A HIGH-RESOLUTION DIGITAL SOIL MAPPING FRAMEWORK FOR NEW BRUNSWICK, CANADA

by

Shane Robert Furze

Master of Environmental Management, University of New Brunswick, 2012
Bachelor of Science in Environmental Science in Biology,
St. Francis Xavier University, 2010

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

Doctor of Philosophy in Forestry

in the Graduate Academic Unit of Forestry and Environmental Management

Supervisor: Paul Arp, PhD, Forestry and Environmental Management

Examining Board: Paul Arp, PhD, Forestry and Environmental Management
Sheng Li, PhD HRA, Forestry and Environmental Management
Joe Nocera, PhD, Forestry and Environmental Management
Raid Al-Tahir, PhD, Geodesy and Geomatics

External Examiner: Margaret Schmidt, PhD, Geography, Simon Fraser University

This dissertation is accepted by

The Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

July, 2018

ABSTRACT

For decades researchers have been studying forest soils and summarizing findings in the form of soil surveys with thematic soil maps depicting soil associations, broad polygons representing groups of individual soil types. With growing availability of high-resolution spatial data, it has become possible to model and map how individual soil properties vary, both spatially and with depth, across the landscape at high resolution. This dissertation demonstrates how this can be accomplished for the Province of New Brunswick (NB), Canada by way of digital soil mapping (DSM) based on (i) existing soil information and related data sets, (ii) principles of soil formation as dictated by location-specific changes in topography, surficial geology, and climate. For this purpose, existing elevation data sets were fused via error reduction procedures to generate a comprehensive province-wide digital elevation model (DEM) at 10m resolution. The resulting DEM was then used to delineate a variety of data sets detailing spatial variations in topography, hydrology, and climate. Various sources of spatial geology depictions were combined by way of similarities in classifications resulting in re-delineations of landform and lithological attributes. In combination, the data layers generated were used to determine how specific soil properties (n = 12,058) vary, both spatially and with increasing depth, across the province at 10m resolution. These determinations were made possible by way of machine-based random forest regression modelling.

This dissertation provides details in terms of how (i) a province-wide soil database was generated from existing soil survey reports, (ii) how missing soil data were substituted through the process of pedotransfer function development and analysis, (iii) how the province-wide DEM layers were fused, and (iv) how the DSM procedure was formulated and executed. The soil properties selected for modelling and mapping purposes refer to soil depth, drainage, bulk density, texture, coarse fragment content, and soil organic matter content. In turn, these properties, in combination with spatial data sets (topography, geology, and climate), can be used to model and map other soil variables such as, e.g., pH, soil water retention at field capacity and permanent wilting point, and cation exchange capacity.

DEDICATION

I dedicate this dissertation to future digital soil mapping research and to those who embark on the adventure of soil studies. May this research be of some assistance in better understanding soil property variations and methods to model soil properties.

ACKNOWELDGEMENTS

I would like to thank Dr. Paul Arp and the members of the Forest Watershed Research Team, namely Marie France Jones, Mark Castonguay, and Jae Ogilvie for their technical assistance, brainstorming, and ongoing support. I would like to sincerely thank Dr. Arp for the creativity and passion he instills while providing ongoing support, wisdom, and insights.

I would also like to thank the members of NB Department of Energy and Resource Development, including Chris Hennigar, Toon Pronk, Serge Allard, and Will Gilmore for not only providing me with pertinent data, but also updates and feedback.

I would like to thank those who helped fund this research: J.D. Irving, NSERC, NBIF, and AWARE. This financial support allowed me to continue my research and explore many possibilities while providing me with excellent networking experiences.

Finally, I would like to thank my family and Stacey for their endless support and encouragement during difficult times and throughout this entire process.

SF

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF ACRONYMS, UNITS, AND ABBREVIATIONS

Units

°C – Celsius, measurement of temperature

Atm – Atmospheres, unit of pressure

Bar – Bar, unit of pressure

cm – Centimeter, unit of length

eq – Equivalents, measure of concentration

g – Gram, unit of mass

ha – Hectares, unit of area

kg – Kilogram, unit of mass

kPa – Kilopascal, unit of pressure

meq – Milliequivalents, measure of concentration

mm – Millimetre, unit of length

Chemical Formulas

H_2O – Water

Ca – Calcium

$CaCl_2$ – Calcium Chloride

K – Potassium

K_2O – Potassium Oxide

Mg – Magnesium

NaCl – Sodium Chloride

Nh_4OAc – Ammonium Acetate

Statistics

MAE – Mean Absolute Error, unit of statistical importance

RMSE – Root Mean Square Error, unit of statistical importance

Geospatial Data

ASTER – Advanced Spaceborne Thermal Emission and Reflectance

CDED – Canadian Digital Elevation Data

DEM – Digital Elevation Model

DSM_x – Digital Surface Model

DTM – Digital Terrain Model

DTW – Cartographic Depth-to-water table index

LiDAR – Light Detection and Ranging

IDW – Inverse Distance Weighted, interpolation technique

SRTM – Shuttle Radar Topography Mission

Soils

BS – Base Saturation, unitless

CEC – Cation Exchange Capacity, measured in meq/100g

CF – Coarse Fragment content, measured in %

Col – Colluvial parent material

Db – Bulk Density, measured in g/cm^3

Dp – Particulate Density, measured in g/cm^3

DSM – Digital Soil Mapping

FC – Field Capacity, measured as % at -33kPa

FSNB – Forest Soils of New Brunswick

GLMA – Glaciomarine-derived parent material

MDEP – Mode of deposition of parent material

NSDB – National Soils Database

OC – Organic carbon, measured in %

OM – Organic Matter content, measured in %

PTF – Pedotransfer Function

PWP – Permanent Wilting Point, measured as % at -1500kPa

SNB – Soils of New Brunswick

SOM – Soil Organic Matter content, measured in %

Res – Residual parent material

Government Agencies

CANSIS – Canadian Soil Information Service

NBDEM – New Brunswick Department of Energy and Mines

NBDNR – New Brunswick Department of Natural Resources

NBSNB – Service New Brunswick

NrCAN – Natural Resources Canada

Geographic Locations

NB – New Brunswick, Canada

NS – Nova Scotia, Canada

ON – Ontario, Canada

CHAPTER 1 - INTRODUCTION

1.1. OVERVIEW

Soil is a complex and essential, yet non-renewable, natural resource which plays a critical role in policy and resource management (Adhikari *et al.*, 2012; Cambule *et al.*, 2013; Poggio *et al.*, 2013). As such, this resource is gaining increasing attention with growing concerns around climate change adaptation and how soil properties vary with changing environments (Grimm and Behrens, 2010; Häring *et al.*, 2012; Poggio *et al.*, 2013). Of particular importance, especially for precision agriculture and forestry, is knowing how soils and their properties vary across the landscape and how these variations influence root development and the movement, and storage, of soil carbon, nutrients, minerals, and water (Florinsky *et al.*, 2002; Smith *et al.*, 2006; Carré *et al.*, 2007; Keys, 2007; Grimm and Behrens, 2010; Häring *et al.*, 2012; Weil and Brady, 2017).

Soil formation relies on five primary factors: parent material, relief (topography), vegetation, climate, and time (Jenny, 1941; Klingebiel *et al.*, 1988; Valladares and Hott, 2008; Adhikari *et al.*, 2012; Cambule *et al.*, 2013). Therefore, understanding the variation in these factors for any given location can provide a glimpse into soil conditions, and in turn, may allow for better understanding the relationships between soil properties and vegetative productivity. Conventionally, soils are depicted as choropleths, i.e., tessellated boundaries that hierarchically group soils by similarities in landforms (i.e. soil associations, distinguished by morphological features and mineralogical composition) as illustrated in Fig. 1.1 (Colpitts *et al.*, 1995; Fahmy *et al.*, 2010). This form of mapping

represents a generalized simplification of surveyed soil properties (McBratney *et al.*, 2000a; MacMillan *et al.*, 2005; Mora-Vallejo *et al.*, 2008). The resulting map units, however, are too broad to reflect many of the soil variations within these units as affected by elevation, drainage, type of vegetation cover, surface exposure, water flow, and gravitation influences (Odgers *et al.*, 2014).

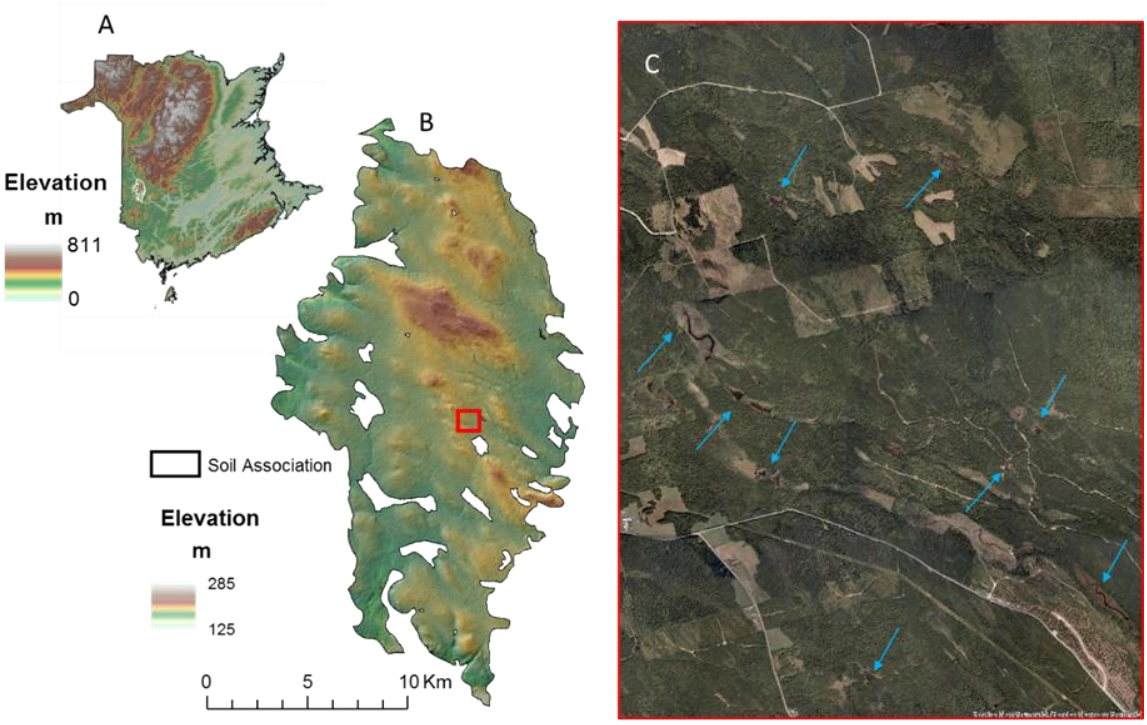

Figure 1.1. Visual example of one of Holmesville soil associations within NB with extent outlined in white (**A**), range of elevation within the association (**B**), and range in varying land types (organic) and land management practices (both agriculture and forestry) within small subset of association overlain on GeoNB basemap imagery (**C**). This one association covers 35,683ha (0.6% of NB).

Traditionally surveyed and delineated soil associations are limited, specifically due to (Pitty, 1979; Moore *et al.*, 1993; Zhu and Mackay, 2001; Heung *et al.*, 2014; Odgers *et al.*, 2014):

1. Soil surveys were developed for assessing land uses instead of soil property variability.
2. Soil properties vary continuously and do not abruptly change at defined boundaries (Simbahan *et al.*, 2006; Odgers *et al.*, 2014).
3. Relationships between soil properties and landform/landscape position were often missed (MacMillan *et al.*, 2005).
4. Locations of individual soil types within each, and across, soil associations remain unknown with reference to actual landform extent and topographic delineations (Dobos *et al.*, 2000),
5. Survey objectives were limited in various ways: time and available financial support, lack of survey-supporting data layers, and non-standardized data collection and laboratory procedures (Park *et al.*, 2001).

With increasing availability of high-resolution datasets within geographic information systems, pertaining to, e.g., digital elevation models (DEMs) (with resolutions ranging between 50 - 100cm) and DEM-derived secondary models (topographic and climate models), it is possible to ascertain how soil properties, those of which influence the rooting medium, vary spatially across the landscape, from upper reaches of hill crests to valley bottoms and stream banks, with properties focusing on soil depth, texture, organic matter content (SOM), bulk density (Db), coarse fragment (CF) content, pH, cation exchange capacity (CEC), and moisture retention at field capacity (FC) and permanent wilting point (PWP). This is feasible by way of digital soil mapping (DSM):

1. Developing pedotransfer functions (PTFs) (Bouma, 1989), mathematical relationships between soil properties, and
2. Producing spatially-continuous soil attribute maps by modeling relationships between geographically-explicit soil profiles across the landscape with data sets representing surficial geology, climate, and topographic information (from DEMs).

Digital soil mapping (DSM) is a growing science directed toward modeling soil properties spatially across the landscape at a continuous extent by resolving the issue of soil boundary discretization and omission of soil property inter-variability per soil association (McBratney *et al.*, 2003; Smith *et al.*, 2006). This approach is based on three fundamental principles:

1. pedometrics, mathematical and statistical approaches to modeling soil properties and relationships (Florinsky, 2012),
2. soil-landscape relationships (Gerrard, 1981; Birkeland, 1999; McBratney *et al.*, 2000; MacMillan *et al.*, 2005; Barka *et al.*, 2011), and
3. soil formation and soil forming factors (Jenny, 1941; Birkeland, 1999; Wu *et al.*, 2008; Adhikari *et al.*, 2012).

With DSM, soil properties from field-collected soil samples can be statistically compared to DEM-derived topographic derivatives, climatic and geological datasets from ancillary data sources, and remote sensing techniques (Moore *et al.*, 1993b; Odeh *et al.*, 1994; Gessler *et al.*, 1995), following the framework of the well-known soil-formation

model, 'CLORPT', as outlined by Jenny (1941). With advancements in soil science and geospatial analyses, McBratney *et al.* (2003) introduced 'SCORPAN' to also incorporate a spatial component, neighborhood (N), into the modeling framework. The SCORPAN model suggests that a soil property at a given spatial location, $Soil$, is a function of other soil properties at the same spatial location (S), climate (C), organisms (O), relief (topography) (R), parent material (P), age (A), and neighborhood (N) (McBratney *et al.*, 2003; Florinsky, 2012; Cambule *et al.*, 2013). This model coincides with enhancements in geospatial analyses and the ability to spatially compare soil properties to underlying ancillary data sources for any given location.

This dissertation introduces and explores the development of a DSM framework for NB, Canada, with the purpose of producing continuous soil physical and chemical property maps across the landscape. NB was chosen as a study extent due to the wealth of existing data and the variability in soil forming factors. These factors refer to changing temperature and precipitation conditions from lowlands in the southeast to highlands in the northwest, changing glacial landscapes, and topographic expressions varying from flat lowlands to steeply sloping hillsides (Pronk and Ruitenberg, 1991; Colpitts *et al.*, 1995; Fahmy *et al.*, 2010).

1.2. RESEARCH QUESTIONS

1. To what precision can pedotransfer functions be derived from county-based soil surveys?

2. How can spatial models be derived for use in predicting soil properties at a continuous extent?

3. To what extent can the underlying relationships between soil formation and soil forming factors be modeled by way of digital soil mapping for predicting soil properties spatially?

1.3. OBJECTIVES

The objective of this dissertation was to develop a framework with functioning models for predicting soil physical and chemical properties continuously across the landscape for the province of NB, Canada as a case study. The process to complete this task was applied as follows:

1. County-based soil surveys and current soil maps were utilized to develop an aspatial database, and, in turn, produce pedotransfer functions relating soil attributes to one another.

