topography and thereby influence the formation of soil properties. With this supposition, three environmental characteristics were used to delineate the study area into stratifying units termed soil mapping units (SMU's), which guided the selection of soil sites in representative and unique SMU locations. The SMU's serve as a starting point for site selection and provide an understanding of the connection of soil property formation to environmental characteristics. ArcMap (v.10.5.1) was used to overlay three layers consisting of:

a. land cover, which relates to vegetative growth (Alberta Biodiversity Monitoring Institute, 2010)

b. annual solar radiation, which relates to elevation, slope, and aspect and represents the energy available for plant growth, *and*

c. topographic wetness index (TWI) data that relates to the steepness of slope and the uphill area, representing relative soil moisture conditions.

Both, solar radiation and wetness index were created in ArcMap, using a ten meter spatial resolution digital elevation model (DEM) (Natural Resources Canada, 2016). These layers were combined with the land cover through spatial overlay of the raster datasets using the raster calculator tool. The resulting raster dataset constitutes the mapping units used to delineate important soil forming environments to be sampled. Table 1 illustrates the classification of raster attributes used to create the SMU's. The vegetation dataset contained 10 classes which were given single digit values ranging between 0 and 9. The wetness index contained a range of values from 0 to 26.97. The range of values were divided into three classes (*High, Medium and Low*), where:

i. High, represents known rivers, streams and wetlands

ii. Medium, represents slopes, valleys and meadows and

iii. Low, represents ridges and steep uphill terrain.

The annual solar radiation raster analysis resulted in a range of values from 1.4 W/m^2 in deep valleys which never receive direct sunshine, to over 180 W/m^2 at south-facing slopes at high

elevations. The range was classified into three equal quantiles to divide the study area into regions

of high, medium and low solar radiation.

Table 3.1: Summary of data type, class, break values, and the reclassified value assigned to each
class for SMU Creation.

Data Type	Class	Break Values	Reclassified Value
Land Cover	Water	N/A	0
Land Cover	Rock/Rubble	N/A	1
Land Cover	Exposed Land	N/A	2
Land Use	Developed Land	N/A	3
Vegetation	Shrubland	N/A	4
Vegetation	Grassland	N/A	5
Land Use	Agriculture	N/A	6
Vegetation	Coniferous Forest	N/A	7
Vegetation	Broadleaf Forest	N/A	8
Vegetation	Mixed Forest	N/A	9
Wetness Index	Low	0 - 0.01	10
Wetness Index	Medium	0.01 – 5.5	20
Wetness Index	High	5.5 – 26.9	30
Solar Radiation	Low	1.4 to 126.9 W/m^2	100
Solar Radiation	Medium	126.9 – 140.5 W/m^2	200
Solar Radiation	High	140.5 – 183.2 W/m^2	300

Upon completion of cell classification and raster overlay addition, 90 SMU's were

delineated. A resulting SMU classified as 311 means that the unit is characterized by high solar

radiation (300), with a low wetness (10), and rock and rubble (1), which is most likely on a rocky

ridge facing south. The area of the units was calculated and the top thirteen units, covering 73 %

of the study area, were selected as the priority SMU's to be sampled. Priority was especially set

on the thirteen SMU's to gain understanding of soil formation in relation to environmental

variables in SMU's across many locations, which allowed for the testing of the hypothesis that

soil properties are linked to SMU's. As planning progressed, other SMU's, encountered between

top priority SMU's, were sampled to gain understanding of those units with our preferential

sampling scheme.

Preferential stratified random sampling was implemented in these units to choose

individual soil sampling sites. Preferential sampling has the benefit of reducing time, cost and

effort applied to obtaining soil samples in a large and complex study area, however, preferential sampling is known to add bias and skew inferences if the sampling strategy is not taken into consideration in the modelling stage (Clifford et al., 2012; Diggle, Menezes, & Su, 2010; Shaddick & Zidek, 2014). Stratified random sampling was implemented to sample locations without bias within preferred SMU's. Each sample site (~0.25 m^2) was chosen randomly inside an SMU within reachable distances (~<= 1000 m) from roads, trails and cut lines.

Although there is a tool in ArcMap called *Create Random Points,* which creates a selected number of random point locations within a SMU, this tool was not used, as the tool creates random points for an individual SMU anywhere in the watershed. To get around this problem the SMU polygons would have to be broken apart and each SMU would have to be exported as its shapefile, then random points would have to be created. Instead of conducting this laborious procedure for every SMU and sample site, random points were chosen within a preferred SMU, sometimes in a transect orientation. Figure 3.2 illustrates the soil sample location procedure, where multiple soil sample locations were chosen across multiple SMU's to sample as many unique sites as possible in a transect. Notice that the sites are located off of a cut line and off a road to maximize ease of access and to hike as far as possible in the time available.
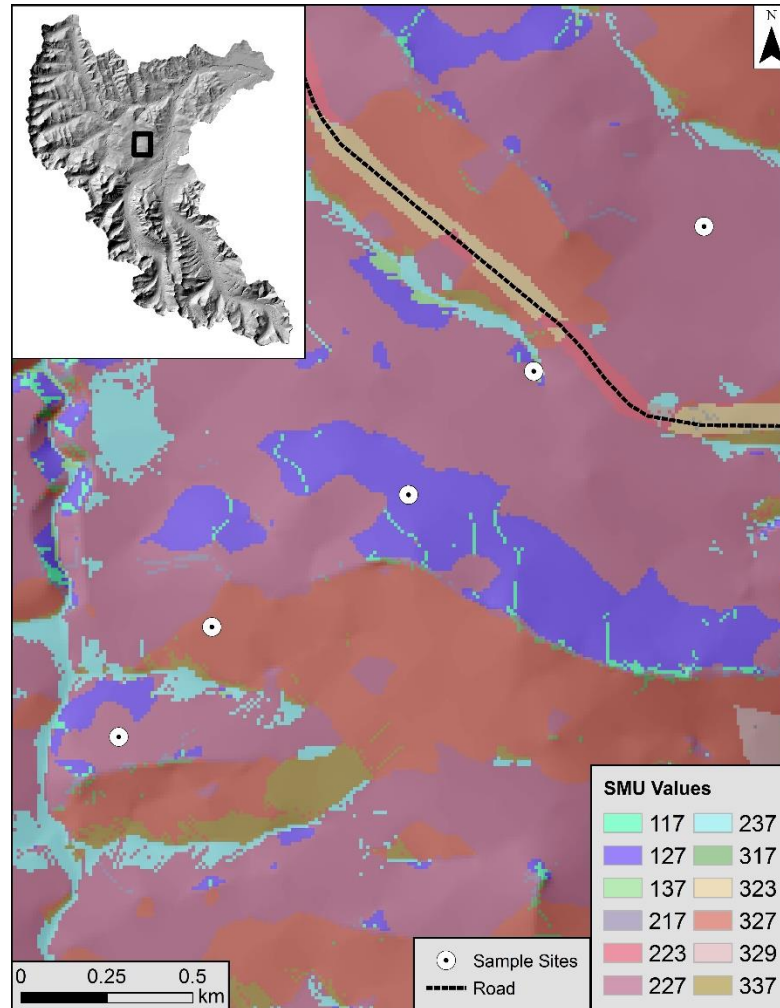
Figure 3.2: Soil sample locations in a transect across multiple SMU's. The inset map shows the location of the transect in the watershed.

An initial goal of the research was to collect 200 samples as proposed, based on the literature review of similar study area size (Afzali et al., 2015; Chai et al., 2008; Stoorvogel, Kempen, Heuvelink, & de Bruin, 2009; Zhang et al., 2012). However during field exploration and data collection it was realized that reaching the goal of 200 samples was not feasible within the research timeframe. Thus, much effort was put into strategic sampling by constructing transects and clusters of locations to gather as many samples as possible up to 200 samples.

3.2.2 Site Description

Post creation of sampling sites allowed for soil sampling site coordinates to be added to a GPS so that sites could be found when hiking to them. During site hikes, general site

characteristics were observed and noted if they were markedly important/different than the rest of the area (examples: drainage depressions/streams, bogs, evidence of fire). Upon arriving at the designated sites, specific site observations such as wind conditions, estimated cloud cover, evidence of rain, air and soil temperature (with thermometer), estimated slope steepness (degree), aspect (N/NE/SE/S/SW/NW), shape of slope (convex or concave), condition of the site floor (presence or absence of deadfall, plowed, grassland), and the dominant vegetation species at the site location were noted. These observations helped in inferring the soil formation characteristics by connecting the soil properties to environmental characteristics. For example, I would infer that a grassed meadow would have a well-developed A horizon due to the dry and sunny conditions preferred by grasses. Once the site was described (see example in Table 3.2), sample collection was conducted.

Table 3.2: Example of field note organization and targeted variables.

Soil ID/Personal	23718, TD, KB, MG
Date/Time	July 14, 2017 / 11:46 am
Weather Conditions	Warm, 0% cloud; Soil Temp: 11 ºC; Air Temp: 18 ºC
Vegetation	Thimble Berry, Western Meadow Rue, Engelmann and White Spruce; Shady location; ~30-40% canopy cover
Slope	Concave, flat (no aspect tilt)
Depth to Restrictive layer and Hits (H)	>=152.5cm & 55H; >=152.5cm & 50H
Litter Depths	Not gathered (needle residue and moss layer only)
BD Hits/Depths	8H; 12, 11.7,12.2, and 12.4cm
Horizonation	A – 6.4, 6.8, and 6.5 cm; B1 – 19, 21.7, and 24 cm; B2 – 21.2, 22.1, 16.3 cm; CB – 11.4, 9.5, 9.7; C – Rest to restrictive depth; Total depth of hole = 76.8 cm
Other	Discontinuity; Charcoal layer present between B1 and B2 6, 3, 2, and 3.4 cm width.

3.2.3 Soil Restrictive Depth Measurements

Soil restrictive depth was chosen because of it may be helpful in understanding water movement in the watershed and assessing potential vegetation change in future research. The restrictive depth was found by driving a 150 cm smooth, pointed steal rod into the ground.

Consistent drops of a 3 pound hammer were made to the top of the rod until the rod would stop

penetrating the soil (Figure 3.3), similar to methods described by Shanley, Hjerdt, McDonnell,

and Kendall (2003). The rod was driven into the ground three times in a triangular pattern around

the sample site to gain an average depth to restrictive layer. The strokes taken to hammer the rod

were recorded for quality control reference, wherein a substantially large or small amount of hits

observed in one location compared to another location would signify a discrepancy in striking

methodology and/or unrepresentative changes in the soil profile (due to large boulders or weak

points in the restrictive layer).

Figure 3.3: Photograph showing the knocking pole and hammer in use [summer 2018].

3.2.4 Soil Horizonation

In order to identify and analyze soil horizons a hole was dug approximately 20 cm wide and down to the top of the C horizon. Changes in colour, texture, and structure were examined to delineate soil horizons by placing a trowel and/or knife into the side of the hole, (Figure 3.4).

Figure 3.4. Soil holes with trowel and knife delineating the soil horizons (Left Site ID: 224; Right Site ID: WCS2).

Notes were taken on the relative changes of the aforementioned properties. Depths of the horizons were measured from the top of the mineral horizon to the bottom and top of each horizon and to the bottom of the hole, allowing for solum depth to be measured. Solum depth was deemed to be important for hydrological modelling. Conducting measurements in this manner allowed for a continuous depth profile to be gathered. The width of each horizon was determined back in the lab. With the horizons delineated and measured, samples of approximately 500 g were taken from each of the A and B horizons for OM and texture analysis in the lab. The holes were filled in according to the horizonation of the profile as precisely and carefully as possible to minimize disturbance and potential hazard to humans and/or animals that may come across the sample site.

3.3 Laboratory Analysis: Methods

3.3.1 Organic Matter by Ignition

Organic matter was chosen because of its connection to water holding capacity and support of vegetation health, important for hydrological and vegetation change research. Organic matter (OM) analysis of the A and B horizons was completed through the adaptation of loss on ignition (LOI) methods outlined by Karam (2008). Soils were oven dried at 105 $^{\circ}$C until the weight of the samples remained constant, signifying that all moisture was removed. Then, 2 g of soil were measured into a 30 milliliters (ml) crucible and placed into a muffle furnace. The furnace was programed to ramp to 370 $^{\circ}$C at 10 $^{\circ}$C per minute, then soaked for one hour to prevent the extreme temperature change that could create excessive volatilization of carbonates. Then the temperature was ramped to 600 $^{\circ}$C and soaked for 6 hours to burn the OM off. The furnace would shut off and allowed to cool for approximately one hour. Once samples were safe to handle, they were removed from the oven and placed into a desiccator to continue to cool for safe weighing.

Upon preliminary data collection and visualization, it was realized that the samples contain more than 2 % OM. Since, hydrometer method for textural analysis would require the removal of OM, therefore 45 grams (g) of soil was used instead of 2 g in the muffle furnace at the temperature and time intervals as explained above. Upon completion of the first 23 samples for textural analysis, it was found that the OM was successfully ignited out of the samples, despite the large amount of soil used. This was evident upon comparison to the 2 g samples (Table 3.3). The hydrometer readings dropped dramatically to the density of the control solution after 24 hours of settling, which confirmed that all the OM was removed from the samples. Therefore, it was decided to adapt the LOI method to ignite 45 g of soil, to remove the organic matter in the soils and to use them in the hydrometer method, thereby increasing efficiency across the LOI and hydrometer methods. The deviation of less than 3 % from the original 2 g method is seen as acceptable in order to gain efficiency of procedures.

Table 3.3: LOI data to analyze how the effectiveness of ignition with 45 g samples.

Sample ID	% OM in 2 g of soil (1)	% OM in 2 g of soil (2)	%OM in 45 g of soil
Sidewinder 1-A A	5.97	5.00	7.87
T&O #11 Bt1	7.00	5.50	6.09
Haig Lake B	14.43	15.00	14.30

(1) and (2) represent two separate ignition tests using subsamples of 2 g from the same site.

3.3.2 Soil Texture

Textural Analysis of A and B horizons was completed through adaptation of hydrometer methods discussed by Kroetch and Wang (2008), and Day (1965). This method was chosen because it is important for hydrological modeling and it can be used for vegetation and other environmental research. The first experiment conducted was without removal of OM. Six samples were oven dried, then 40 g of soil from each sample was measured into six 600 ml beakers. The samples were soaked overnight with 100 ml of sodium hexametaphosphate (SHMP) solution, where 100 ml was measured out of a full prepared solution of 1000 ml distilled water and 50 g of sodium hexametaphosphate ($(NaPO_3)_6$). Initial results suggest that OM did significantly affect the soil texture measurements (Day, 1965; Kroetch & Wang, 2008). For example, when the OM was left intact within the samples, the clay content were always overestimated, the sand content were mostly underestimated and silt content were mostly overestimated with some exceptions (Figure 3.5). The data also support these conclusions. For example, the % sand, silt, and clay were 20.87, 42.21, and 36.93, respectively with OM in the BZ08 A sample, and 80.22, 19.78, and 0 without the OM. Also, the % sand, silt, and clay were 57.23, 37.74, and 5.03 with OM and 57.23, 41.51, and 1.26 without OM in Start Point Pit A sample. As well, the percent sand, silt, and clay were 22.52, 62.48 and 15, respectively with OM in the AUG33 Ah sample, and 63.67, 30.69 and 5.63 without the OM.
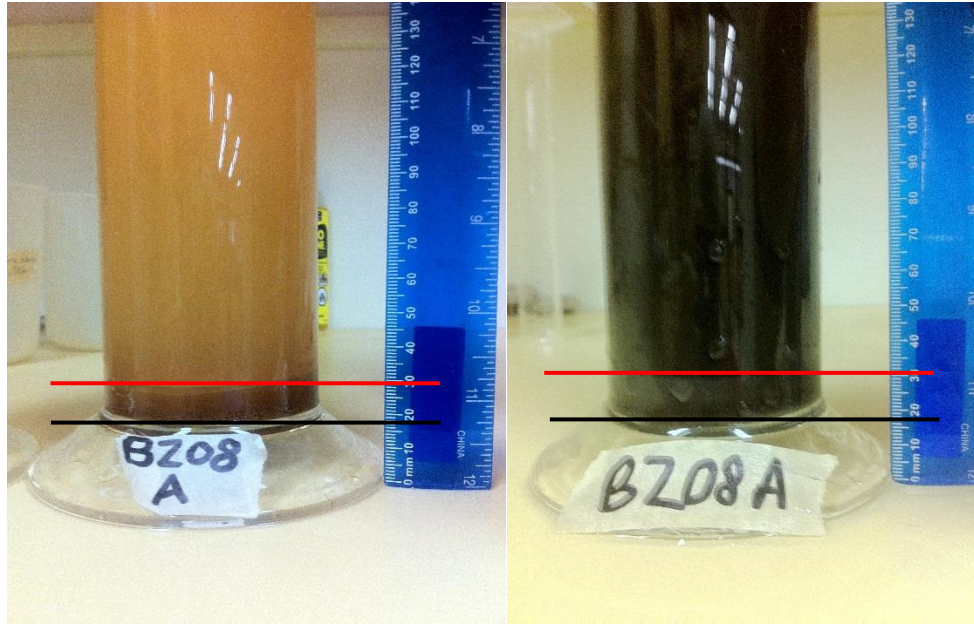
Figure 3.5: Samples of sediment after 7 hours of settling. Left photo of BZ08 A horizon without organic matter, and the right photo is of BZ08 A horizon with organic matter.

Based on these findings, it was decided to burn the OM out of the samples before conducting the hydrometer analysis. To burn off the OM, 45 g of soil was oven dried at 105°C for ~10 hours, then the samples were weighed and placed into a muffle furnace at 370°C for 1 hour, then at 600°C for 6 hours. With the OM burned off, the remaining mineral portion of the soil were transferred into 600 ml beakers and soaked with 100 ml of SHMP solution overnight. Twelve samples, with OM removed, were prepared this way and used in the hydrometer method with 1000 ml of distilled water. Using an ASTM 152H hydrometer, three 40 second readings were taken along with one reading taken at 3, 10, 30, 90, 420, and 1440 minute interval. These readings were plotted in Excel on a logarithmic graph with a logarithmic trend line and its equation to visualize the results. The sand silt and clay were calculated following the methods of Kroetch and Wang (2008). Sand was calculated using Equation 2, clay using Equation 3, and silt using Equation 4.

$$100 - (100/ \text{ ignited soil dry weight}) \text{ x corrected reading at 40 seconds} \qquad \text{Equation 2}$$

$$(100/\text{ignited soil dry weight}) \text{ x corrected reading at 7 hours} \qquad \text{Equation 3}$$

$$100 - (\text{Silt summation} + \text{Clay summation}) \qquad \text{Equation 4}$$

Near complete clay settling is seen in the small difference of settling between the 7 hour and 24 hour readings. Kroetch and Wang (2008), and Day (1965) describe that the 40 second and 7 hour readings can be used without the readings in between or beyond to obtain accurate particle size calculations. It was decided that taking readings at 40 seconds and 7 hours would be the most efficient times to take readings (Table 3.4).

Table 3.4: Hydrometer test without organic matter using 40-second and 7-hour readings.

		Elapsed Time (minutes)	Corrected Reading (g/mL)	Percent Summation	Diameter (mm)
Sample ID	Sidewinder 1-A A HRZ	0.667	8	21.89	0.0593
Date	Dec. 11, 2017	0.667	10	27.36	0.0586
Tested By	Trevor Deering	0.667	11	30.10	0.0583
Hydrometer # (if applicable)	1	Avg. 40 sec.	9.7	26.54	0.0588
Specific Gravity of Solids	2.65	3	9	24.62	0.0278
Dispersing Agent	100 ml HMP	10	6.5	17.78	0.0154
Weight of Soil Sample	41.55 g	30	4	10.94	0.0090
Classification	Loamy Sand	90	3	8.21	0.0053
		420	2.5	6.84	0.0025
		1440	0	0	0.0013

3.4 Lab Methodological Validation

One of my moral standings is that the applied scientific methods must be conducted in a precise and transparent manner. Holding true to this value and recognizing that the LOI and Hydrometer lab methods were extremely important analysis for this research, it was decided that these methods needed to be conducted as precisely as possible. Therefore, the LOI and hydrometer lab methods were investigated as to their precision through testing six samples by Down to Earth Labs (Lethbridge AB), Exova Labs (Edmonton AB), and by me (Soil lab, University of Lethbridge). The hydrometer method of leaving the OM in the soil was used by all three labs. The LOI method was used by all three labs to determine percent organic matter in the

samples, with slightly different adaptations of temperature and weight. Down to Earth Labs ignited their 2 gram soil samples at 370°C for 16 hours, Exova also ignited 2 g at 500°C for 2.5 hours, and I ignited ~45 g at 370°C for 1 hour followed by 600°C for 6 hours. The sand, silt, clay, and percent organic matter results were investigated through analysis of variance (ANOVA) statistical testing using CoStat (CoHort Software, 2018). First, the data were investigated for normality to ensure that the data were normally distributed for ANOVA analysis. The data were found to be close to normal (Table 3.5). Data is normally distributed when (1) the mean and median are similar, (2) the skewness is near zero, and (3) the kurtosis is near zero (Ghasemi & Zahediasl, 2012).

Table 3.5: Normality measures of data for 6 samples tested by three laboratories.

Variable	Skewness	Skew Significance	Kurtosis	Mean	Median
Sand (S)	0.44	1.15	-1.39	43.80	41.80
Silt (Si)	-0.27	1.15	-1.38	36.05	36.42
Clay (cl)	0.90	1.15	-0.77	20.06	12.73
Organic Matter	0.44	1.15	-0.92	7.8	7.36

Skew significance: If skewness is >2*SQRT (6/count) = significantly skewed (Verschuuren, 2013, November 26).

The skew and kurtosis values for all the variables (S, Si, Cl, and OM) are all within the -2 to 2 range using a significance of 0.05 (Ghasemi & Zahediasl, 2012). The ANOVA results, show no significant difference across all variables (See Appendix A, tables A.1-A.4). A larger set of observations would result in a stronger confidence of the distribution, however, testing similar methods across the same samples does result in a valid comparison. These results show that the significance values are all greater than 0.05 and thus are deemed not significant. Finding that my results are close to the other labs gives me confidence that I have conducted laboratory methods properly and that my analytical results are accurate and precise. Moving on from here, removal of the organic matter through LOI can also be confidently relied upon, to use the mineral portion only in the hydrometer method.

3.5 Data Analysis and Methods

3.5.1 Stepwise Linear Regression

Stepwise linear regression (SLR) was the statistical method chosen to analyze relationships between soil properties (dependent variables) and environmental characteristics (independent variables) using IBM SPSS 24.0 (Statistical Package for the Social Sciences, 2017).

Independent variables were chosen based on literature and possible influential contributions to soil property formation (Table 3.6). For example, based on a literature review it was recognized that geology and vegetation type have been rarely included in regression analysis, but because of their potential influence in soil depth and soil texture, these independent variables were added to the regression.

Table 3.6: Dependent and independent variable list with accompanying data type.

Dependent Variables	Data type	Independent variables	Data Type
Restrictive Depth	Continuous	Topographic Wetness Index	Continuous
Mineral BD	Continuous	Slope (%)	Continuous
Litter BD	Continuous	Elevation (m)	Continuous
%S, %Si, %Cl	Continuous	Solar Radiation (W/m^2)	Continuous
A and B horizon depths	Continuous	Aspect (degrees)	Continuous
		Profile Curvature	Continuous
		Plan Curvature	Continuous
		Curvature	Continuous
		Specific Catchment Area	Continuous
		Contributing Area	Continuous
		Landform Position	Categorical
		Slope Position	Categorical
		Surficial Geology	Categorical
		Bedrock Geology	Categorical
		Land Cover Type	Categorical
		Vegetation Density	Categorical
		Vegetation Type	Categorical
		Vegetation Height	Categorical

Before conducting SLR, the data were split into training (75%) and validation (25%) sets. The SLR models were created with the training datasets and tested with the validation datasets. Upon conducting SLR, the most influential independent variables were chosen to create the final regression equations for each dependent variable. The variables used in or taken out of the regression equations were controlled by manipulating the significance thresholds in the SPSS dialog.[1] At each step a variable with the smallest p-value less than the specified threshold of Entry is entered into the model (Significance IN). Also, at each step a variable may be removed if the p-value of the variable becomes greater than the specified threshold of Removal (Significance OUT), according to it level of affect as each step progresses. Nine steps were applied through the implementation of SLR to ensure that data structure and assumptions were proper for SLR analysis (Lund Research, 2018).

1.   A database was established in individual Excel sheets to associate each dependent variable with each independent variable. This ensured ease of use in R programing and SPSS statistics, along with matching of environmental variables to each site where they were measured.

2.   Dependent variables were expressed as continuous numeric units. For example, organic matter content was expressed in percent (between 0 and 100 percent).

3.   Independent variables were expressed as either continuous (e.g. slope, elevation, wetness index) or categorical data (e.g. geological unit, land cover class, slope position). Dummy variables were created for each categorical variable, which resulted in as many Dummy variables as there were categories. For example, as there were 10 land cover classes, 10 Dummy variables were created representing either the presence or absence of that particular land cover class (e.g. coniferous forest) (Suits, 1957).

---

[1] Multiple significance thresholds were ran for all soil property variables, using all independent variables, to find the optimal regression models while implementing all nine steps.

4.   Independence of both response and explanatory variables was assumed to be met because the samples were taken only once, many observations were not taken over time.

5.   Linearity was assumed for the dependent variable with each independent variable, even where small coefficient of determination (R^2) values existed, after testing multiple other trends and finding there to be little difference in R^2. Also, linearity was tested by analyzing the pattern of scatter plots of the dependent variable after computing regressions (the standardized residual value on the y axis and standardized predicted value on the x axis). Inspection for non-curving data to meet the linearity assumption of residual plots.

6.   Homoscedasticity was sought and investigated through visual inspection of scatter plots by adding a best fit line to the residual scatter plot (standardized predicted value – x-axis; standardized residual – y-axis), seeking a flat line, with a low R^2, and no errors that get larger in one direction. If residuals showed heteroscedasticity it was recognized that the model may introduce uncertainties when predicting soil properties using validation data.

7.   Multicollinearity of data was avoided by using the Variance Inflation Factor (VIF) collinearity diagnostics where the VIF should be between 1 and 10. Also, the Pearson correlation matrix was investigated to ensure variables had correlations less than 0.80. If these values are not met the variables do present collinearity and correlation and were removed.

8.   Normality of the residuals between the dependent variables and the independent variables were investigated through interpretation of histograms and Probability-Probability plots (PP-plots) of model residuals.

9.   SLR was conducted using SPSS and the major contributing environmental variables were identified to build final regression models that were used to model dependent variables of the watershed.

3.5.2 Spatial Modelling of Soil Properties

Once regression modelling, with training datasets, was completed, performances were compared between two of the best models and a final regression model was chosen. The equations were then used in ArcMap to predict soil properties across the entire Westcastle watershed because the independent variables were available for the whole watershed. The equation to predict each soil property was written into the Raster Calculator tool using the con (condition) command and the independent variable rasters. Once the soil properties were predicted, validation analysis was conducted, visually and statistically. The performance of two of the best regression models was compared, both visually and statistically. RMSE (root mean squared error), MAE (mean absolute error), r-square (R^2), and average standard error are used to examine the performance of regression models (Equations 5, 6, 7 and 8) (Patil & Singh, 2016).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(E_i - M_i)^2}{n}} \qquad \text{Equation 5}$$

$$\text{MAE} = \sum_{i=1}^{n}\frac{|E_i - M_i|}{n} \qquad \text{Equation 6}$$

$$R^2 = \left(\frac{1}{n-1}\sum_{i=1}^{n}\frac{(M_i - \bar{M})(E_i - \bar{E})}{S_M S_E}\right)^2 \qquad \text{Equation 7}$$

$$\text{Avg. Std. Error} = \frac{\sum SE}{n} \qquad \text{Equation 8}$$

where E_i represents the computed values, M_i the measured values and n the number of data used in the model (Equation 5). Also, where $\bar{M}$ represents the mean of the measured values, and $\bar{E}$ represents the mean of the computed values. And S_M represents the sum of the measured values and S_E represents the sum of the computed values.

Chapter 4: Results

4.1 Introduction

The presentation of lab results, stepwise regression, and mapping analysis of the distribution of restrictive soil depth, soil texture, and percent organic matter are provided in this chapter. Lab analysis results include all measurements taken to compile the soil depth and soil texture data. Stepwise linear regression using a wide range of independent variables was implemented to arrive at optimal regression equations for computing each dependent variable. This formed the basis to map the distribution of the dependent soil variables at the watershed scale. Lastly, modelled soil properties are compared to SOILGRIDS and detailed soil survey data.