2. Available DEMs for NB were amalgamated resulting in a new DEM via open-sourced DEM fusion. In turn, DEM-derived topographic and hydrographic datasets were developed.

3. Continuous climate data sets pertaining to temperature and precipitation were developed for NB by way of comparing historical records for spatially explicit weather stations from NB and Ontario to underlying DEM and spatial location.

4. Existing surficial geology delineations were updated via similarity modeling comparing different data sets representing parent material mode of deposition and primary lithology.

5. Spatial database of soil properties was developed via amalgamating field-collected soil profiles from numerous sources (n = 12,058).

6. Statistical comparison of soils information in spatial database to underlying geologic, topographic, and climatic datasets, and conventional soil maps was conducted using Random Forest machine learning algorithm with model results utilized to develop spatial continuum maps for specific soil properties.

7. By mapping important soil properties such as depth, bulk density (Db), texture (% sand, silt, and clay), coarse fragment content (% CF), and soil organic matter content (% SOM), model results can be expanded with pedotransfer functions to predict additional soil properties spatially across landscape of New Brunswick at 10m resolution.

1.4. SUBJECT MATTER

This dissertation is completed in "article" format with the body consisting of four distinct chapters (i.e., Chapters 2 through 5), as outlined below. The final chapter (Chapter 6) provides a summary of the material, overall conclusions from this research, and recommendations for future research.

Chapter 2 introduces and describes the development of an aspatial database from the amalgamation and harmonization of historical county-based soil surveys.

Chapter 3 presents the development of soil physical and chemical PTFs, derived from the aspatial database from Chapter 2. Soil physical PTFs were compared to other published PTFs to test the overall performance.

Chapter 4 describes the development of a method to systematically and comprehensively reduce DEM errors across New Brunswick based on fusing province-wide DEM layers from various sources, and calibrating the result using select LiDAR-generated DEMs for final elevation calibration.

Chapter 5 introduces the development of a spatial database from field-collected samples and resulting DSM spatial models by comparing this database to underlying topographic, geological, and climatic data sets. This chapter also describes how these data sets were adjusted prior to spatial modeling. With this, soil properties of drainage, horizon depths, sand, silt, clay, CF, Db, and SOM were modeled in continuum format at 10m resolution.

Chapter 6: provides an overall summary of this dissertation, a statement of original contributions, and recommendations for further development and research.

CHAPTER 2 – AMALGAMATION AND HARMONIZATION OF SOIL SURVEY REPORTS INTO A MULTI-PURPOSE DATABASE: AN EXAMPLE

Shane Furze, Paul Arp

Faculty of Forestry and Environmental Management

University of New Brunswick, Fredericton, NB, Canada, E3B 6C2

Foreword:

The following chapter is an article submitted to the Open Journal of Soil Science. It was submitted on May 31st, 2018.

Citation:

Furze, S. and Arp, P.A. 2018. Amalgamation and Harmonization of Soil Survey Reports into a Multi-Purpose Database: An Example. *Open Journal of Soil Science* (submitted).

2.1. ABSTRACT

This article describes procedures used to generate an aspatial, terminologically-consistent, province-wide database for forest soils from soil survey reports, with New Brunswick, Canada, serving as an example. The procedures involved summarizing existing soil information into soil associations via similarities in landform and parent material. Consistent soil associations were developed, and pedologically-correct horizon sequences were assigned to each soil associate within each soil association, all done with reference to soil-forming factors. These factors refer to: (i) soil parent materials as classified by mode of deposition and lithology, (ii) topographic surface expressions, (iii) soil drainage, and (iv) dominant vegetation type. Additionally, each associate was characterized with horizon–specific physical and chemical soil properties distinguished by depths. An

amalgamated database containing 106 soil associations, 243 soil associates (differing within the soil associations by drainage only), and 522 soil profiles was developed.

Key Words: Soil surveys, amalgamation, associations, associates, parent materials, landforms, lithology, drainage, profiles, horizons, properties.

2.2. INTRODUCTION

Soils are part of the natural environment consisting of complex interactions between living organisms and soil forming factors pertaining to geology, climate, topography, organisms, and time (Jenny, 1941; Birkeland, 1999; Adhikari *et al.*, 2012). As such, soils vary spatially in type (soil profile) and spatial extent. This is generally reflected by most soil survey reports which group surveyed soil units into landform- and lithology-defined soil associations (McKeague and Stobbe, 1978). Cross-referencing these reports to one another, however, revealed numerous inconsistencies in terms of naming and labelling similar soils types and horizons. This is in part due to changing soil classification and mapping protocols over a period of roughly 70 years. In detail, some of the inconsistencies refers to:

1. Changes in soil survey methods, including sampling strategies, laboratory analyses, and quantitative units for reporting results (McKeague, 1978; Group, 1981; Working Group on Soil Survey Data, 1982; Guertin *et al.*, 1984);

2. Incompleteness in terms of reporting small to large scale variations in soil associates, varying in scale and resolution, with main focus on agricultural lands

(Pitty, 1979; Zhu and Mackay, 2001; Adhikari *et al.*, 2012; Odgers *et al.*, 2014). Past survey practices generally addressed soil variations at the 1:10,000 scale (and often coarser), and were therefore generally mute about small scale variations. As a result, spatial pedological variations which influence crop, forest productivity and root growth via nutrient and water retention, remained unrecognized (Parr *et al.*, 1992; Southorn, 2003; Keys, 2007; Taylor *et al.*, 2013).

3. Implied differences in soil association conditions and extent across arbitrary survey boundaries.

Typically, each soil survey report included 3 sections (e.g., Poitras map unit, 2615 ha; Langmaid et al. 1980):

Section 1: an overview of soil associates including information regarding association, landform, lithology, vegetation, drainage, and topography,

Section 2: a profile description for sampled soil associates which included field-based measurements for horizons, depths, texture, CF content, structure, root presence, mottling (if applicable), and pH, (Fig. 2.1):

The Poitras soils, which dominate this map unit, are the very poorly drained members of the Holmesville catena. These soils occur in depressions and are classified as Orthic Gleysols. The silt loam phase occupies 1526 ha, and the fine sandy loam phase occupies 206 ha. The vegetation consists of black spruce, eastern white cedar, alder, balsam fir, running club moss, spinulosa wood-fern, goldthread, sphagnum mosses, tree moss, haircap moss, and sedges. The description of a moist undisturbed profile follows:

Horizon	Depth (cm)	
L	Trace	Leaf litter, mostly moss and twigs; pH 4.2.
F	5-3	Brown semidecomposed litter and roots, felted; pH 4.2.
H	3-0	Black well-decomposed organic material; moderate, medium granular; pH 4.2.
Aeg	0-25	Light gray (10YR 7/1); gravelly loam; many, coarse, prominent very pale brown (10YR 7/4) mottles; massive; plastic, sticky; abrupt, smooth to slightly wavy boundary; pH 4.7.
Bg	25-55	Light olive gray (5Y 6/2); loam; many, medium, prominent yellowish brown (10YR 5/6) mottles; some of these are concretions 2-4 mm in diameter; strong, coarse, granular; plastic, slightly sticky; clear, wavy boundary; pH 5.5.
Cg1	55-80	Light olive gray (5Y 6/2); gravelly loam; common, medium, prominent brownish yellow (10YR 6/6) mottles; very weak, platy; plastic, slightly sticky; pH 5.4.

Figure 2.1. Example of information obtained from soil surveys and utilized in developing the database, including general information (section 1) (**top paragraph**) and field-based measurements for each horizon (Section 2) (**bottom descriptions**) separated by dotted line. The example provided represents the Poitras soil associate retrieved from Langmaid et al. (1980).

Section 3: an overview of lab-measured physical and chemical properties by soil associate (e.g., % carbon, % sand, silt and clay, Db, field capacity (FC), permanent wilting point (PWP)), (Fig. 2.2):

Horizon	Depth cm	pH	% Total C	% Total N	Cation exchange capacity, meq/100 g	Exchangeable cations, meq/100 g			Base saturation %
						Ca	Mg	K	
									Poitras
L-H	5-0	4.2	41.40	1.35	222.88	6.80	2.80	1.27	39.8
Aeg	0-25	4.7	0.85	0.09	13.79	0.52	0.04	0.05	4.9
Bg	25-55	5.5	0.30	0.03	5.36	0.52	0.02	0.05	17.0
Cg1	55-80	5.4	0.12	0.03	4.80	0.48	0.01	0.06	11.5
Cg2	80-106	5.5	0.03	0.03	5.36	0.80	0.01	0.06	16.2
Cg3	106-132	5.5	0.21	0.03	6.57	0.68	0.00	0.06	11.3
Cg4	132+	5.6	0.15	0.03	4.92	0.68	0.00	0.06	15.0

Figure 2.2. Example of information obtained from soil surveys and utilized in developing the database, including lab-measured, horizon-specific soil physical and chemical properties. The example provided represents the Poitras soil associate retrieved from Langmaid et al. (1980).

The rationale for a province-wide compilation of soil survey reports arose from the need for understanding where and how soils respond, locally and regionally, to intensifying land uses and overall climate-change expectations. As such, unified soil databases per province would assist in, e.g., (i) estimating survey-based soil carbon storage, and (ii) determining how soil carbon storage would change within existing and proposed changes to land management and related climate-change adaptation scenarios (Carré *et al.*, 2007; Aksoy *et al.*, 2009; Grimm and Behrens, 2010; Poggio *et al.*, 2013). These changes would vary from abandoning farm fields, converting natural stands into agricultural land, expanding build-up areas through urban sprawl to land reclamation operations including wetland restoration and afforestation.

The objective of this article was to develop a seamless database by amalgamating and harmonizing existing soil survey reports for NB as a case study. This objective was accomplished by:

1. Compiling all the existing soil survey information into one consistent and seamless database,

2. Unifying the classifications and descriptions assigned to the surveyed soil associations, soil associates, and soil-forming factors, namely, parent material (by mode of deposition and lithology), topography, vegetation cover type, and drainage, and

3. Standardizing the soil associate names by profile and property descriptions, with emphasis on horizon labels and properties, notably horizon depth, soil texture, SOM content, CF content, Db, and soil moisture retention at FC and PWP.

All of this was completed in reference to:

1. the standardized soil surveying terminology of the Mapping System Working Group (1981) and the Expert Committee on Soil Survey (1982),

2. sampling and analytical techniques described by McKeague (1978) and Guertin *et al.* (1984),

3. the National Soils Database ("NSDB", retrieved from the CANSIS website at http://sis.agr.gc.ca/cansis/nsdb/index.html), and

4. the spatial soil distribution and representation as established by the province-wide soil mapping approximations also retrieved from the CANSIS website, i.e.,

    a. "Soils of New Brunswick: the Second Approximation" ("SNB", Fahmy *et al.*, 2010),
    b. "Forest Soils of New Brunswick" ("FSNB", Colpitts *et al.*, 1995)

2.3. COMPILATION OF SOIL SURVEY REPPORTS

The New Brunswick soil survey reports were retrieved from the publications section of the Canadian Soil Information System (CANSIS) as available from Agriculture

and Agri-Food Canada website at http://sis.agr.gc.ca/cansis/publications/surveys/nb/index.html. Each survey was downloaded and assessed to determine the extent of data availability. Table 2.1 summarizes these findings and identifies which reports were utilized in developing the amalgamated soil database and which provide spatial coverage. Some of the reports needed to be excluded either due to the omission of horizon-specific data, or insufficient information about the horizon-specific soil forming process, i.e., "A_1" instead of "Ae" or “Ah”. Also omitted were surveys specifically dealing with small sections of agricultural lands. Of the available soil surveys (both included and excluded from the database), only 53.9% of NB has survey coverage with the spatial coverage of the surveys utilized in this study only representing 35.5% of NB (Fig. 2.3).

An overview of the soil associates and soil associations is provided in Appendix I. From this, an aspatial database was developed, with specific attention given to soil associations, soil associates, soil classification (subgroup, great group, and order), dominant vegetation type, topography (landform, slope position, slope steepness, aspect), soil parent materials (lithology and mode of deposition), drainage, stoniness (amount of exposed coarse fragments at the surface), and rockiness (amount of exposed bedrock at the surface). Next, all soil horizon data pertaining to horizon depth, density, texture, organic matter content, coarse fragment content and other physical and chemical properties were entered into the database in an associate-by-associate manner. When inconsistencies arose, either varying analytical procedure or units of measurement, all values were retained but placed in separate fields to decide which of these entries best

reflect theoretical expectations. The procedures to do so including estimating omitted values and standardizing across different analytical techniques are described in Chapter 3.

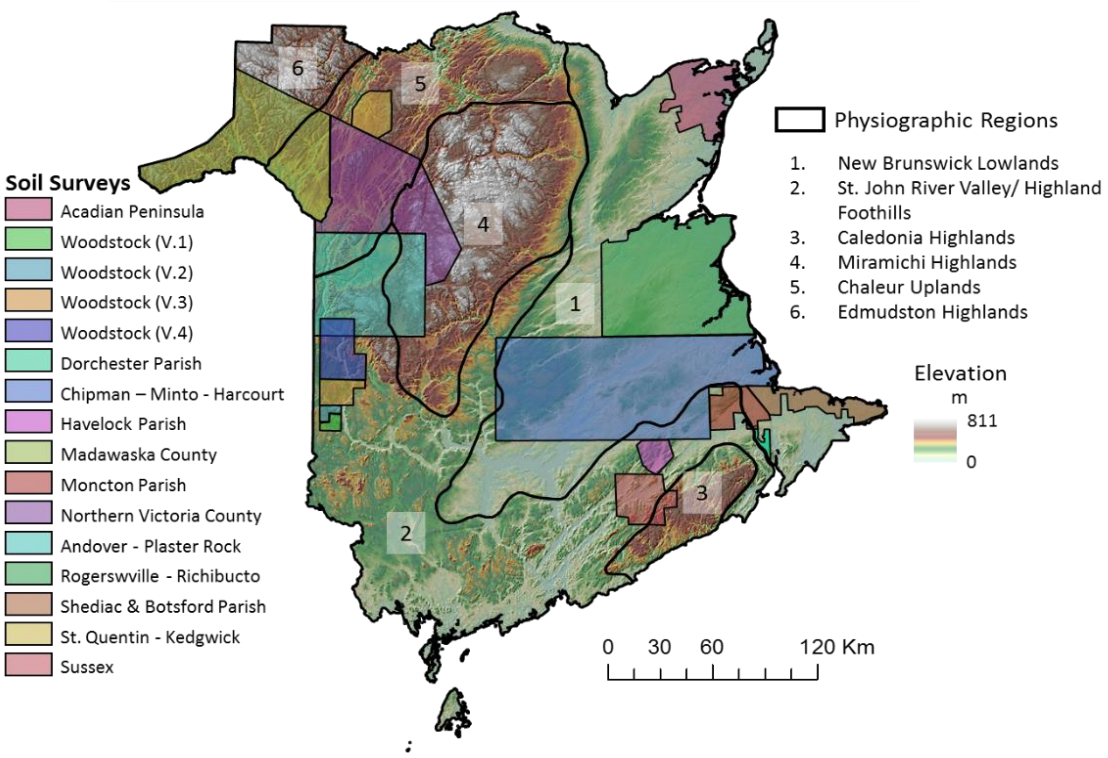

Figure 2.3. Spatial coverage of soil surveys utilized for developing database. Also included is elevation and physiographic regions to represent how coverage varies in each region. Physiographic regions retrieved from Colpitts *et al.* (1995).

Table 2.1. Overview of available soil surveys for New Brunswick, Canada, with publication year, scale, whether it includes data on a horizon-by-horizon basis, whether it was utilized in this study, and if spatial coverage is available. Some reports did not provide a spatial data layer but do provide images. These have been georeferenced (GR) but not digitized.

Soil Survey	Year Published	Scale	Horizon - Specific (Y / N)	Utilized	Digitized (Y / N)	Spatial Coverage (% of NB)
Fredericton - Gagetown	1940	95,040	N		Y	4.28
Sussex Area	1986	20,000	Y	X	N (GR)	0.25
Chipman - Minto - Harcourt	1992	50,000	Y	X	Y	9.98
Woodstock - Florenceville (Vol. 1)	1989	20,000	Y	X	Y	0.12
Woodstock - Florenceville (Vol.2)	1992	20,000	Y	X	Y	0.11
Woodstock - Florenceville (Vol. 3)	1996	20,000	Y	X	Y	0.39
Woodstock - Florenceville (Vol. 4)	2001	20,000	Y	X	Y	0.98
Moncton Parish	1993	20,000	Y	X	N (GR)	1.10
Shediac and Botsford Parishes	1996	20,000	Y	X	Y	1.03
Dorchester Parish	1998	20,000	Y	X	Y	0.10
Acadian Peninsula	2000	20,000	Y	X	Y	1.50
Woodstock Area	1944	63,360	Y	X	N	
Andover - Plaster Rock	1963	63,360	Y	X	Y	4.60
Southern Northumberland County	1964	31,680	N		N	-
Northern Victoria County	1976	63,360	Y	X	Y	4.60
Madawaska County	1980	50,000	Y	X	Y	4.78
Rogersville - Richibucto Region	1983	50,000	Y	X	Y	5.81
Blackbrook Watershed	1993	10,000	Y	X	N	-
St. Quentin - Kedgwick	1982	50,000	Y	X	Y	0.62
Southeastern New Brunswick	1950	126,720	N		N	-
Havelock Parish	1980	10,000	Y	X	Y	0.36
Central and Northern New Brunswick	2005	250,000	N		N	-
Southwestern New Brunswick	1953	156,720	N		Y	12.90
Agriculture Canada Benton Ridge Potato Breeding Substation	1992	5,000	N		N	-
Agriculture Canada Research Station, Fredericton	1984	4,800	N		N	-
Senator Herve J. Michaud Experimental Farm Agriculture Canada, Buctouche	1983	3,000	N		N	-
Mount Carleton Provincial Park	1972	-	N		N	-
Lepreau Provincial Park	1973	-	Y		N	-

2.4. DATA HARMONIZATION AND AMALGAMATION PROCEDURES

The compilation of the surveyed soil data was guided by the comprehensive soil association overviews for New Brunswick SNB and FSNB reports, and by the cross-referenced aspatial listing of soil association names, profile descriptions and soil horizon

properties within the Canada-wide NSDB database. This compilation resulted in a database consisting of 2,490 rows of data with coverage for individual properties outlined in Table 2.2. Of the 522 soil profiles, 500 contain soil classifications and 507 contain drainage classifications.

Table 2.2. Overview of measured soil properties within amalgamated database with overall completeness.

Soil Property	# of Horizons	% Complete
Horizon Depth (cm)	2490	100.00
Coarse Fragment Content (%)	885	35.54
Texture (% sand, silt, and clay)	1306	52.45
pH (H_2O)	1535	61.64
pH ($CaCl_2$)	647	25.98
Bulk Density (g/cm^3)	938	37.67
Organic Matter Content (%)	121	4.90
Base Saturation (%)	466	18.71
Cation Exchange Capacity	659	26.47
Field Capacity (-33kPa)	846	33.98
Permanent Wilting Point (-1500kPa)	738	29.64
Ca (meq/100g)	976	39.20
Mg (meq/100g)	960	38.55
K (meq/100g)	973	39.08

Soil Names and Inconsistencies. The amalgamated database was organized by soil association name, each with its own soil associates and horizon sequences. In this, the well-drained soil associate members of each soil association carry the name of the soil association. Naming inconsistencies occurred and were resolved using the SFNB and SNB reports as guiding authority, as follows:

1. Some of the soil associations were referred to as a complex between two associations, i.e., "Baie du Vin - Galloway", "Barrieau - Buctouche", and “Parleeville - Tobique” because of similar soil forming factors and soil properties.

For these instances, only one name was retained based on descriptions of parent material lithology and mode of deposition.

2. In some cases, only the names of the soil associations were provided although profiles were provided for different drainage classes. For these, new drainage-related soil associate names were assigned based on Table 6 ("Correlation of New Brunswick Soil Series/Associations with Forest Soil Units") of the SNB reports.

3. Within some reports, horizon and depth specifications by soil associate (section 2, Fig. 2.1) were inconsistent with their listing at the end of the reports (section3, Fig. 2.2). This was most prevalent in the Northern Victoria and St. Quentin soil surveys. To correct this, the measured properties were kept separate from the general data and were entered at the end of the database.

4. Also inconsistent were the amounts of data provided for each soil type. For example, only general information was provided for some soil types (sections 1 and 2) while measured soil properties (section 3) were omitted. This resulted in some soil associates lacking horizon-specific property measurements.

Drainage. Soil drainage was classified from very poor (wetlands and organic soils) to rapidly- and excessively-drained (coarse-textured, upper slope positions), as outlined by the Expert Committee on Soil Survey (1982). The procedure in Fig. 2.4 was used to determine if soil drainage was correctly classified for each soil associate in terms of soil horizon sequence. This procedure ensured that:

1. the soil associate names within each association were consistent with the drainage expectations based on Table 4 in the SNB report, entitled “New Brunswick mineral soil catenas”;

2. well- to rapidly-drained members occur on upper slope and hill-crest positions, imperfect- to moderately well-drained members occur on the lower slopes, and very poor- to poorly-drained members along toe slopes and in depressions.

The drainage classifications derived were generally consistent with the original survey drainage assignments. In cases where the original drainage classifications provided broad ranges, i.e., "VP-I", the middle drainage class (P) was retained. The drainage assignment procedure in Fig. 2.4 was used to ensure proper drainage classification for all samples, particularly for those of which were assigned broad classes, e.g., “VP-R”, or specified neighboring drainage classes, e.g., “VP-P”.

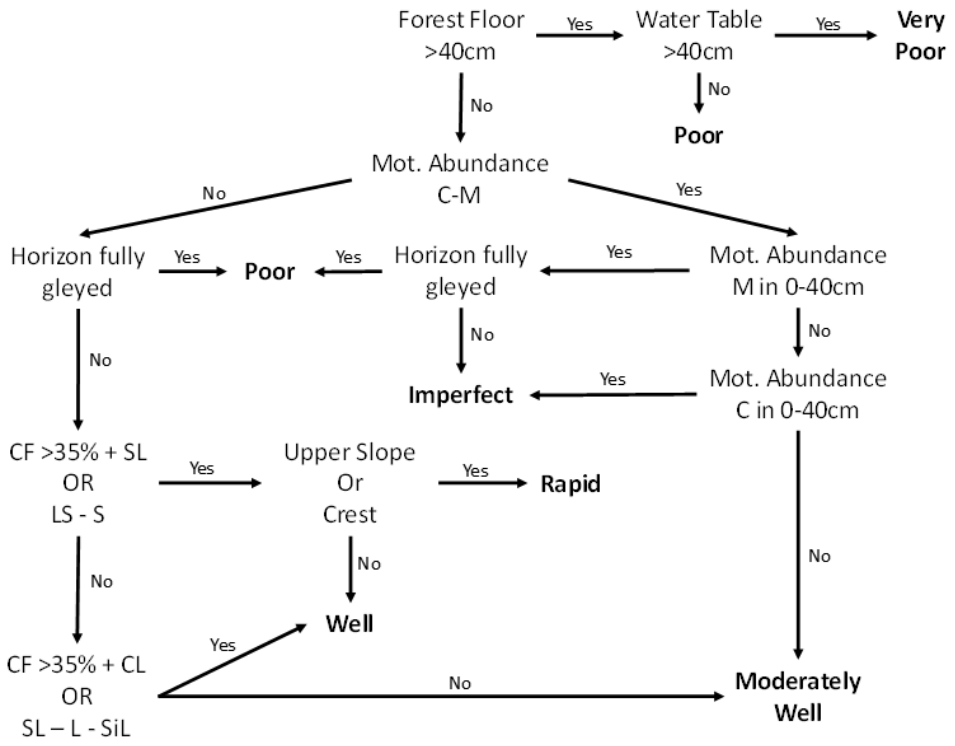

Figure 2.4. Visual representation of model developed to assign drainage regime to aspatial database. Resulting drainage classes are bolded.

Soil Classification. Each soil profile was placed within the Canadian Soil Classification (Soil Classification Working Group, 1998) context by specifying its belonging to Soil Order, Great Group, and Subgroup. Once completed, abbreviations and rankings for soil classifications, stoniness, rockiness, and drainage were assigned to every soil associate (where applicable). Table 2.3 provides an overview of the % distribution of database entries by soil classification.

Table 2.3. Overview of soil orders separated by great group and subgroup within database with overall representation of great groups provided.

Order	Great Group	Subgroup	Number of Profiles	% Total
Podzol	Ferro-Humic	Gleyed	1	4.79
		Gleyed Fragic	1	
		Orthic	23	
	Humic	Orstein	1	0.19
	Humo-Ferric	Fragic	4	41.95
		Gleyed	29	
		Gleyed Luvic	1	
		Gleyed Mini	1	
		Gleyed Orthic	18	
		Gleyed Sombric	1	
		Mini	1	
		Orstein	7	
		Orthic	155	
		Sombric	2	
Brunisol	Dystric	Eluviated	11	6.51
		Gleyed	7	
		Gleyed Eluviated	7	
		Orthic	9	
	Eutric	Gleyed Eluviated	3	0.77
		Orthic	1	
	Melanic	Gleyed Eluviated	1	1.53
		Gleyed	4	
		Orthic	3	
	Sombric	Gleyed Eluviated	1	4.02
		Gleyed	7	
		Orthic	13	
Luvisol	Gray	Brunisolic	7	18.39
		Dark	5	
		Gleyed Brunisolic	10	
		Gleyed	8	
		Gleyed Podzolic	20	
		Orthic	3	
		Podzolic	43	
	Gray Brown	Brunisolic	1	0.38
		Gleyed	1	
Regosol	Humic	Gleyed	3	0.57
	Regosol	Cumulic	1	1.72
		Gleyed Cumulic	2	
		Gleyed	4	
		Orthic	2	
Gleysol	Gleysol	Fera	3	4.02
		Orthic	16	
		Rego	2	
	Humic	Orthic	20	4.21
		Rego	2	
	Luvic	Fera	5	4.60
		Fragic	2	
		Humic	1	
		Orthic	16	
Organic	Fibrisol	Typic	5	2.49
		Terric Mesic	3	

		Terric Humic	2	
		Mesic	1	
		Terric	2	
	Mesisol	Typic	6	2.11
		Terric	3	
		Terric Fibric	1	
		Terric Humic	1	
	Humisol	Typic	3	1.72
		Terric	2	
		Terric Fibric	2	
		Terric Mesic	2	

According to this compilation, Podzols (46.9% of database) have a much higher representation than any other order (representing almost half of the database), followed by Luvisols (18.8%), Brunisols and Gleysols (12.8%), and Regosols (2.3%). Organic soils (6.4%) were both surveyed and included in the database but were omitted from the analyses by focusing on mineral soils only.

Soil Forming Factors. Soil parent materials were classified by:

1. landforms which vary by mode of deposition (how the material was deposited geologically) and
2. lithology (referring to the chemical and physical makeup of the deposited material).

Differences in surficial geology descriptions, both in mode of deposition and lithology, occurred within the same soil association names when cross-referencing the SNB (Table 6), FSNB (Tables 2 and 5), and NSDB reports. This was corrected as follows: if three or more sources (including soil surveys, SNB, FSNB, and NSDB) provided the same mode of deposition for an individual soil association, then that mode of deposition was assigned

to that association. Any remaining inconsistencies were addressed by determining the expected mode of deposition by surface expressions (topography), coarse fragment content, and horizon sequences. Together, this cross-referencing resulted in 21 unique modes of deposition (Table 2.4), with some associations having two distinct modes of deposition overlaying one another (i.e. glaciomarine/basal). In such instances, the top parent material will have the dominant influence on soil formation and development.

Table 2.4. Summary of updated parent material modes of deposition within aspatial database including the quantity of associations within each mode of deposition.

Mode of Deposition	Number of Associations	% of Total
Residual	4	3.60
Residual and Colluvium	1	0.90
Colluvium and Water Re-worked Till	4	3.60
Ablation/ Residual	9	8.11
Ablation	14	12.61
Ablation/ Basal	2	1.80
Basal	29	26.13
Basal/ Residual	1	0.90
Glaciomarine/ Basal	5	4.50
Glaciomarine	5	4.50
Glaciomarine/ Marine	1	0.90
Marine	3	2.70
Marine/ Basal	1	0.90
Glaciofluvial and Marine	6	5.41
Glaciofluvial	10	9.01
Alluvium and Glaciofluvial	2	1.80
Ancient Alluvium	1	0.90
Alluvium	3	2.70
Lacustrine	1	0.90
Glaciolacustrine	1	0.90
Organic	8	7.21

Not yet included in the Table 2.3 description are landforms that specifically refer to, e.g., valley trains, glaciofluvial outwash plains, drumlins, moraines, and eskers. Additionally, glacial tills (ablation and basal) lack information on depth of deposits. Additional

information pertaining to these deposits and depths can be obtained from Rampton (1984).

With respect to soil lithology specifications, there were inconsistencies as well. For example, the parent material of the Baie du Vin association was labelled as: "*acidic GLFL or MA sand, petrologically similar to underlying sandstone bedrock, and rich in biotite*". The Galloway soil units, stated to have the same lithology, was labelled: "*acidic, petrologically similar to the underlying sandstone bedrock and rich in biotite*". These inconsistencies were addressed through re-labelling and by updating lithology by dominant rock types, grain sizes, and mineral hardness (based on Mohs hardness scale, retrieved from http://rocks.comparenature.com/). This was followed by (i) providing binary descriptors for sedimentary, igneous, and metamorphic parent materials per soil association, and (ii) by the ranking of (a) rock type weatherability (ease of physical and chemical breakdown of parent material) and (b) fertility (mainly Ca, and Mg richness of the weathering parent material) (Table 2.5). Weatherability and fertility were based on Table 4 in the FSNB report.

Some soil associations (Bellefleur, Bottomland, Bransfield, Chockpish, Gulquac, Lower Ridge, St. Charles, and Wakefield) could not be identified as forest soil associations in the FSNB and SNB reports. For these associations, weatherability and fertility assignments could therefore not be determined. This conforms with the previous statements in which inconsistencies occur with soil mapping initiatives.

Table 2.5. Overview of weatherability and fertility rankings assigned to common rock types, retrieved from Figure 4 (The relative weatherability and fertility of the common rock types found in New Brunswick) from FSNB.