4.2 Sampling and Lab results

Overall, 131 soil sample sites were visited throughout the watershed (Figure 4.1). Results reveal that measurements of restrictive soil depth were taken at 91 % of sites and that A horizon texture samples were taken at 83 % of sites. Texture samples of the B horizons were taken at 53 % of sampling sites. The restrictive soil depth measurements were missed at some sites, due to oversight of field procedures or it was not used at C horizon or road cut sites, whereas horizon textures were dependent on the presence of the horizon. Descriptive statistics of field and laboratory results (Table 4.1) are important for addressing objective 1 and working through objective 2. Mean restrictive soil depth of 74 cm was measured, which is slightly deeper than pre-field work expectations. Soil texture analyses revealed a high sand content in both the A and B horizons, as is clearly evident in the soil textural triangle (Figure 4.2), where the samples are greatly clustered in the sand, loamy sand, and sandy loam textural classes. Organic matter content was measured in each of the texture samples and, as expected, with ~3 % higher organic matter in the A horizon versus the B horizon was observed (Table 4.1). The depth of soil horizons and the solum depth (A plus B horizon) were measured at each of the 131 sites. Descriptive statistics (Table 4.1) show that 118 depth measurements were gathered, with a mean soil restrictive depth

of 74 cm, a maximum soil restrictive depth of 182 cm, and a minimum depth of 3 cm. In order to analyze the relationships between the dependent variables (soil properties) and the independent variables (environmental properties), frequency histograms of the dependent variables, scatter plots with linear trend lines and their coefficients of determination, and boxplots were made. Plots of restrictive soil depth and a histogram of solum depth (Fig. 4.3 - 4.5) are shown as examples (see Fig. B.1 - B.21 in Appendix B for other soil property plots).
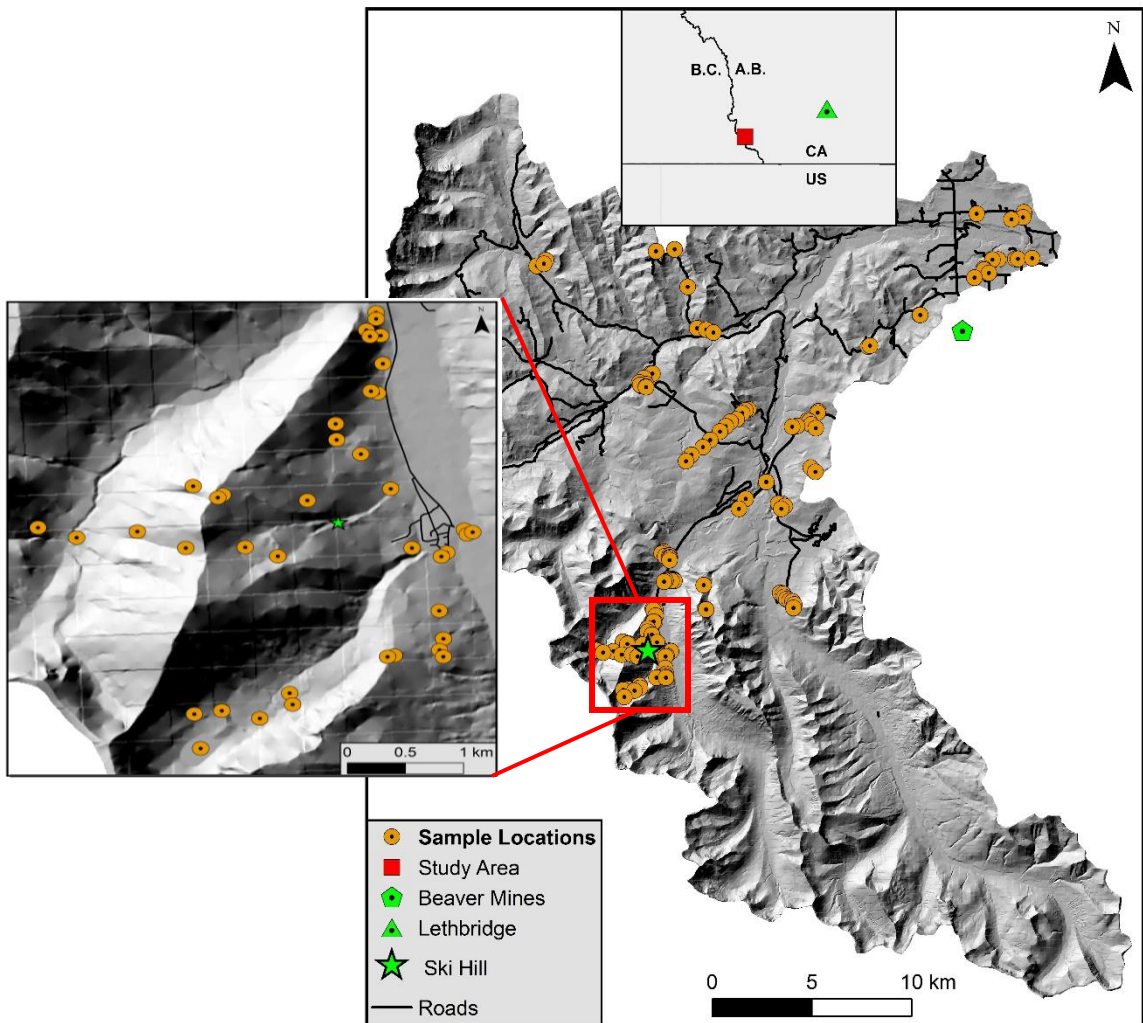
Figure 4.1: Soil sample site locations.

Table 4.1: Descriptive statistics of field and laboratory property measurements.

Variable	N	Mean	SD	Min	Max	25th Percentile	50th Percentile	75th percentile
Soil restrictive depth (m)	118	74.00	41.25	3.0	182	41.9	63.4	104.4
A HRZ %Sand	104	67.3	14.8	31.5	97.6	56.0	67.6	77.4
B HRZ %Sand	69	68.6	16.0	28.9	97.1	58.8	68.1	81.2
A HRZ %Silt	104	28.2	13.4	0.0	62.3	20.1	26.6	38.2
B HRZ %Silt	69	26.5	14.5	0.0	66.2	16.2	26.3	37.1
A HRZ %Clay	104	4.4	3.2	0.0	17.8	2.3	3.7	6.0
B HRZ %Clay	69	4.8	3.7	0.0	22.8	2.3	4.7	5.9
A HRZ %OM	104	9.3	5.5	2.4	47.0	5.9	8.2	11.3
B HRZ %OM	69	6.1	3.6	2.1	23.3	3.7	5.9	6.8
Solum Depth (cm)	131	36.1	29.9	0.0	115.6	10.5	30.0	54.9
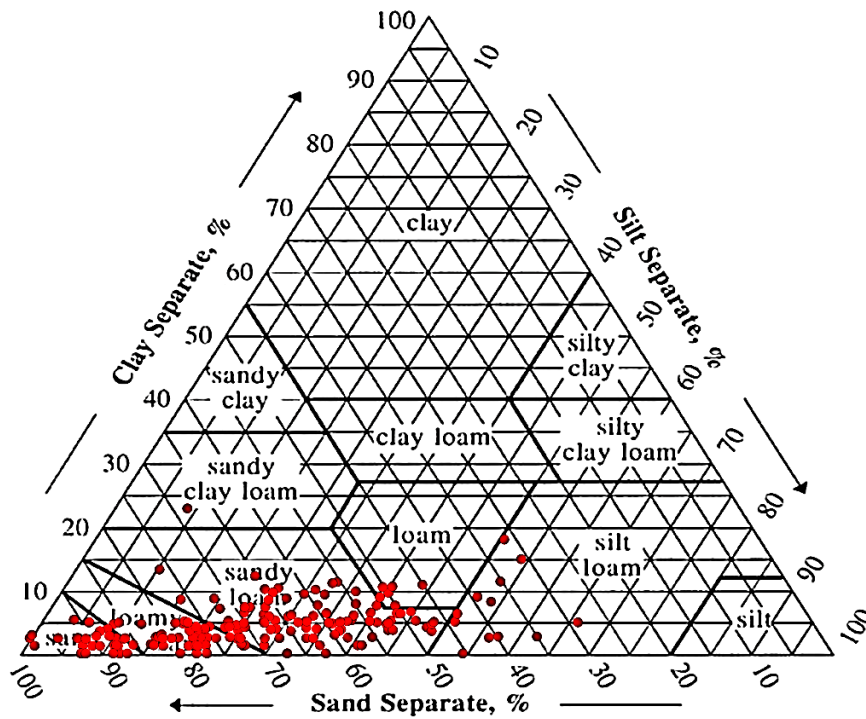
Figure 4.2: Soil textural triangle of A and B horizon texture classes (Natural Resources Conservation Service, 2018).
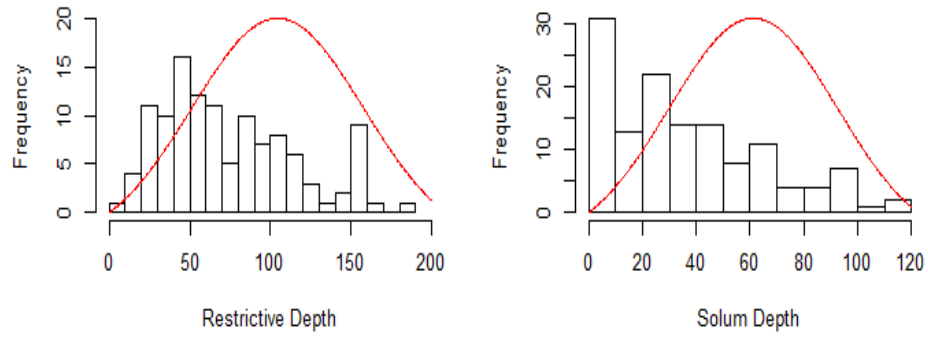
Figure 4.3: Histogram distribution of soil restrictive depth and solum depth (A+B horizon width).
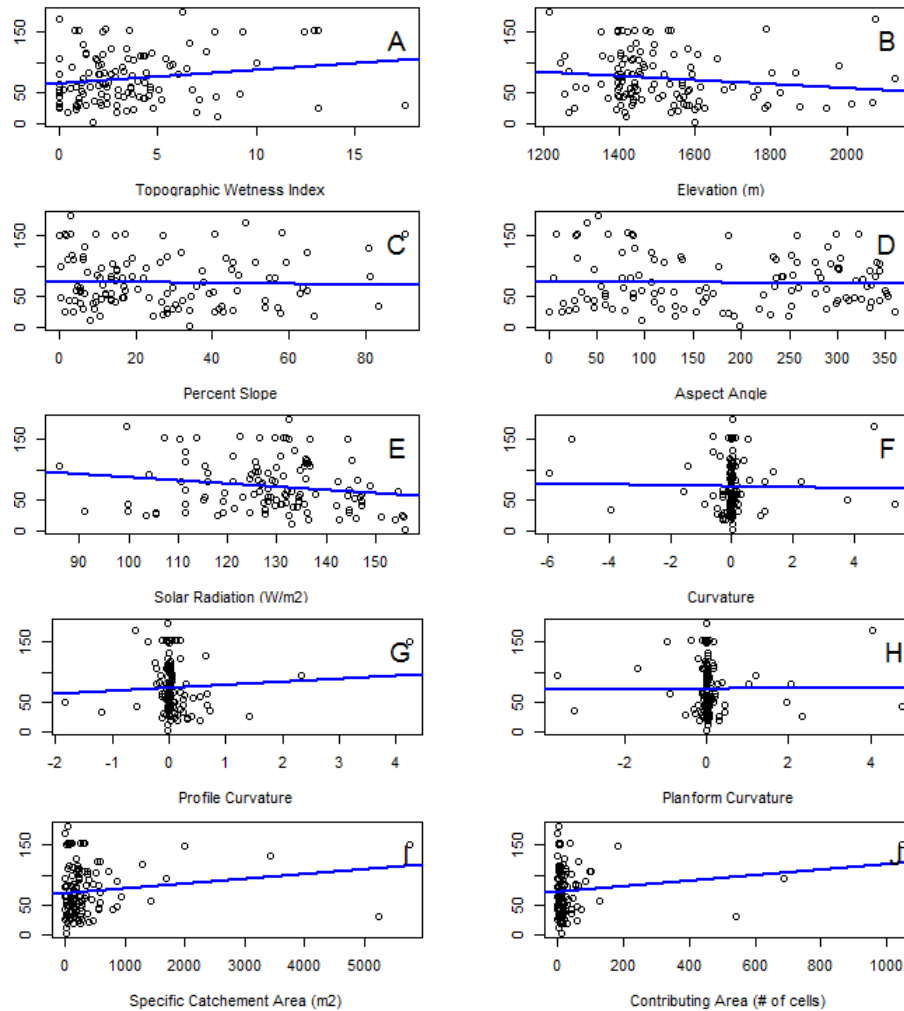
Figure 4.4: Scatter plots of soil restrictive depth (on the y-axis) and all the continuous environmental variables on (the x-axis). TWI (R^2 =0.0277), Elevation (R^2 =0.0190), Slope (R^2 =0.00077), Aspect (R^2 =0.0017), Solar Radiation (R^2 =0.0288), Curvature (R^2 =0.00052), Profile Curvature (R^2 =0.00427), Planform Curvature (R^2 =9.96e-5), Specific Catchment Area (R^2 =0.0236), and Contributing Area (R^2 =0.0198) m^2.
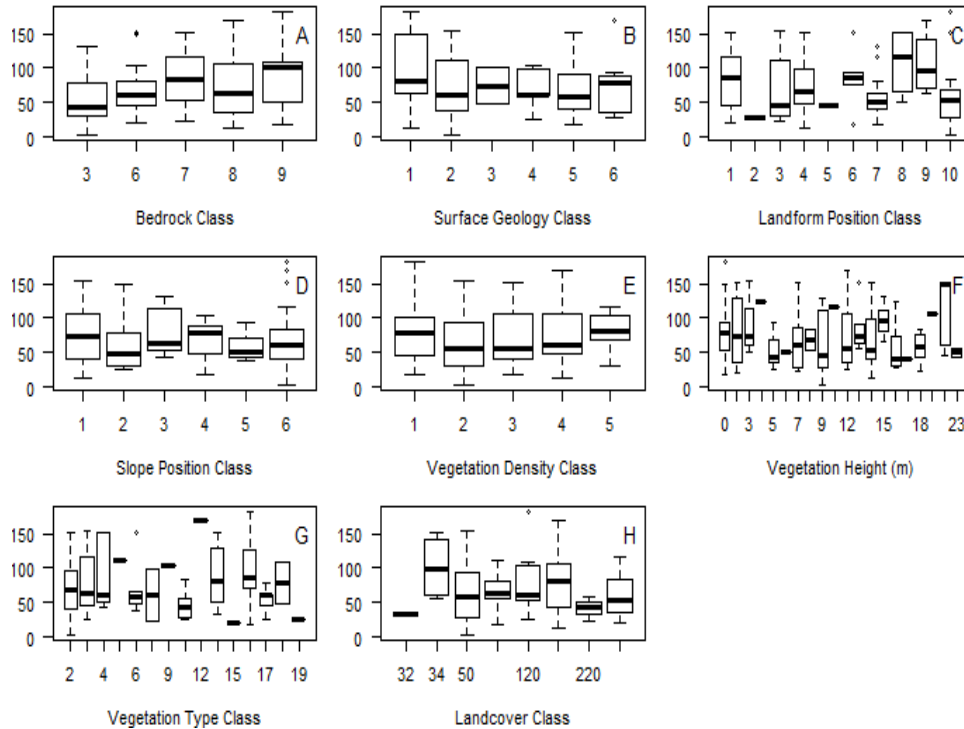
Figure 4.5. Boxplots of soil restrictive depth (cm) on the y-axis and the other variables on the x-axis.

4.3 Spatial Distribution of the Soil Properties

4.3.1 Soil Restrictive Depth Relationships with Independent Variables

Soil depth was measured in centimeters from the top of the soil to the maximum penetrated depth using the knocking pole method. For the stepwise regression analysis, the restrictive soil depth was the dependent variable, and multiple environmental variables were the independent variables (Table B.1, Appendix B). All linear relationships between the soil depths and independent variables were very weak. Figure 4.4 shows the two strongest linear relationships of soil restrictive depth being solar radiation ($R^2 = 0.029$) and TWI ($R^2 = 0.028$). The third most influential linear relationship is specific catchment area ($R^2 = 0.024$), followed by contributing area ($R^2 = 0.020$) and elevation ($R^2 = 0.019$). The other continuous variables show even lower R^2 values. By implementing one-hot coding of categorical variables to "dummy variables", larger R^2 relationships were found. For example, the independent variables "Surface Geology type 3" ($R^2 = 0.051$), "Vegetation type 12" ($R^2 = 0.046$), "Vegetation Height 21 m" ($R^2 = 0.038$), and

"Vegetation type 10" ($R^2 = 0.037$) had higher R^2 values than the highest R^2 relationship of all other terrain variables. These relationships are considered when analyzing regression models and for logical relationships.

4.3.2 Spatial Autocorrelation Analysis of Soil Properties

4.3.2.1 Restrictive Soil Depth

Spatial autocorrelation was investigated to assess its helpfulness in understanding soil property spatial variation and for its potential for use in future research considerations. Spatial autocorrelation of soil restrictive depth samples was analyzed by using the Spatial Autocorrelation (Moran's I) tool in ArcMap and the GS+ program to analyze semi-variance of soil properties. The Moran's I report indicates the soil restrictive depth measurements are spatially random (Table 4.2), based on a z-score of 0.275, therefore measurements differ from each other with no clustering due to intrinsic environmental aspects.

Table 4.2. Moran's I summary report of the restrictive soil depth.

Moran's Index:	**0.092**
Expected Index:	-0.007
Variance:	0.130
z-score:	0.275
p-value:	0.783
Distance Threshold	1989.512 meters

A variogram was constructed with GS+ with non-transformed data (Figure 4.6), resulting in a spherical model, the optimized model. A nugget of 75.0 and a sill of 1656 m were auto-calculated and the resulting spherical model was fitted with an R^2 of 0.080. The large nugget value indicates that measurements taken close together exhibit large variation from each other. So the measurements were not affected by sampling sites that were close together. Also, the small R^2 value means the variation in the data is large. Therefore, independence of samples is achieved for regression analysis. Patterns that we find in conducting step-wise regression can be attributed to the interaction of environmental variables and results will not be biased by counting correlated sites more than once.
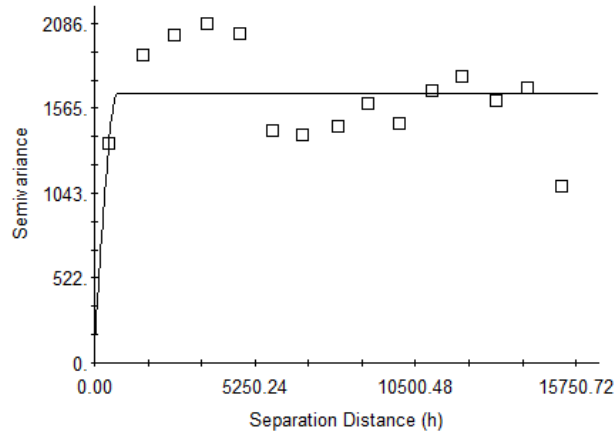
Figure 4.6: Spherical variogram model of soil restrictive depth.

4.3.2.2 Soil Textures, OM, and Solum Depth

The Moran's I spatial autocorrelation statistic was calculated for the A and B horizon textures (Table 4.3) and assessed for spatial clustering of the samples that would imply significance of spatial patterning of the soil properties.

Table 4.3: Moran's I spatial autocorrelation report summaries for soil pit A and B horizon samples.

Property Type	Moran's I Index	Variance	z-score	p-value	Dispersed/Random/Clustered
A HRZ % Sand	0.217	0.006	2.876	0.004*	Clustered
A HRZ % Silt	0.162	0.006	2.181	0.029^	Clustered
A HRZ % Clay	0.175	0.006	2.376	0.018^	Clustered
A HRZ % OM	0.005	0.005	0.206	0.837^	Random
B HRZ % Sand	0.267	0.009	2.906	0.004*	Clustered
B HRZ % Silt	0.236	0.009	2.592	0.010*	Clustered
B HRZ % Clay	0.100	0.008	1.237	0.212^	Random
B HRZ % OM	0.165	0.008	1.976	0.048^	Clustered
Solum Depth (cm)	1.288	0.142	3.435	0.00059*	Clustered

Note: A horizon sand and B horizon sand and silt are clustered with less than a 1% likelihood* of being caused by random chance, whereas A horizon silt and clay and B horizon organic matter are clustered with less than a 5% likelihood^ of being caused by random chance.

All the A horizon percent texture measurements are patterned in a spatially clustered manner, whereas A horizon organic matter is spatially random (Table 4.3). On the other hand, B horizon percent sand, silt and organic matter are patterned in a spatially clustered manner with percent clay being spatially random (Table 4.3). With these results the decision was made to assess spatial autocorrelation with variogram analysis, done in the same manner as soil restrictive depth, because the positive values and clustering suggest that samples are similar in distinct areas of the study area. Table 4.4 and Figures 4.7 – 4.11 show results of the variogram calculations and plotting using the GS+ program. Each variable was determined to have low to moderate spatial dependency. The moderate spatial dependency variables area promising for use in geostatistical methods. Geostatistical methods are not suitable for the low spatially dependent properties. Therefore, it was decided to use only stepwise linear regression for all the variables to simplify statistical analysis and to meet time constraints of this research project.

Table 4.4: Results of variogram calculations to assess spatial dependence of texture and organic matter data.

Property Type	Model Type	R^2	Nugget Value (C_o)	Sill (C_o+C)	Range (m)	Proportion $C/(C_o+C)$ (%)	Spatial Dependency Level
A HRZ % Sand	Spherical	0.14	6.7	212.0	890	96.84	Low
A HRZ % Silt	Spherical	0.13	8.5	179.2	830	95.26	Low
A HRZ % Clay	Exponential	0.09	1.8	11.2	510	83.87	Low
A HRZ % OM	Gaussian	0.55	19.9	53.1	11,080	62.54	Moderate
B HRZ % Sand	Gaussian	0.14	141.3	282.7	3,100	50.02	Moderate
B HRZ % Silt	Gaussian	0.12	114.8	229.7	3,020	50.02	Moderate
B HRZ % Clay	Exponential	0.02	2.0	14.0	370	85.46	Low
B HRZ % OM	Spherical	0.17	1.4	14.8	2,670	90.52	Low
Solum Depth (cm)	Exponential	0.68	503	1261	17,710	60.11	Moderate
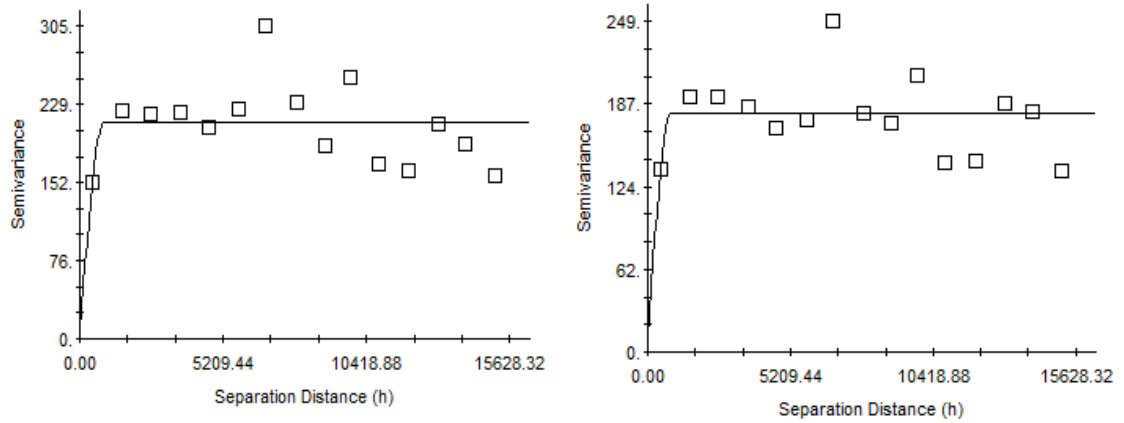
Figure 4.7: Spherical variograms of A horizon percent sand (left) and percent silt (right).
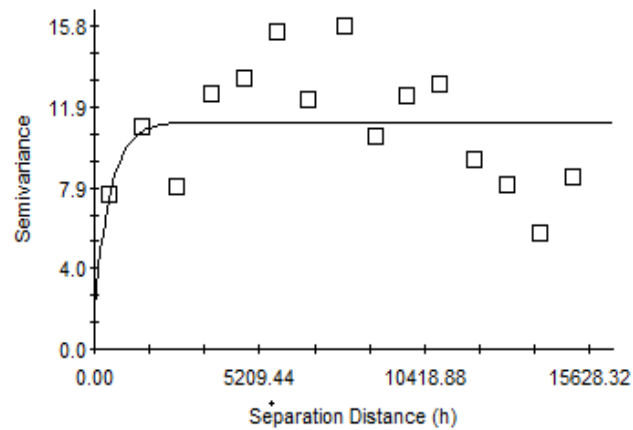
Figure 4.8: Exponential variogram of A horizon percent clay.
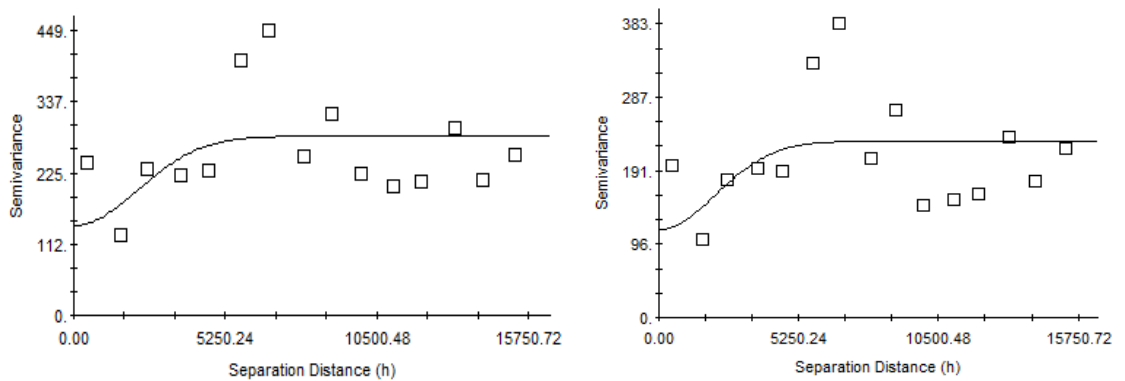
Figure 4.9: Gaussian variograms of B horizon percent sand (left) and percent silt (right).
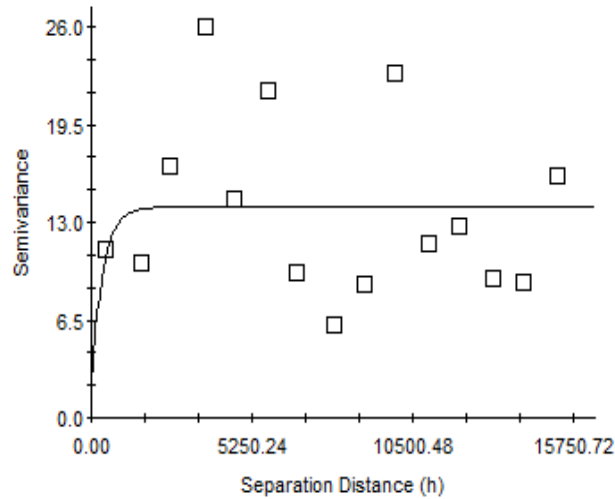
Figure 4.10: Exponential variogram of B horizon percent clay.
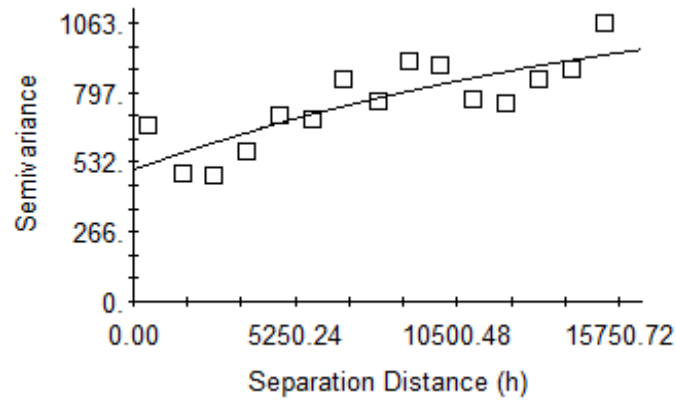
Figure 4.11: Exponential variogram of solum depth (cm).

4.4 Model Analysis

4.4.1 Soil Restrictive Depth Stepwise Regression

With relationships explored through descriptive statistics, stepwise linear regression was employed. After it was determined that there are some variables that had skewed distributions, it was decided that the box cox transformation function in R would be performed on the data (Equation 5) and used in stepwise regression, with the SPSS program, to explore model performance with transformed data.

$$t(x) = \frac{(x^\lambda - 1)}{\lambda}$$
(Equation 5)

where t(x) represents the transformed variable value, x represents the original variable value, and λ represents the transformation parameter (Tesfa, Tarboton, Chandler, & McNamara, 2009). Regressions with non-transformed data and variations of non-transformed and transformed data were compared based on their R^2 performance. Using the Shapiro Wilk Normality Test, variables were transformed and tested for normality. It was found that only soil restrictive depth, slope, and specific catchment area variables became normal upon transformation. Due to these results many regression models were performed whereby both transformed and non-transformed data variations were used. It was found that not transforming the data resulted in the highest R^2 value of 0.856 (Table 4.5). A PP-plot of the modelled residuals revealed that the estimated soil depth values were normally distributed. Therefore, data input into the stepwise regression model were not transformed for the soil restrictive depth variable modelling, neither for modelling the remaining soil properties.

Categorical variables were tested against continuous terrain variables to assess their influence in predicting soil restrictive depth through inspection of R^2 values (Table 4.6). Continuous terrain variables consisted of elevation, slope, aspect, TWI, plan curvature, profile curvature, curvature, solar radiation, specific catchment area, and contributing area. It was found that vegetation height, vegetation density, and vegetation type had the largest R^2 value (0.601) when used with terrain variables. Of the other variables used by their own, slope and landform position, and then land cover offered the next higher R^2 values (0.267 and 0.223). Slope and landform position and land cover type were used separately with terrain variables and vegetation variables. Slope and landform position improved the R^2 value (0.267) by 0.061 over land cover (0.223). Slope and land cover position improved the R^2 value of terrain variables and vegetation variable combination by 0.610. Therefore, it was recognized that some variables are more influential than others, namely that terrain variables have a very small influence when compared to vegetation variables.

Table 4.5: R squared training results for non-transformed and transformed data. Significance level in = 0.5 and significance level out = 0.99.

Variables Used	R Squared Values	Adj. R^2
All Variables Non-Transformed	0.856	0.664
All Continuous Variables Transformed	0.846	0.657
Max Depth Non-Transformed / All Continuous Independent Variables Transformed	0.849	0.657
Max Depth Non-Transformed / Slope and Specific Catchment Area Transformed Only	0.835	0.690
Max Depth Transformed / Slope and Specific Catchment Area Transformed Only	0.843	0.657

Table 4.6: Comparison of training regression equations comparing terrain variables (TWI, elevation, % slope, solar radiation, aspect) with other categorical variables individually and combined. Significance level in = 0.5 and significance level out = 0.99.

Variables in Regression	R Squared Values	Adj. R^2
Terrain Variables Alone	0.118	0.05
Vegetation Variables Alone	0.568	0.358
Terrain Variables + Surface and Bedrock Geology	0.194	0.092
Terrain Variables + Landcover Type	0.223	0.125
Terrain Variables + Landform and Slope Positions	0.267	0.145
Terrain Variables + Vegetation Density, Height, & Type	0.601	0.449
Terrain Variables + Landcover Type and Vegetation Density, Height, & Type	0.710	0.549
Terrain Variables + Landform & Slope positions		
Landform & Slope positions and Vegetation Density, Height, & Type	0.740	0.554

A final regression equation was chosen for mapping the entire watershed on the basis of maximizing the R^2 performance, minimizing the adjusted R^2, and maintaining normal distribution of the predicted values (visual P-P plot analysis of Figure 4.12 and 4.13). Also, steps 5, 6, 7, 8 and 9 from the methods were also followed. Two regression models were chosen that met the steps, one with 17 variables (0.10 Sig. IN and 0.99 Sig. OUT) and one with 40 variables (0.50 Sig. IN and 0.51 Sig. OUT) (Table 4.7). Reviewing the two regression models to meet step 7, a weak correlation (0.930) was found between vegetation height 1 (VIF = 10.095) and vegetation type 14

(VIF = 8.806) for the 40 variable regression model, but this was deemed acceptable. Tables 4.8

and 4.9 show the model summaries produced with SPSS as illustrations of the statistics performed

and analyzed for soil restrictive depth and all other soil properties. Predictions based on validation

data were made using both regression models, and the validation statistics were used with training

data statistics to select the final regression model (Table 4.10).