Rock Type	Relative Weatherability 0-1 (slow-fast)	Relative Fertility 0-1 (poor-rich)
Quartz-Pebble Conglomerate	0.04	0.2
Felsic Volcanic (Rhyolite)	0.07	0.19
Felsic Pebble Conglomerate	0.14	0.17
Schists	0.15	0.43
Metaquartzites	0.18	0.44
Gneiss	0.20	0.42
Granodiorites	0.20	0.54
Quartz Diorites	0.23	0.61
Granites	0.28	0.31
Diorites	0.28	0.55
Alkali Granites	0.34	0.25
Gabbros	0.33	0.65
Polymictic Metaconglomerates	0.39	0.37
Mafic Volcanic (Basalt)	0.38	0.72
Quartzose Sandstone	0.42	0.40
Metasandstone	0.47	0.55
Polymictic Conglomerates	0.54	0.31
Slates	0.53	0.61
Metasiltstones	0.58	0.66
Lithic Sandstones	0.62	0.39
Metawackes	0.64	0.71
Feldspathic Sandstones	0.66	0.47
Mudstones	0.73	0.53
Calcareous Sandstones	0.77	0.78
Calcareous Slates	0.80	0.80
Calcareous Siltstones	0.84	0.84
Calcareous Mudstones	0.87	0.87
Limestones	0.90	0.94
Argillaceous Limestones	0.95	0.79

Topography. Topographic surface expression descriptors by soil association also varied by survey report, being absent in some of the reports. When present, topographic surface expressions ranged from flat (or domed) for organic soils to strongly rolling and hilly on dense igneous parent materials in the New Brunswick Highlands. Although included in the database, little emphasis was placed on topographic expressions because

the expressions vary by resolution, intensity and frequency of changing topographic positions, slopes, and geographic regions, resulting in inconsistent descriptions between reports. A consistent measure for each association referred to average slope position and slope percent, but this information was only provided for 40% of the reports. To remedy, soil association boundaries can be re-defined spatially via unified landform and lithology classifications and topographic expressions can be classified via digital terrain modeling as described in Chapter 4.

Vegetation. Some surveys listed the presence of dominant overstory species, generally within the vicinity of the soil sampling points. These specifications were entered into the database in the form of binary fields referring to dominance of shade tolerant hardwoods, softwoods, and mixedwoods within the overstory canopy. Also, where forest floor data were provided, forest floor thickness was assigned to each horizon sequence. If spatial coordinates were available for the soil associates then, to some extent, vegetation types could be re-assigned to the associates via forest inventory data providing information on pre-harvest conditions.

Horizon Descriptions and Depths: Considerable effort was placed on ensuring that the horizon classification within the surveys was consistent with those outlined in the "Canadian System of Soil Classification". Horizons were also generalized by master horizon (forest floor, A, B, and C) and by the first subscript for each master horizon, i.e., Ae, Ah, Bf (represents the dominant process influencing the soil). Table 2.6 outlines the range of master horizon descriptors encountered. The additional horizon specifications

such as g, c, x, j, t, etc., were entered into the database as binary fields (0 when absent, 1 when present). Profiles where soil horizon descriptors where marked by "?" or "or" were re-labeled through cross-referencing with other similar soil profiles. Horizon depths (many were originally specified in inches) were re-assigned in cm.

Table 2.6. Variability in soil horizon classification encountered within aspatial database, separated by master horizons, followed by primary subscripts, resulting in 180 unique soil horizons.

Master Horizon	Primary Subscript	Variations
O (Organic)	Of	Of, "Of, Om", Of_1, Of_2
	Om	Om, Om_1, Om_2, Om_3, Om_4, Om_5
	Oh	Oh, Oh_1, Oh_2, Oh_3, Oh_4
	Ol	
	Oco	
LFH (Forest Floor)	L	L, LF, LF_2, LFH
	F	F, F_1, F_2, FH
	H	H, H_2, HC
A	Ae, Aeg	Ae, Ae_1, Ae_2, Aeh, Aej, Aeg, Aeg_1, Aeg_2, Aegj, 2Aeg, Aexjg, Aejg
	Ah, Ahg	Ah, Ah_1, Ah_2, Ahb, Ahg, Ahgj
	Ahe, Aheg	Ahe, Aheg, Ahejg, Ahegj
	Ap, Apg	Ap, Apg, Apgj
B	Bf, Bfg	Bf, Bft, Bftg, Bfc, "Bf, Bfj", "Bf, Bm", Bf_1, Bf_2, Bf_3, Bf_4, Bfj, Bfg, Bfg_1, Bfg_2, Bfgj, Bfjg, $Bfjg_2$, $Bfjg_3$, Bfjgj, Bfjg, Bfcg, Bfcjg, Bfgcj, Bfjgc, Bfjgjc
	Bfh, Bfhg	Bfh, Bfht, Bfhc, Bfh_2, Bfhg, Bfhgj, $Bfhgj_2$
	Bh	Bh, Bhcg
	Bg	Bg, Bg_1, B_2, Bgj, 2Bg, Bgc, Bgx, Bgf, Bgfcc
	Bhf, Bhfg	Bhf, Bhfg, Bhfgj, $Bhfg_2$, Bhfjg
	Bm, Bmg	Bm, Bm_1, Bm_2, 2Bm, Bmg, Bmx
	Bt, Btg	Bt, "Bt, C", Bt_1, Bt_2, Bt_3, Bt_4, Btj, 2Bt, $2Bt_2$, 2Btj, Btg, Btg_1, Btg_2, Btgj, "Btgj or Cgj", "Btgj, Cgj", $Btgj_1$, $Btgj_2$, $Btgj_3$, Btgk, Btjg, $Btjg_1$, $Btgj_2$, 2Btg, 2Btg2, 2Btgj, 2Btjgj, Btxg, Btxjgj
BC		BC, BCgj, BCx
C	C, Cg	C, C_1, C_2, C_3, C_4, C_5, 2C, $2C_1$, $2C_2$, $2C_3$, 3C, Cg, Cg_1, Cg_2, Cg_3, Cg_4, Cg_5, Cgj, Cgj_1, Cgj_2, Cgj_3, 2Cg, $2Cg_1$, $2Cg_2$, $2Cg_3$, 2Cgj, 3Cg, 3Cgj, 4Cg, 4Cgj, 5Cg, 6Cg,
	Ck, Ckg	Ck, Ckg, Ckg_1, Ckg_2, Ckg_3, Ckgj, 2Ckgj
	Cx, Cxg	Cx, 2Cxj, Cxgj
R	R	

Soil Properties. For each horizon, surveyed values for soil properties referring to soil texture, structure, Db, SOM content, CF content, and water retention (at both FC and PWP) were entered into the database, using the following procedures.

Texture: Soil texture information, where available, was entered into the database in two forms:

1. Texture classes as assigned from texture triangle, as outlined in the Canadian Soil Classification System (Fig. 2.5), and

2. Proportions of sand, silt, and clay within the fine-earth fraction as percentages with the summation equaling 100% (although not always the case within the database).

Some texture descriptions provided broad ranges, i.e., “SiL-LS”. For these cases, texture class and percentage of sand, silt, and clay were assigned by choosing the center point within these classes on the texture triangle. Additionally, some texture classifications provided did not fall within the realm of the texture triangle, i.e. “G” (gravel) and “SG” (sandy gravel). This occurred for 37 samples, of which, could not be provided sand, silt, and clay contents. Although not in the texture triangle, these classifications were retained. Also, most texture classifications were assigned including a modifier, i.e., “vfSL” (very fine sandy loam). These modifiers were retained, but texture classes without modifiers were placed in a separate column.

For some horizons, only a texture class was provided, this was typically the case when general descriptions (section 2) did not coincide with specific horizon properties (section 3) within the survey. With these horizons, the percentage of sand, silt, and clay were left absent. For the cases where only the percentage of sand, silt, and clay were

present (without assigned texture class), an automated model (outlined in Table 2.7) was derived to determine the texture class based on these percentages. This was used to fill in the voids where the texture class was absent, resulting in 86.6% of the database with texture measurements.

Table 2.7. Logical Rule statements applied in ascending order to determine proper soil texture class based on texture triangle as outlined in Soil Classification Working Group (1998).

Rule	Output Class	Output Abbreviation
Clay ≥ 60	Heavy Clay	HC
(60 – Sand) < Clay ≤ 40 & Sand ≤ 45	Silty Clay	SiC
Clay ≥ 40 & Sand < 45 & Clay ≥ (60 – Sand)	Clay	C
Clay ≥ 35 & Sand ≥ 45	Sandy Clay	SC
Clay ≥ 28 & Sand ≤ 20	Silty Clay Loam	SiCL
27.5 < Clay < 40 & 20 < Sand < 45	Clay Loam	CL
Clay ≤ 12 & Clay < (20 – Sand)	Silt	Si
Clay < (50 – Sand)	Silt Loam	SiL
Clay ≤ (2 · (Sand – 70))	Sand	S
Clay ≤ (Sand – 70)	Loamy Sand	LS
Clay ≤ 20 & Sand ≥ 53 OR Clay ≤ 7	Sandy Loam	SL
Clay ≥ (73 – Sand)	Sandy Clay Loam	SCL
Else	Loam	L

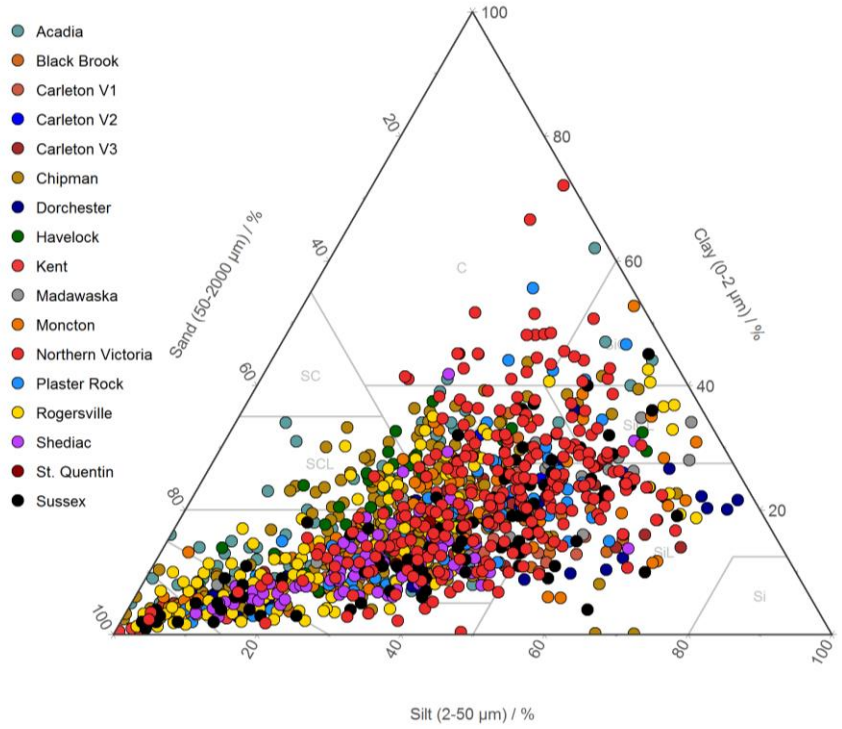

Figure 2.5. Texture composition of soil samples for each soil survey utilized in developing the aspatial database overlain on texture triangle.

Coarse Fragments: CF content was provided for 36% of the mineral horizons. Some of these included ranges, i.e., "40-60%, while others provided qualitative descriptions, i.e., "few" or "some". With the ranges, the middle values were assigned. Additionally, coarse fragment content was also included as part of the horizon texture description, e.g., “gravelly sandy loam”. For these cases, the suggestions of the Expert Committee on Soil Survey (1982) were adopted as follows: <15% CF by volume was assigned to “Non”, 15-35% not assigned an adjective, 30-60% assigned to “Very”, and >60% assigned to “Extremely”. Where the texture modifiers were not consistent with the

CF% values (i.e., “very gravelly sand” with a CF% of 10%), and were also not consistent with landform expectations, Table 2.8 was assessed to aid in providing CF estimates.

Table 2.8. Overview of CF content for each parent material mode of deposition found within aspatial database. Note that some modes of deposition lack CF content data while others have large variations in values.

Mode of Deposition	Coarse Fragment Content				
	Min.	Max.	Mean	SD	Sample Size
Residual	10	40	24.2	12.8	6
Residual + Colluvium	-	-	-	-	0
Ablation	5	80	24.7	16.6	76
Ablation/ Basal	0	30	11.9	7.5	94
Ablation/ Residual	10	70	47.1	16.0	24
Alluvium	0	0	0	0	14
Alluvium + Glaciofluvial	-	-	-	-	0
Ancient Alluvium	0	0	0	0	6
Basal	1	60	15.5	9.8	371
Basal/ Residual	-	-	-	-	0
Colluvium + Water-reworked Till	-	-	-	-	0
Glaciofluvial	1	80	35	22.9	45
Glaciofluvial + Marine	0	35	4.6	7.4	106
Glaciolacustrine	-	-	-	-	0
Glaciomarine	0	12.5	3.6	3.4	55
Glaciomarine/ Basal	0	45	4.7	8.6	54
Glaciomarine/ Marine	0	0	0	0	8
Lacustrine	0	1	0.3	0.3	14
Marine	0	0	0	0	11

Due to the omission of samples with measured CF values for some modes of deposition, values could not be generalized. Alternatively, values could also be inferred from general parent material descriptions provided for each soil association. In principle, these values could be assigned to the soil horizons. However, the specified ranges were, in most cases, too broad to be consistent with actual CF% survey values.

Soil Structure: Soil structure was provided as a description with three components, shape, size, and distinctness. Therefore, each of these three components

were assigned to the database. An overview of soil structures can be found in “Manual for Describing Soils in the Field: 1982 Revised” by Expert Committee on Soil Survey (1982). It was noticed that terminology for structureless soils were used interchangeably, namely "single grain", "loose", "amorphous", and "massive"; therefore, these were grouped into two classes (“massive” for amorphous and massive descriptions and “single grain” for the remaining two). Soil structure information was provided for 73.6% of the database.

Organic Matter Content: SOM content was provided in four formats, % organic matter, % carbon, and loss on ignition (LOI) at 450°C and 850°C. Soil surveys for Plaster Rock and Northern Victoria Counties provided both % carbon and LOI at 450°C (328 samples, 13% of database). Kent County was the only report to record LOI at 850°C (11 samples, 0.4% of database) and did not record % carbon for comparison. Due to the lack of samples and omission of carbon values for comparison, readings for LOI at 850°C were omitted. The % carbon readings were converted to % organic matter via Eq. 1.

$$\%SOM = \%C \cdot 1.72 \quad (1)$$

where $\%SOM$ is % soil organic matter, $\%C$ is % carbon and 1.72 is the conversion factor since SOM is composed of 58% carbon (Romano and Palladino, 2002; Pollacco, 2008; Chaudhari *et al.*, 2013; Poggio *et al.*, 2013). Standardizing these measurements into % SOM resulted in 1,202 samples with measurement (48.3% of database).

Soil Density: For particle density (Dp) and Db, box plots were used per horizon label to quickly determine the extent to which density outliers were present (Fig. 2.6). For

each outlier, the original report was reviewed to determine if a data entry mistake had occurred. It was ensured that the ranges of the density values were generally consistent with soil texture, SOM, and soil depth expectations, with additional considerations to distinguish density in compacted versus non-compacted soils. Data entry with obvious data errors (e.g., soil Dbs greater than densities for silicate rocks) were deleted. Db determinations were generally sparse, with 937 of the horizon- based data entries (38% of database).

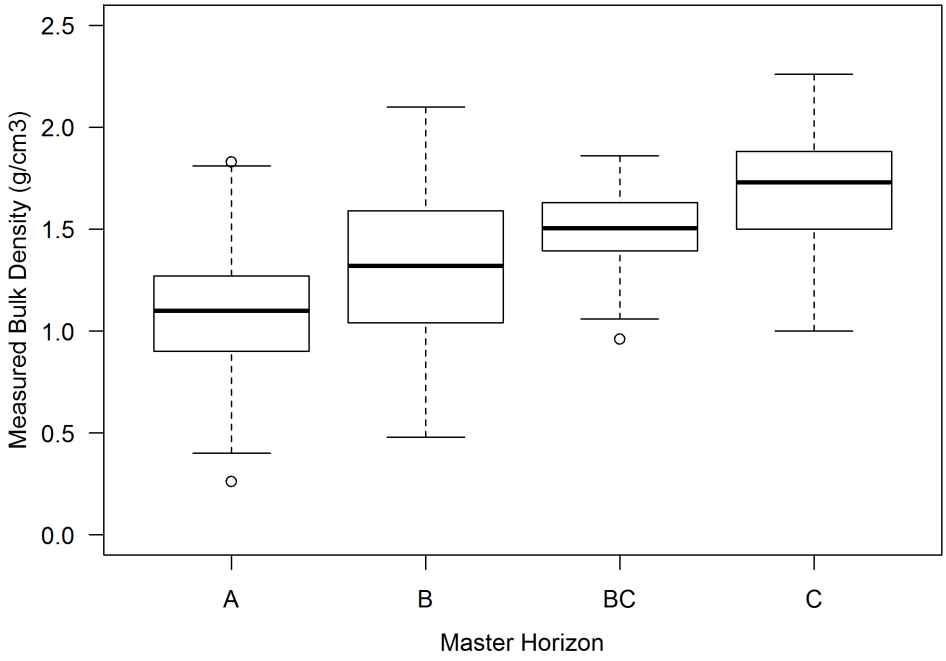

Figure 2.6. Visual comparison of range in Dbs associated with each master horizon to assess presence (or absence) of outliers.

Water Retention: In terms of water retention (the ability to access moisture under different pressure gradients), depending on the report, values were provided in bars (bar), atmospheres (atm), and kilopascals (kPa) with values measured in both volumetric and gravimetric form. Together, ten reports provided water retention measures in

gravimetric form while five reports provided measures in. Two reports did not specify whether the measures were gravimetric or volumetric. Gravimetric water retention values were recorded and converted to volumetric form via multiplication with Db. The different units of measurement were then amalgamated and adjusted to represent water retentions in kPa: -33kPa for water retention at FC, and -1,500kPa for water retention at PWP. Also, with water retention at FC, moistures were provided under the title "Moisture Equivalent" which is FC values measured via a specific test.

Additional moisture measurements included water % at 0cm, water holding capacity, maximum water holding capacity, and water retention at saturation, 10cm, 50cm, 100cm, -100kPa, -400kPa, hygroscopic moisture, available water, and moisture percentage. Emphasis was placed on moisture retention at FC and PWP due to the influence of these pressures on rooting. Once combined, water retention at FC had 836 samples measured (34.0% of database) whereas PWP had 743 samples measured (29.8%).

Aspatial Overview. The soil database, amalgamated from 17 soil surveys for NB, Canada, is intended to provide a comprehensive overview of forest soil conditions across NB. As such, this database contains information for 106 soil associations, 243 soil associates, and 522 soil profiles, each with their own soil horizon specifications as outlined above. Through careful cross-referencing, all data entries were examined to ensure they coincided with soil association and horizon-specific expectations as outlined in the “Canadian System of Soil Classification” (Soil Classification Working Group, 1998).

Most soil associates had more than one profile described, depending on frequency of occurrence, i.e., some reports listed the same soil association even though reports occurred in different geographic locations. For example, the Holmesville soil association occurred in eight soil surveys. As a result, the well-drained soil associate (also called Holmesville) had 18 profiles within the database. Its moderately-well drained associate, Johnville, had 12 profiles within the database, followed by the poorly-drained associate, Poitras, also with 12 profiles. It was common for the broadest soil associations to occur within different surveys, and therefore, had multiple profiles within the database. In contrast, some of the less-common soil associations lacked a single soil profile altogether (i.e., Aulac, Babineau, Becaguimec, Belledune, Big Bald Mountain, Blackland, Caissie, Bottomland, Catamaran, Clearwater, Escuminac, Jacquet River, Kingston, Research Station, and Tetagouche).

The outcome of all considerations and actions is presented in Appendix II, completed in reference to the spatial context of the SNB soil coverage for New Brunswick. This representation reveals (i) that not all SNB-mapped soil associations occur within the original soil survey reports (i.e., Becaguimec, Big Bald Mountain, Catamaran, Jacquet River, Kingston, Popple Depot, and Tetagouche), and (ii) that some of the surveyed soil associations are not spatially represented in the existing SNB delineation, as outlined in Table 2.9. In total, the soil association coverage in Appendix II corresponds to the 71,450 km^2 land base of New Brunswick. The total provincial area amounts to 72,907 km^2 of which 1,458 km^2 is water. The organic soil coverage at 11.38% amounts to 8,131 km^2.

Table 2.9. Surveyed soil associations not currently spatially covered in province-wide SNB soil association map.

Soil Associations			
Aldouane	Flemming	Kouchibouguac	Riley Brook
Anagance	Fundy	Lord and Foy	Salem
Baie du Vin	Galloway	Lower Ridge	Shemogue
Bellefleur	Green River	Maliseet	St. Charles
Benedict	Green Road	Monquart	Sussex
Big Hole	Gulquac	Mount Hope	Tobique
Boston Brook	Harquail	Parsons Brook	Tormentine
Bransfield	Island Lake	Guimond River	Upper Caraquet
Bretagneville	Jardine	Petitcodiac	Violette
Caraquet	Jeffries Corner	Queenville	Wakefield
Chockpish	Kingsclear	Quisbis	
Dorchester	Knightville	Richibucto	

Although amalgamated and harmonized, the aspatial database remains incomplete in terms of measurement gaps for horizon-specific physical and chemical properties, as outlined in Table 2.2. Chapter 3 addresses and, where feasible, fills some of these gaps via the development of PTFs by way of linear multivariate regression analyses. Decision trees can be used to fill in remaining data gaps.

2.5. CONCLUDING REMARKS

Creation of the aspatial database expedites PTF development for gap-filling and summarizing and quantifying both soil association and associates via similarities and differences. Once complete, this will enable spatially re-digitizing the updated database using already-existing soil association delineations, followed by revising these to ensure topographic mapping consistencies. For example, all digitized floodplain-derived soil associations must conform with the extent of topographically delineated floodplains, all residual soil units need to coincide with topographically delineated ridge tops and steep slopes. Additionally, all organic soil units need to fall within topographically delineated

depressions and wet areas next to streams, rivers, lakes and coastal shores (Murphy *et al.*, 2008, 2009). Finally, all soil map units need to reflect landscape features as defined by digitally-delineated landforms such as eskers, kames, till plains, drumlins, and moraines, and need to further reflect the lithology origin of these formations (e.g., calcareous versus siliceous).

The topographically-corrected soil delineations, with the amalgamated and harmonized database, provide many possible applications. For example:

1. The correct placement of each soil association and associated soil properties will assist in refining the moisture regime classification within these associations based on topographic location. For example, clay-textured soil associations reduce water infiltration into the soil, thus having ephemeral flow channel networks that reach further upslope to ridges than what would be the case for coarse textured associations. The latter would require larger upslope flow accumulations for stream flow initiation.

2. Improved soil mapping will allow for better soil erosion estimation, because the soil descriptors needed to calculate soil erosion potentials correspond with local variations in topography.

3. The revised mapping will improve the quantification of soil organic matter, because soil organic matter accumulations correlate positively with soil rooting depth as affected by soil lithology, drainage, and compaction.

4. Since vegetation growth and health is closely related to soil fertility (availability of soil nutrients) and water availability, the revised soil map will allow for better decision making in site selection for crop production whether in forestry or agriculture.

5. Since soil and vegetation type (and vegetation structure) vary across landscapes, the improved soil mapping will find much use in conservation and reclamation practices.

6. Given that there is a high network of paved and unpaved roads and trails permeating New Brunswick, the revised map will provide much needed information about their locations across drainage-challenged soils.

CHAPTER 3 – FROM SOIL SURVEYS TO PEDOTRANSFER FUNCTION DEVELOPMENT AND PERFORMANCE ASSESSMENT

Shane Furze, Paul Arp

Faculty of Forestry and Environmental Management

University of New Brunswick, Fredericton, NB, Canada, E3B 6C2

Foreword:

The following chapter is an article submitted to the Open Journal of Soil Science. It was submitted on May 31st, 2018.

Citation:

Furze, S. and Arp, P.A. 2018. From Soil Surveys to Pedotransfer Function Development and Performance Assessment. *Open Journal of Soil Science* (submitted).

3.1. ABSTRACT

This article presents the development of pedotransfer functions (PTFs) for soil physical and chemical properties such as bulk density, texture (sand, silt, and clay), coarse fragment content, pH, soil organic matter content, cation exchange capacity, field capacity, and permanent wilting point for varying soil types and drainage conditions. Pertinent data for PTF development resulted from the amalgamation of county-based soil surveys for the province of New Brunswick, Canada into a harmonized aspatial database for both soil associations and soil types. It was ensured that PTF outputs coincided with realistic thresholds, namely, bulk density does not exceed 2.4g/cm^3, soil organic matter does not exceed 100%, combination of sand, silt and clay does not exceed 100% for fine earth mineral fraction, permanent wilting point does not exceed field capacity, and all

values are positive. PTF development for sand, silt, clay, organic matter content, bulk density, cation exchange capacity, field capacity, and permanent wilting point resulted in capturing between 60 to 75% of the total soil property variation with PTF models validated by via comparison with published PTF equations.

Key Words: Province-wide database, physical and chemical soil properties, pedotransfer functions.

3.2. INTRODUCTION

Due to the high cost associated with intensive soil sampling and subsequent laboratory analyses required to develop high-resolution soil maps, there is a need for algorithms to model and predict soil properties that are difficult, or too costly, to measure (Moore *et al.*, 1993; McBratney *et al.*, 2002). These algorithms, known as pedotransfer functions (PTFs) are a feasible alternative allowing for the prediction of soil properties based on a few easy-to-measure variables. PTFs, a term introduced by Bouma (1989) are not a new science and have been used vastly in the past for predicting soil properties, for example, a review of hydraulic PTFs is provided by Wösten *et al.* (2001), and two general PTF reviews provided by McBratney *et al.* (2002) and Nanko *et al.* (2014). Many PTFs have been developed for soil physical and chemical attributes with soil organic matter (SOM), soil texture, structure, and bulk density (Db) as the most commonly used predictors (Moore *et al.*, 1993)**.**

PTF development requires a data set that is heterogenous in terms of soil types and site conditions, particularly varying parent materials, topographies, flora, and physiographic regions (varying mesoclimates), since these variables have been proven as dominant soil forming factors (Jenny, 1941; McBratney *et al.*, 2003; Florinsky, 2012; Heung *et al.*, 2014). Therefore, soil property variation will be dependent upon these factors. Developing PTFs over a wide range of soil conditions is more likely to result in higher predicting accuracy (Pollacco, 2008) with a readily-available source of information in the form of conventional soil surveys (McBratney *et al.*, 2000, 2002; Balland *et al.*, 2008). In New Brunswick, Canada, there is a wealth of information available in conventional county-based soil surveys.

The objectives of this article are:

1. to determine how the survey-compiled data for individual soil physical chemical variables relate to one another through PTF analysis;

2. to use the resulting best-fitted functions for filling the data gaps among the compiled data;

3. to evaluate the general validity of the resulting functions based on theoretical limits and their predictive correspondences with respect to literature equivalents.

3.3. METHODS

Prior to statistical analyses, the NB soil database (Chapter 2) had undergone rigorous amalgamation and harmonization procedures regarding detailed soil horizon and property assessments by soil associations and soil associates (Chapter 2). This included determining that general soil property range expectations were not exceeded. For example, the summation of specified sand, silt and clay % values (soil texture) cannot exceed 100%, soil Db should not exceed 2.4g/cm^3, all percentages (CF, SOM, carbon (C), and base saturation) must remain between 0 and 100%, soil pH values should vary at most from 2.5 to 8. For some soils, only qualitative descriptions were provided for CF content while others were provided with CF ranges (e.g., CF = "30 - 50").

Table 3.1 provides data frequencies, means and ranges for each database-listed soil variable prior to methodology and unit standardization. This standardization followed the variable-by-variable soil survey guidelines and specifications by McKeague (1978), the Mapping System Working Group (1981), the Expert Committee on Soil Survey (1982), Guertin *et al.* (1984), and GlobalSoilMap (http://www.isric.org/documents/document-type/globalsoilmap-specifications-v24-07122015).

Table 3.1. Overview of soil physical and chemical attributes within the amalgamated database prior to attribute standardization and data filling. Bolded variables represent those of which are modeled in this study whilst non-bolded represent those requiring additional standardization.

Attribute	Sample Size	% of total	Mean	Min	Max	Standardized
Sand (%)	1306	52.5	47.4	1.2	100	
Silt (%)	1306	52.5	34.3	0	76.1	
Clay (%)	1306	52.5	18.3	0	72.1	
CF (%)	885	35.5	14.18	0	80.0	
Db (g/cm^3)	938	37.7	1.4	0.11	2.8	
C (%)	1082	44.0	1.4	0	20.1	OM (%)
OM (%)	121	4.9	1.7	0	12.4	OM (%)
pH (H_2O)	1757	68.6	5.2	2.9	8.1	pH (H_2O and CaCl2)
pH($CaCl_2$)	689	26.9	4.6	2.9	7.5	pH (H_2O and CaCl2)
CEC (meq/100g)	659	26.5	15.0	1.0	73.9	
FC (%, gravimetric)	678	27.2	27.0	2	94	FC (volumetric and gravimetric)
FC (%, volumetric)	168	6.5	26.5	6	43	FC (volumetric and gravimetric)
PWP (%, gravimetric)	455	18.3	9.9	1	58	PWP (volumetric and gravimetric)
PWP (%, volumetric)	283	11.4	8.7	1	28	PWP (volumetric and gravimetric)
Base Saturation (%)	466	18.7	23.0	0	100	
Ca (meq/100g) (NH_4OAc)	990	38.7	2.5	0	65	Ca (meq/100g) (NH_4OAc)
Ca (kg/ha) (NH_4OAc)	67	2.6	458.3	44.8	2847	Ca (meq/100g) (NH_4OAc)
Ca (meq/100g) (NaCL)	92	3.6	2.4	0.1	12.2	Ca (meq/100g) (NH_4OAc)
Mg (meq/100g) (NH_4OAc)	974	38.0	0.5	0	8.2	Mg (meq/100g) (NH_4OAc)
Mg (kg/ha) (NH_4OAc)	67	2.6	76.6	0	336.3	Mg (meq/100g) (NH_4OAc)
Mg (meq/100g) (NaCL)	92	3.6	0.7	0	6.3	Mg (meq/100g) (NH_4OAc)
K (meq/ 100g) (NH_4OAc)	987	38.6	0.3	0	6.6	K (meq/ 100g) (NH_4OAc)
K_2O (kg/ha) (NH_4OAc)	67	2.6	97.1	22.4	336.3	K (meq/ 100g) (NH_4OAc)
K (meq/100g) (NaCL)	92	3.6	0.2	0	2.5	K (meq/ 100g) (NH_4OAc)

Once standardized, both simple and multiple linear regression analyses were conducted on each attribute to test the feasibility of modeling soil physical and chemical properties. Only the Table 3.1 variables in bold were subject to this analysis due to the lack of a consistent procedures for standardizing base cation measurements. The resulting PTF equations were evaluated in terms of their best-fitted intercept, regression coefficients, t- and p-values. The overall goodness-of-fit was expressed by the adjusted

coefficient of variation (R^2), the root mean square error (RMSE), and mean absolute error (MAE), and was visualized by plotting actual versus best-fitted data values. The PTF equations were subsequently used to predict the values of bolded variables across the database for general confirmation, and this was also done for the same data using published PTF equations as a method of validation.