Table 4.7: Systematic increase of significance inclusion of stepwise regression to investigate model performance with increasing training variables. All variables were not transformed.

Sig. IN	Sig. OUT	R^2	Adj. R^2
0.05	0.1	0.506	0.450
0.05	0.2	0.506	0.450
0.05	0.5	0.506	0.450
0.05	0.99	0.508	0.446
0.1	0.2	0.544	0.480
0.1	0.5	0.544	0.480
0.1	**0.99**	**0.621**	**0.542**
0.2	0.99	0.729	0.626
0.3	0.99	0.788	0.665
0.4	0.99	0.830	0.676
0.5	0.99	0.856	0.664
0.5	**0.51**	**0.825**	**0.704**
0.6	0.99	0.862	0.645
0.7	0.99	0.864	0.629
0.8	0.99	0.865	0.598
0.9	0.99	0.865	0.544
No Restriction		0.865	0.511

Note: Bold values represent the chosen regression models. Also, Sig. IN = the p value cut off threshold for the input of independent values into the regression model and Sig. OUT = the p-value cut off threshold for the independent values to be taken out of the model.

Table 4.8: Model summary of 17 variable soil restrictive depth regression model.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.231[a]	.053	.044	40.769	.053	5.478	1	97	.021	
2	.329[b]	.108	.090	39.773	.055	5.921	1	96	.017	
3	.409[c]	.167	.141	38.644	.059	6.689	1	95	.011	
4	.459[d]	.211	.177	37.822	.043	5.174	1	94	.025	
5	.508	.258	.218	36.859	.048	5.977	1	93	.016	
6	.540	.292	.246	36.202	.034	4.407	1	92	.039	
7	.590	.348	.298	34.941	.056	7.760	1	91	.007	
8	.630	.397	.343	33.789	.049	7.309	1	90	.008	
9	.663	.440	.384	32.731	.043	6.914	1	89	.010	
10	.693	.480	.421	31.714	.040	6.799	1	88	.011	
11	.713	.508	.446	31.024	.028	4.956	1	87	.029	
12	.727	.528	.462	30.581	.019	3.544	1	86	.063	
13	.739	.546	.476	30.174	.018	3.332	1	85	.071	
14	.760	.577	.507	29.280	.032	6.272	1	84	.014	
15	.771	.594	.521	28.867	.017	3.418	1	83	.068	
16	.780	.608	.532	28.531	.014	2.966	1	82	.089	
17	.788[q]	.621	.542	28.220	.013	2.817	1	81	.097	2.016

a. Predictors: (Constant), VegetationTypeNumber=12.0

b. Predictors: (Constant), VegetationTypeNumber=12.0, Landform position=8.0

c. Predictors: (Constant), VegetationTypeNumber=12.0, Landform position=8.0, VegetationTypeNumber=10.0

d. Predictors: (Constant), VegetationTypeNumber=12.0, Landform position=8.0, VegetationTypeNumber=10.0, VegetationTypeNumber=16.0

q. Predictors: (Constant), VegetationTypeNumber=12.0, Landform position =8.0, VegetationTypeNumber=10.0, VegetationTypeNumber=16.0, height_allign=15.0, cadem_50mb, height_allign=3.0, Solar.Rad..w.m2., Landform position=1.0, height_allign=16.0, VegetationTypeNumber=14.0, densit_allign=1.0, LandcoverNumber=6.0, VegetationTypeNumber=5.0, LandcoverNumber=9.0, height_allign=13.0, VegetationTypeNumber=2.0

Table 4.9: Model summary of 40 variable soil restrictive depth regression model.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.231[a]	.053	.044	40.769	.053	5.478	1	97	.021	
2	.329[b]	.108	.090	39.773	.055	5.921	1	96	.017	
3	.409[c]	.167	.141	38.644	.059	6.689	1	95	.011	
4	.459	.211	.177	37.822	.043	5.174	1	94	.025	
56	.903	.816	.700	22.836	.005	1.668	1	60	.201	
57	.903	.815	.703	22.730	-.001	.435	1	60	.512	
58	.905	.819	.704	22.693	.004	1.200	1	60	.278	
59	.906	.821	.703	22.723	.003	.845	1	59	.362	
60	.906	.821	.707	22.564	-.001	.165	1	59	.686	
61	.907	.823	.706	22.612	.002	.745	1	59	.392	
62	.908[bj]	.825	.704	22.693	.002	.578	1	58	.450	2.191

Note: Not all of the models are given because that would take up a lot of space and the ones in between are not necessary to show the optimal model performance outcomes.
a. Predictors: (Constant), VegetationTypeNumber=12.0
b. Predictors: (Constant), VegetationTypeNumber=12.0, Landform position=8.0
c. Predictors: (Constant), VegetationTypeNumber=12.0, Landform position=8.0, VegetationTypeNumber=10.0
d. Predictors: (Constant), VegetationTypeNumber=12.0, Landform position=8.0, VegetationTypeNumber=10.0, VegetationTypeNumber=16.0
bj. Predictors: (Constant), VegetationTypeNumber=12.0, VegetationTypeNumber=16.0, height_allign=15.0, cadem_50mb, height_allign=3.0, Solar.Rad..w.m2., Landform position=1.0, VegetationTypeNumber=14.0, height_allign=13.0, height_allign=21.0, LandcoverNumber=6.0, VegetationTypeNumber=5.0, height_allign=4.0, LandcoverNumber=8.0, LandcoverNumber=4.0, Slope position=6.0, Bedrock Geology=7.0, Bedrock Geology=8.0, height_allign=1.0, height_allign=17.0, height_allign=5.0, Landform position=2.0, Bedrock Geology=9.0, Surface Geology=3.0, densit_allign=2.0, LandcoverNumber=5.0, Surface Geology=5.0, Landform position=9.0, Landform position=3.0, densit_allign=4.0, twi4_0f, Landform position=7.0, VegetationTypeNumber=2.0, height_allign=23.0, Bedrock Geology=3.0, Slope.Percent, Landform position=10.0, Landform position=5.0, height_allign=19.0, height_allign=18.0
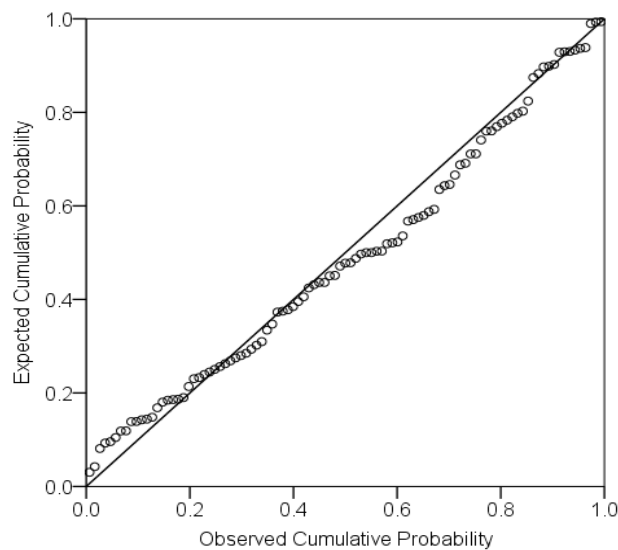
Figure 4.12: P-P plot of regression analysis based on sig. level in of 0.10 and Sig. level out of 0.99 (17 variables). Observed Cumulative Probability on the x-axis are the cumulative probability

of the Studentized residuals and the Expected Cumulative Probability is the normalized probability of the studentized residuals.
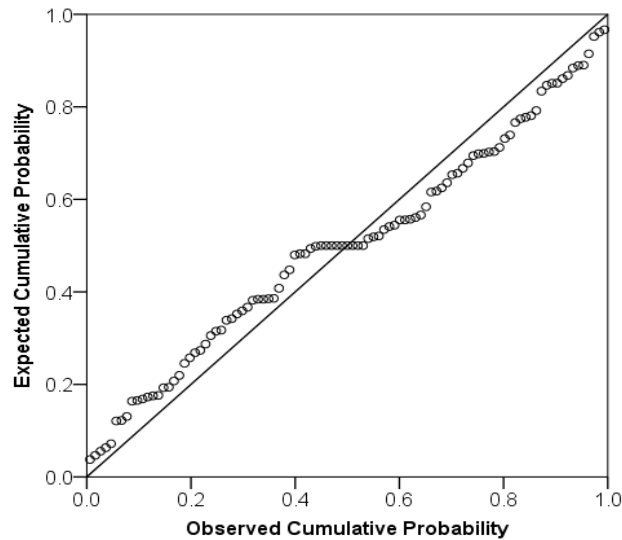
Figure 4.13: P-P plot of regression analysis based on sig. level in of 0.50 and Sig. level out of 0.51 (40 variables). Observed Cumulative Probability on the x-axis are the cumulative probability of the Studentized residuals and the Expected Cumulative Probability on the y-axis are the normalized probability of the Studentized residuals.

Table 4.10: Model performance statistics of the validation datasets to assess prediction maps.

	Training Model 17 Variables	Validation Model 17 Variables	Training Model 40 Variables	Validation Model 40 Variables
R^2	0.621	0.006	0.825	0.012
RMSE	25.53	54.44	17.37	66.18
MAE	20.12	42.63	13.37	52.10
Average Standard Error	11.17	11.73	14.10	24.21

The top 10 variables of the 17 variable model are: Vegetation Type 12 (Alpine larch), Landform Position 8 (local ridges, hills in valleys), Vegetation Type 10 (Engelmann spruce), Vegetation Type 16 (Herbaceous Grassland), Height 15 m, Elevation, Height 3 m, Solar Radiation, Landform Position 1 (canyons, deeply incised streams), and Height 16 m. Half of the variables in the model are categorical vegetation variables and half are continuous and categorical terrain variables.

4.4.2 Soil Texture Stepwise Regression Modelling

Samples were collected from both the A and the B horizons of sample sites. Percent sand, silt, and clay were measured in the lab for each horizon, and they were individually modelled with stepwise regression methods. The selection of the regression equations was based on model performance statistics in the same fashion as for soil restrictive depth.

The residual scatter plots for the percent sand models show that when the significance level for variables to enter the model is set to 0.45 the normality of the residuals is significantly degraded. Despite the fact that the R^2 increases and the RMSE and MAE decrease with increased significance settings using the percent sand training data (Table 4.11), when the percent sand validation data (Table 4.12) is used to predict soil texture the performance statistics support lower significance settings for variables to be entered and/or removed from a the modelling process. Therefore, Sand 1 model was chosen for mapping. Percent silt and clay training and validation model analysis (Table 4.12) follow the same logic as percent sand, therefore Silt 1 and Clay 1 regression models were used for mapping. Table 4.13 presents the top 10 variables of each model, used to assess the logic of the models in predicting soil properties because the top ten variables have a strong influence on prediction power of the models.

Table 4.11: A horizon percent sand, silt, and clay content training models (N = 75).

Model Name	Sig. IN	Sig. OUT	R^2	Adj. R^2	RMSE	MAE	Average Standard Error
Sand 1	0.35	0.5	0.758	0.594	7.23	5.72	5.79
Sand 2	0.45	0.5	0.879	0.721	4.62	3.32	5.78
Silt 1	0.35	0.5	0.759	0.612	6.55	5.10	4.97
Silt 2	0.35	0.36	0.845	0.713	5.25	3.93	4.77
Clay 1	0.15	0.2	0.845	0.779	1.34	0.96	0.82
Clay 2	0.25	0.3	0.886	0.812	1.15	0.74	0.88

Table 4.12: Model performance statistics, for the A horizon validation data (N = 29).

Variable	R^2	RMSE	MAE	Average Standard Error
Sand 1	0.078	23.95	18.99	8.09
Sand 2	0.035	59.84	40.37	11.37
Silt 1	0.006	22.33	18.33	6.26
Silt 2	0.015	49.89	34.70	8.92
Clay 1	0.020	4.34	3.51	0.86
Clay 2	0.002	5.04	3.93	0.96

Table 4.13: Top 10 variables in the A horizon stepwise linear regression models.

Model Name		Top 10 Explanatory Variables
Sand 1	1	VegetationDensity5 (71-100% crown closure)
	2	VegetationType5 (Balsam fir)
	3	Landcover6 (Grassland)
	4	Landform9 (Local ridges, hills in valleys)
	5	Landcover7 (Agriculture)
	6	VegetationHeight21 (21 m)
	7	Aspect
	8	Landcover9 (Broadleaf)
	9	ContributingArea
	10	SlopePosition5 (Upper slope)
Silt 1	1	VegetationDensity5 (71-100% crown closure)
	2	VegetationHeight0 (0 m)
	3	Landcover9 (Broadleaf)
	4	Elevation
	5	Aspect
	6	BedrockGeo8 (Argillite, limestone, and dolostone)
	7	VegetationHeight21 (21 m)
	8	VegetationHeight9 (9 m)
	9	SlopePosition5 (Upper slope)
	10	Landcover7 (Agriculture)
Clay 1	1	VegetationType5 (Balsam fir)
	2	VegetationType10 (Engelmann spruce)
	3	SlopePosition4 (Middle slope)
	4	Landcover5 (shrubland)
	5	VegetationDensity1 (0-5 % Crown closure)
	6	VegetationType9 (Jack pine, White-bark pine)
	7	VegetationType15 (Open shrubland)
	8	SurfaceGeo1 (fluvial deposits)
	9	VegetationType17 (Perennial forage crops)
	10	VegetationHeight11 (11 m)

Note: These variables are the first 10 variables entered into the stepwise regression model, thereby making them the 10 most important variables in the model to account for variability.

Texture data from the B horizons were modelled and models were chosen based on the same criteria as for the A horizon textures. Sand 4 was chosen over Sand 3 (Table 4.14) because Sand 4 performs with a higher R^2 value and lower RMSE and MAE values with the training data, and the average standard error was only slightly larger. With the validation data (Table 4.15), Sand 4 performs slightly better with a larger R^2 value, but has only slightly larger RMSE, MAE, and average standard errors. The residual P-P plot of Sand 4 did show patterns of normality as did the Sand 3 model, however, the Sand 5 model P-P plot did not follow a normally distributed pattern. Sand 5 was rejected because the residuals were not normally distributed, even though the R^2 of the training data was high and the performance statistics were acceptable (Table 4.14). Also, the performance statistics of the validation set for Sand 5 show poor performance due to high RMSE and MAE, significantly worse than the other models (Table 4.15). Silt 3 was chosen over Silt 4 because the R^2 is higher and the RMSE and MAE perform better with the training data (Table 4.14), with only a slightly larger mean standard error. Silt 3 performs closely to Silt 4 with validation data (Table 4.15). The P-P plots of Silt 3 and 4 both show normality, so Silt 3 was chosen based on its better performance. Clay 3 was chosen over Clay 4 because, even though both performed closely with training data (Table 4.14) in all statistical tests, Clay 3 performed substantially better with validation data in statistical tests (Table 4.15). Table 4.16 presents the top 10 variables of the models.

Table 4.14: B horizon percent sand, silt and clay content regression model performance statistics of training data (N = 52).

Model Name	Sig. IN	Sig. OUT	R^2	Adj. R^2	RMSE	MAE	Average Standard Error
Sand 3	0.10	0.15	0.788	0.722	7.31	5.77	3.89
Sand 4	0.15	0.20	0.848	0.778	6.19	4.67	3.94
Sand 5	0.20	0.30	0.996	0.983	1.04	0.670	1.79
Silt 3	0.15	0.20	0.848	0.785	5.64	4.53	3.61
Silt 4	0.10	0.50	0.753	0.700	7.19	5.95	3.28
Clay 3	0.15	0.20	0.817	0.755	1.64	1.16	3.61
Clay 4	0.20	0.25	0.827	0.761	1.60	1.11	2.15

Table 4.15: Model performance statistics, for the B horizon validation data (N = 17).

Variable	R^2	RMSE	MAE	Average Standard Error
Sand 3	0.003	20.03	15.19	3.92
Sand 4	0.016	21.01	16.58	4.07
Sand 5	0.017	34.92	26.68	3.34
Silt 3	0.018	18.25	14.70	4.71
Silt 4	0.021	17.03	13.66	3.60
Clay 3	0.151	4.95	3.89	0.15
Clay 4	0.139	27.63	22.73	3.27

Table 4.16: Top 10 variables in the B horizon texture stepwise linear regression models.

	1	VegetationHeight21 (21 m)
	2	BedrockGeo3 (Sandstone and mudstone)
	3	Aspect
	4	VegetationHeight12 (12 m)
	5	Landcover10 (Mixed)
Sand 4	6	SlopePosition4 (Middle slope)
	7	TWI
	8	VegetationDensity3 (31 to 50 %)
	9	VegetationHeight8 (8 m)
	10	VegetationType6
	1	VegetationHeight21 (21 m)
	2	BedrockGeo3 (Sandstone and mudstone)
	3	Aspect
	4	VegetationType2 (Lodgepole pine)
	5	BedrockGeo6 (Shale, siltstone, and sandstone)
Silt 3	6	VegetationHeight12 (12 m)
	7	SlopePosition2 (Lower slope)
	8	VegetationType6 (Trembling aspen)
	9	Landcover10 (Mixed)
	10	SlopePosition5 (Upper slope)
	1	Landcover2 (Rock/Rubble)
	2	VegetationHeight21 (21 m)
	3	VegetationType17 (Perennial forage crops)
	4	Landcover10 (Mixed)
	5	VegetationType8 (Douglas fir)
Clay 3	6	Landcover9 (Broadleaf)
	7	VegetationHeight11 (11 m)
	8	VegetationHeight12 (12 m)
	9	VegetationHeight23 (23 m)
	10	VegetationType18 (Annual Crops)

4.4.3 Organic Matter and Solum Depth Stepwise Regression Modelling

One hundred and seventy three soil samples were burned in a muffle furnace and weighed to analyze the percent organic matter in the A and B horizons, which prepared the samples for hydrometer textural analysis methods. There are 173 organic matter samples due to the combination of A and B horizons found at the 131 sample sites. Solum depth, the sum of A and B horizons at sample locations, was also calculated for all 131 sample sites. Model selection was carried out in the same manner as with the soil restrictive variables through analysis of training and validation statistics (Table 4.17 and 4.18). The A horizon organic matter model 1 (AOM 1) was chosen as the best model while maintaining normal residuals (Table 4.17). The B horizon organic model 2 (BOM 2) was chosen as the best model and Solum model 1 (Solum 1) was chosen (Table 4.17). The models for OM prediction performed well with high R^2 values, whereas solum depth had much lower R^2 values for both models in the training data (Table 4.17) and had similar R^2 values in the validation data (Table 4.18).

Table 4.17: Training model statistics for A and B horizon percent organic matter and solum depth.

Model Name	Sig. IN	Sig. OUT	R^2	Adj. R^2	RMSE	MAE	Average Standard Error
AOM 1	0.30	0.50	0.887	0.773	1.35	1.03	1.34
AOM 2	0.20	0.50	0.764	0.677	1.94	1.48	1.10
BOM 1	0.05	0.10	0.839	0.809	1.25	0.96	0.50
BOM 2	0.10	0.20	0.862	0.829	1.15	0.877	0.54
Solum 1	0.30	0.31	0.528	0.432	20.51	15.49	8.80
Solum 2	0.50	0.51	0.643	0.460	17.83	12.49	12.41

Note: AOM is for A horizon Organic Matter, BOM is for B horizon Organic Matter, and Solum is for Solum depth.

Table 4.18: Validation model statistics for A and B horizon percent OM and solum depth.

Variable	R^2	RMSE	MAE	Average Standard Error
AOM 1	0.022	10.60	7.53	1.94
AOM 2	0.013	10.02	6.52	1.22
BOM 1	0.015	5.34	3.43	0.59
BOM 2	0.016	5.39	3.44	0.60
Solum 1	0.037	34.18	26.83	8.47
Solum 2	0.042	39.89	32.24	14.36

Table 4.19: Top 10 variables in the organic matter and solum stepwise linear regression models.

AOM 1	1	SurfaceGeology4 (Glaciolacustrine Deposits)
	2	BedrockGeology7 (Shale, siltstone, sandstone, and limestone)
	3	VegetationType8 (Douglas fir)
	4	Landcover9 (Broadleaf)
	5	VegetationHeight11 (11 m)
	6	VegetationDensity2 (6 to 30 %)
	7	LandformPosition3 (upland drainages, headwaters)
	8	TWI
	9	BedRockGeology9 (Sandstone and shale)
	10	SurfaceGeology3 (Glaciofluvial Deposits)
BOM 2	1	ContributingArea
	2	BedrockGeology7 (Shale, siltstone, sandstone, and limestone)
	3	VegetationType20 (Infrastructure)
	4	VegetationHeight14 (14 m)
	5	LandformPosition6 (Open slopes)
	6	LandformPosition10 (Mountain tops, high ridges)
	7	VegetationDensity2 (6 to 30 %)
	8	LandformPosition7 (Upper slopes, mesas)
	9	SurfaceGeology2 (Colluvial Deposits)
	10	SurfaceGeology3 (Glaciofluvial Deposits)
Solum 1	1	SurfaceGeology4 (Glaciolacustrine Deposits)
	2	Aspect
	3	VegetationType18 (Annual crops)
	4	SlopePosition3 (Flat Slopes)
	5	VegetationHeight23 (23 m)
	6	BedrockGeology6 (Shale, siltstone, and sandstone)
	7	VegetationType15 (Open shrubland)
	8	LandformPosition8 (Local ridges, hills in valleys)
	9	VegetationHeight16 (16 m)
	10	VegetationType6 (Trembling aspen)

4.5 Watershed Predictions

4.5.1 Soil Restrictive Depth

Two soil restrictive depth equations (Table 4.7) were input into the Raster Calculator of ArcMap and the prediction rasters were calculated and analyzed. Both models over-predict restrictive soil depth values in areas of extreme environmental sites, such as very steep slopes at very high elevations, which were not sampled during this study (Figure. 4.13). For example, low elevation, high TWI, low slope, and high solar radiation areas are generally over-estimated. Also, high elevation, south facing, high solar radiation, steep slope, and low vegetation areas are

generally under-estimated. The predictions using 40 variables appeared to overestimate in higher

elevations, but predicted well in the lower elevations, in particular in the far northeast cluster of

sampling points in agricultural areas. In comparison, the predictions using 17 variables seemed to

estimate shallower soil depths in the lower elevation sites and they predicted soil depths in the

upper elevations more closely to the observed soil restrictive depths. These observations occur

over the watershed with some exceptions, specifically in the red highlighted locations (Figure

4.14 map B), where the 40 variable model estimates shallower soil restrictive depth in lower

elevations than the 17 variable model.
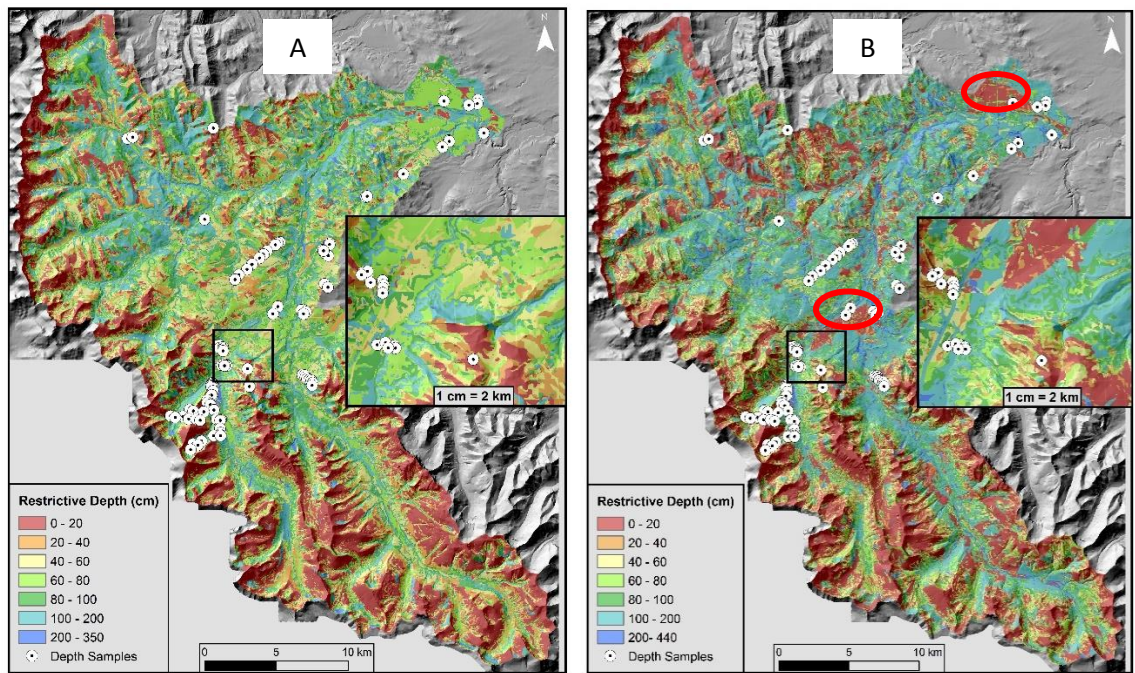
Figure 4.14. Restrictive soil depth predictive maps, A) 17 variables; B) 40 variables. Red circle
signify exceptions to the overall patterns.

Based on model performances (Table 4.7 and 4.10) the model using 17 variables was

more reliable than the 40 variable model. Even though the model with 17 variables had a smaller

R^2 value in training and validation sets and had larger RMSE and MAE, it had smaller ASE

(Table 4.7 and 4.10). Also, the 17 variable model predicts restrictive soil depths more closely to

the knocking pole method depth measurements than the 40 variable model. Deep soil profiles

exist in the upper North East corner of the study area (Figure 4.14) due to grassland vegetation

cover and flat slopes, however, the knocking pole method results in shallow measured soil

restrictive depth measurements due to restrictions by clay content, compaction, and stoniness or a

combination of all three. The 17 variable model predicts closer to measured depths in the upper

North East corner of the study area (average 75 cm) and in other valley areas. Therefore the 17

variable model performs more closely to apparent environmental relationships.

4.5.2 A and B Horizon Texture

   Similar to soil restrictive depth linear model calculations, texture calculation for A and B

horizon sand and silt results in over-estimations above 100% and underestimations below 0%.

This is also due to un-sampled locations that had higher or lower environmental variable

measurements which lead to over- or under-estimations. Percent clay predictions also under-

predict, but did not predict over 100%. Instead, predictions were calculated to a maximum of

30%. The maximum percent clay in the A and B horizon, measured in the lab, was 17.9% and

22.8% respectively, which suggests that the maximum clay predictions are acceptable. Predictions

over 100% are made to be 100% and predictions under 0% are made to be 0%. The resulting

maps (Fig. 4.15 - 4.20) were prepared and analyzed for patterns. These patterns are discussed in

depth in the next chapter, however, some logical large scale patterns are easily recognized. For

example, percent sand in the A horizon (Fig. 4.15) shows decreasing sand in lower elevations, in

valleys, and on less steep slopes. These patterns introduce confidence in the regression equations

due to the recognition of logical patterns.
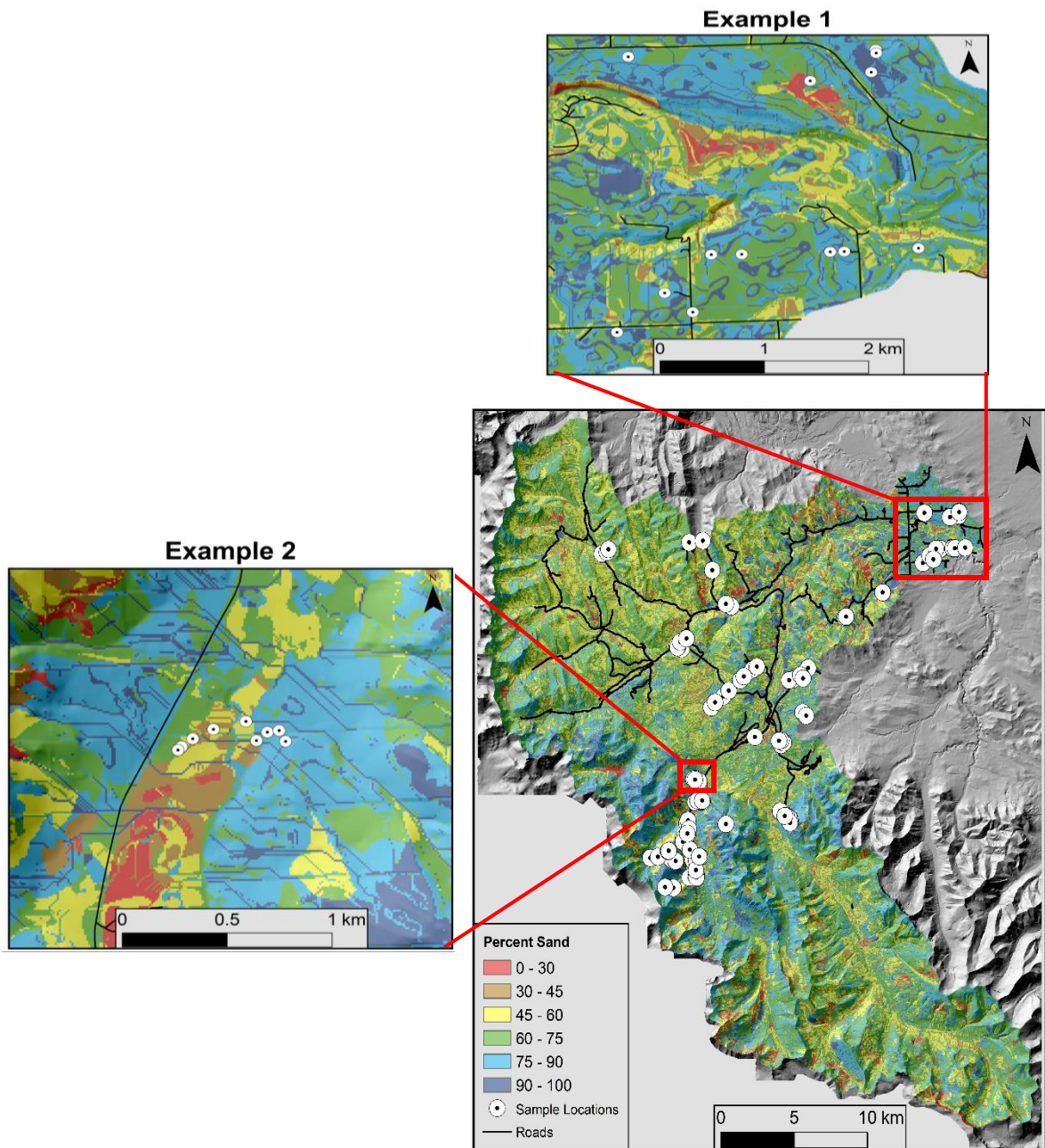
Figure 4.15: Map of percent sand in the A horizon with sample locations. Example 1 is of a low elevation and less steep slope area and Example 2 is of a valley area.