3.4. RESULTS AND DISCUSSION

Soil Texture: Soil texture refers to the proportion of sand, silt, and clay in the fine earth fraction (<2mm) of soil (Weil and Brady, 2017), as classified by way of the texture triangle (Fig. 3.1; (Working Group on Soil Survey Data, 1982). In general, soil texture influences many soil physical and chemical processes and resulting attributes pertaining, e.g., to the extent of organic matter accumulation, which, in turn, contributes to soil structure by increasing soil porosities while lowering soil Db (Birkeland, 1999; Gessler *et al.*, 2000; McBratney *et al.*, 2000; Azlan *et al.*, 2013; Chaudhari *et al.*, 2013)**.** Clay in combination with fully decomposed SOM (humus) contributes to the soil colloidal fraction (Pitty, 1979; Fullen *et al.*, 2007), and, hence, improves moisture and nutrient retention (Anderson, 1988; Pitty, 1979).

The frequency of each texture class within the database was determined with results listed in Table 3.2. As seen above, pure silts, sandy clays, and the heaviest clays (clay percentage exceeding 80%) are absent from the soil surveys. This may be a result of surveying emphasis placed on agricultural soils, which would avoid surveying heavy textured wetlands, lacustrine and estuary deposits. Also missing from the data entries are

nearly pure sand- and silt-textured soils, as these are typically uncommon in forested landscapes. Some of the absent textural variations are likely due to:

1. translocation of fine-textured minerals (silts and clays) from steep slopes to low-lying positions and depressions (Simonson, 1959), and
2. non-linear trends with respect to soil depth due to secondary clay enrichments (i.e., Luvisols) and differences in soil parent materials (i.e., ablation over basal till), and vertical variations in alluvial fine to coarse deposits.

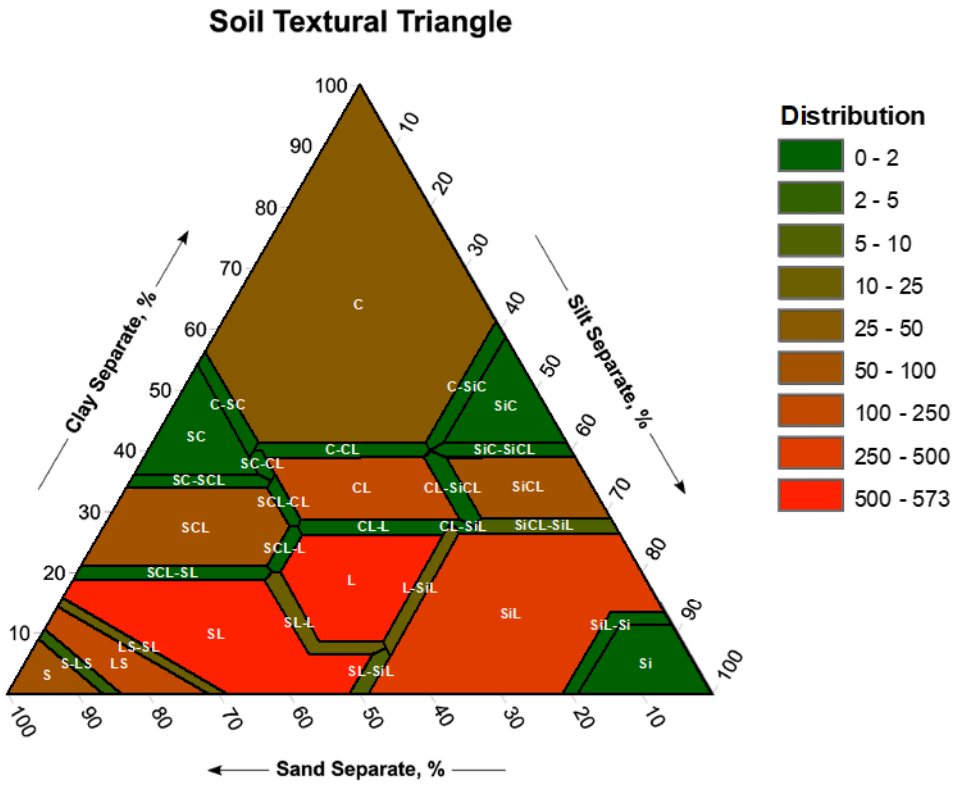

Figure 3.1. Representation of frequencies of each soil texture class with intermediates between classes overlain on soil texture triangle. Classes with no occurrence include silt (Si), sandy clay (SC), and heavy clay (uppermost portion of clay class, >80% clay).

Table 3.2. Frequency distribution of texture classes and ranges found within aspatial database.

Texture Class	Count	Texture Class	Count
Clay (C)	37	Loam (L)	573
Clay – Sandy Clay (C-SC)	0	Silt Loam – Loam (SiL-L)	16
Sandy Clay (SC)	0	Sandy Clay Loam – Sandy Loam (SCL – SL)	0
Clay – Silty Clay (C-SiC)	2	Clay – Clay Loam (C-CL)	0
Silty Clay (SiC)	0	Loamy Sand – Sandy Loam (LS-SL)	14
Silty Clay Loam (SiCL)	75	Loamy Sand (LS)	157
Clay Loam – Silty Clay Loam (CL-SiCL)	2	Sand – Loamy Sand (S-LS)	3
Silty Clay – Silty Clay Loam	1	Sand (S)	85
Sandy Clay Loam – Clay Loam (SCL-CL)	1	Clay Loam – Silt Loam (CL-SiL)	0
Clay Loam – Loam (CL-L)	1	Silt Loam (SiL)	329
Clay Loam (CL)	190	Silt Loam – Silt (SiL-Si)	1
Sandy Clay – Sandy Clay Loam (SC-SCL)	0	Silt (Si)	0
Sandy Clay – Clay Loam (SC-CL)	0	Sandy Loam – Silt Loam (SL-SiL)	10
Sandy Clay Loam (SCL)	53	Sandy Loam (SL)	519
Sandy Clay Loam – Loam (SCL-L)	0	Sandy Loam – Loam (SL-L)	14
Silty Clay Loam – Silt Loam (SiCL-SiL)	7		

Since the sand, silt and clay percentages within the mineral fraction of fine earth (i.e., gently crushed air-dried soil passing through a 2mm sieve) must add up to 100%, only two of these variables need to be modelled in order to fill the textural data gaps within the amalgamated and harmonized database. Since the texture-influencing variables within this database refer to soil depth, mode of surface deposition, and lithology (Anderson, 1988; Bui *et al.*, 2006), the missing database entries for the sand and silt fractions were obtained by evaluating the following generalized regression formulation:

$$Sand\ or\ Silt\ (\%) = f(MDEP + \log(Depth) + Mineral\ Hardness + Rock\ Type), n = 1{,}138 \quad (1)$$

Where *Depth* is the depth (cm) measured at the middle of the horizon from the bottom of the forest floor, *Mineral Hardness* is a numerical value determined by the Moh's hardness scale and assigned to the dominant mineral types of each parent material. If two rock types were present within the same parent material, then the hardness was averaged. *MDEP* (Mode of Deposition) and *Rock Type* are categorical variables describing the parent material's mode of deposition and dominant lithology, respectively. The formulated least-squares fitting results are listed, variable-by-variable, in Tables 3.3 and 3.4, represented with significant intercept, regression coefficients, t- and p-values, and with model performances (measured vs. fitted) presented in Fig. 3.2.

Table 3.3. Least squares modeling results for sand by soil depth, and parent material mode of deposition and lithology (rock type and mineral hardness), including best-fitted PTF intercept, regression coefficients, and associated t- and p-values.

Mode of Deposition	Estimate	t- value	p- value	Rock Type(s)	Estimate	t-value	p-value
Ablation/ Basal	-19.0730	2.863	0.0043	*Conglomerate, sandstone*	40.4028	8.166	<0.0001
Alluvium + Glaciofluvial	29.9414	10.913	<0.0001	*Conglomerate, sandstone, mudstone*	21.5430	3.018	0.0026
Colluvium + Water re-worked Till	10.9177	3.540	0.0004	*Sandstone*	16.9196	3.069	0.0022
Glaciofluvial	22.4670	10.486	<0.0001	*Sandstone, shale*	26.9470	5.928	<0.0001
Glaciofluvial + Marine	12.8926	3.900	0.0001	*Shale*	33.2131	4.436	<0.0001
Glaciomarine	-26.2909	6.399	<0.0001	*Shale, mudstone*	25.6190	2.792	0.0053
Glaciomarine/ Basal	37.7626	-7.024	<0.0001	*Shale, sandstone, conglomerate*	28.2841	4.931	<0.0001
Glaciomarine/ Marine	44.1617	5.542	<0.0001	*Shale, slate, quartzite*	27.4443	4.822	<0.0001
Lacustrine	31.2234	2.666	0.0078	*Slate, argillite, quartzite*	27.1734	5.058	<0.0001
Marine	34.9532	2.984	0.0029	**Other Variables**	**Estimate**	**t-value**	**p-value**
Residual + Colluvium	-15.2584	-2.865	0.0042	*Intercept*	-59.0462	-3.417	0.0007
				Log_{10} Depth	6.4952	7.631	<0.0001
				Mineral Hardness	15.3375	4.811	<0.0001

Table 3.4. Best-fitted model results for % silt, following the framework of the sand model with best-fitted PTF intercept, regression coefficients, and associated t- and p-values.

Mode of Deposition	Estimate	t- value	p- value	Rock Type(s)	Estimate	t-value	p-value
Ablation/ Basal	-17.5526	-3.520	0.0045	*Conglomerate, sandstone*	-30.8564	-8.331	<0.0001
Alluvium + Glaciofluvial	-12.2780	-5.978	<0.0001	*Conglomerate, sandstone, mudstone*	-19.7362	-3.694	0.0002
Colluvium + Water Re-worked Till	-6.0150	-2.605	0.0093	*Sandstone*	-13.6167	-3.299	0.0009
Glaciofluvial	-16.1226	-10.502	<0.0001	*Sandstone, shale*	-22.2808	-6.547	<0.0001
Glaciofluvial + Marine	-12.1065	-4.893	<0.0001	*Shale*	-30.9917	-5.529	<0.0001
Glaciomarine	-13.1527	-7.411	<0.0001	*Shale, mudstone*	-32.8401	-4.781	<0.0001
Glaciomarine/ Basal	11.2622	4.019	<0.0001	*Shale, sandstone, mudstone*	-12.6824	-1.966	0.0496
Glaciomarine/ Marine	-29.5591	-5.794	<0.0001	*Shale, sandstone, conglomerate*	-26.7660	-6.234	<0.0001
Lacustrine	-30.9938	-4.267	0.0028	*Shale, slate, quartzite*	-20.4679	-4.803	<0.0001
Residual	-17.1535	-1.956	0.0397	*Slate, argillite, quartzite*	-25.0773	-6.235	<0.0001
Residual + Colluvium	8.2288	2.060	<0.0001	**Other Variables**	**Estimate**	**t-value**	**p-value**
				Intercept	114.9107	8.882	<0.0001
				Log_{10} Depth	-6.4266	-10.086	<0.0001
				Mineral Hardness	-10.4199	-4.366	<0.0001

Of the variables outlined in Tables 3.3 and 3.4, some modes of deposition and rock types were not significant and removed from the analysis. For the sand model, these included (with p-value) *Ablation/ Residual* (0.7887), *Basal* (0.4804), and *Residual* (0.6086) for mode of deposition, and *Shale, Sandstone, Mudstone* (0.3059), *Rhyolite, andesite, granite* (0.8152), *Quartzite, sandstone* (0.1958), *Granite, gneiss, quartzite* (0.7728), and *Granite, gneiss, basalt, felsite* (0.8725) for rock types. The sand versus silt intercept and regression coefficients, as to be expected, carry opposite signs except for the *Ablation/Basal* (both negative) and *Glaciomarine/Basal* (both positive) entries. The former is due to elevated clay content, and the latter due to sandy beach deposits with

low clay. Generally, sand % increases as silt content decreases, this is also reflected by the increasing regression coefficients that signal increasing sand and decreasing silt with increasing soil depth and mineral hardness, as to be expected. The results in Tables 3.3 and 3.4 are due to the extent of glacial activity in NB, with 90% of the province influenced by glaciation (Pronk and Ruitenberg, 1991).

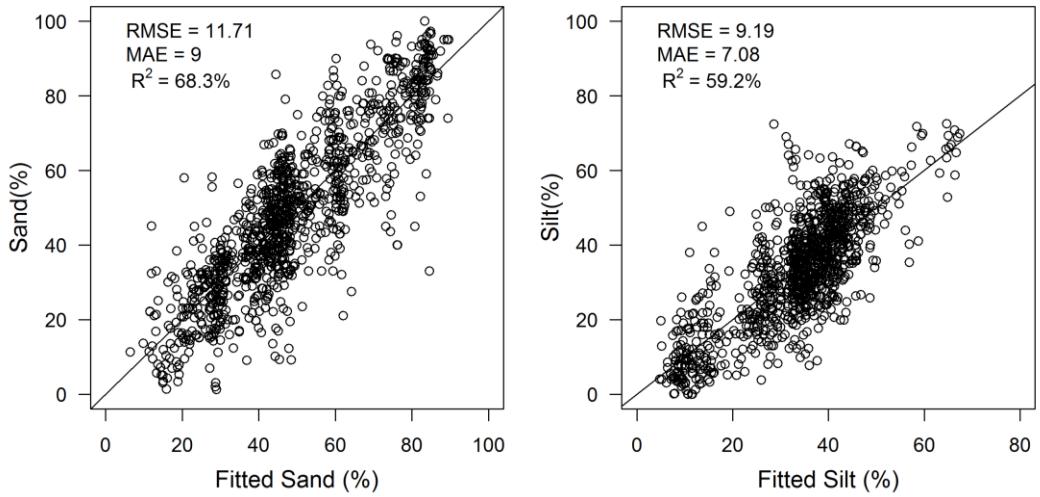

Figure 3.2. Performance plots (measured vs. fitted) for the sand model (left) and silt model (right) including RMSE, MAE, and R^2(adj.) as numerical performance indicators.

Repeating the analysis for clay, by switching clay percentage for sand or silt in Eq. 1 captured 51% of the variation in clay content. Hence, the combination of the silt and sand models are more effective at 59 and 68% (respectively) than using the combined 51 to 68% sand – clay or silt - clay regression results for estimating the missing database entries for textural composition.

In terms of literature comparison, there is no information about texture-informing PTFs. This is because soil texture specifications are generally used to quantify other sparsely sampled soil variables. Where geo-referenced texture data are not available,

DEM-derived topographic datasets have been used for computerized training purposes to estimate texture variations across landscapes (Zhao *et al.*, 2009; Dobarco *et al.*, 2016). In principle, this can also be done for this case study, but the amalgamated database provides only 223 texture entries with coordinate information.

Coarse Fragment Content: CF content refers to the amount of rock fragment in a soil, measured as a percentage. The amount of rock fragments in any soil depends on both topography and parent material (both mode of deposition and lithology). As parent material weathers, unconsolidated material will move depending on topographic position and on deposition event (glaciation, gravity, floods) (Pitty, 1979; Birkeland, 1999; Weil and Brady, 2017). As such, CF is more influenced by soil forming factors, namely topography, geology, and climate, than other soil properties (although chemical weathering is influenced by soil pH and moisture content). Due to this, CF is typically not predicted by way of PTF and seldom assessed in digital soil mapping. Comparing CF to soil forming factors within the database yielded:

$$CF = f(log_{10}(depth) + MDEP + Rock\ Type), n = 857 \quad (2)$$

Least squares fitting results are outlined in Table 3.5 and Fig. 3.3. As seen, surficial geology is the dominant influence on CF. Within the database, adequate information on topography is lacking, therefore, could not be included in predictions although believed to strongly influence CF.

Table 3.5. Best-fitted CF model results with best-fitted PTF intercept, regression coefficients, and associated t- and p-values.

Mode of Deposition	Estimate	t-value	p-value	Rock Type(s)	Estimate	t-value	p-value
Ablation/ Basal	-10.2375	-6.056	<0.0001	*Sandstone and Shale*	-6.0412	-3.619	0.0003
Ablation/ Residual	16.5645	5.224	<0.0001	**Other Variables**	**Estimate**	**t-value**	**p-value**
Alluvium	-25.4281	-7.953	<0.0001	*Intercept*	7.6088	3.174	0.0016
Ancient Alluvium	-25.1229	-5.731	<0.0001	*Log_{10} Depth*	10.4534	11.284	<0.0001
Basal	-5.9695	-3.352	<0.0001				
Glaciofluvial	9.9784	4.903	<0.0001				
Glaciofluvial + Marine	-19.3878	-9.224	<0.0001				
Glaciomarine	-20.1031	-11.769	<0.0001				
Glaciomarine/ Basal	-18.7318	-9.887	<0.0001				
Glaciomarine/ Marine	-24.4282	-6.610	<0.0001				
Lacustrine	-22.9514	-7.175	<0.0001				
Marine	-25.7216	-7.404	<0.0001				

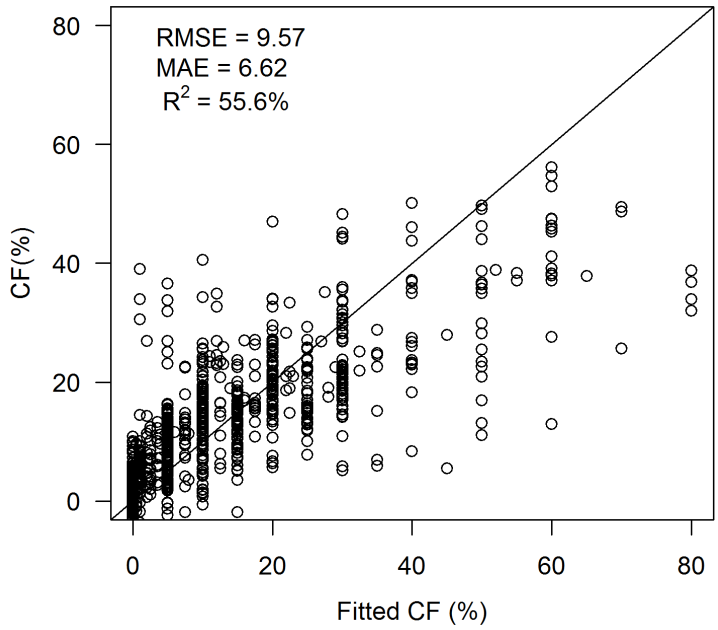

Figure 3.3. Performance plots (measured vs. fitted) for CF model including RMSE, MAE, and R^2(adj.) as numerical performance indicators.

pH: pH is a measure of the hydrogen ion concentration in solution and represents the level of acidity (or akalinity) within a soil (Weil and Brady, 2017). This is a function of microbial activity, moisture content and retention (topography), and parent material weatherability and fertility. pH influences the rate of OM decomposition (by affecting microbial populations) and influences the availability of SOM and base cations for cation exchange and root uptake (Pitty, 1979; Bui *et al.*, 2006)**.** pH controls the retention of base cations via reduction reactions, with low pH reducing base cations into soluble form, which are lost with downward and laterally moving water. Therefore, pH has a strong influence on the cation exchange capacity (CEC) of a soil (Pitty, 1979; Weil and Brady, 2017).

As part of the soil surveying procedures, pH was measured for each soil horizon via two methods depending on moisture conditions: from dry ($CaCl_2$) to moist/ wet (H_2O). Typically, these determinations are well correlated to one another (as seen in Table 3.6 and Fig. 3.4) such that:

$$pH(H_2O) = f(Intercept + pH(CaCl_2))\ n = 464 \quad (3)$$

$$pH(CaCl_2) = f\big(Intercept + pH(H_2O)\big) n = 464 \quad (4)$$

Table 3.6. Model standardization performance for pH measured in both H_2O (left) and $CaCl_2$ (right) including estimate, t-value, and p-value.

pH (H_2O)	Estimate	t- value	p- value	pH ($CaCl_2$)	Estimate	t-value	p-value
Intercept	0.7924	10.53	<0.0001	*Intercept*	-0.1934	-2.401	0.0167
pH ($CaCl_2$)	0.9719	59.04	<0.0001	*pH (H_2O)*	0.9087	59.043	<0.0001

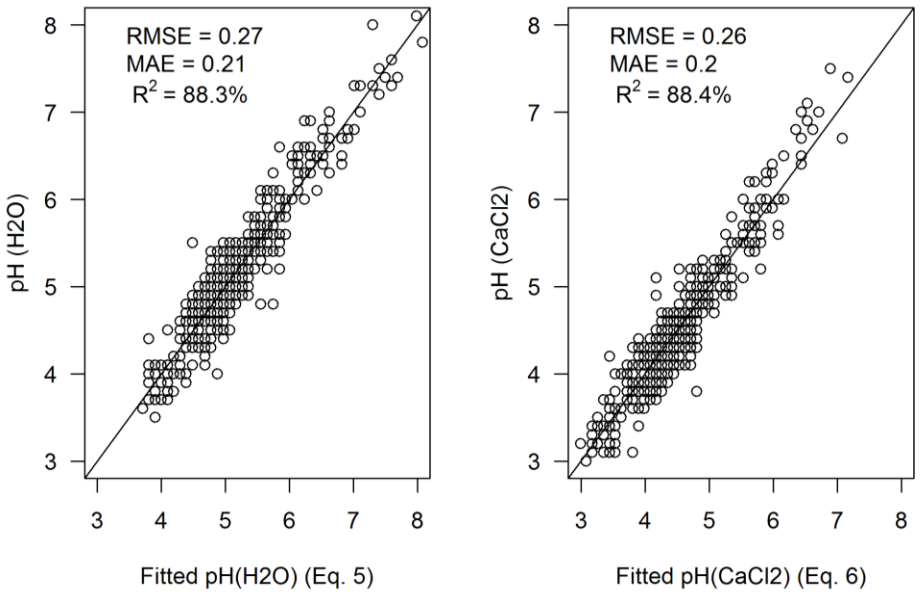

Figure 3.4. Performance plot (measured vs, fitted) for standardization of pH (H_2O) (**left**) and pH ($CaCl_2$) (**right**) including RMSE, MAE, and R^2(adj.) for numerical determinations of performance.

Many published PTFs exist for standardizing pH and an overview of such can be found in Appendix C, Table 11 (Example regression equations for converting values of pH between different methods) of Specifications Version 1 of GlobalSoilMap.net products. A difficulty with soil survey data is that the surveys typically do not specify the exact procedure applied when determining physical and chemical properties. The reports typically refer to a manual that specifies numerous sampling techniques. Alternatively, some of the published PTFs require additional information pertaining to chemical properties and these typically are not measured for each sample. Equations 3 and 4 were incorporated into the database to create two continuous measures of soil pH.

Soil Organic Matter: Soil organic matter (SOM) is composed of organic residues at various stages of decomposition and is one of the key properties influencing soil fertility

and quality (Whitbread, 1992; Bui *et al.*, 2006; Fullen *et al.*, 2007; Azlan *et al.*, 2013). SOM acts as a nutrient pool, aids in soil aggregation (thus improves aeration), decreases Db and facilitates infiltration (when not dry) while increasing water holding capacity. SOM is part of the colloidal fraction of soil (along with clay particles), aids in increasing the CEC for the retention of Ca, Mg, and K ions, and releases nitrogen, phosphorus, and sulfur to the soil solution as part of the decomposition process (Pitty, 1979; Whitbread, 1992; Birkeland, 1999; Kerry *et al.*, 2012; Azlan *et al.*, 2013). SOM is a function of biomass introduction and microbial decomposition which is dominantly controlled by pH, temperature, and moisture (Bui *et al.*, 2006). Thus, SOM concentrations are a result of topography, climate, vegetation and weathering of parent material (Azlan *et al.*, 2013).

It is common to include the weight percentage of SOM when analyzing soil attribute relationships in relation with soil texture, depth and CF content (Balland *et al.*, 2008; Pollacco, 2008). As such, SOM has been determined through loss ignition at 500°C, through loss by wet-chemical oxidation, and/or direct elemental C analysis. The conversion factor from soil C to SOM takes the form of SOM = 1.72 C%. This implies that 58% of SOM is organic C (Romano and Palladino, 2002; Pollacco, 2008; Chaudhari *et al.*, 2013; Poggio *et al.*, 2013). Although this percentage can vary (Pribyl, 2010), it nevertheless serves as a means to convert all soil C to SOM. Through preliminary analyses, missing SOM data gaps were filled using the expression:

$$\log_{10}(SOM\%) = f(\log_{10}(Depth) + FC + h + f), n = 606 \quad (5)$$

Where FC is field capacity (gravimetric moisture retention, %, at -33kpa), $Depth$ is depth of the center of the horizon from the surface (cm), and f and h are binary for whether the horizon is enriched in iron or humus (respectively). Alternatively, excluding FC from the analysis yields results with a slightly reduced performance, but increased sample size.

$$\log(SOM\%) = f(Intercept + \log_{10}(Depth) + f + h), n = 1{,}177 \qquad (6)$$

It is arguable that SOM is more easily measured in a laboratory setting than is field capacity (FC), therefore, FC should be excluded when modeling SOM. Table 3.7 represents the coefficients (with estimates, t-values, and p-values) for both SOM models with comparison of fitted plots outlined in Fig. 3.5.

Table 3.7. Comparison of best-fitted model results for the two SOM PTFs, including intercept, regression coefficients, and associated t- and p-values.

Model	Variable	Estimate	t- value	p- value
All Variables (Eq.5)	*Intercept*	0.2095	3.381	<0.0001
	Log_{10}(Depth)	-0.4428	-15.105	<0.0001
	FC	0.0120	9.276	<0.0001
	f	0.4744	18.159	<0.0001
	h	0.4737	9.586	<0.0001
Exclude FC (Eq.6)	*Intercept*	0.6178	17.09	<0.0001
	Log_{10} (Depth)	-0.5475	-24.31	<0.0001
	f	0.5222	24.76	<0.0001
	h	0.6239	15.46	<0.0001

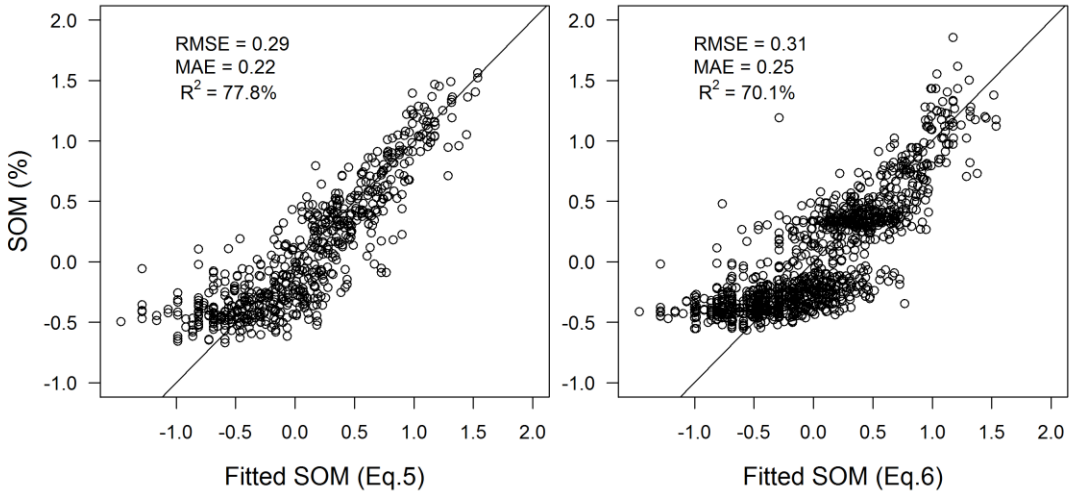

Figure 3.5. Comparison of PTF-derived performance plots (measured vs. fitted) for the OM model including all variables (**left**) and excluding FC (**right**), including RMSE, MAE, and R^2(adj.) as numerical performance indicators.

Bulk Density: Bulk density (Db) is the mass per unit volume of dry soil, including pore space, measured in g/cm^3 (Birkeland, 1999; Weil and Brady, 2017). Db is therefore strongly related to soil texture, soil structure, SOM content, and soil depth (Pitty, 1979; Chaudhari *et al.*, 2013, Nanko *et al.*, 2014). These, in turn, are dependent on parent material, vegetation, and topography. Typically, Db in solid rock ranges between 2.5-3g/cm^3, whereas mineral horizons remain within the 0.5 < Db < 2.5 g/cm^3 range (Birkeland, 1999). With increasing Db, soil infiltration, permeability and root penetration decrease with root extension ceasing when Db > 1.5 g/cm^3 (Pitty, 1979).

Typically, Db is determined by extracting a known volume of soil using, e.g., a non-compressing soil corer, and drying and weighing the extracted soil. Since Db varies by

texture and SOM, it is important to convert the sand, silt, clay, and SOM percentages of the fine earth particle fraction into actual per soil weight percentages such that:

$$Sand\% + Silt\% + Clay\% = 100 \quad (7a)$$

$$Sand_W\% + Silt_W\% + Clay_W\% + SOM_W\% = 100 \quad (7b)$$

$$Sand_W + Silt_W + Clay_W + SOM_W = 1 \quad (7c)$$

With this, preliminary analysis for predicting Db yielded:

$$Db = f(Intercept + \log_{10}(Depth) + \log_{10}(SOM_W) + gley + Basal + \left(\frac{Ablation}{Residual}\right) + f + FC), n = 653 \quad (8)$$

Where SOM_W is soil organic matter (%) as particle fraction, *gley* is binary for whether the horizon is gleyed, f is binary for whether horizon is enriched in iron, and *Basal* and *Ablation/ Residual* are binary for whether the parent material mode of deposition is one of those two forms. Excluding information on parent material (if data on parent material is unavailable) yielded:

$$Db = f(Intercept + \log_{10}(Depth) + \log_{10}(SOM_W) + gley + f + FC), \mathrm{n} = 65 \quad (9)$$

Alternatively, omitting FC while including parent material returned:

$$Db = f(Intercept + Log\log_{10}(Depth) + gley + f + \log_{10}(SOM_W) + \left(\frac{Ablation}{Residual}\right) +$$

$$Basal + GLMA + \left(\frac{Ablation}{Basal}\right) + (COL + WWTL)), \mathrm{n} = 795 \qquad (10)$$

With *GLMA* (glaciomarine), *Ablation/Residual*, *Ablation/Basal*, *COL + WWTL* (colluvium and water re-worked till), and *Basal* representing additional landform-based predictors. Results for Eqs. 8 through 10 are outlined in Table 3.8 and Fig. 3.6.

Table 3.8. Comparison of best-fitted model results for the three Db PTFs, including intercept, regression coefficients, and associated t- and p-values.