A pattern was noticed in the A horizon silt map (Fig. 4.16), slightly similar to the geology of the watershed. The geology pattern also shows up in the percent sand and silt in the B horizon clearly.
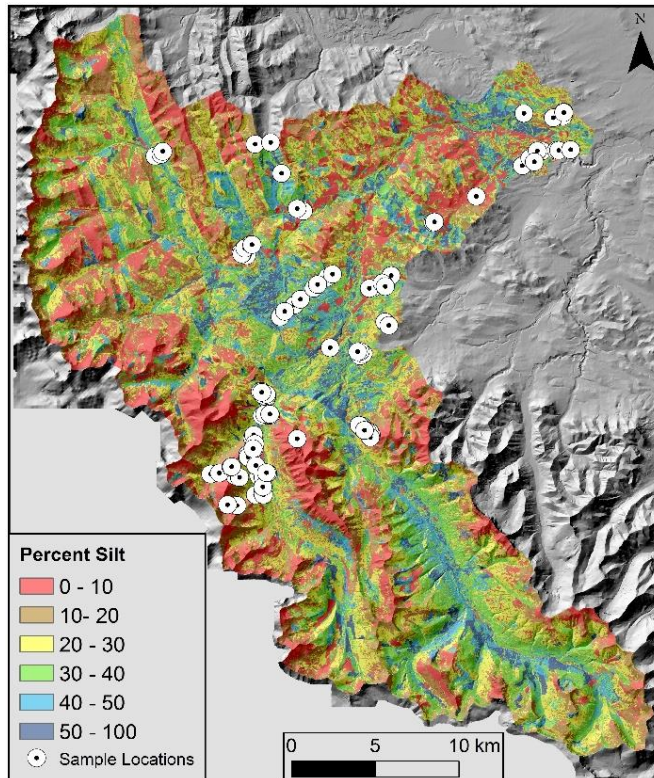
Figure 4.16: Map of percent silt in the A horizon with sample locations.
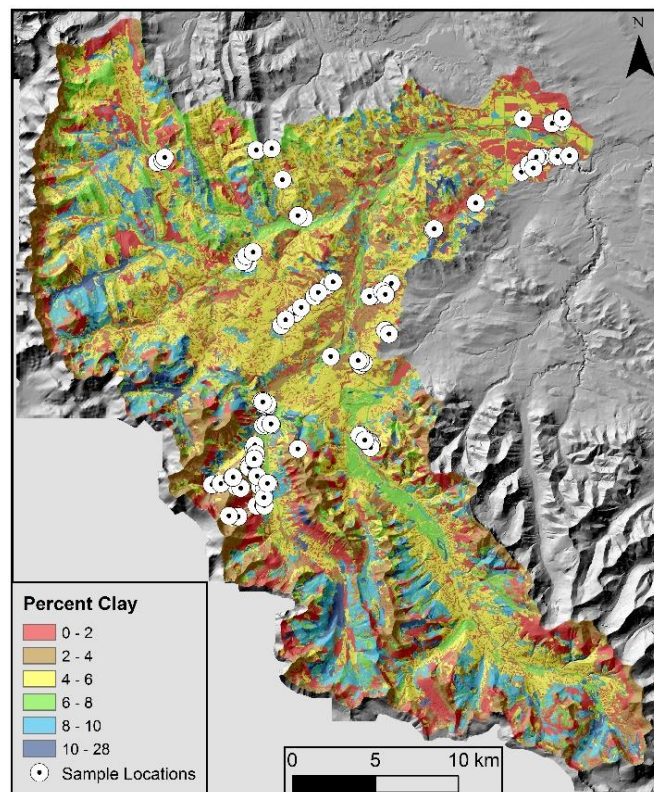
Figure 4.17: Map of percent clay in the A horizon with sample locations.
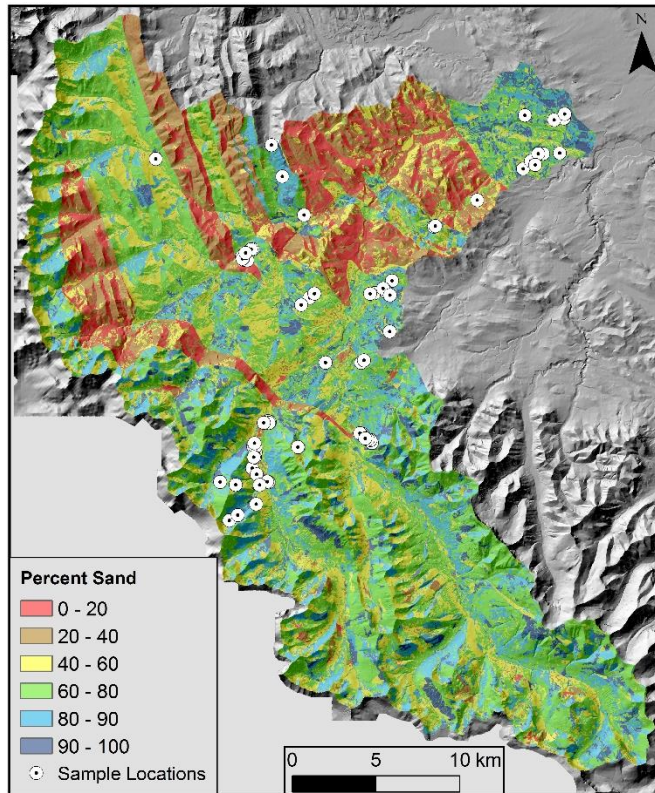
Figure 4.18: Map of percent sand in the B horizon with sample locations.
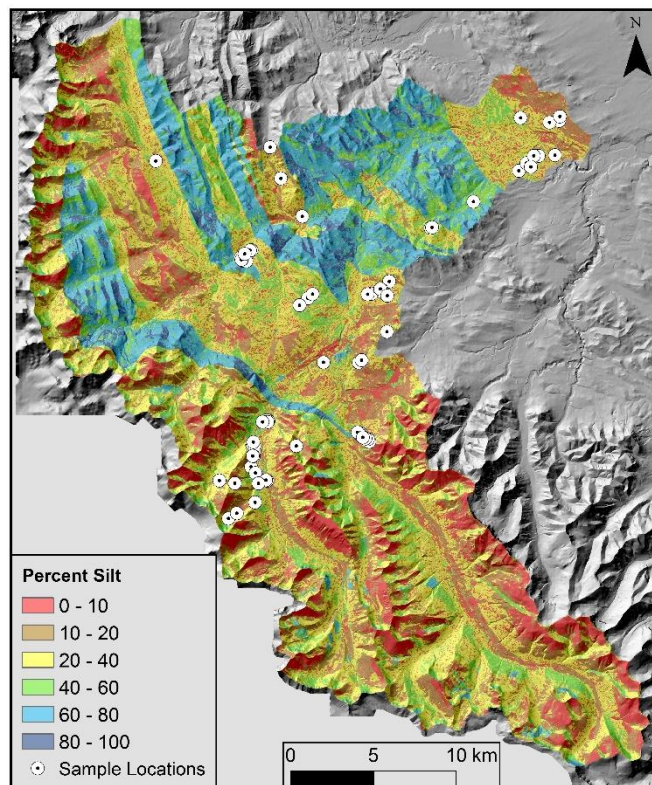
Figure 4.19: Map of percent silt in the B horizon with sample locations.
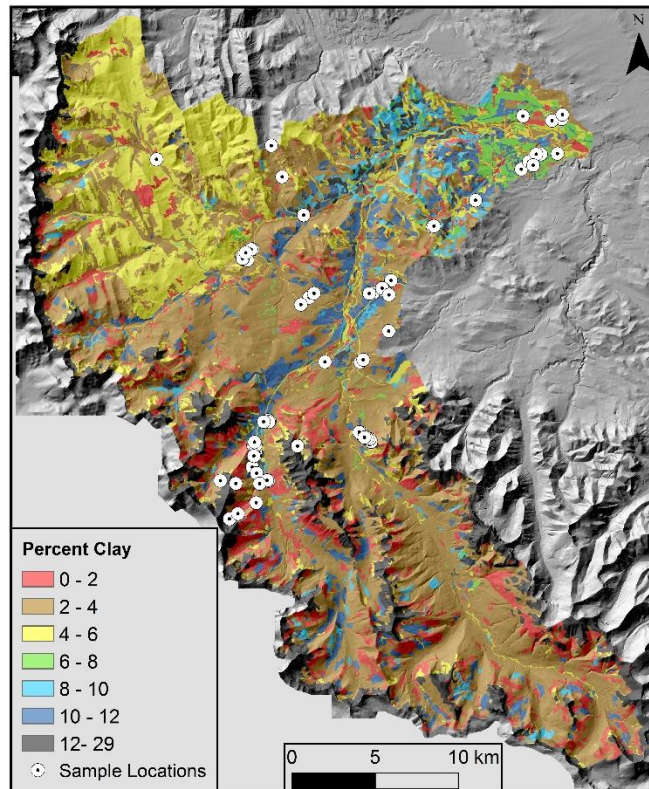
Figure 4.20: Map of percent clay in the B horizon with sample locations.

4.5.3 A *and B horizon Organic Matter and Solum Depth*

The model Solum 1 is deemed optimal and was used despite of its low R^2 value to predict soil depth in the watershed, with an understanding that the model is not error free. Spatial patterns of the predictions show patterns similar to the texture predictions. Surficial geology is a strong variable in the spatial predictions of A horizon OM (Fig. 4.21). B horizon OM (Fig. 4.22) does not show any strong spatial patterning upon visual inspection except for the small light blue patch in the North of the watershed that can be attributed to surficial geology. South and East facing slopes, which receive relatively high solar radiation and have low TWI, appear to have had strong relationships with solum depth (Fig. 4.23). These results indicate logical spatial predictions that are useful and are discussed in Chapter 5.
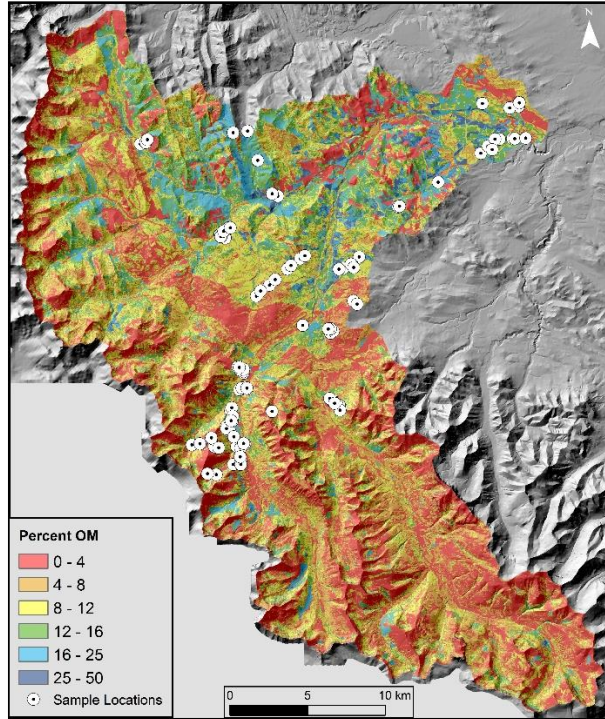
Figure 4.21: A horizon percent organic matter map with soil sample sites.
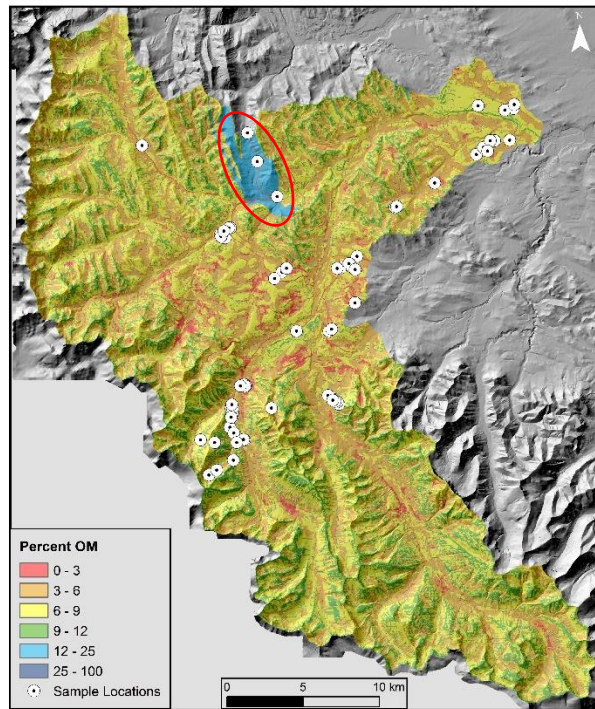
Figure 4.22: B horizon percent organic matter map with soil sample sites.
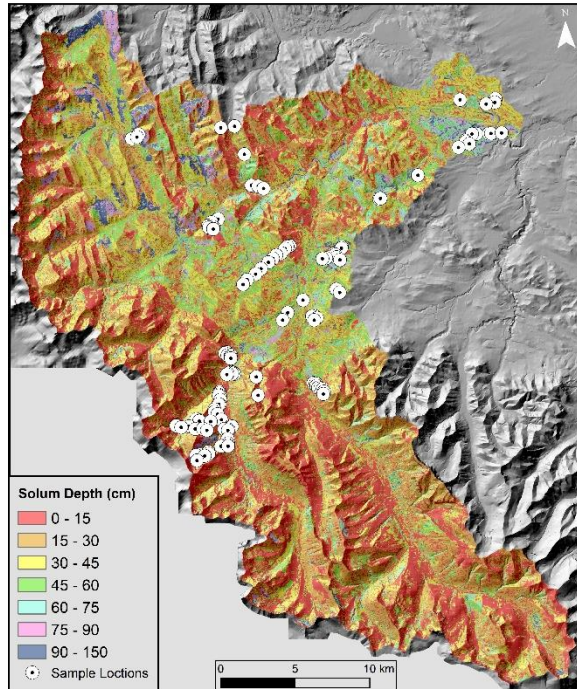
Figure 4.23: Solum depth prediction map with sample sites.

4.6 Comparison with Other Data

There are two soil datasets that are available at different extents for the West Castle watershed. First, the DSS soil information data set, included in the AGRASID database, was downloaded and covers a small part of the watershed in the North East. Crop production and cattle grazing operations dominate the land use types in the area. Second, data were downloaded from SOILGRIDs global soil mapping website. Information available for comparison of both datasets includes soil texture (S, SI, and Cl %) and organic carbon (% by weight). Soil textures were directly comparable to each dataset, however, both dataset reports organic carbon (SOC) as percent, therefore percent organic matter was adjusted by multiplying sampled data of this research by a scaling factor. The scaling factor of 0.54 is based on an average of a typical 48-60% range of soil organic carbon that is in soil organic matter (Tan, 2005). A and B horizon data from this research was averaged and the averaged data was compared with the survey and SOILGRIDS data, because soil survey data contains average texture and OM measurements. SOILGRID's data, even though they are available for multiple depths, do not differ more than 0-5% across any

70

of the layers. Therefore, the averaged data were also compared to the SOILGRID's data. Tables

4.20 and 4.21 present the differences between sampled data and the survey and SOILGRID's data

(Collected data subtracted from the other data). Positive numbers mean that sampled and

predicted layers were greater than the other data and negative values mean that the sampled and

predicted data were less than the other data. It can be observed that the sand sampled SOC data

are all much larger than the surveyed data (Table 4.20). The percent silt data overall, is close to

the soil survey data but it is all lower as seen with negative values (Table 4.20). The clay data are

all much smaller than the survey data as seen with the large negative values (Table 4.20). Texture

data collected through this research differs significantly from survey data, in the sand and clay

fractions especially.

Table 4.20: Comparison of sampled data and soil survey data.

Site ID	Av. S (1)	Av. Si (1)	Av. Cl (1)	Av. SOC (1)	S (2)	Si (2)	Cl (2)	SOC (2)	S Diff	Si Diff	Cl Diff	SOC Diff	Area of soil polygons (Ha)
BZ07	79	18	2	3	37	36	27	4	42	-18	-25	0	1681
T&O#11	66	26	8	5	51	32	17	3	15	-6	-9	2	1681
T&O#12	57	35	8	4	51	32	17	3	6	3	-9	1	947
Rod	61	31	8	5	51	32	17	3	10	-1	-9	2	947
2361	91	8	1	2	60	25	15	3	31	-17	-14	0	947
BR1	70	28	3	4	64	21	15	4	6	7	-12	0	291
BZ03	75	20	5	5	35	30	35	4	40	-10	-30	1	428
BZ02	76	18	6	5	35	30	35	4	41	-12	-29	1	392
MDS1	81	12	7	8	35	30	35	4	46	-18	-28	4	392
WCW_DM 1	79	17	4	5	35	30	35	4	44	-13	-31	1	392
BZ01	83	15	2	6	35	30	35	4	48	-15	-33	2	494
BZ08	74	24	2	5	21	49	30	4	53	-25	-28	1	494
BZ05	83	16	1	6	40	30	30	1	43	-14	-29	5	494
BOBG1 A	77	20	3	5	40	30	30	1	37	-10	-27	4	161
BZ06 A	78	21	1	4	37	36	27	4	41	-15	-26	1	2247
BZ04 A	77	21	2	4	35	30	35	4	42	-9	-33	0	2247
								Av.	**34**	**-11**	**-23**	**2**	

Note: (1) is for data collected through this research and (2) is for the soil survey data for
comparison. "Av." is the average difference in % between the collected data and the survey data.
"S" is for % sand, "Si" is for % silt, "Cl" is for % clay, and "Ha" is for Hectares.

Mapping and analyzing the differences between sampled data and SOILGRID's data

reveals that percent sand is over-predicted relative to amounts of SOILGRID's, with most values

between -6 and 46%, which is a large spread between the data (Table 4.21 and Fig. 4.22). Silt

differences reveal that predicted percent silt is mostly less than SOILGRID's data, with most

values between -32 and 0% (Table 4.21 and Fig. 4.23). Also, percent clay is predicted at lower

values than SOILGRIDs with a majority of values between -14 and -2% (Table 4.21 and Fig.

4.24), which is a smaller spread than sand and silt. Percent SOC predicted values show a narrow

range of +/- 5% difference from SOILGRIDs data. SOC, therefore, is similar to SOILGRIDs data.

The importance of these comparisons are discussed in the next chapter.

Table 4.21: Comparison of sampled data versus SOILGRIDs data.

Variable Type	Sampled – SOILGRIDs Minimum	Sampled – SOILGRIDs Maximum
Sand	-43	67
Silt	-50	51
Clay	-26	11
SOC	-21	39

Figure 4.24: Map of the percent sand difference between sampled data and SOILGRIDs data.

72

Figure 4.25: Map of the percent silt difference between sampled data and SOILGRIDs data.

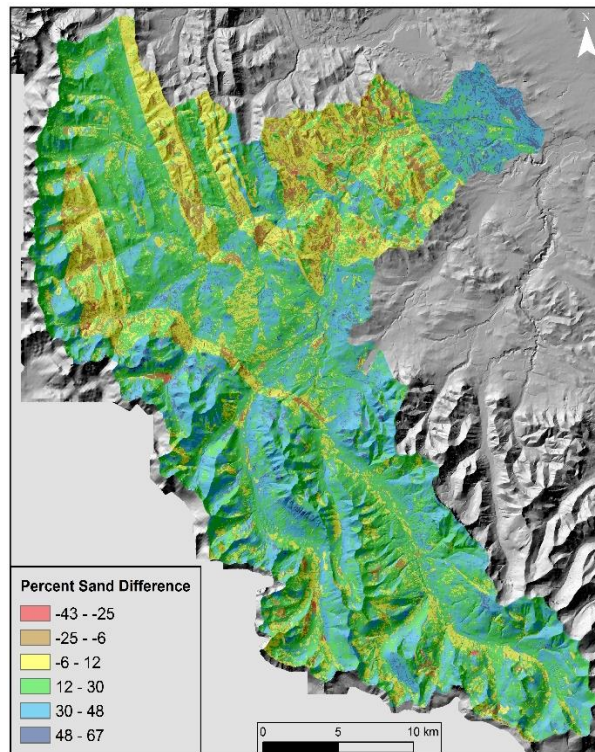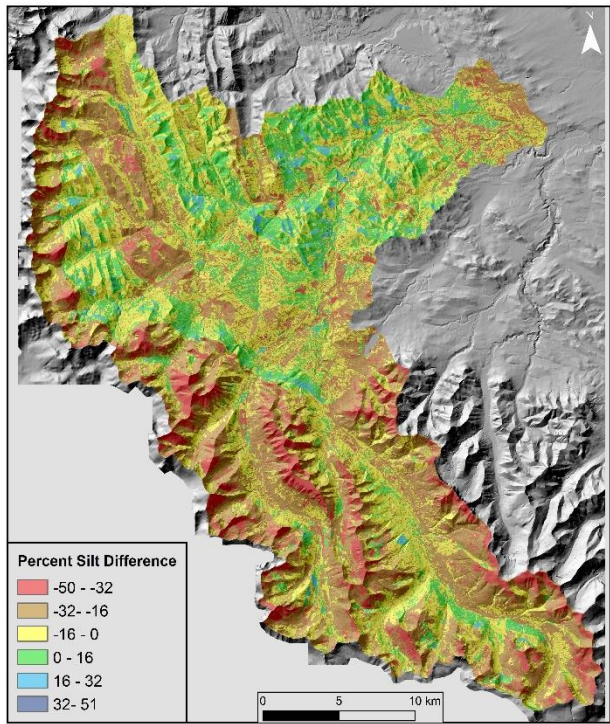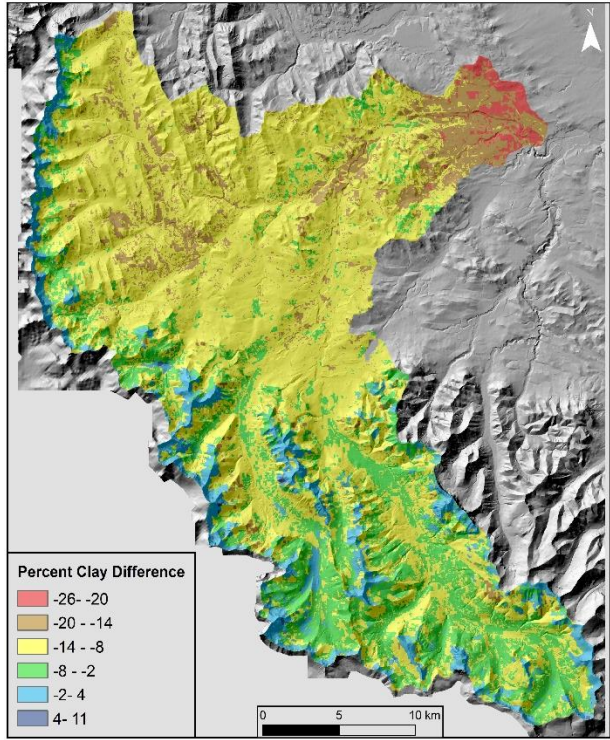

Figure 4.26: Map of the percent clay difference between sampled data and SOILGRIDs data.
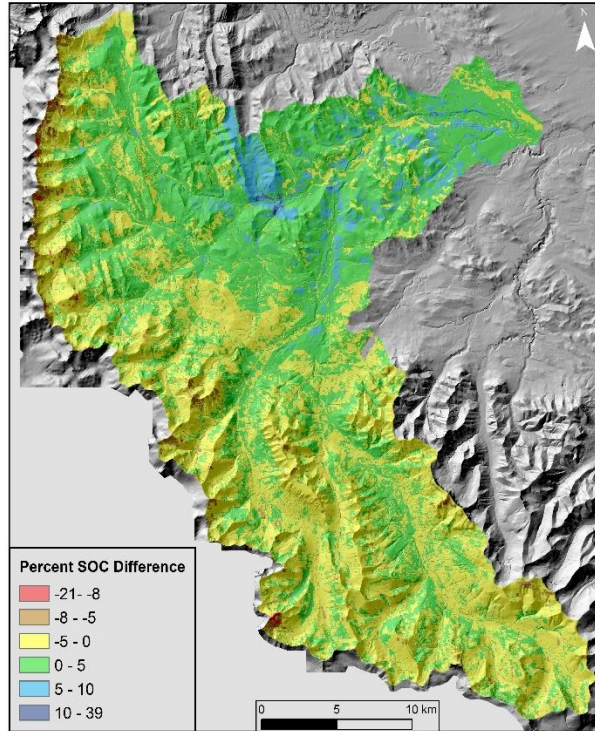
Figure 4.27: Map of the percent SOC difference between sampled data and SOILGRIDs data.

Chapter 5: Discussion and Conclusion

5.1 Objective One: Soil Sampling and Laboratory Analysis

Soil information was gathered through random stratified sampling of 131 sites in the West Castle watershed. Soil lab results were collected through implementation of accepted lab methods, closely following methods in *Soil Sampling and Methods of Analysis* by Carter and Gregorich (2008). Soil texture and SOM were compared to independent commercial soil laboratories to ensure organic matter and textural laboratory methods were conducted accurately. This information has greatly improved our understanding of soils in the study area. Soil restrictive depth sampling resulted in an average of 74 cm with a minimum of 3 cm and a maximum of 182 cm, and 50 % of the data deeper than 63 cm. This information exceeded pre-field operation expectations that the average soil restrictive depth would be approximately 50 cm, whereas solum depth met expectations with an average depth of 36 cm were deeper than expected. This is of great importance for water storage understanding because, in theory, deeper soils have the capacity to hold more water than shallower soils (Shukla, 2014, p. 95). No distinct spatial patterns are seen by viewing the data in relation to its spatial location alone. Further investigation of the spatial patterns by modelling soil data with linear regression allowed for large scale interpretation of spatial patterns.

Texture analyses for both A- and B-horizons were carried out for 173 sites, due to the combination of A and B horizons that were sampled throughout the 131 sampling sites. Descriptive statistics for the A horizon reveal a mean percent sand of 67.3, a minimum of 31.5%, a maximum of 97.6% and a 25th percentile of 56.0% (Table 4.1). These results suggest that the soils are predominantly sandy soils. The mean percent silt was determined to be 28.3%, with a minimum of 0%, a maximum of 62.3%, and a 75th percentile of 38.2%. Further, mean percent clay was 4.4 percent, with a minimum of 0%, a maximum of 17.9 percent, and 75th percentile of 6.1%. These data all lead to the conclusion that the A horizon soils of the watershed are generally high in sand content and low in clay content, as is supported by the soil textural triangle (Figure

4.2). The question of whether the soil texture was dominantly sandy or silty is answered with this data as being dominantly sandy, confirming early preliminary field scouting activities where it was thought that the soils of the watershed would be highly sandy. We can conclude that the sandy soils would likely be well drained and may have a lower water holding capacity as opposed to soil with a larger clay fraction.

B horizon sand measurements were processed for 69 sites. Mean percent sand in the B horizon is 68.6%, with a minimum 28.9%, a maximum of 97.1%, and a 25th percentile of 58.8%. Silt mean percent was 26. 6%, with a minimum of 0 percent, a maximum of 63. 3%, and a 75th percentile of 37.1%. Mean clay content was 4.4%, with a minimum of 0%, a maximum of 17.9% percent, and a 75th percentile of 6.1%. Based on the data, it can be conclude that the B horizon is also highly sandy, with similar overall descriptive statistics results. From the A and B horizon data description it appears that the soil profiles are highly sandy in nature.

Assessing the descriptive statistics for OM results (Table 4.1), it is clear that there is more OM in the A horizon than in the B horizon because all descriptive statistics are greater for the A horizon. This suggests that the water holding capacity is increased due to the OM content, thereby slightly offsetting the poor capacity of the sandiness of the soils. It was thought that the OM content would be low in the mountainous areas, however, it was realized that the mountainous soils contain much OM. This could be due to the large areas of forests that accumulate much litter from deciduous and coniferous leafy and woody deposits. This material is deposited each fall and decomposes into more stable soil organic carbon elements. Arguably a large influence on the OM is the highly productive forest vegetation which is supported by a complex root system leading to large amounts of root break down into OM additions when roots decompose (Fisher et al., 2000, pp. 161-172).

5.2 Objective Two: Statistical Modelling and Spatial Patterns

5.2.1 Spatial Autocorrelation Inferences

Investigation of Moran's I spatial autocorrelation statistic revealed that several soil variables were significantly clustered (Table 4.3), therefore, further investigation of autocorrelation was considered to understand spatial patterns of the soil properties. Semi-variogram analysis was conducted and it was found that all variables revealed weak spatial autocorrelation. Geostatistical analysis, such as kriging was not carried out due to these results. The preferential stratified random sampling method does not strongly support the use of geostatistical methods because sites were gathered without consideration of distance between sites. Also, the sampling method did not allow for complete study area coverage, thereby geostatistical analysis would not allow for complete analysis of the entire watershed. Organic matter in the A horizon and solum depth showed medium spatial dependency and should be highly considered for used in geostatistical analysis in future research that does consider sampling methods for geostatistical analysis from the start.

5.2.2 Soil Restrictive Depth Regression Modelling: SLR

Implementation of SLR allowed for the calculation of largescale soil property predictions throughout the entire watershed. These predictions allow for the extrapolation of the measured soil properties to the scale of the watershed so that the data can be used on a large scale. Also, of great importance, predictions allow for the assessment of soil property spatial distribution and recognition of logical spatial relationships that are established and explainable by regression models.

Soil restrictive depth modelling using 17 environmental variables was chosen as the best model for further exploration. Statistical performance of the training and validation data sets (Table 4.10) reveals that the R^2 value of the validation data, using the training data formula, was very low (0.006). Also, the MAE is extremely large (42.63 cm), twice as large as the training data (20.12 cm). The average standard error is similar between the data sets. Kuriakose et al. (2009)

found that land cover variables were better able to model soil restrictive depth than terrain variables, which is corroborated through the findings of this research project (Table 17). Kuriakose et al. (2009) were able to explain nearly half (R^2 of 0.41) of the validation soil restrictive depth measurements with only land cover variables and just slightly more with all validation variables (R^2 of 0.47). Regression modelling resulted in a lesser ability to explain soil restrictive depth using only vegetation variables of the validation set (0.031) and using all variables (0.002) compared to Kuriakose et al. (2009) who had high R^2 validation values (0.44) using reveres stepwise regression. A conclusion drawn from these statistical results is that there is large inherent variability in the soil properties which the regression model has difficulty explaining. This is likely the result of the limited number of soil samples that could be sampled and analyzed during this research project.

The top ten variables of the chosen model with 17 variables reveals half of the variables important to predicting soil restrictive depth are vegetation variables. Depth of soil is a determining factor for the type and health of forest vegetation, especially tree type, that will grow in any particular area (Fisher et al., 2000). Soil depth controls moisture storage, rooting depth, anchoring of the plants, and supply/storage of nutrients (Fisher et al., 2000). Vegetation roots also contribute to soil properties. Tree roots hold soil together with the small roots and infection of fugal mycelia, which both release biochemicals into the soil. The roots, fungus, and biochemicals all hold together soil particles leading to better soil structure and decrease soil erosion (Fisher et al., 2000). Increased stability of soil through vegetation roots does allow soil to build up more steadily than soil that has less root mass. In the West Castle watershed, the vegetation variables are both a function of the soil depth and they aid in soil depth accumulation over time with addition of organic matter and breakdown of the soils. In the case of restrictive depth, the vegetation is strongly a result of soil properties. For example, one sample was gathered in Alpine larch vegetation growing in the second deepest soil restrictive depth location (169.5 cm). Due to the deep soil restrictive depth and only one sample taken within Alpine larch, the Alpine larch

78

became very important in the training regression, with the largest R^2 of 0.054. Engelmann spruce and Grassland vegetation were found at many sites, so their appearance in the regression equation was expected and confirmed confidence in the equation(s). Vegetation heights of 3, 15, and 16 m correspond to shrub and tree species of many of the sites. Variables in the regression are logical and explainable in most cases, further investigation and validation would help improve understanding of variables and their relationships with soil properties.

5.2.3 Soil Texture Regression Modelling

A horizon texture validation data shows low R^2 values for all three sand, silt and clay models: Sand 1 (0.078), Silt 1 (0.006), and Clay 1 (0.020). Also, RMSE and MAE statistics are high for all three models. Texture validation statistics for the B-horizon texture models result in similar low R^2 values: Sand 4 (0.016), Silt 3 (0.018), and Clay 3 (0.151). The models Sand 4 and Silt 3 RMSE and MAE validation statistics (from validation data predictions) are much larger than their training statistics, whereas Clay 3 RMSE and MAE validation statistics are only slightly larger than Clay 3 training statistics. B horizon model Clay 4 and 3 result in the highest and best R^2 values and lowest RMSE and MAE of texture data. These results lead to the same conclusion as for soil restrictive depth, that the models had troubles in predicting soil property variations, most attributable to the variation inherent in the watershed. This becomes apparent with the validation predictions. Highest confidence is placed on the B horizon Clay 4 regression model due to the best validation statistics.

Similar to the soil restrictive depth results, the variables entered into the stepwise regression models for A and B horizon textures are largely made of vegetation and land cover types. 80 to 100 % of the variables being categorical type variables. Therefore, similar to soil restrictive depth, vegetation and land cover are seen as indicators of soil texture more so than drivers. Between the A horizon models, Sand 1 and Silt 1 had 6 variables in common, whereas Clay 1 has only one variable in common with Sand 1 only.   Clay 1 shows middle slopes to be more important than upper slopes, which is logical in the sense that clay particles tend to

accumulate on lower and less steep slopes where water transports and accumulates clay particles (Breemen & Buurman, 1998). Between the B horizon models, Sand 3, Silt 3, and Clay 3 had three variables in common. Also, the models Sand 3 and Silt 3 of the B horizon share three variables in common.

Duchaufour (1998) explain that soil texture is highly influenced by water accumulation and movement of soil particles are a results of multiple weathering agents: water, oxygen, mineral acids, and organic acids. There are 4 main processes that each of the agents feed into: hydrolysis, oxidation, hydration, and dissolution. Texture is highly influenced by these processes, strongly dependent on parent material, and also influenced by biochemical reactions (Duchaufour, 1998). The movement of water is expected to transport small particles, especially clay particles, accumulating small particles in low areas (concave), non-steep areas and leaving behind larger particle in convex and steep slope areas (Ließ, Glaser, & Huwe, 2012).

Between A horizon models, Sand 1 and Silt 1, six of the top ten variables are similar but in a differing order: 71 to 100% crown closure, Balsam fir, agriculture land cover, 21 m vegetation height, aspect, Broadleaf land cover, and upper slopes. Vegetation Type 5 (Balsam fir) is one variable common between Sand 1 and Clay 1 models. The high crown closure is not surprising because the coefficient was negative, meaning higher crown closure would lead to less sand. Higher crown closure has a healthy root system, deep litter layer, much woody debris on the floor, and large vegetation area that all impede the movement of water down slope, holding finer soil particles in place. Also, the vegetation aids biochemical breakdown of soil particles and decreases evapotranspiration from the forest floor leading to higher soil moisture, aiding in the breakdown of soil particles. Balsam fir type vegetation has a large negative coefficient meaning that all other vegetation types will have larger percent sand. There are very few Balsam fir zones in the watershed, and only one sample was taken on Balsam fir vegetation type zone. Therefore, the importance of the variable was inflated and was included into the Sand 1 and Clay 1 models because of the statistical significance, rather than a strong causal relationship. A clear pattern of

sandy soils is evident in the agriculture land use zones of the study area, which is used in the sand and silt models logically with the large positive coefficient. The vegetation height of 21 m is similar to the crown closure logic, whereas broadleaf land cover has a negative coefficient which is not expected. It would be expected that broadleaf vegetation, which adds much organic matter to the soil, holds finer soil particles in place with its extensive roots, and therefore would be expected to hold much moisture, would aid in the breakdown and accumulation of finer soil particles resulting in a positive coefficient such as the crown closure and vegetation height variables.

B horizon models had 3 variables common between Sand 4, Silt 3 and Clay 3, whereas Sand 4 and Silt 3 share 6 variables in common (Fig. 4.16). Like the A horizon texture models, vegetation variables make up much of the top 10 variables. Vegetation height of 21 m, sandstone and mudstone (Bedrock 3), and aspect are all included in Sand 4 and Silt 3 in the same order, with opposite coefficient signs. Vegetation height of 21 m has a large negative coefficient in Sand 4, whereas Silt 3 has a large positive coefficient, meaning sand decreases whereas silt increases with vegetation reaching a height of 21 m. Bedrock 3 (Sandstone and mudstone) is the second variable in both Sand 4 and Silt 3. Sand 4 has a large negative coefficient and Silt 3 has a large positive coefficient, meaning that Bedrock 3 has less sand and more silt. Bedrock geology appears as a very important explanatory variable of Sand 4 and Silt 3 (Fig. 4.16). Sand has a small positive aspect coefficient, whereas silt has a small negative coefficient, meaning that both sand and silt percentages are somewhat related to aspect. Clay 3 model is comprised of all vegetation and land cover properties. Land cover 2 (Rock and Rubble) was added as the most important variable in the model with a positive coefficient, meaning clay will increase with rock and rubble land cover. This explanatory variable is not logical, as it would be expected that rock and rubble type land cover would not contain much clay and therefore not result in a positive relationship. Three of the four vegetation heights had negative coefficients and all are not easily interpretable (+21, -11, -12, and -23 m). The three negative coefficient heights suggest that large heights correspond to smaller

amounts of clay, contradictory to this is the vegetation height of 21 m which suggests that large height corresponds to more clay. Large vegetation heights would be logical in that large vegetation would have large root systems to hold soil in place and allow clay particles to build up. Vegetation type 17 (perennial crops), Landcover 10 (mixed), and Landcover 9 (broadleaf) most appropriately are explained by decreased erosion, decreased evaporation, and lower slopes that decrease loss of clay particles and aid in particle break down. Those variables have positive coefficients, so as those variables are encountered the clay content increases. Vegetation type 18 (annual crops) had a negative coefficient, meaning as this variable was encountered it can be expected that clay content would be less, possibly due to loss by erosion or that chosen sample sites happened to have higher sand content.

Further research into the relationships of texture with explanatory variables would improve knowledge of the spatial distribution of soils in the West Castle watershed.

5.2.4 Organic Matter and Solum Depth Modelling

Validation of the optimized OM and solum regression models results lead to the same conclusions as for the previously discussed soil properties. Validation statistics show poor performance, being attributed to the variation inherent in the soil properties gathered. Results indicate that the BOM 1 model performs better with much lower RMSE and MAE statistics, leading to more confidence in the model compared to the other two models.

Comparing the top ten variables of AOM 1 and BOM 2 reveals three of the same variables, Shale, siltstone, sandstone, and limestone bedrock geology, 6 to 30 % vegetation density, and glacio-fluvial surface geology deposits. Interestingly glacio-lacustrine was one variable common between the AOM 1 and Solum 1 models. Similar to all soil property models is the large number of categorical variables in the models. The models AOM 1 and BOM 1 use many geology and slope and landform position variables. Glacio-lacustrine surface geology deposits in the AOM 1 model is logical because grass, shrubland, and agricultural land cover types occur in the same areas, which lead to large amounts of organic matter accumulation. Also,

82

generally in glacio-lacustrine zones, the overland flow of water is slow due to low slope, which is supported by TWI inclusion in the model. TWI is included in the model with a negative coefficient which means that increased TWI indicates lower percent organic matter. This is not logical because some of the largest OM samples were taken in and near the glacio-lacustrine geology, so it would be thought that increased TWI would increase the OM content.

Four of the top ten variables in the Solum 1 model are vegetation variables and nine are categorical variables. Glacio-lacustrine bedrock geological deposits has a large positive coefficient, meaning deep solum depths are strongly related to glacio-lacustrine deposits. Glacio-lacustrine deposits are formed through large accumulations of sediments that often allow for grassland growth. Grassy vegetation can lead to increased soil horizonation due to their high rooting density and depth. For example, Chernozemic soils often develop on Glacio-lacustrine deposits and can develop deep A and B horizons (Soil Classification Working Group, 1998). This bedrock geology type inclusion in the model is very logical and increases confidence for the model. Annual crops (Vegetation type 18) had a high positive coefficient and also follow the logic of Glacio-lacustrine deposits. Slope Position 3 (flat slopes) has a large negative coefficient in the model, meaning flat slopes will decrease the solum depth. This is contradictory, as it would be thought that flat slopes would allow for soil particle accumulation and vegetation that would produce deeper A and B horizons. The Vegetation Type 6 property (Trembling aspen) has a large positive coefficient in the model, which is logical because trembling aspen root systems are very extensive and much OM accumulates each year when leaves are dropped, all leading to large A and B horizons, especially the A horizon.

In conclusion, solum depth and organic matter models contain both logical and illogical explanatory variables. The relationships presented through the models can guide further investigations of organic matter and solum depth studies to confirm and better understand the relationships between these soil properties and environmental properties.

5.3 Objective Three: Comparison with Other Data

5.3.1 Comparison of Sampled Data to Soil Survey and SOILGRIDs Data

Soil survey textural data and organic carbon data were compared to sampled data. Soil texture were able to be compared straight away with the survey data because they are both in percent by weight. Soil organic matter (SOM) collected through this research were scaled down by a theoretical amount of 54% to convert SOM to soil organic carbon (SOC). This theoretical scale manipulates the data and adds some uncertainty into the data. Yet, this was done so that the data could be compared to the survey data that was measured as SOC.

Soil texture comparisons reveal that sand was much higher, silt was slightly lower, and clay was much lower than the survey data. These differences are most attributable to two reasons: 1) the surveyed data were likely collected using different pre-treatments to remove OM and other substance for textural analysis and 2) the survey data may have used modelled expectations of soil texture in the area based on soil profiles similar to few pedons analyzed in the area. Point one was difficult to confirm, as the laboratory methodological information is not explicitly given in much of the documentation of the soil surveys. It is most likely the case that pre-treatments of the soil samples, grinding of soils, and removing OM and other compounds differ from this research. Too much grinding can lead to breakdown of rock into sand size particles and pre-treatments to remove soil compounds has been found to cause large differences when differing pre-treatments are used (Keller & Gee, 2006). Lab analysis of this research project included gridding of soils with a non-rubber tipped pestle, which could have inflated sand fractions due to rock breakdown. Also, the samples were heated to 600°c to remove OM from the samples, this pre-treatment may have not allowed for proper separation of soil particles leading to higher amounts of sand. These studies may not be properly comparable without further information of soil survey methods. For point 2 it was found that only a limited amount of pedons were investigated in or near the Beaver mines area (Agriculture and Agri-Food Canada, 2016), therefore the conclusion is that models were used to classify the texture of the soil units. This leads to the acceptance that point number 2

is the most likely culprit of the differences between the data sets. The textures in the survey data seem to be averages of the soil texture class given to the units, supporting the conclusion that model values were used. The conclusion drawn from all these points is that more confidence is placed on the texture data collected through the efforts of this thesis, because samples were actually analyzed in the lab and taken in many differing environmental and soil type locations. The sample locations and lab methods of this thesis work are known and explicitly presented to all, giving rise to more certainty, transparency, and understanding by those who are interested in this research.

Comparison with SOILGRIDs data reveals that percent sand is mostly lower from SOILGRIDs, whereas silt and clay were mostly higher. One important critique of SOILGRIDs is that the percent sand, silt, and clay is concluded to be the same through all depth increments. In contrast, the data gathered by this thesis specifically gathered data for A and B horizons and found some differences between those horizons. The methods are slightly different between this research and SOILGRIDs, but it is concluded that SOILGRIDs data generalizes data to lower soil depths, whereas this research project does not. Measuring soil properties in depth increments does standardize the data collection and reporting, however, it blurs the understanding of individual horizons especially if the horizons are not measured. Similar to comparison with the survey data this thesis explicitly provides location and methodology, whereas this information is not readily accessible through SOILGRIDs. Therefore, more confidence is placed on the results presented through this thesis.

Soil organic carbon between this research project, the soil survey, and SOILGRIDs data all reveal similar SOC measurements of +/-5 %. One interesting pattern recognized in comparison with SOILGRIDs was that SOILGRIDs data was much larger than this research data in the North East section of the study area, in the area covered by the soil survey data. Therefore, some of the SOILGRIDs data overestimates SOC. In light of these results, much confidence is placed in all three data sources, concluding that the small difference is not a coincidence and all of these data

sources are accurate. The greatest confidence is placed on the SOC data retrieved by this research project due to the small differences between the data sources.

5.4 Applications

Soil properties researched through this research project could aid in modelling hydrological responses resulting from climate changes, specifically where physical models use soil data. The ACRU agro-hydrological modelling system is one such model that uses soil information. Soil information is highly important in the ACRU model due to the multi-layer soil water budget module (Aduah, Jewitt, & Toucher, 2017). In reviewing some papers that use the ACRU model it was recognized that soil properties are often lacking in study areas, and when soil properties were lacking default or inferred values from proxies were used (Aduah et al., 2017; Warburton, Schulze, & Jewitt, 2010). Having data that was sampled and measured in the area can help to accept or reject the soil data produced through proxy and default data as being near the sampled data or not. Soil data derived through this thesis could be used to model hydrological responses and the model could be compared to models using soil data derived from proxies. In the end, hydrological modelling could be improved either by confirming models that have been made already with limited soil data or by making new models with the surveyed soil information. Kienzle et al. (2012) and Bonifacio (2016) suggest that improved soil information could increase credibility of ACRU model outputs, strengthening confidence that the soil data of this thesis could aid in hydrological modelling efforts in the West Castle watershed.

Having gathered soil information for the mountainous region, the data could help to inform land cover change research. As described in Chapter 2, climate induced land cover change research could benefit from soil information because soil is an important factor in determining suitable locations for vegetation to grow (Zolkos et al., 2015). Soil moisture, soil temperature, and soil nutrients are explained as being important properties that aid tree growth (Butler, Malanson, Walsh, & Fagre, 2007). Geomorphic variables (Macias-Fauria & Johnson, 2013), on their own or coupled with soil property data, that influence these soil properties are important to understanding

forest habitat changes. Soil texture and depth properties could aid in understanding soil moisture and temperature potential that influences vegetation movement/position. For research conducted in the West Castle watershed, there have not been sufficient soil property data to aid in understanding vegetation movement. McCaffrey (2018) studied the changing of alpine treeline ecotone in the West Castle watershed and used CTI (compound topographic index) as a proxy for soil accumulation, which has also been used by other studies. Having soil depth information could be used in place of or in tandem with CTI, directly providing soil information that was not available. All of these points mean that the soil information gathered for the West Castle watershed could be very important in the study of forest vegetation changes occurring due to climatic changes.

5.5 Future Research Considerations

Future research of soil properties in the West Castle watershed would enhance the understanding of the spatial distribution of soil properties. Due to the scope of this research project soil pH, hydraulic conductivity, CEC and many other soil properties were not analyzed. Also, many environmental variables were not taken into consideration in the modelling processes: distance to stream, rainfall, temperature, and NDVI are just a few. Therefore, other soil and environmental properties could be analyzed to increase our understanding of the watershed. In pointing those properties out, it should also be recognized that including too many properties in a study at once increases the complexity of field operational constraints. Focus could be placed on the properties analyzed in this study to further improve our understanding of those properties.

Three main consideration of soil property research in mountainous regions are recommended: 1) sampling environmental variables not strongly covered by this study; 2) focus on vegetation type, height, and density for sampling and statistical analysis; and 3) a more patterned sampling approach could be adopted for the effective implementation of geostatistical and other modelling techniques.

Some environmental properties were only sampled a couple of times. Larch vegetation type, found in high elevations, is one example that would be important to sample in other locations (a Larch stand is known to be on the East slope of Syncline Mountain). Sampling in another Larch forest would allow for the investigation into whether or not very deep soil profiles do indeed form due to Larch vegetation or if it was an anomaly due to other forces. This would improve understanding of the relationship of restrictive depth and solum depth to Larch vegetation. Balsam fir vegetation types, higher elevations, and steeper slopes could also be sampled more to improve understanding of soil properties and their environmental relationships. Other areas could be considered for sampling due to the results of this research project.

Vegetation properties should be highly considered for stratifying a sampling regime due to the findings that suggest vegetation properties are highly important proxies for soil property occurrence in the watershed. Also, landform position, surface geology, and bedrock geology were important in models and should be used for stratifying soil mapping units. Categorical variables in general were highly important in all of the models, and, therefore, it is recommended that statistical analysis that may be better able to use categorical variables should be considered. Classification and regression trees could be one method to be considered.

Consideration would be placed on the use of geostatistical and other spatial modelling methods for extrapolating soil properties in the study area. For geostatistical methods to be utilized throughout the watershed, a patterned sampling regime would need to be implemented. A Latin hypercube, grid stratified random sampling, or more rigorous stratified random sampling should be considered. In this study, a stratified random preferential sampling regime was used and this did not allow for total spatial coverage when using geostatistical methods, and may have induced much bias into the semivariogram models. Diggle et al. (2010) conclude that preferential sampling design added severe bias attributable to reduced range of measurements, and that preferential sampling can lead to misleading inferences. These conclusions are relevant for other statistical methods too. A lattice type sampling regime with additional samples taken in regions of

extreme conditions (very steep slopes for example) is concluded to be effective (Diggle et al., 2010). This would allow for representative sampling of the full spatial extent of the watershed. Issues of accessibility would still be present, which guided this thesis. A more sophisticated sampling regime that increases the effectiveness of preferential sampling (Clifford et al., 2012) and/or incorporates operational constraints into other sampling strategies (Roudier et al., 2012) could help improve efficient and effective sampling strategies in mountainous regions.

References

Aduah, M. S., Jewitt, G. P. W., & Toucher, M. L. W. (2017). Assessing suitability of the acru hydrological model in a rainforest catchment in ghana, west africa. *Water Science, 31*(2), 198-214. https://doi.org/10.1016/j.wsj.2017.06.001

Afzali, A., Varvani, J., & Jafarinia, R. (2015). Using geostatistical method for prediction the spatial variability of soil texture and its effect on environment (case study: Farahan plain of Markazi province, Iran). *Journal of Biodiversity and Environmental Sciences, 6*(3), 300-306.

Agriculture and Agri-Food Canada. (1998, June 25, 2013). Canada land inventory, national soil data base. Retrieved from http://sis.Agr.Gc.Ca/cansis/publications/maps/index.Html.

Agriculture and Agri-Food Canada. (2016). *National pedon database summary layer.* Retrieved from https://open.Canada.Ca/data/en/dataset/6457fad6-b6f5-47a3-9bd1-ad14aea4b9e0.

Alberta Agriculture and Forestry. (2018). *Alberta soil infromation viewer.* Accessed march 20, 2018 from https://soil.Agric.Gov.Ab.Ca/agrasidviewer/.

Alberta Biodiversity Monitoring Institute. (2010). *Abmi wall-to-wall land cover map version 2.1.* [ESRI file geodatabase] Accessed in 2017 from http://www.abmi.ca.

Anderson, D. W., & Smith, C. A. S. (2011). A history of soil classification and soil survey in canada: Personal perspectives. *Canadian Journal of Soil Science, 91*(5), 675-694. doi:10.1139/CJSS10063

Boettinger, J. L., Howell, D. W., Moore, A. C., Hartemink, A. E., & Kienast-Brown, S. (2010). *Digital soil mapping: Bridging research, environmental application, and operation*: Springer Science & Business Media.

Bonifacio, C. M. T. (2016). *Developing acru utilities for modelling future water availability: A case study of the oldman reservoir watershed, Alberta.* Lethbridge, Alta: University of Lethbridge, Dept. of Geography.

Botero-Acosta, A., Chu, M. L., Guzman, J. A., Starks, P. J., & Moriasi, D. N. (2017). Riparian erosion vulnerability model based on environmental features. *Journal of Environmental Management, 203*, 592-602. https://doi.org/10.1016/j.jenvman.2017.02.045

Brady, N. C., & Weil, R. R. (2008). *The nature and properties of soils* (14th ed.). Upper Saddle River, N.J: Pearson Education.

Breemen, N. v., & Buurman, P. (1998). *Soil formation.* Dordrecht: Kluwer Academic Publishers.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. doi:10.1023/a:1010933404324

Brevik, E. C., Calzolari, C., Miller, B. A., Pereira, P., Kabala, C., Baumgarten, A., & Jordán, A. (2016). Soil mapping, classification, and pedologic modeling: History and future directions. *Geoderma, 264, Part B*, 256-274. http://dx.doi.org/10.1016/j.geoderma.2015.05.017

Brevik, E. C., & Hartemink, A. E. (2010). Early soil knowledge and the birth and development of soil science. *CATENA, 83*(1), 23-33. https://doi.org/10.1016/j.catena.2010.06.011

Butler, D. R., Malanson, G. P., Walsh, S. J., & Fagre, D. B. (2007). Influences of geomorphology and geology on alpine treeline in the american west—more important than climatic influences? *Physical Geography, 28*(5), 434-450. doi:10.2747/0272-3646.28.5.434

Canada - Alberta Environmentally Sustainable Agriculture Agreement (CAESA). (2018). *Agrasid version 3.0: Soil landscapes user's manual*. Retrieved from http://www1.Agric.Gov.Ab.Ca/$department/deptdocs.Nsf/all/sag3254#whatis.

Carter, M. R., & Gregorich, E. G. (2008). *Soil sampling and methods of analysis* (2nd ed.). Boca Raton, Fla: CRC Press.

Certini, G., & Ugolini, F. C. (2013). An updated, expanded, universal definition of soil. *Geoderma, 192*, 378-379. https://doi.org/10.1016/j.geoderma.2012.07.008

Chai, X., Shen, C., Yuan, X., & Huang, Y. (2008). Spatial prediction of soil organic matter in the presence of different external trends with REML-EBLUP. *Geoderma, 148*(2), 159-166. http://dx.doi.org/10.1016/j.geoderma.2008.09.018

Chang, O., & Lipson, H. (2018). Neural network quine. *arXiv preprint arXiv:1803.05859*.

Chiles, J.-P., & Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty, second edition* (2nd ed.). Hoboken, N.J: John Wiley & Sons.

Churchman, G. J. (2010). The philosophical status of soil science. *Geoderma, 157*(3), 214-221. https://doi.org/10.1016/j.geoderma.2010.04.018

Clayton, J. S., Ehrlich, W. A., Cann, D. B., Day, J. H., & Marshall, I. B. (1977). *Soils of canada*. Ottawa, Canada: Supply and Services Canada.

Clifford, D., Kuhunert, P., Dobbie, M., Baldock, J., Harch, B., McKenzie, N. J., . . . McBratney, A. B. (2012). The effect of preferential samping on sampling variance. *In minasny, b., malone, b.P. & mcbratney, a.B. (eds.), digital soil assessments and beyond* (pp. 223-226). london, UK: CRC Press.

CoHort Software. (2018). *Costat 6.45 data manipulation and statistical analysis program*. Accessed february 2018 from http://cohort.Com/downloadcostatpart2.Html

Cornelissen, T., Diekkruger, B., & Bogena, H. R. (2016). Using high-resolution data to test parameter sensitivity of the distributed hydrological model hydrogeosphere. *Water, 8*(5), 21. doi:10.3390/w8050202

Corwin, D., & Lesch, S. (2003). Application of soil electrical conductivity to precision agriculture. *Agronomy journal, 95*(3), 455-471.

Cutforth, H. W., McConkey, B. G., Woodvine, R. J., Smith, D. G., Jefferson, P. G., & Akinremi, O. O. (1999). Climate change in the semiarid prairie of southwestern saskatchewan: Late winter–early spring. *Canadian Journal of Plant Science, 79*(3), 343-350. doi:10.4141/P98-137

Davie, T. (2008). *Fundamentals of hydrology*. London:Routledge.

Day, P. R. (1965). Particle fractionation and particle-size analysis. *Methods of soil analysis. Part 1. Physical and mineralogical properties, including statistics of measurement and sampling*(methodsofsoilana), 545-567.

De'ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology, 81*(11), 3178-3192. doi:10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

Debella-Gilo, M., & Etzelmüller, B. (2009). Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in gis: Examples from Vestfold county, Norway. *CATENA, 77*(1), 8-18. http://dx.doi.org/10.1016/j.catena.2008.12.001

Dent, D., & Young, A. (1981). *Soil survey and land evaluation*. London;Boston;: Allen & Unwin.

Diggle, P. J., Menezes, R., & Su, T. l. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 59*(2), 191-232. doi:10.1111/j.1467-9876.2009.00701.x

Dingwall, N., & Potts, C. (2016). Are categorical variables getting lost in your random forests? (vol. 2018): Roam analytics. Retrieved from https://roamanalytics.Com/2016/10/28/are-categorical-variables-getting-lost-in-your-random-forests/.

Diochon, A., Basiliko, N., Krzic, M., Yates, T. T., Olson, E., Masse, J., . . . Kumaragamage, D. (2016). Profiling undergraduate soil science education in canada: Status and projected trends. *Canadian Journal of Soil Science, 97*(2), 122-132. doi:10.1139/cjss-2016-0058

Duchaufour, P. (1998). *Handbook of pedology: Soils, vegetation, environment* (5th ed.). Rotterdam;Brookfield, VT;: Balkema.

Esri. (2018). *World basemap imagery*. Accessed march 6, 2018 from arcmap 10.5.

Estrada, J. G., Ahneman, D. T., Sheridan, R. P., Dreher, S. D., & Doyle, A. G. (2018). Response to comment on "predicting reaction performance in c–n cross-coupling using machine learning". *Science, 362*(6416).

Ezeaku, P. I., & Anikwe, M. A. (2006). A model for description of water and solute movement in soil-water restrictive horizons across two landscapes in south eastern nigeria. *Soil science, 171*(6), 492-500.

Fischer, G., Tubiello, F. N., van Velthuizen, H., & Wiberg, D. A. (2007). Climate change impacts on irrigation water requirements: Effects of mitigation, 1990–2080. *Technological Forecasting and Social Change, 74*(7), 1083-1107. https://doi.org/10.1016/j.techfore.2006.05.021

Fisher, R. F., Binkley, D., & Pritchett, W. L. (2000). *Ecology and management of forest soils* (Vol. 3rd --). New York;Toronto;: John Wiley.

Flores, K. O., Butaslac, I. M., Gonzales, J. E. M., Dumlao, S. M. G., & Reyes, R. S. (2016). *Precision agriculture monitoring system using wireless sensor network and raspberry pi local server.* Paper presented at the Region 10 Conference (TENCON), 2016 IEEE.

Florinsky, I. V., & Florinsky, I. V. (2012). The dokuchaev hypothesis as a basis for predictive digital soil mapping (on the 125th anniversary of its publication). *Eurasian soil science, 45*(4), 445-451. doi:10.1134/S1064229312040047

Forkuor, G., Hounkpatin, O. K. L., Welp, G., & Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in south-western burkina faso: A comparison of machine learning and multiple linear regression models. *PLoS ONE, 12*(1), e0170478. doi:10.1371/journal.pone.0170478

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 *Digital soil mapping* (pp. 325-335): Springer.

Gessler, P. E., Moore, I. D., McKenzie, N. J., & Ryan, P. J. (1995). Soil-landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems, 9*(4), 421-432. doi:10.1080/02693799508902047

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486-489. doi:10.5812/ijem.3505

Government of Alberta. (2013). *Bedrock geology of Alberta*. Alberta Geological Survey. Accessed March 19, 2018 from https://geology-ags-aer.opendata.arcgis.com/datasets/bedrock-geology-of-alberta-dig-2013-0019.

Government of Alberta. (2014). *Watersheds of Alberta (goa)*. Edmonton: Alberta Environment and Parks, Government of Alberta. Accessed March 6, 2018 from https://geodiscover.alberta.ca/geoportal/catalog/search/resource/details.page?uuid=%7B55B83D9D-C44F-4C33-8F2C-62A4D2E59C4B%7D.

Government of Canada. (2013). Soil landscapes of canada version 3.2.   Retrieved from http://sis.agr.gc.ca/cansis/nsdb/slc/v3.2/index.html

Government of Canada. (2016). *Detailed soil survey compilations. Retrieved from http://sis.Agr.Gc.Ca/cansis/nsdb/dss/v3/index.Html*.

Government of Canada. (2017). *Canadian climate normal 1981-2010: Beaver mines station data*. Retrieved from http://climate.weather.gc.ca/climate_normals/results_1981_2010_e.html?searchType=stnProv&lstProvince=AB&txtCentralLatMin=0&txtCentralLatSec=0&txtCentralLongMin=0&txtCentralLongSec=0&stnID=2366&dispBack=0.

Grunwald, S. (2005). *Environmental soil-landscape modeling: Geographic information technologies and pedometrics*: CRC Press.

Grunwald, S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma, 152*(3), 195-207. https://doi.org/10.1016/j.geoderma.2009.06.003

Grunwald, S. (2010). Current state of digital soil mapping and what is next. *In digital soil mapping* (pp. 3-12): Springer.

Hanjra, M. A., & Qureshi, M. E. (2010). Global water crisis and future food security in an era of climate change. *Food Policy, 35*(5), 365-377. https://doi.org/10.1016/j.foodpol.2010.05.006

Hartemink, A. (2008). Soils are back on the global agenda. *Soil Use and Management, 24*(4), 327-330.

Hartemink, A., & McBratney, A. (2008). A soil science renaissance. *Geoderma, 148*(2), 123-129. https://doi.org/10.1016/j.geoderma.2008.10.006

Herbst, M., Diekkrüger, B., & Vereecken, H. (2006). Geostatistical co-regionalization of soil hydraulic properties in a micro-scale catchment using terrain attributes. *Geoderma, 132*(1–2), 206-221. http://dx.doi.org/10.1016/j.geoderma.2005.05.008

Heung, B., Bulmer, C. E., & Schmidt, M. G. (2014). Predictive soil parent material mapping at a regional-scale: A random forest approach. *Geoderma, 214-215*, 141-154. https://doi.org/10.1016/j.geoderma.2013.09.016

Hitziger, M., & Lieb, M. (2014). Comparison of three supervised learning methods for digital soil mapping: Application to a complex terrain in the ecuadorian andes. *Applied and Environmental Soil Science, 2014*, 12. doi:10.1155/2014/809495

Huang, P. M., Li, Y., & Sumner, M. E. (2012). *Handbook of soil sciences: Properties and processes* (2nd ed.). Boca Raton, Fla - eBook: CRC Press.

Hudson, B. D. (1994). Soil organic matter and available water capacity. *Journal of Soil and Water Conservation, 49*(2), 189-194.

Jenny, H. (1980). *The soil resource: Origin and behavior* (Vol. 37). New York: Springer-Verlag.

Jiang, R., Gan, T. Y., Xie, J., Wang, N., & Kuo, C.-C. (2017). Historical and potential changes of precipitation and temperature of Alberta subjected to climate change impact: 1900–2100. *Theoretical and Applied Climatology, 127*(3), 725-739. doi:10.1007/s00704-015-1664-y

Jin, X., Zhang, L., Gu, J., Zhao, C., Tian, J., & He, C. (2015). Modelling the impacts of spatial heterogeneity in soil hydraulic properties on hydrological process in the upper reach of the heihe river in the qilian Mountains, northwest china. *Hydrological Processes, 29*(15), 3318-3327. doi:10.1002/hyp.10437

Karam, A. (2008). Chemical properties of organic soils. In M. R. Carter & E. G. Gregorich (Eds.), *Soil sampling and methods of analysis* (2nd ed., pp. 331 - 340). Boca Raton, FL: CRC Press, Taylor and Francis Group.

Keller, J. M., & Gee, G. W. (2006). Comparison of american society of testing materials and soil science society of america hydrometer methods for particle-size analysis. *Soil Science Society of America Journal, 70*(4), 1094-1100. doi:10.2136/sssaj2005.0303N

Kidd, D., Malone, B., McBratney, A., Minasny, B., & Webb, M. (2015). Operational sampling challenges to digital soil mapping in tasmania, australia. *Geoderma Regional, 4*, 1-10. https://doi.org/10.1016/j.geodrs.2014.11.002

Kienzle, S. W., Nemeth, M. W., Byrne, J. M., & MacDonald, R. J. (2012). Simulating the hydrological impacts of climate change in the upper north saskatchewan river basin, Alberta, canada. *Journal of Hydrology, 412–413*, 76-89. http://dx.doi.org/10.1016/j.jhydrol.2011.01.058

Kroetch, D., & Wang, C. (2008). Particle size distribution. In M. R. Carter & E. G. Gregorich (Eds.), *Soil sampling and methods of analysis* (2nd ed., pp. 713-725). Boca Raton, Fla: CRC Press, Taylor and Francis Group.

Kuriakose, S., Devkota, S., Rossiter, D., & Jetten, V. (2009). Prediction of soil depth using environmental variables in an anthropogenic landscape, a case study in the western ghats of kerala, india. *CATENA, 79*(1), 27-38. https://doi.org/10.1016/j.catena.2009.05.005

Laboski, C., Dowdy, R., Allmaras, R., & Lamb, J. (1998). Soil strength and water content influences on corn root distribution in a sandy soil. *Plant and Soil, 203*(2), 239-247.

Lemmen, D. S., Warren, F. J., & Lacroix, J. (2007). *From impacts to adaptation : Canada in a changing climate 2007 : Synthesis*: Natural Resources Canada, Ottawa, ON (Canada);.

Ließ, M., Glaser, B., & Huwe, B. (2012). Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and random forest models. *Geoderma, 170*, 70-79. https://doi.org/10.1016/j.geoderma.2011.10.010

Lund Research. (2018). *Multiple regression analysis using spss statistics* Retrieved from https://statistics.laerd.com/spss-tutorials/multiple-regression-using-spss-statistics.php#procedure

Macias-Fauria, M., & Johnson, E. A. (2013). Warming-induced upslope advance of subalpine forest is severely limited by geomorphic processes. *Proceedings of the National Academy of Sciences*, 201221278. doi:10.1073/pnas.1221278110

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*(1–2), 3-52. http://dx.doi.org/10.1016/S0016-7061(03)00223-4

McCaffrey, D. R. (2018). *Assessing historic change in subalpine forest : A case study in the west castle watershed.* (Dissertation/Thesis), Lethbridge, Alta. : Universtiy of Lethbridge, Department of Geography. Retrieved from http://opus.uleth.ca/handle/10133/5086

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

McKenzie, N., & Ryan, P. (1999). Spatial prediction of soil properties using environmental correlation. *Geoderma, 89*(1–2), 67-94. http://dx.doi.org/10.1016/S0016-7061(98)00137-2

Menezes, M. D. d., Silva, S. H. G., Mello, C. R. d., Owens, P. R., & Curi, N. (2014). Solum depth spatial prediction comparing conventional with knowledge-based digital soil mapping approaches. *Scientia Agricola, 71*, 316-323.

Miller, D. A., & White, R. A. (1998). A conterminous united states multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth Interactions, 2*(2), 1-26. doi:10.1175/1087-3562(1998)002<0001:acusms>2.3.co;2

Minasny, B., & McBratney, A. B. (2006). A conditioned latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences, 32*(9), 1378-1388. https://doi.org/10.1016/j.cageo.2005.12.009

Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma, 264, Part B*, 301-311. http://dx.doi.org/10.1016/j.geoderma.2015.07.017

Moore, I. D., Gessler, P. E., Nielsen, G. A., & Peterson, G. A. (1993). Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal, 57*(2), 443-452. doi:10.2136/sssaj1993.03615995005700020026x

Mosleh, Z., Salehi, M. H., Jafari, A., Borujeni, I. E., & Mehnatkesh, A. (2016). The effectiveness of digital soil mapping to predict soil properties over low-relief areas. *Environmental monitoring and assessment, 188*(3), 195.

Motha, R. P., & Baier, W. (2005). Impacts of present and future climate change and climate variability on agriculture in the temperate regions: North america. *Climatic Change, 70*(1-2), 137-164.

Naiman, R. J., & Décamps, H. (1997). The ecology of interfaces: Riparian zones. *Annual Review of Ecology and Systematics, 28*(1), 621-658. doi:10.1146/annurev.ecolsys.28.1.621

Natural Resources Canada. (2016). *Canadian digital elevation model*. Government of Canada. Accessed in 2017 from http://maps.canada.ca/czs/index-en.html.

Natural Resources Conservation Service. (2018). *Soil texture calculator.* Retrieved from https://www.Nrcs.Usda.Gov/wps/portal/nrcs/detail/soils/survey/?Cid=nrcs142p2_054167.

Nikiforuk, W. L. (1998). *Soil inventory project procedures manual*. Retrieved from http://www1.Agric.Gov.Ab.Ca/$department/deptdocs.Nsf/all/sag6170 (pp. 103).

Oltean, G. S. (2016). Linking the depth-to-water topographic index to soil moisture on boreal forest sites in Alberta. *Forest science, 62*(2), 154-165. doi:10.5849/forsci.15-054

Patil, N. G., & Singh, S. K. (2016). Pedotransfer functions for estimating soil hydraulic properties: A review. *Pedosphere, 26*(4), 417-430. https://doi.org/10.1016/S1002-0160(15)60054-6

Piedallu, C., Gégout, J.-C., Bruand, A., & Seynave, I. (2011). Mapping soil water holding capacity over large areas to predict potential production of forest stands. *Geoderma, 160*(3), 355-366. https://doi.org/10.1016/j.geoderma.2010.10.004

Roecker, S., Howell, D., Haydu-Houdeshell, C., & Blinn, C. (2010). A qualitative comparison of conventional soil survey and digital soil mapping approaches *Digital soil mapping* (pp. 369-384): Springer.

Rood, S. B., Pan, J., Gill, K. M., Franks, C. G., Samuelson, G. M., & Shepherd, A. (2008). Declining summer flows of Rocky Mountain rivers: Changing seasonal hydrology and

probable impacts on floodplain forests. *Journal of Hydrology, 349*(3–4), 397-410. http://dx.doi.org/10.1016/j.jhydrol.2007.11.012

Roudier, P., Beaudette, D., & Hewitt, A. (2012). A conditioned latin hypercube sampling algorithm incorporating operational constraints. *Digital soil assessments and beyond*, 227-231.

Ryan, P. J., McKenzie, N. J., O'Connell, D., Loughhead, A. N., Leppert, P. M., Jacquier, D., & Ashton, L. (2000). Integrating forest soils information across scales: Spatial prediction of soil properties under australian forests. *Forest Ecology and Management, 138*(1–3), 139-157. http://dx.doi.org/10.1016/S0378-1127(00)00393-5

Sauchyn, D., Barrow, E., Hopkinson, R., & Leavitt, P. (2002). Aridity on the canadian plains. *Géographie physique et Quaternaire, 56*(2-3), 247-259. https://doi.org/10.7202/009109ar

Schindler, D. W., & Donahue, W. F. (2006). An impending water crisis in canada's western prairie provinces. *Proceedings of the National Academy of Sciences, 103*(19), 7210-7216.

Seibert, J., Stendahl, J., & Sørensen, R. (2007). Topographical influences on soil properties in boreal forests. *Geoderma, 141*(1), 139-148. https://doi.org/10.1016/j.geoderma.2007.05.013

Shaddick, G., & Zidek, J. V. (2014). A case study in preferential sampling: Long term monitoring of air pollution in the uk. *Spatial Statistics, 9*(Supplement C), 51-65. https://doi.org/10.1016/j.spasta.2014.03.008

Shanley, J. B., Hjerdt, K. N., McDonnell, J. J., & Kendall, C. (2003). Shallow water table fluctuations in relation to soil penetration resistance. *GroundWater, 41*(7), 964-972.

Shukla, M. (2014). *Soil physics: An introduction*. Boca Raton: CRC Press.

Smith, E. G., Janzen, H. H., Scherloski, L., Larney, F. J., & Ellert, B. H. (2016). Long-term (47 yr) effects of tillage and frequency of summerfallow on soil organic carbon in a dark brown chernozem soil in western canada. *Canadian Journal of Soil Science, 96*(4), 347-350. doi:10.1139/cjss-2015-0120

Soil Classification Working Group. (1998). *Canadian system of soil classification, 3rd ed*: NRC Research Press.

Statistics Canada. (2016). *Canada census boundary*. Accessed March 6, 2017 from http://www.statcan.gc.ca.

Steila, D., & Pond, T. E. (1989). *The geography of soils: Formation, distribution, and management* (2nd -- ed.). Totowa, N.J: Rowman & Littlefield.

Stoorvogel, J. J., Kempen, B., Heuvelink, G. B. M., & de Bruin, S. (2009). Implementation and evaluation of existing knowledge for digital soil mapping in senegal. *Geoderma, 149*(1–2), 161-170. http://dx.doi.org/10.1016/j.geoderma.2008.11.039

Suits, D. B. (1957). Use of dummy variables in regression equations. *Journal of the American Statistical Association, 52*(280), 548-551. doi:10.2307/2281705

Tan, K. H. (2005). *Soil sampling, preparation, and analysis*: CRC press.

Tanzeeba, S., & Gan, T. Y. (2012). Potential impact of climate change on the water availability of south saskatchewan river basin. *Climatic Change, 112*(2), 355-386. doi:10.1007/s10584-011-0221-7

Tesfa, T. K., Tarboton, D. G., Chandler, D. G., & McNamara, J. P. (2009). Modeling soil depth from topographic and land cover attributes. *Water Resources Research, 45*(10). doi:doi:10.1029/2008WR007474

University of Saskatchewan. (n.d.). Soils of canada: About orders. Retrieved from http://www.Soilsofcanada.Ca/orders/orders.Php.

Verschuuren, G. (2013, November 26). Statistical testing for normality in excel [video file]. Retrieved from https://www.Youtube.Com/watch?V=8exzrb9trzg.

Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., & Liu, D. L. (2018). High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern australia. *Science of The Total Environment, 630*, 367-378. https://doi.org/10.1016/j.scitotenv.2018.02.204

Warburton, M. L., Schulze, R. E., & Jewitt, G. P. (2010). Confirmation of acru model results for applications in land use and climate change studies. *Hydrology and Earth System Sciences, 14*(12), 2399.

Were, K., Bui, D. T., Dick, Ø. B., & Singh, B. R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an afromontane landscape. *Ecological Indicators, 52*, 394-403. https://doi.org/10.1016/j.ecolind.2014.12.028

Wiesmeier, M., Barthold, F., Blank, B., & Kögel-Knabner, I. (2011). Digital mapping of soil organic matter stocks using random forest modeling in a semi-arid steppe ecosystem. *Plant and Soil, 340*(1), 7-24. doi:10.1007/s11104-010-0425-z

Wösten, J. H. M., Pachepsky, Y. A., & Rawls, W. J. (2001). Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology, 251*(3), 123-150. https://doi.org/10.1016/S0022-1694(01)00464-4

Yang, L., Zhu, A. X., Qi, F., Qin, C. Z., Li, B., & Pei, T. (2013). An integrative hierarchical stepwise sampling strategy for spatial sampling and its application in digital soil mapping. *International Journal of Geographical Information Science, 27*(1), 1-23. doi:10.1080/13658816.2012.658053

Zhang, S., Huang, Y., Shen, C., Ye, H., & Du, Y. (2012). Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. *Geoderma, 171–172*, 35-43. http://dx.doi.org/10.1016/j.geoderma.2011.07.012

Zhu, Q., & Lin, H. S. (2010). Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes. *Pedosphere, 20*(5), 594-606. https://doi.org/10.1016/S1002-0160(10)60049-5

Zolkos, S. G., Jantz, P., Cormier, T., Iverson, L. R., McKenney, D. W., & Goetz, S. J. (2015). Projected tree species redistribution under climate change: Implications for ecosystem vulnerability across protected areas in the eastern united states. *Ecosystems, 18*(2), 202-220.

Appendix A: Results to Support Methodological Decisions.

Table A.1: ANOVA results for percent Sand.

	Sum of Squares	Df	Mean Square	F	P
Between Groups	14.19	2	7.10	0.65	0.54
Within Groups	3, 548. 91	5	709.78		
Total[a]	17, 3671.58	17			

[a]*Total is the total sum of squares for all 18 samples.*

Table A.2: ANOVA results for percent Silt.

	Sum of Squares	Df	Mean Square	F	P
Between Groups	3.42	2	1.71	0.18	0.84
Within Groups	1, 166.47	5	233.29		
Total	1, 265.61	17			

Table A.3: ANOVA results for percent Clay.

	Sum of Squares	Df	Mean Square	F	P
Between Groups	16.19	2	8.09	2.08	0.18
Within Groups	2, 258.89	5	451.78		
Total	2, 313.97	17			

Table A.4: ANOVA results for percent Organic Matter.

	Sum of Squares	Df	Mean Square	F	P
Between Groups	16.66	2	8.33	3.87	0.057
Within Groups	248.19	5	49.64		
Total	286.38	17			
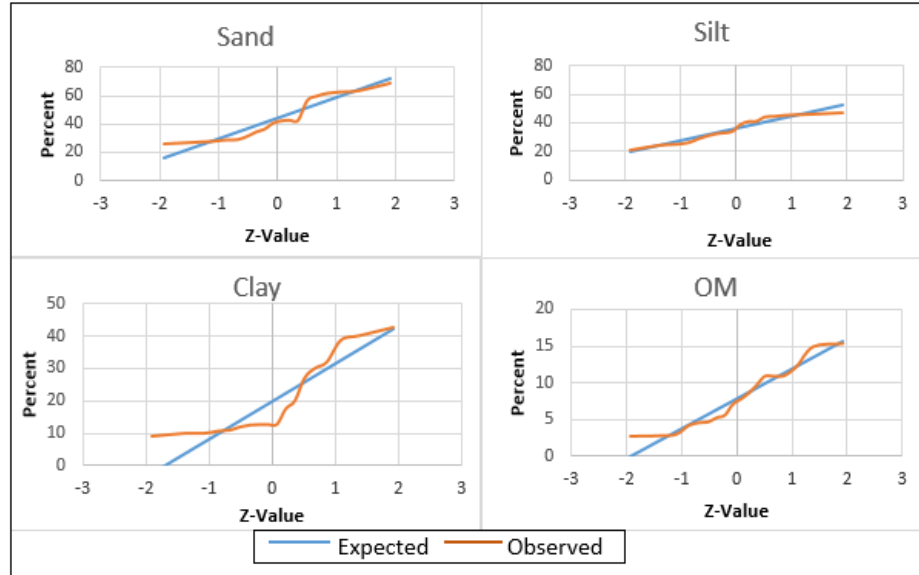
Figure A.1: Q-Q plots of texture and organic matter values encompassing data from all three labs.

Appendix B: Categorical Variable Description and Descriptive Plots.

<p align="center">Table B.1: Categorical variable descriptions.</p>

Independent variables	Class Number	Class Name
Landform Position	1	Canyons, deeply incised streams
	2	Midslope drainages, shallow valleys
	3	Upland drainages, headwaters
	4	U-shaped valleys
	5	Plains
	6	Open slopes
	7	Upper slopes, mesas
	8	Local ridges, hills in valleys
	9	Midslope ridges, small hills in plains
	10	Mountain tops, high ridges
Slope Position	1	Valley
	2	Lower slope
	3	Flat slope
	4	Middle slope
	5	Upper slope
	6	Ridge
Surficial Geology	1	Fluvial Deposits
	2	Colluvial Deposits
	3	Glaciofluvial Deposits
	4	Glaciolacustrine Deposits

Table B.1:	Continued.	
	5	Moraine
	6	Bedrock
Bedrock Geology	1	Basalt
	2	Limestone, shale, dolomite, and siltstone
	3	Sandstone and mudstone
	4	Limestone, dolostone, and shale
	5	Dolomite and siltstone
	6	Shale, siltstone, and sandstone
	7	Shale, siltstone, sandstone, and limestone
	8	Argillite, limestone, and dolostone
	9	Sandstone and shale
	10	Argillite, siltstone, sandstone, and limestone
Land Cover Type	1	Water
	2	Rock/Rubble
	3	Exposed Land
	4	Developed
	5	Shrubland
	6	Grassland
	7	Agriculture
	8	Coniferous
	9	Broadleaf
	10	Mixed
Vegetation Density	1	0 to 5 %
	2	6 to 30 %
	3	31 to 50 %
	4	51 to 70 %
	5	71 to 100 %
Vegetation Type	1	Other
	2	Lodgepole pine
	3	Alpine fir
	4	White spruce
Vegetation Type	5	Balsam fir
	5	Balsam fir
	6	Trembling aspen
	7	Balsam poplar
	8	Douglas fir
	9	Pine
	10	Engelmann spruce

Table B.1:	Continued.	
	11	Black spruce
	12	Alpine larch
	13	Limber pine
	14	Closed Shrubland
	15	Open Shrubland
	16	Herbaceous Grassland
	17	Perennial forage crops
	18	Annual crops
	19	Industrial
	20	Infrastructure
	21	Herbaceous Forbs
	22	Anthro-SubDiv-Rec
	23	Surface Mines
	24	Anthro-Veg+lines
	25	Anthro-Veg+wells
	26	Closed Shrubland/Rough Pasture
	27	Anthro-Gravel Pits
	28	Bryophytes
	29	Farmsteads
Vegetation Height (m)	0 to 28	Categorized in 1 m height classes
	34	Height class on its own
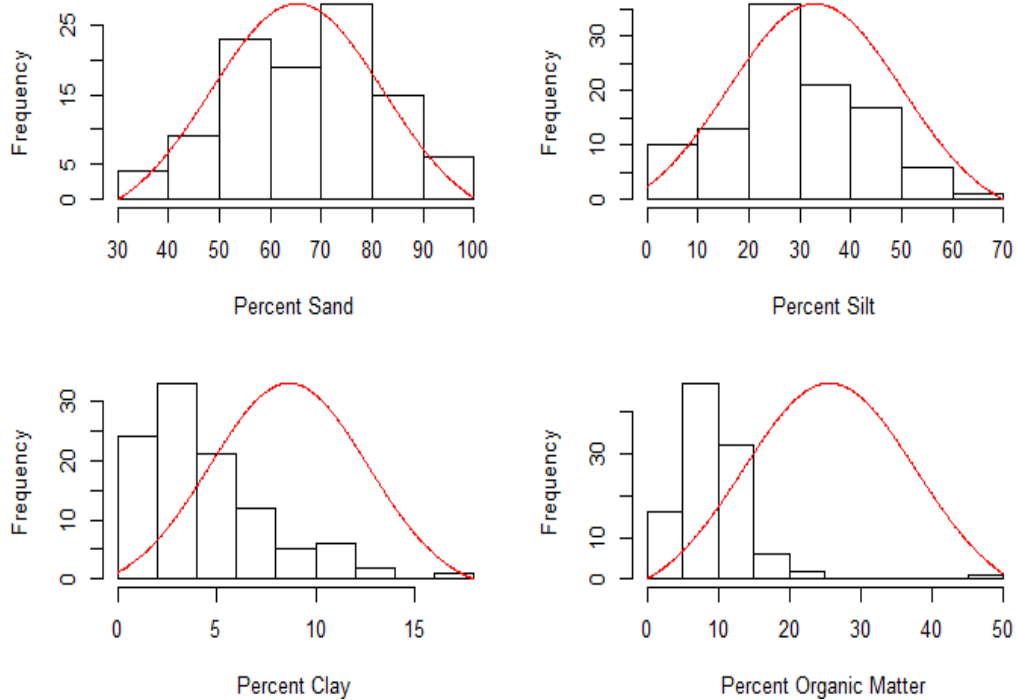
Figure B.1: A horizon soil variable histograms to view the distribution of samples.
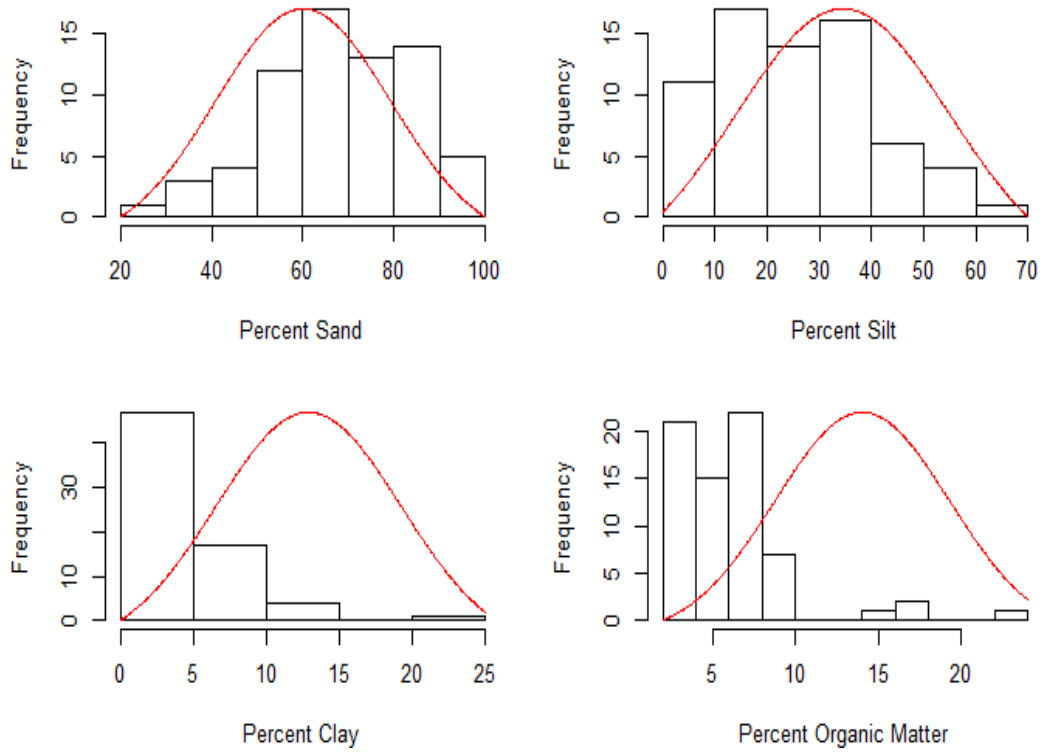
**Figure B.2:** B horizon soil variable histograms to view the distribution of samples.
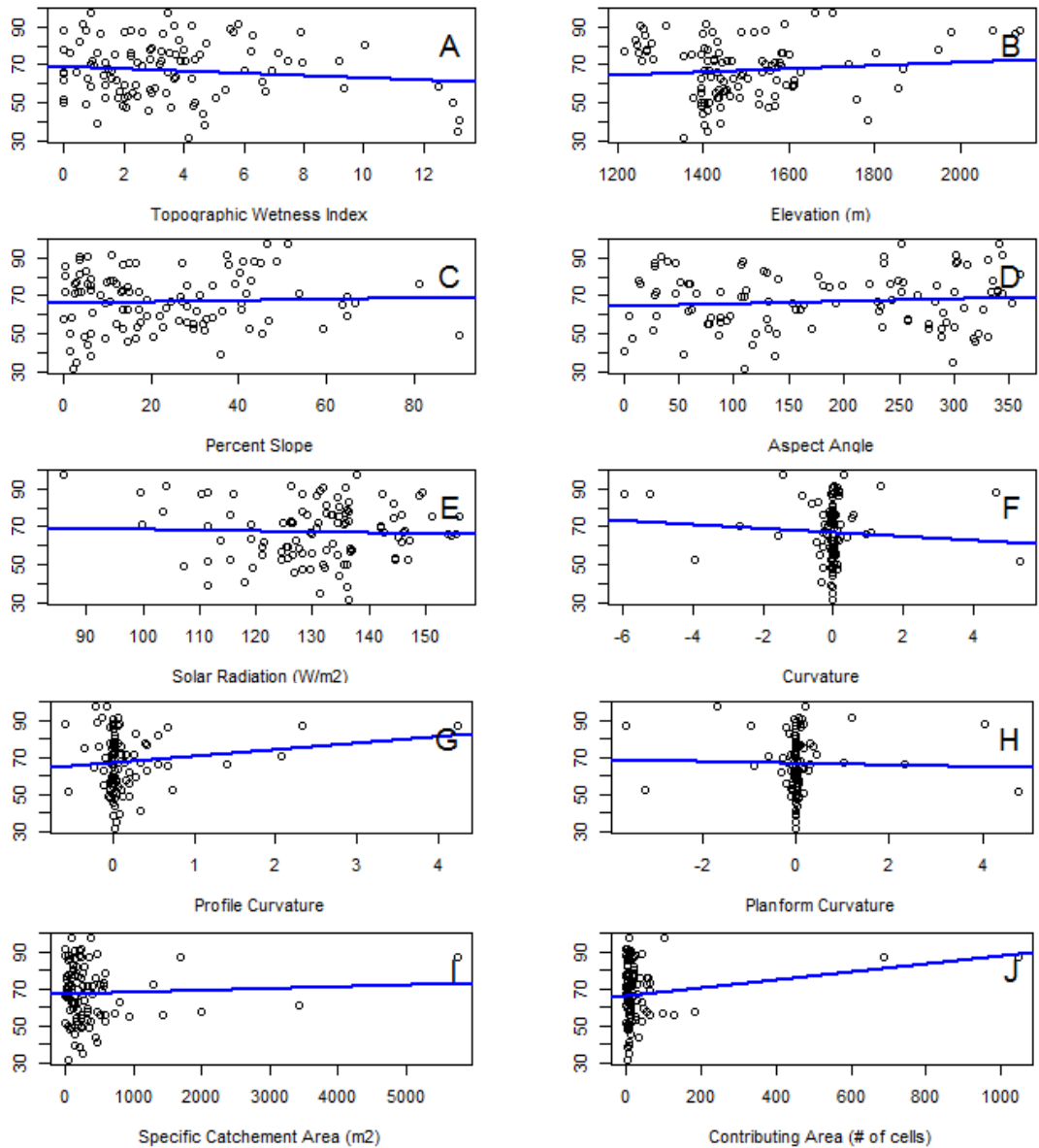
Figure B.3: Scatter plots of A horizon percent sand (on the y-axis) and all the continuous environmental variables on (the x-axis). TWI (R^2 =0.0159), Elevation (R^2 =0.0109), Slope (R^2 =0.0020), Aspect (R^2 =0.0088), Solar Radiation (R^2 =0.0019), Curvature (R^2 =0.0069), Profile Curvature (R^2 =0.0181), Planform Curvature (R^2 =0.0008), Specific Catchment Area (R^2 =0.0024), and Contributing Area (R^2 =0.0325).
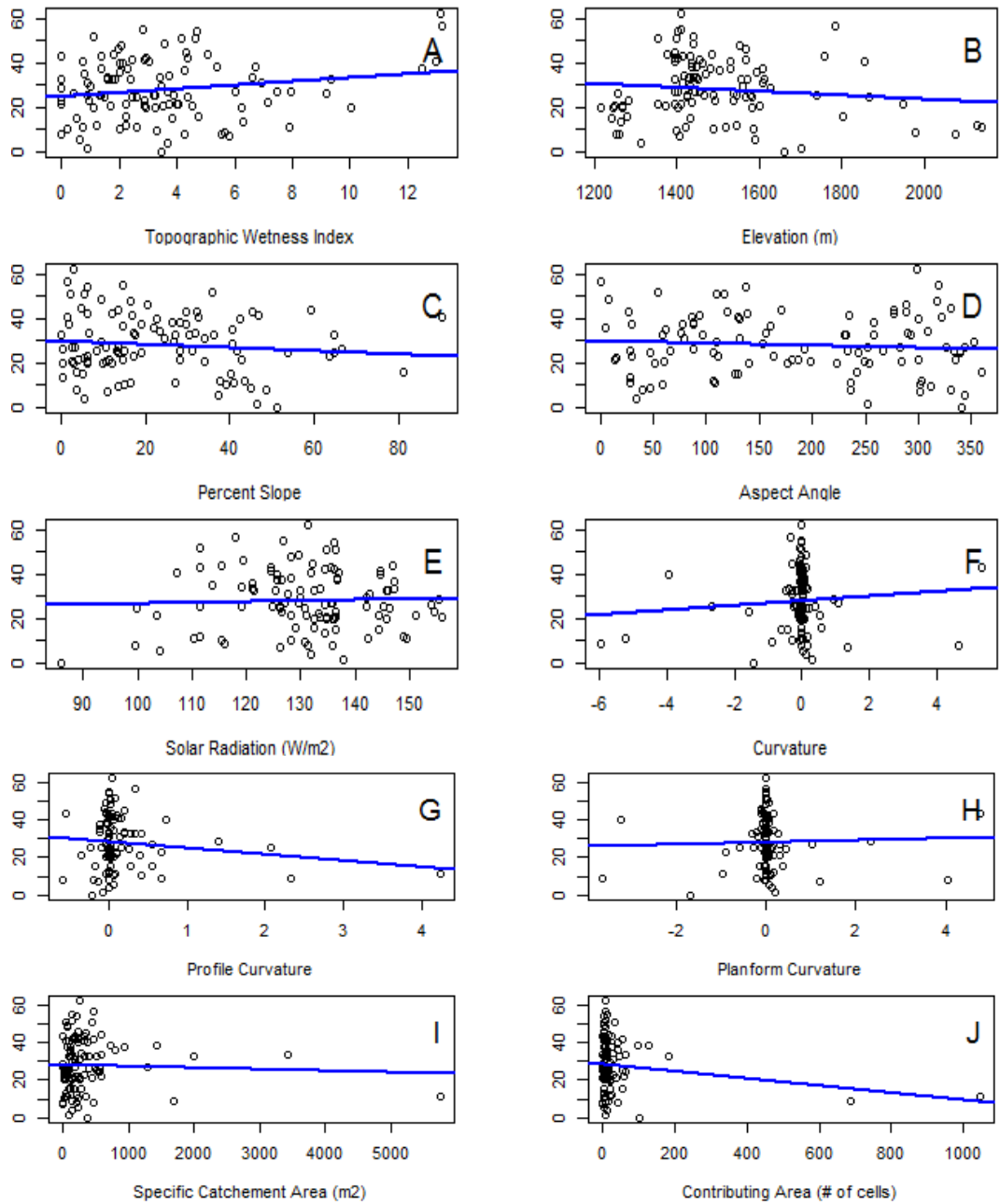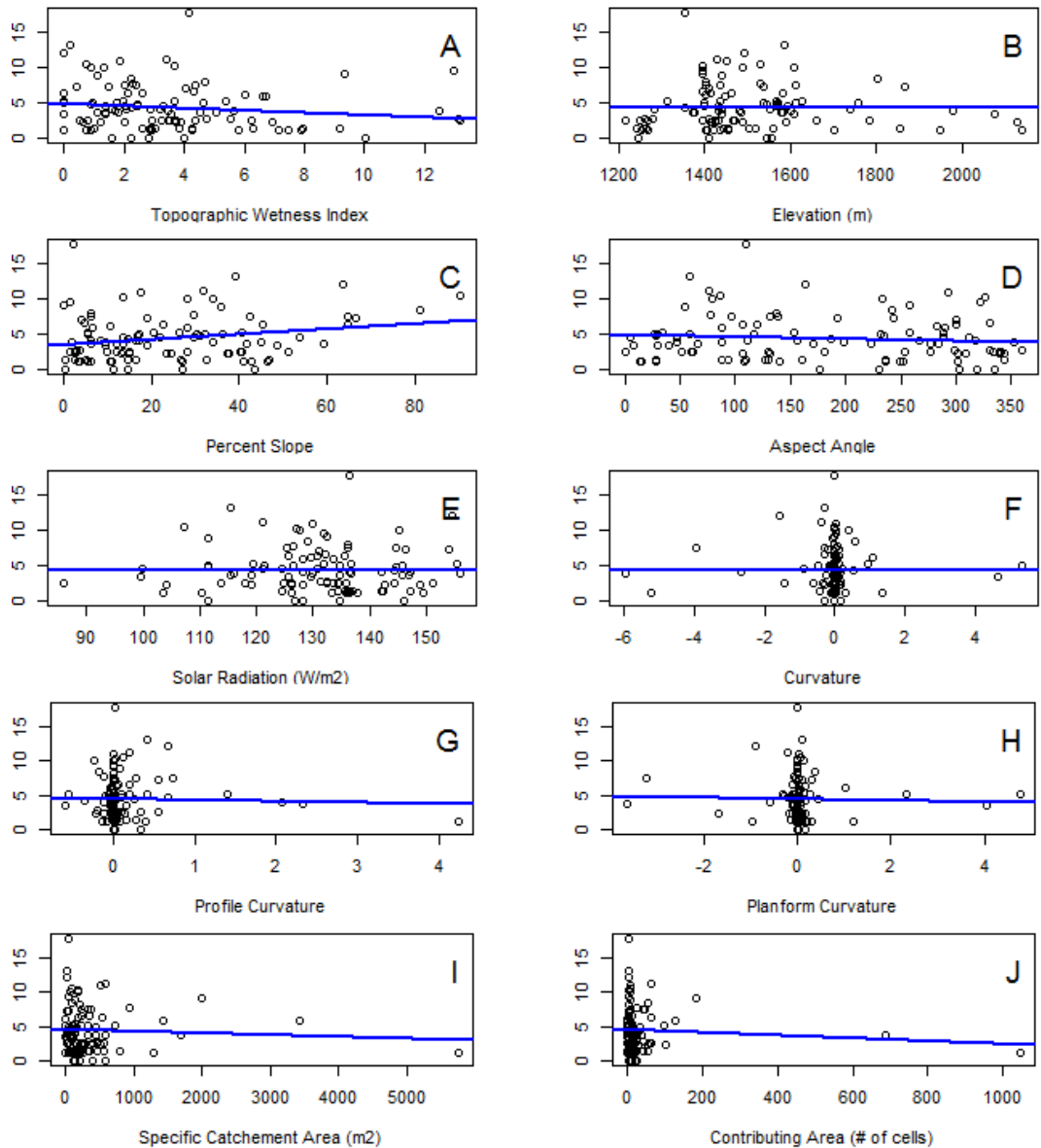
Figure B.4: Scatter plots of A horizon percent silt (on the y-axis) and all the continuous environmental variables on (the x-axis). TWI (R^2 =0.0307), Elevation (R^2 =0.01307), Slope (R^2 =0.01050), Aspect (R^2 =0.00633), Solar Radiation (R^2 =0.00219), Curvature (R^2 =0.00858), Profile Curvature (R^2 =0.01963), Planform Curvature (R^2 =0.00139), Specific Catchment Area (R^2 =0.00166), and Contributing Area (R^2 =0.03194).
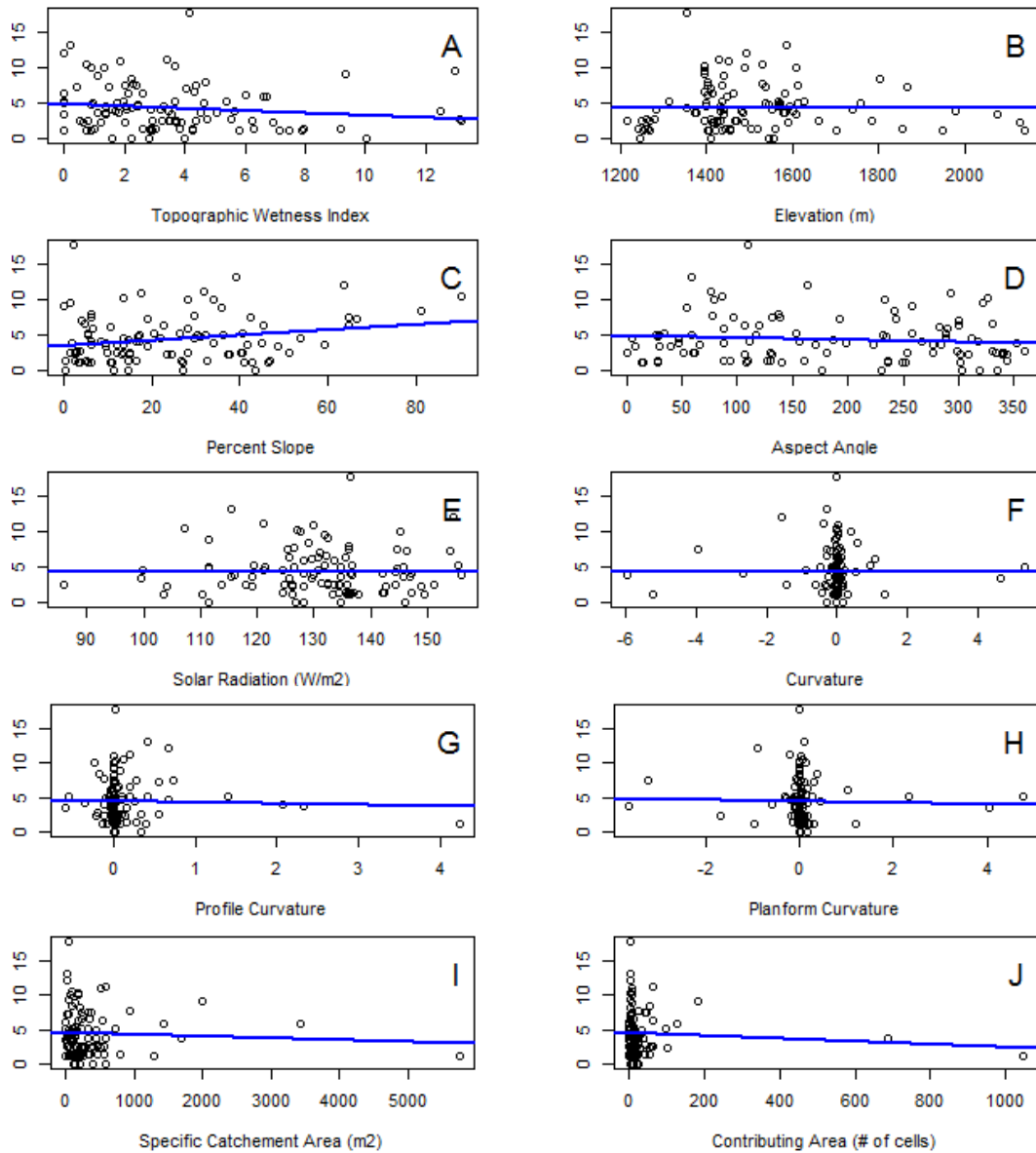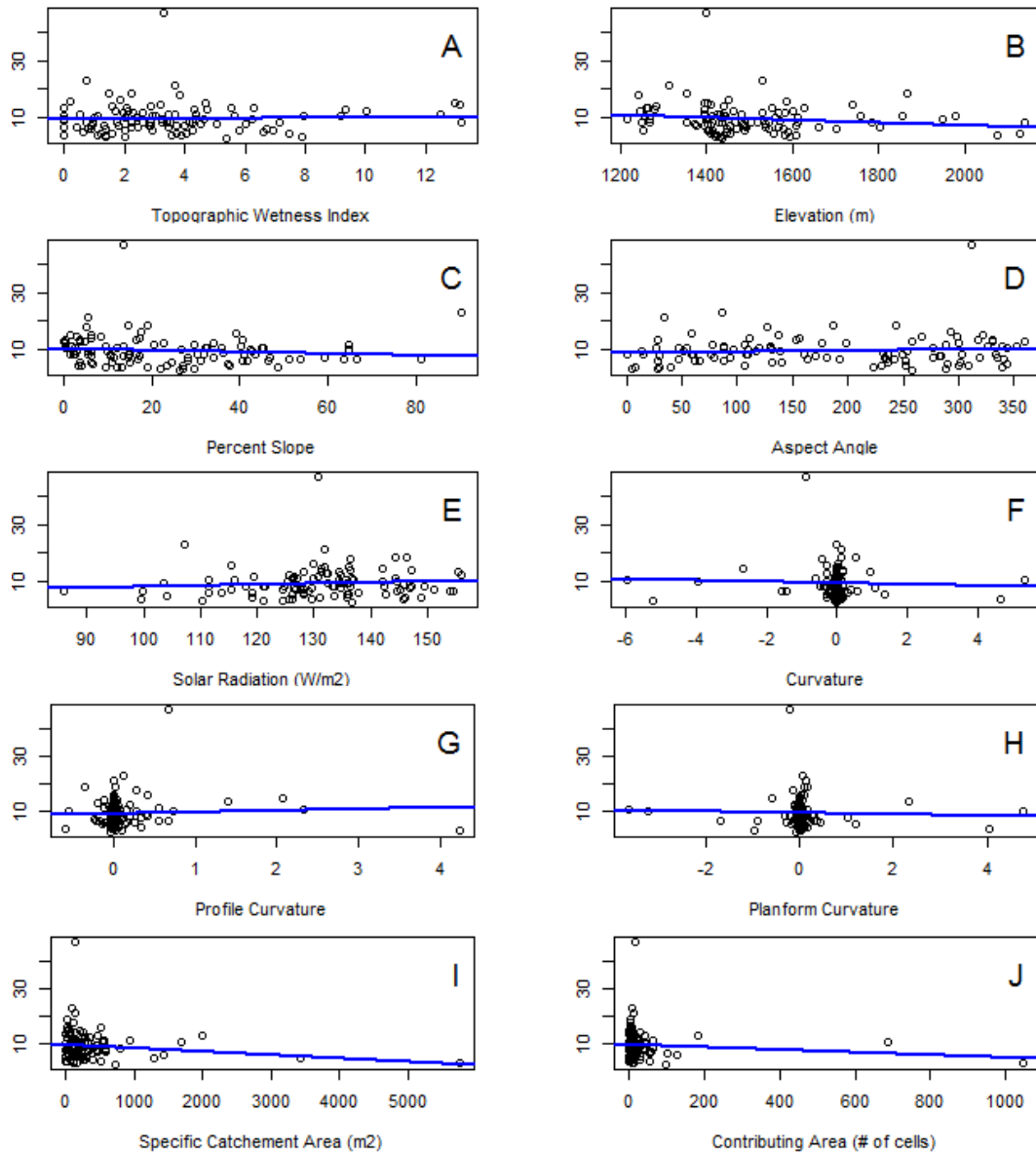
Figure B.5: Scatter plots of A horizon percent clay (on the y-axis) and all the continuous environmental variables on (the x-axis). TWI (R^2 =0.0238), Elevation (R^2 =2.5604e-08), Slope (R^2 =0.0488), Aspect (R^2 =0.0092), Solar Radiation (R^2 =1.9180e-06), Curvature (R^2 =3.4066e-05), Profile Curvature (R^2 =0.0008), Planform Curvature (R^2 =0.0007), Specific Catchment Area (R^2 =0.0028), and Contributing Area (R^2 =0.0059).
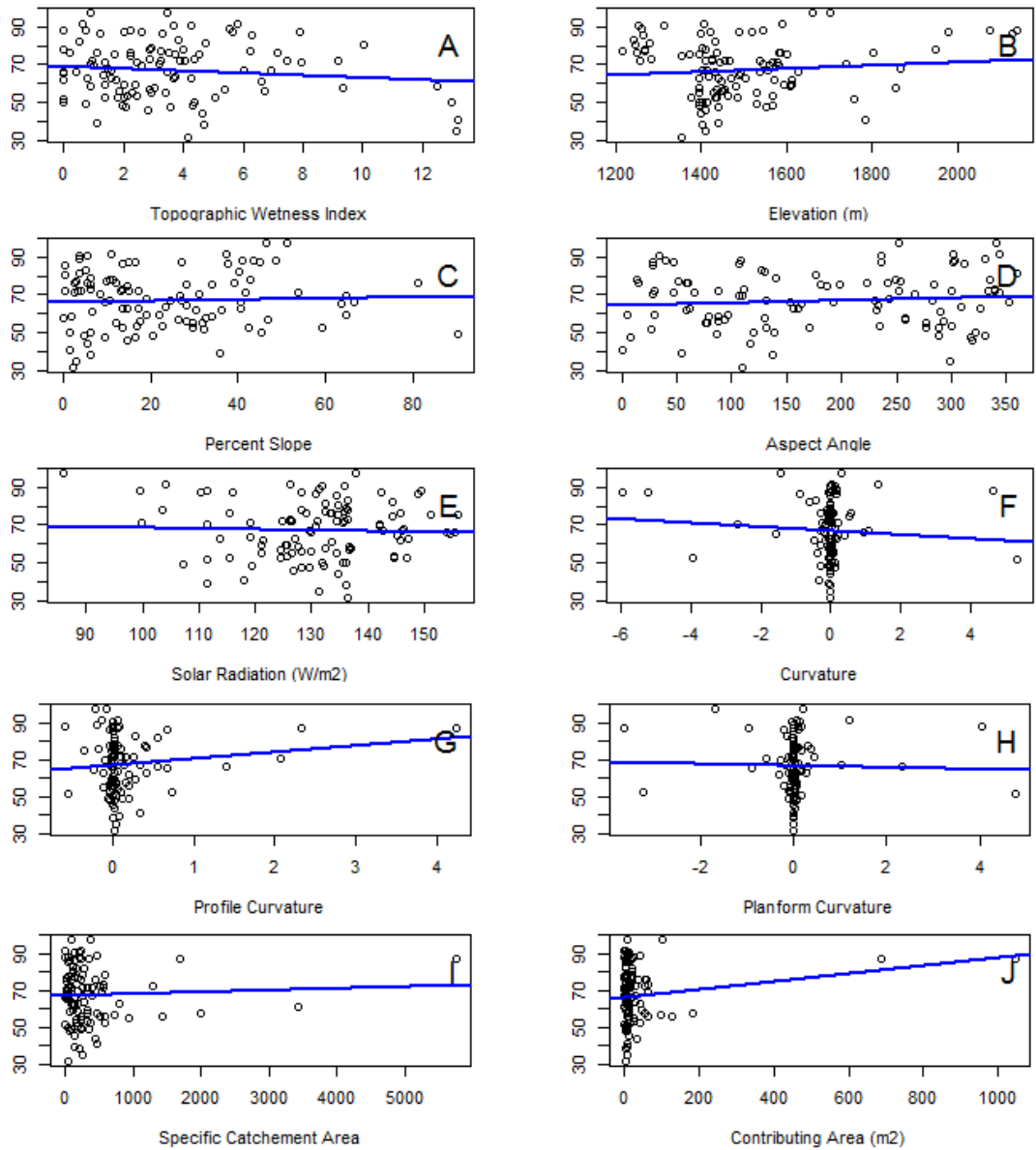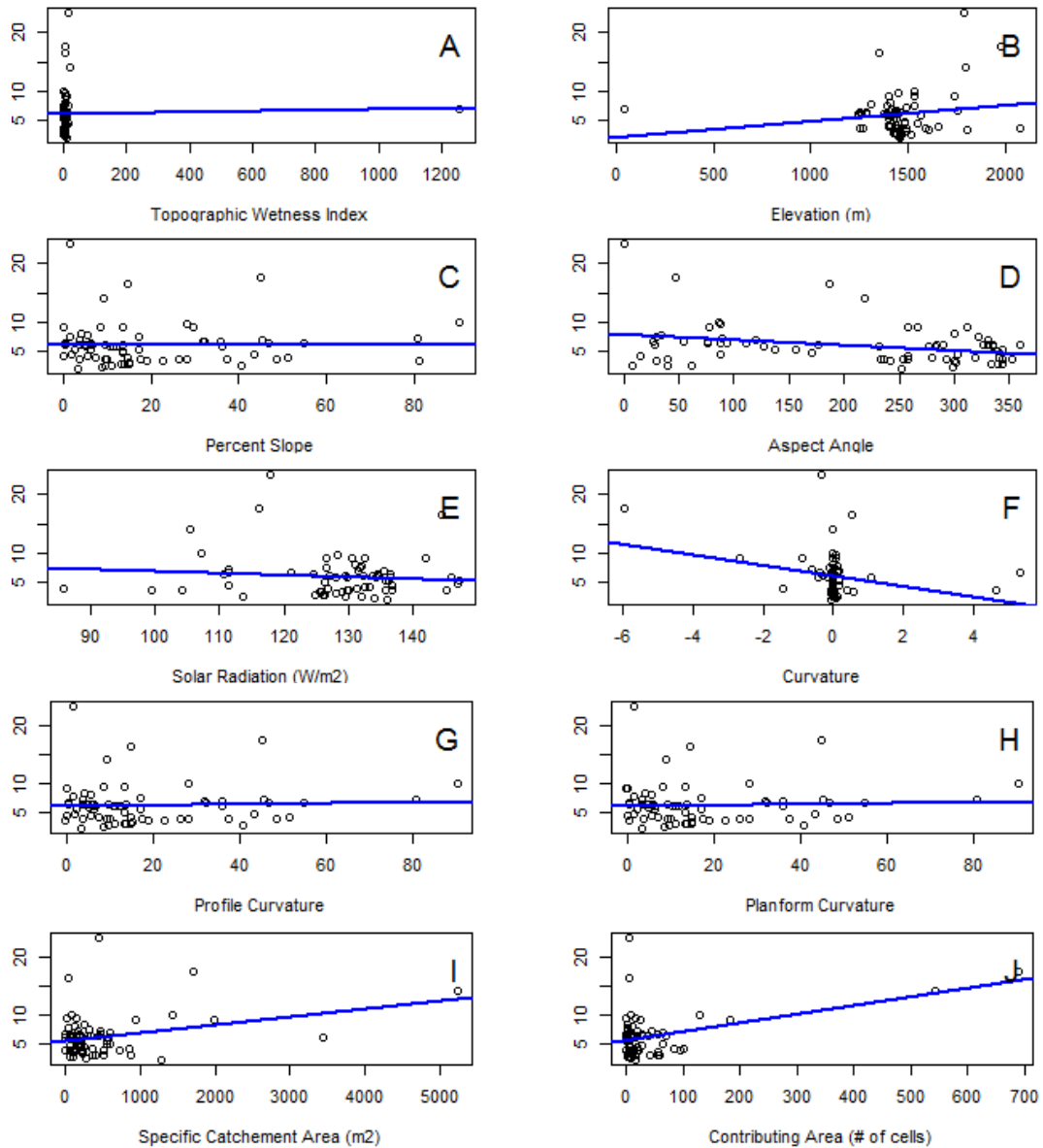
Figure B.6: Scatter plots of B horizon percent sand (on the y-axis) and all the continuous environmental variables on (the x-axis). TWI (R^2 =0.0159), Elevation (R^2 =0.0109), Slope (R^2 =0.0020), Aspect (R^2 =0.0088), Solar Radiation (R^2 =0.0019), Curvature (R^2 =0.0069), Profile Curvature (R^2 =0.0181), Planform Curvature (R^2 =0.0008), Specific Catchment Area (R^2 =0.0024), and Contributing Area (R^2 =0.0325).
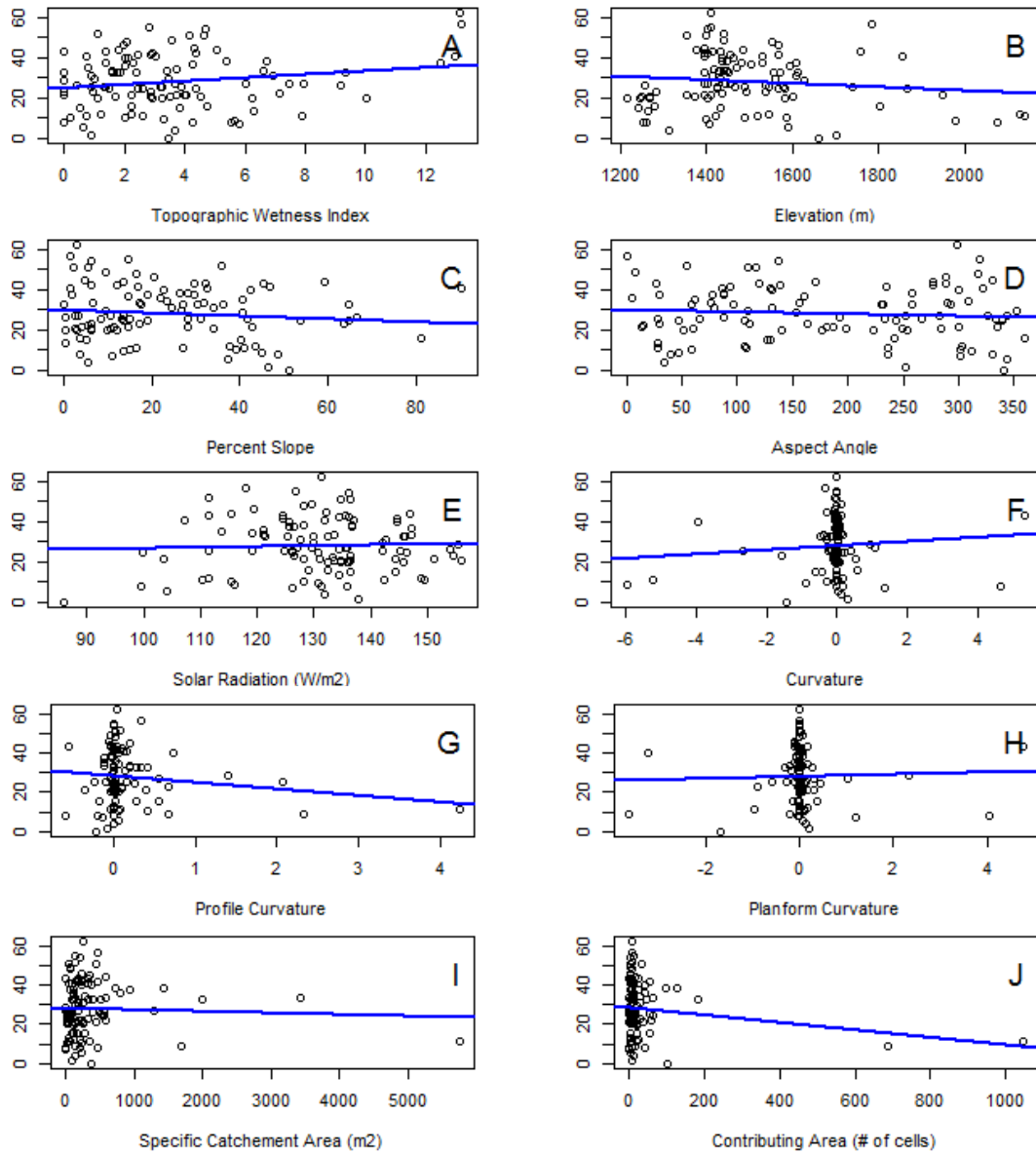
Figure B.7: Scatter plots of B horizon percent silt (on the y-axis) and all the continuous environmental variables on (the x-axis). TWI (R^2 =0.0307), Elevation (R^2 =0.0131), Slope (R^2 =0.0105), Aspect (R^2 =0.0063), Solar Radiation (R^2 =0.0022), Curvature (R^2 =0.0086), Profile Curvature (R^2 =0.0196), Planform Curvature (R^2 =0.0014), Specific Catchment Area (R^2 =0.0017), and Contributing Area (R^2 =0.0319).
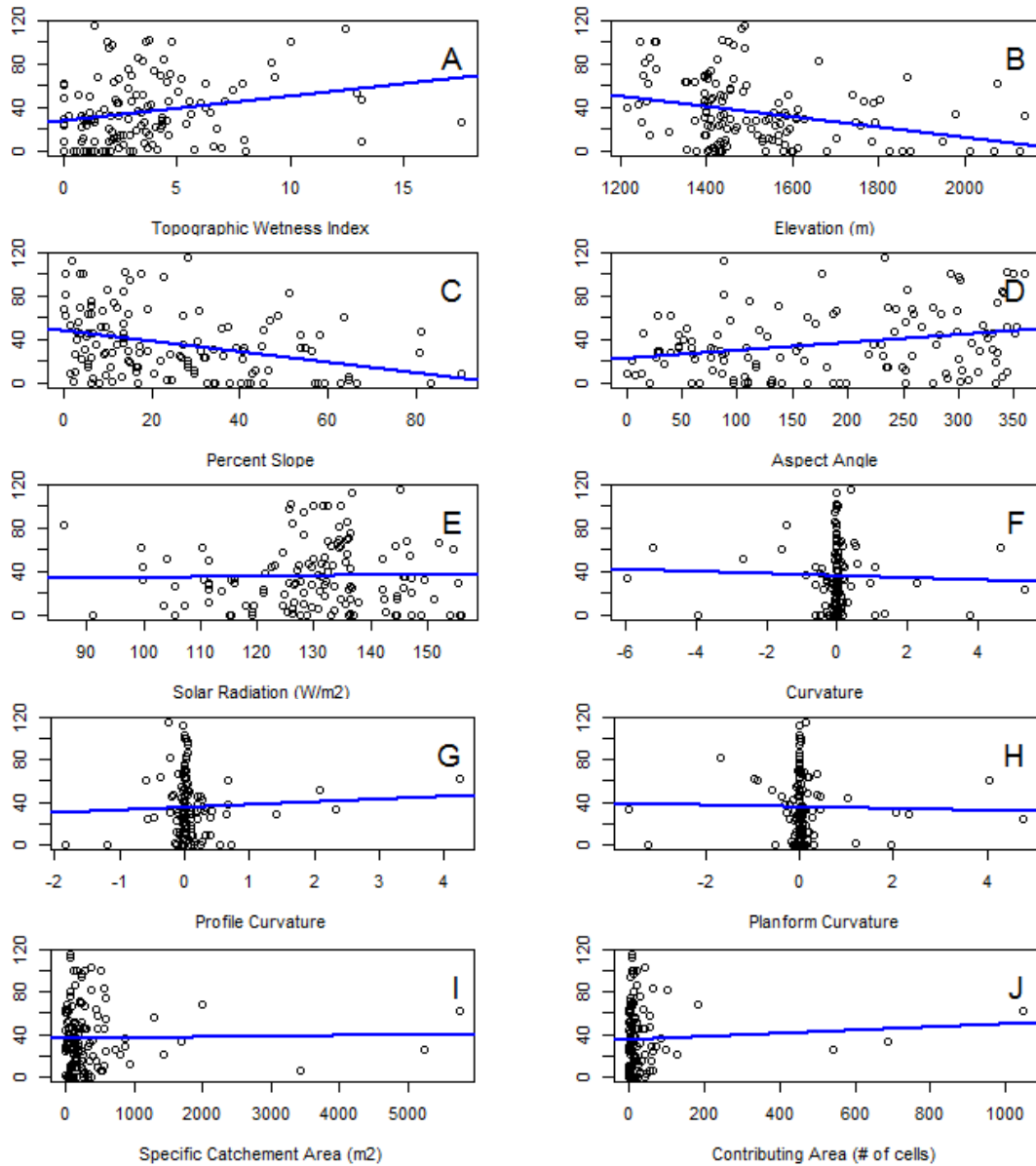
Figure B.8: Scatter plots of B horizon percent clay (on the y-axis) and all the continuous environmental variables on (the x-axis). TWI (R^2 =0.0238), Elevation (R^2 =2.5604e-08), Slope (R^2 =0.0488), Aspect (R^2 =0.0092), Solar Radiation (R^2 =1.9180e-06), Curvature (R^2 =3.4066e-05), Profile Curvature (R^2 =0.0008), Planform Curvature (R^2 =0.0007), Specific Catchment Area (R^2 =0.0028), and Contributing Area (R^2 =0.0059).
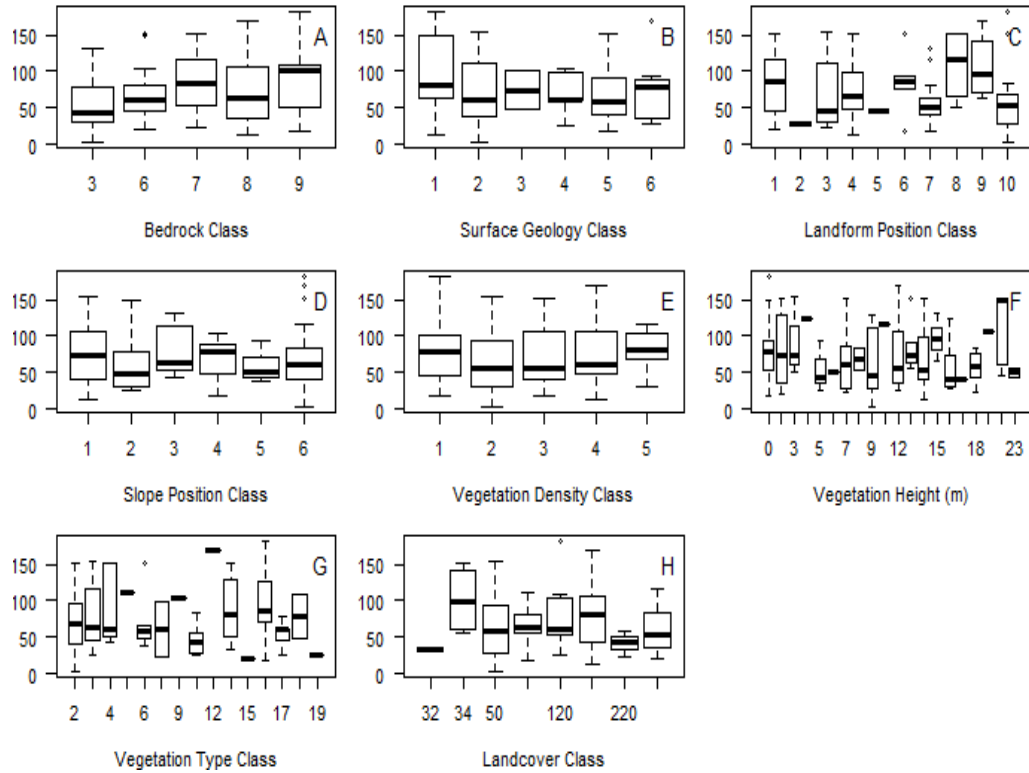
Figure B.9: Scatter plots of A horizon percent organic matter (on the y-axis) and all the continuous environmental variables on (the x-axis). TWI (R2 =0.0027), Elevation (R2 =0.0199), Slope (R2 =0.0097), Aspect (R2 =0.0065), Solar Radiation (R2 =0.0083), Curvature (R2 =0.0028), Profile Curvature (R2 =0.0028), Planform Curvature (R2 =0.0016), Specific Catchment Area (R2 =0.0257), and Contributing Area (R2 =0.0103).

Figure B.10: Scatter plots of B horizon percent organic matter (on the y-axis) and all the continuous environmental variables on (the x-axis). TWI (R^2 =0.0016), Elevation (R^2 =0.0300), Slope (R^2 =0.0002), Aspect (R^2 =0.0805), Solar Radiation (R^2 =0.0113), Curvature (R^2 =0.0904), Profile Curvature (R^2 =0.0022), Planform Curvature (R^2 =0.0021), Specific Catchment Area (R^2 =0.0916), and Contributing Area (R^2 =0.2014).

Figure B.11: Scatter plots of solum depth (cm) (on the y-axis) and all the continuous environmental variables on (the x-axis). TWI (R^2 = 0.0502), Elevation (R^2 = 0.0970), Slope (R^2 = 0.1110), Aspect (R^2 = 0.0753), Solar Radiation (R^2 = 0.0006), Curvature (R^2 = 0.0013), Profile Curvature (R^2 = 0.0021), Planform Curvature (R^2 = 0.0004), Specific Catchment Area (R^2 = 0.0005), and Contributing Area (R^2 = 0.0035).

Figure B.12: Boxplots of soil restrictive depth (cm) on the y-axis and the other variables on the x-axis.
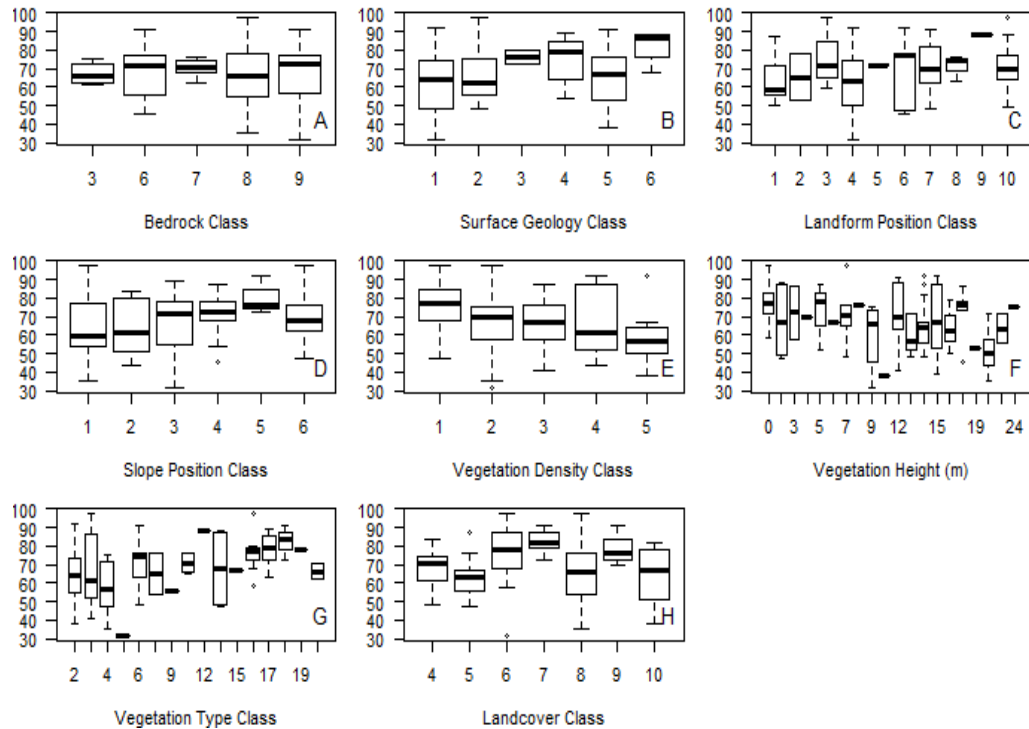
Figure B.13: Boxplots of A horizon percent sand texture on the y-axis and the other variables on the x-axis.
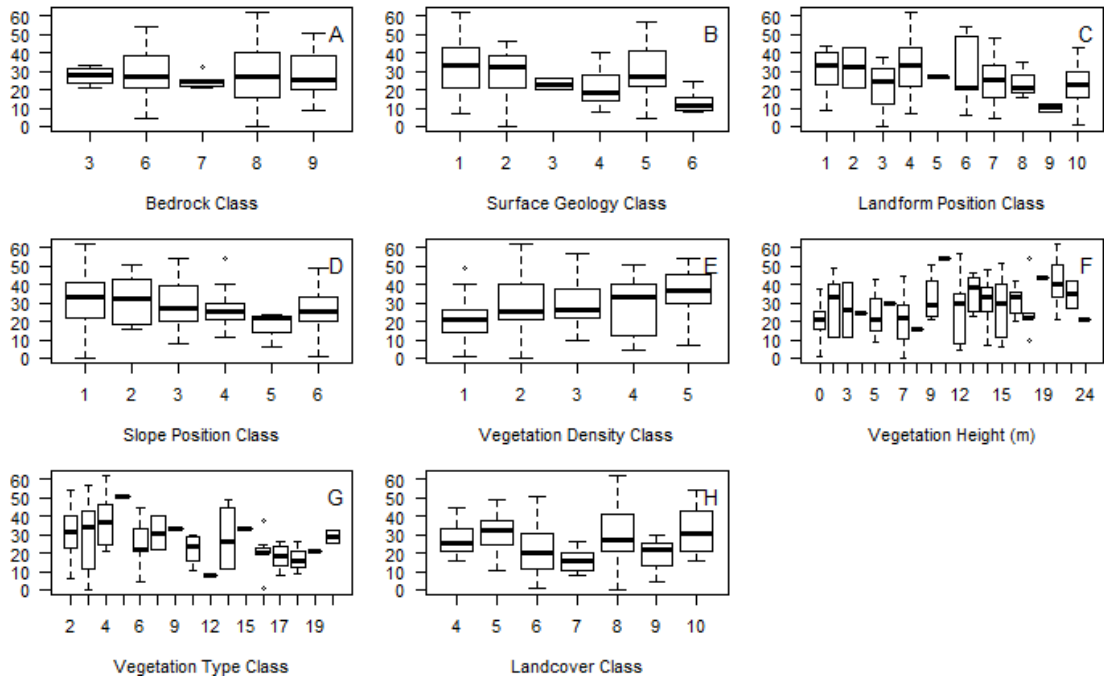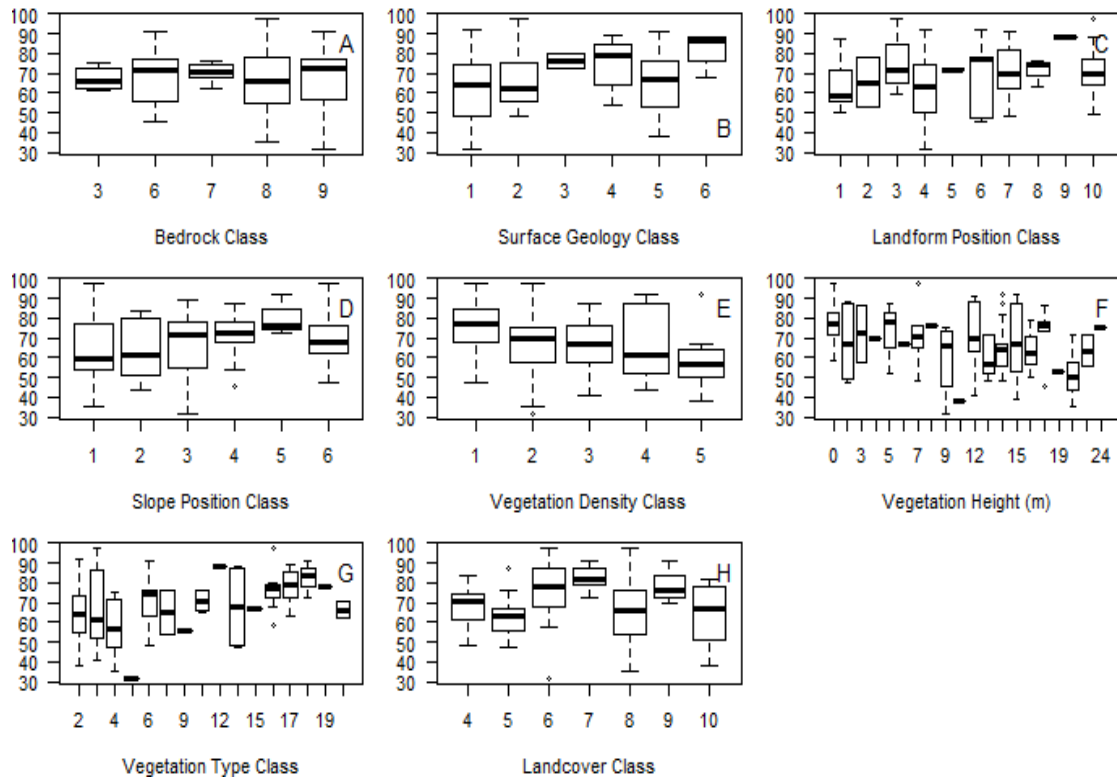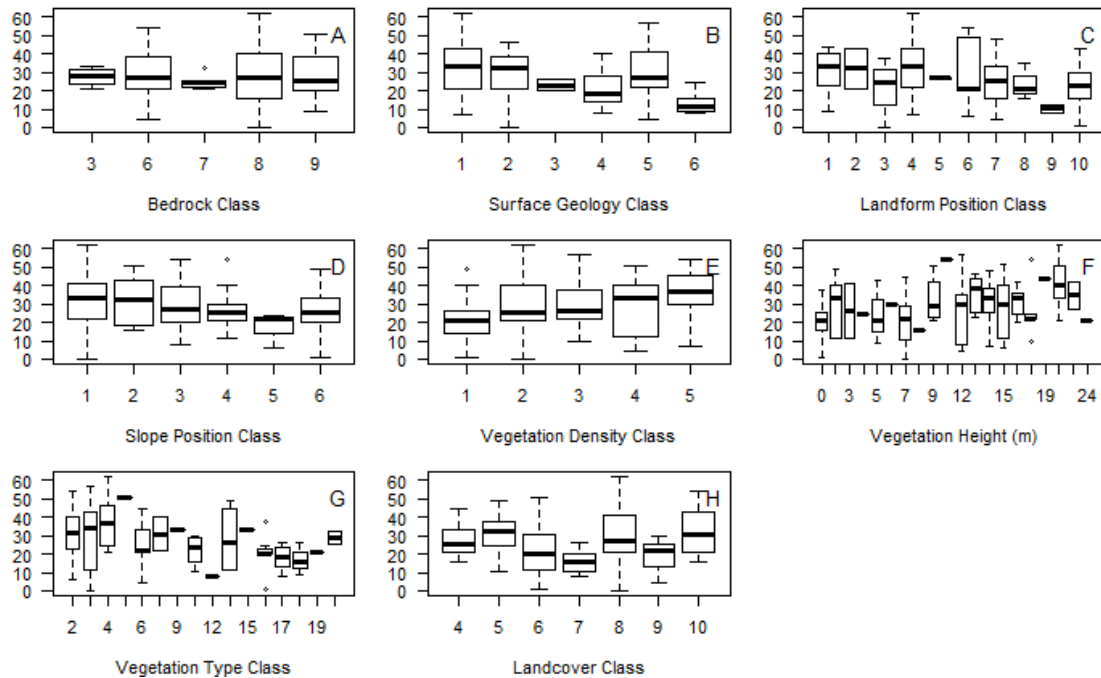
Figure B.14: Boxplots of A horizon percent silt texture on the y-axis and the other variables on the x-axis.
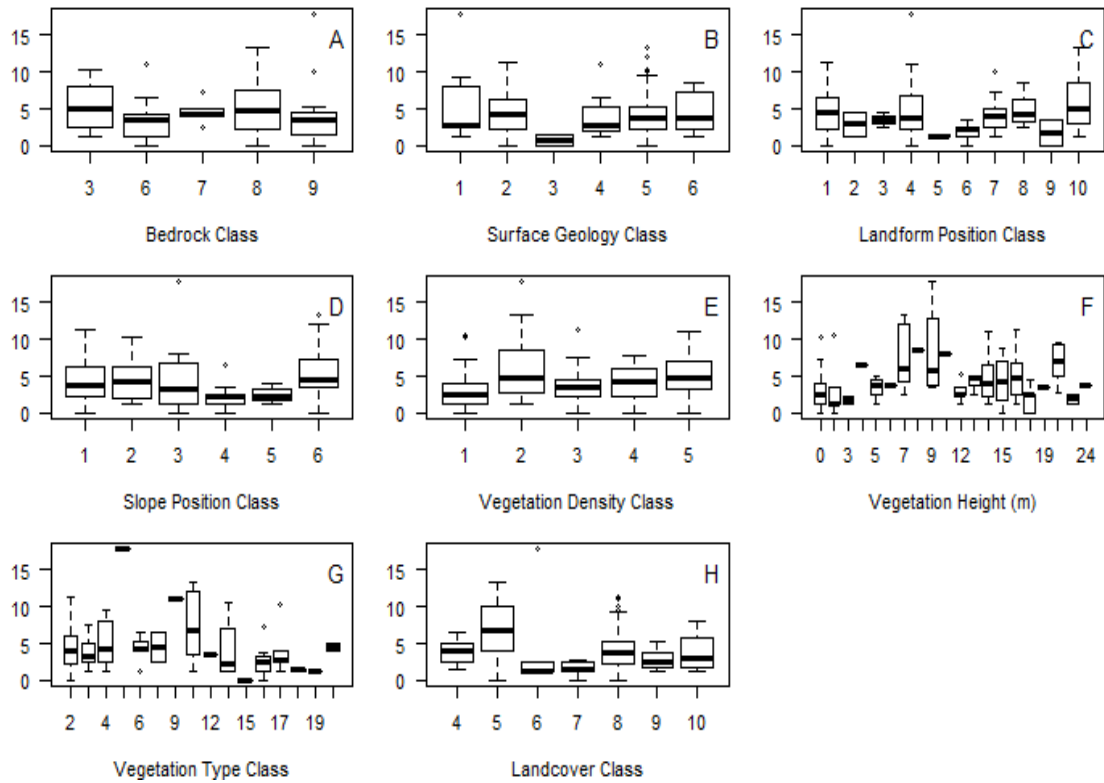
Figure B.15: Boxplots of A horizon percent clay texture on the y-axis and the other variables on the x-axis.

Figure B.16: Boxplots of B horizon percent sand texture on the y-axis and the other variables on the x-axis.

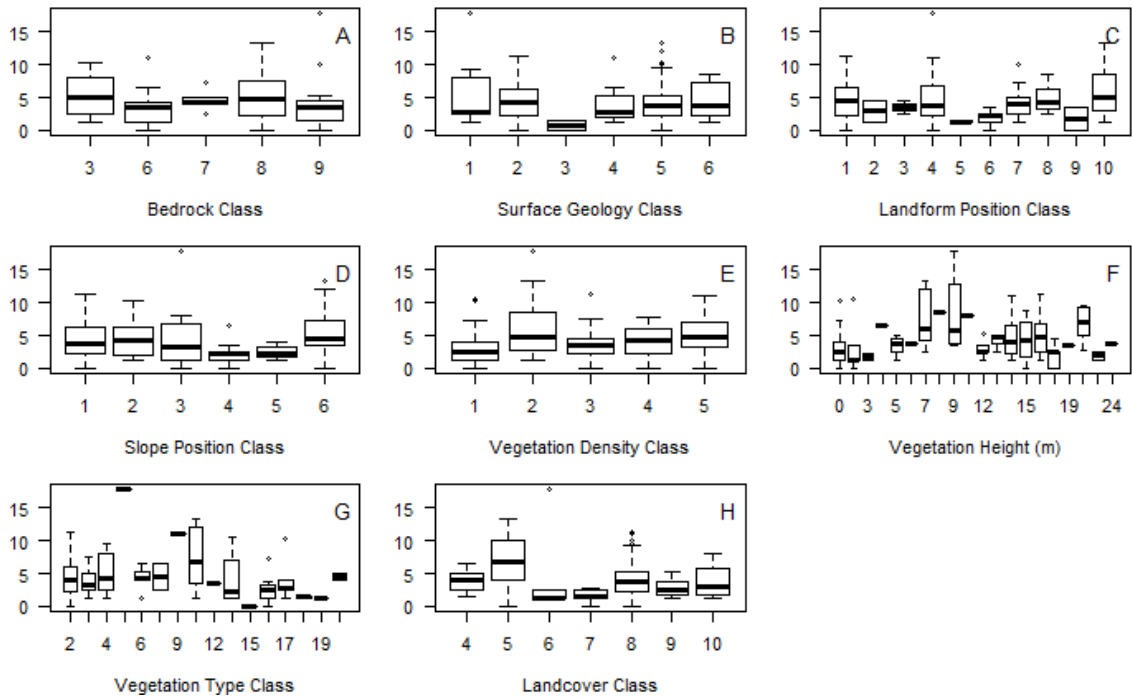**Figure B.17:** Boxplots of B horizon percent silt on the y-axis and the other variables on the x-axis.

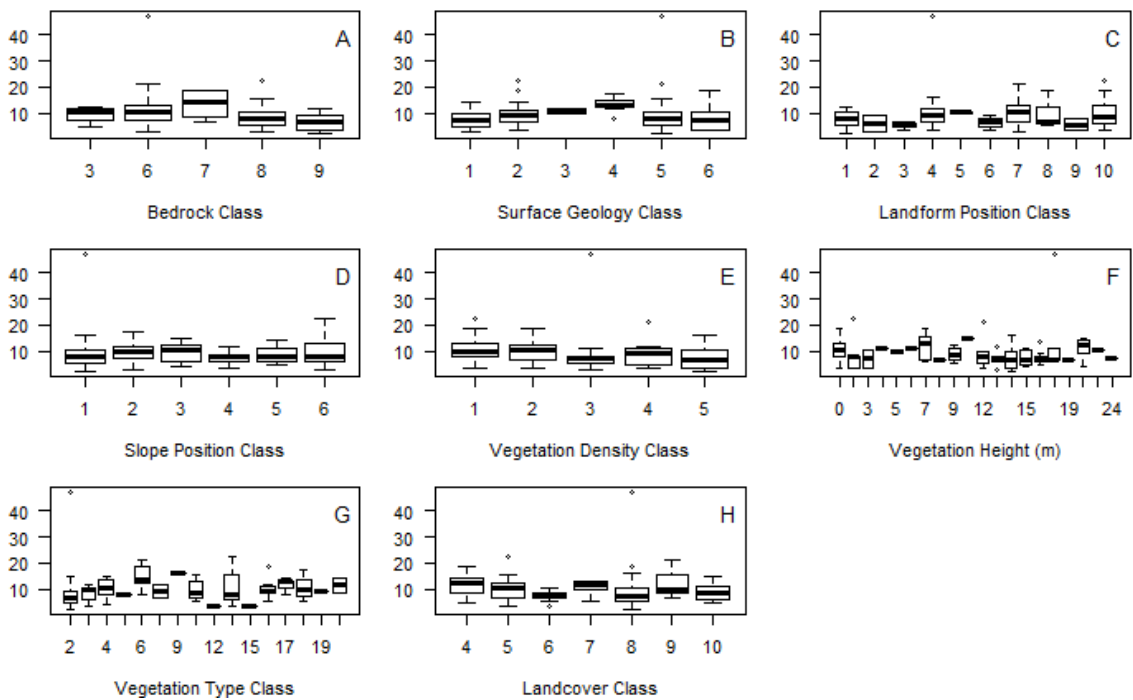**Figure B.18:** Boxplots of B horizon percent clay on the y-axis and the other variables on the x-axis.

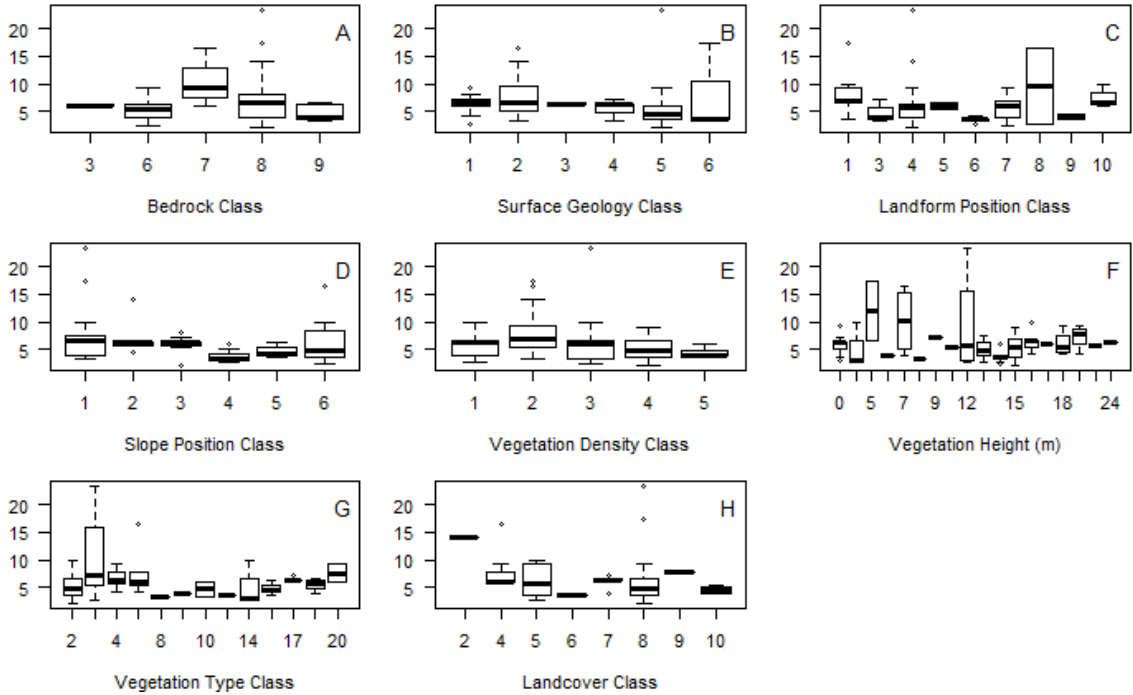**Figure B.19:** Boxplots of A horizon percent OM on the y-axis and the other variables on the x-axis.

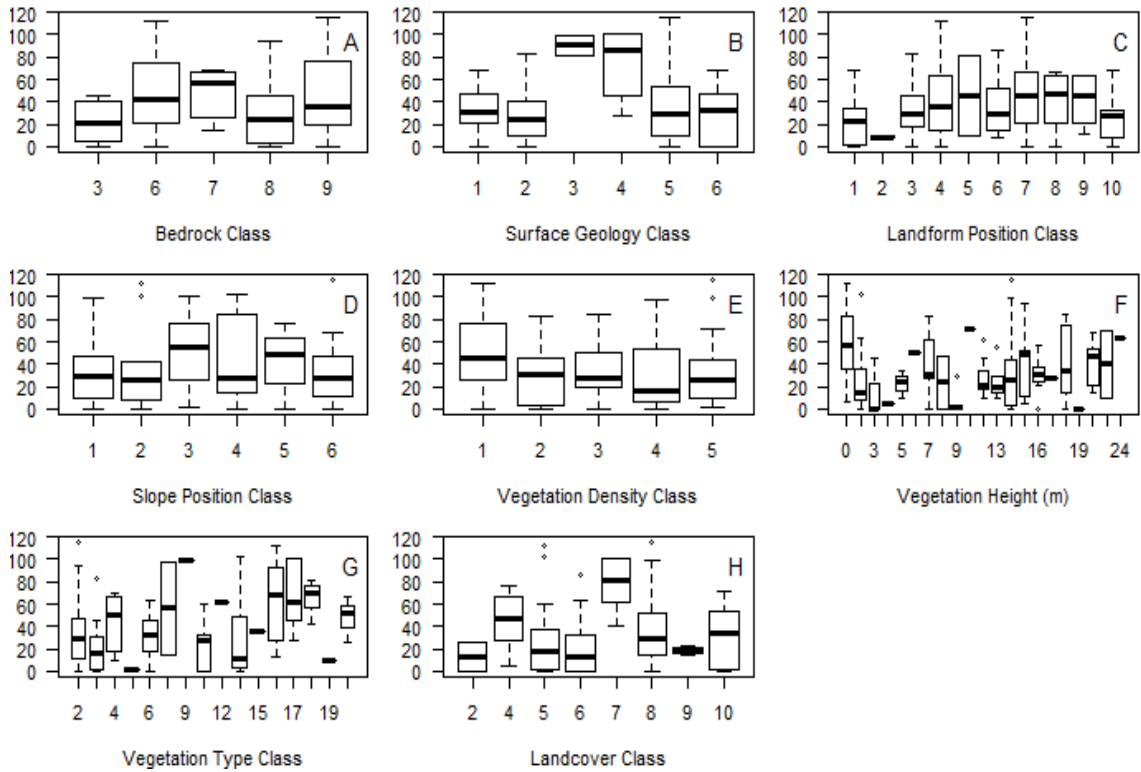**Figure B.20:** Boxplots of B horizon percent OM on the y-axis and the other variables on the x-axis.

Figure B.21: Boxplots of solum depth (cm) on the y-axis and the other variables on the x-axis.