Model	Variable	Estimate	t- value	p- value
All Variables (Eq.8)	*Intercept*	0.8359	15.608	<0.0001
	Log_{10} (Depth)	0.2592	13.228	<0.0001
	Log_{10} (SOM_W)	-0.1599	-7.044	<0.0001
	gley	0.1375	8.739	<0.0001
	Basal	0.1526	11.401	<0.0001
	Ablation/Residual	0.1774	5.888	<0.0001
	f	-0.1141	-6.092	<0.0001
	FC	-0.0096	-12.364	<0.0001
Exclude Parent Material + Include FC (Eq.9)	*Intercept*	0.8222	13.933	<0.0001
	Log_{10} (Depth)	0.2474	11.479	<0.0001
	Log_{10} (SOM_W)	-0.1957	-7.942	<0.0001
	gley	0.1639	9.621	<0.0001
	f	-0.1035	0.0206	<0.0001
	FC	-0.0083	-9.805	<0.0001
Include Parent Material + Exclude FC (Eq.10)	*Intercept*	0.3018	8.690	<0.0001
	Log_{10} (Depth)	0.2898	15.308	<0.0001
	gley	0.1303	8.253	<0.0001
	f	-0.0927	-5.091	<0.0001
	Log_{10} (SOM_W)	-0.2838	-13.977	<0.0001
	Ablation/ Residual	0.1534	4.750	<0.0001
	Basal	0.1698	9.081	<0.0001
	GLMA	0.1879	4.252	<0.0001
	Ablation/ Basal	0.1334	4.476	<0.0001
	COL + WWTL	0.1288	3.365	0.0008

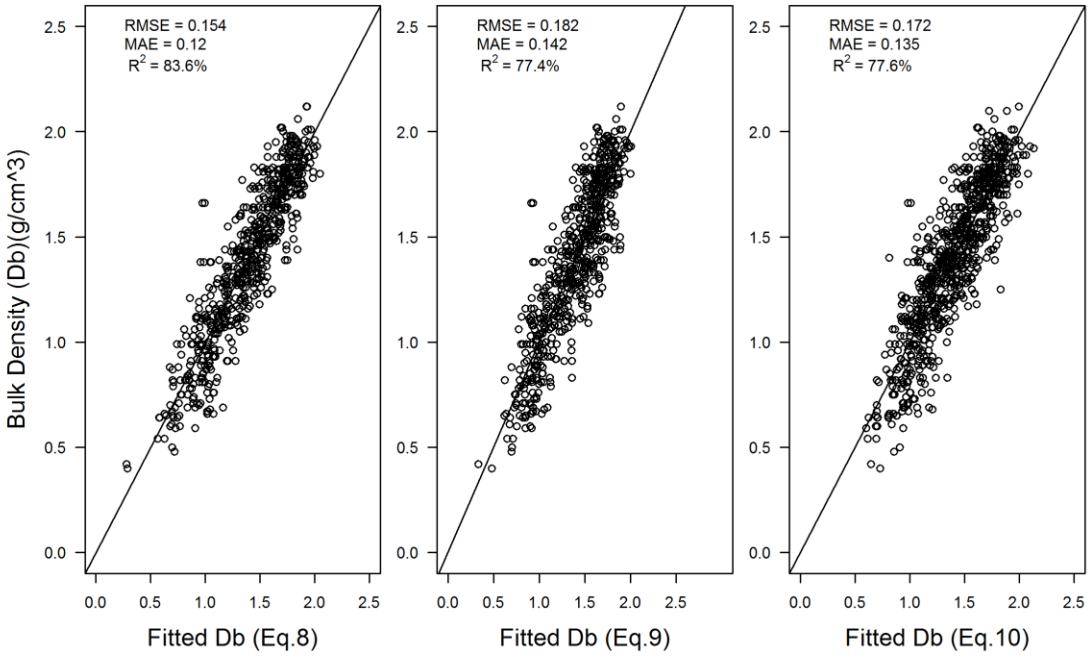

Figure 3.6. Comparison of PTF-derived Db performance plots (actual vs. fitted) for all variables (**left**), excluding parent material (**center**) and excluding FC (**right**). Fitted values were compared to measured values to determine performance.

The results outlined in Table 3.8 and Fig. 3.6 shows that the model with all variables (Eq. 8; R^2 = 83.6%) fits best, followed by the model excluding parent material (Eq. 10; R^2 = 77.6%), then excluding FC (Eq.9; R^2 = 77.4%).

The overall results are consistent with those of Pitty (1979), Heuscher *et al.* (2005), Jutras and Arp (2010), Chaudhari *et al.* (2013), Nanko *et al.* (2014) in that depth and SOM_W dominate Db predictions, with Db increasing with depth, and decreasing with increasing SOM due to the structure- building influence of SOM. Among the parent materials, Db values are higher for basal tills due to the compacted nature of these tills. In addition, the ablation till overlying residual formations have higher Db values due to the coarse nature of these parent materials. The same is noted via Eq. 10 for the coarse-textured

glaciomarine, ablation over basal till, and the colluvium and water re-worked till combinations. While sandy soils are expected to also have lower Db values, this trend is camouflaged by the negative correlation between Db and FC, whereby FC decreases with increasing sand content.

The results outlined in Eq. 9 follow a similar trend to that of Balland *et al.* (2008) for New Brunswick and Nova Scotia soils.

$$Db = \frac{1.23+(Dp-1.23)(1-exp(-0.0106 \cdot Depth))}{1+6.83 \cdot SOM_W\%} \quad (11)$$

where *Dp* is the particle density (g/cm^3). The Db model by Benites et al. (2007) based on 19,651 Brazilian soil samples also indicated that SOM (in the form of % organic carbon, OC) is a dominant Db predictor, as follows:

$$Db = 1.660 - 0.318(OC\%)^{0.5} \quad (12)$$

Comparing the results from Eq. 9 to those of Balland and Benites yield similar results (Fig. 3.7).

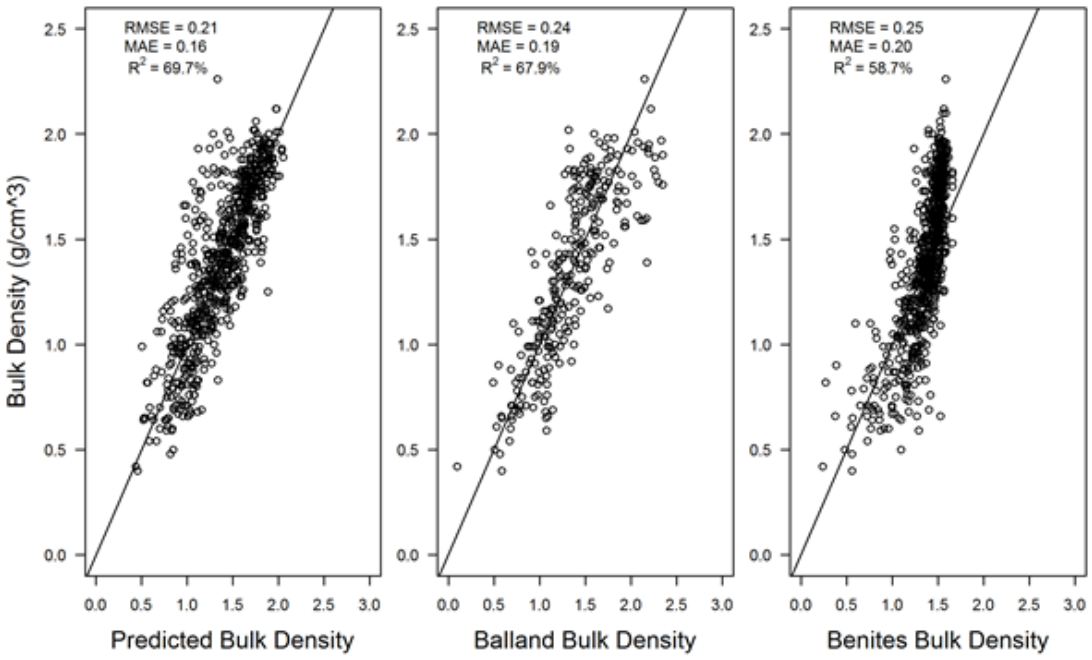

Figure 3.7. Comparison of actual vs. predicted Db values (Eq. 9) (**left**) with Db predictions from Balland *et al*. (2008) (**center**) and Benites *et al*. (2007) (**right**) when compared to measured Db in database.

All three models have similar predictive capabilities with the newly derived PTF performing best, followed by Balland Db, then Benites Db. All three results show the influence of SOM on determining bulk density but excluding depth in the Benites model likely results in the slightly poorer performance.

Field Capacity: water retention within a soil is significantly important to both soil property variation and productivity potential for vegetation (Pitty, 1979; Dobos *et al.*, 2001; Zhu and Mackay, 2001). The amount of water retained in a soil affects both nutrient and oxygen availability to roots (Pitty, 1979). Within a soil sample, water is acted on by different forces, with gravity driving water downward, cohesion attracting water to other water molecules, and adhesion (capillarity) moving water laterally and against gravity via electrostatic charges on soil particles. This results in negative forces responsible for

retaining water within soils, resulting in higher pressures (Pitty, 1979; Birkeland, 1999; Weil and Brady, 2017). FC (%) is regarded as the upper limit of available water for plants (low pressure gradient, -33kPa) and is generally attained 24 hours post soil-saturating precipitation event (Salter and Williams, 1965a, 1965b; Weil and Brady, 2017)**.** Modeling FC yielded:

$$FC_W = f(Intercept + Sand\% + SOM\% + \left(\frac{Ablation}{Basal}\right) + \left(\frac{Ablation}{Residual}\right) + (Alluvium\ and\ GLFL) + GLFL\ and\ MA + GLMA + Db + Mineral\ Hardness), n = 657 \quad (13)$$

Where FC_W is gravimetric moisture content (%) at -33kPa, *and Ablation/Basal*, *Ablation/Residual*, *Alluvium and GLFL*, *GLFL and MA*, *GLMA* are parent material modes of deposition (GLFL = Glaciofluvial, MA = Marine). Excluding parent material returned:

$$FC_W = f(Intercept + Sand\% + SOM\% + Db), n = 670 \quad (14)$$

while excluding Db resulted in:

$$FC_W = f(Intercept + Sand\% + OM\%), n = 710 \quad (15)$$

The best-fitted results so obtained are outlined in Table 3.9 and plotted in Fig. 3.8. Since FC has been reported in gravimetric as well as volumetric (FC_V) data, the FC analysis was done for FC_W because of small FC_V sample size (n = 168). Where Db was entered, FC_V values were converted to FC_W by setting $FC_W = FC_V$ / Db.

Table 3.9. Comparison of best-fitted model results for the three gravimetric FC PTFs, including intercept, regression coefficients, and associated t- and p-values.

Model	Variable	Estimate	t- value	p- value
All Variables (Eq. 13)	*Intercept*	44.9953	23.562	<0.0001
	Sand%	-0.2211	-13.367	<0.0001
	OM%	0.7081	10.060	<0.0001
	Ablation/Basal	-4.0956	-3.331	<0.0001
	Ablation/Residual	5.3602	5.124	<0.0001
	Alluvium and GLFL	-7.3702	-5.267	<0.0001
	GLFL and MA	-5.7806	-4.356	<0.0001
	GLMA	-7.0349	-4.292	<0.0001
	Db	-14.4725	-17.263	<0.0001
	Mineral Hardness	1.2884	5.227	<0.0001
No Parent Material (PM) (Eq. 14)	*Intercept*	51.4764	37.737	<0.0001
	Sand%	-0.2402	-20.634	<0.0001
	OM%	0.7097	9.673	<0.0001
	Db	-14.3442	-17.962	<0.0001
No PM, No Db (Eq. 15)	*Intercept*	29.4223	40.700	<0.0001
	Sand%	-0.2429	-17.160	<0.0001
	OM%	1.5607	22.900	<0.0001

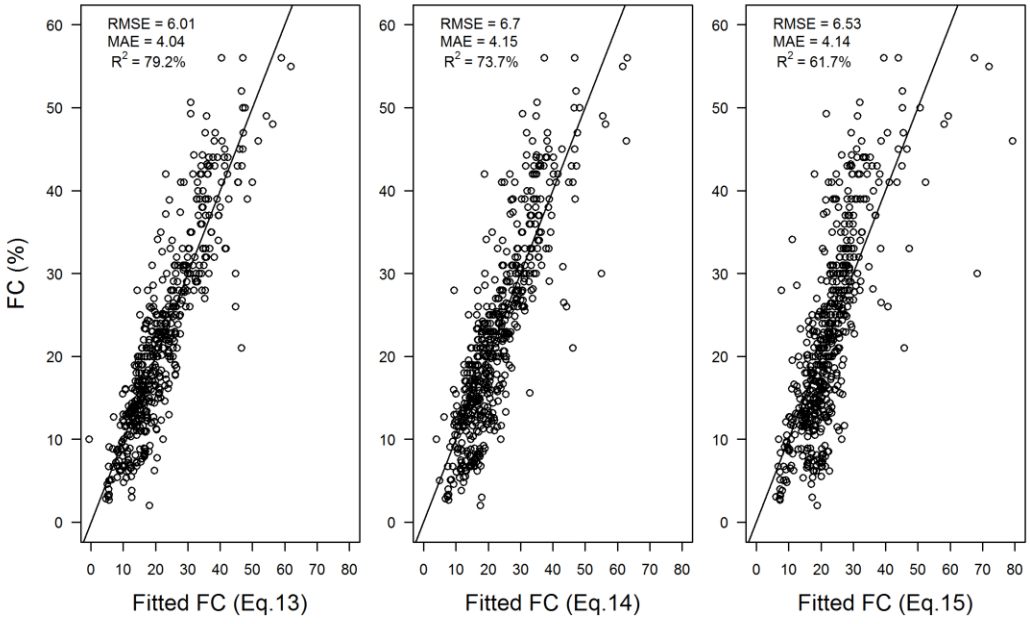

Figure 3.8. Comparison of PTF-derived FC performance plots (actual vs, fitted) for all variables (**left**), excluding parent material (**center**), and excluding both parent material and Db (**right**). Fitted values were compared to measured values to determine performance.

These results follow general expectations such that FC_W increases with increasing SOM and decreases with increasing sand content and Db. Excluding Db from the analysis reduced the goodness-of-fit considerably. Including the entries for coarser soil parent material also lowered FC_W. FC increases with increasing mineral hardness and where the soil parent material refers to ablation till on residuals is unexpected but dropping the soil parent material and mineral hardness entries altogether decreased the goodness-of-fit marginally, from R^2 = 79.2% to 73.7%. This drop would even be smaller by excluding mineral hardness and the ablation till on residuals combination only.

The results for FC (gravimetric and volumetric) are consistent with literature in that soil texture and SOM are important factors when predicting water retention (Pitty, 1979; McBratney *et al.*, 2002; Balland *et al.*, 2008; Pollacco, 2008; Ostovari *et al.*, 2015). When including Db in the model, predicted FC_W values became less than 0, otherwise FC_W values remained above 0. The results from Eq. 13 (all variables) were compared with published PTFs for gravimetric (FC_W) and volumetric (FC_V) field capacities, as follows:

1. Balland *et al.* (2008):

$$FC_W = SP_W\%\left(1 - exp\left(\frac{-0.588(1-Sand_W\%)-1.73\cdot SOM_W\%}{SP_W\%}\right)\right) \quad (16)$$

Where $SP_W\%$ is moisture content at saturation point.

2. Beke and MacCormick (1985):

$$FC_V = 64.37 - 0.5(Sand\%) - 0.23(Silt\% \cdot Db) \quad (17)$$

3. Rawls and Brakensiek (1982):

$$FC_V = 0.2576 - 0.002(Sand\%) + 0.0036(Clay\%) + 0.0299(SOM\%) \quad (18)$$

4. Gupta and Larson (1979):

$$FC_W = 0.003075(Sand\%) + 0.005886(Silt\%) + 0.008039(Clay\%) + 0.002208(SOM\%) - 0.14340(Db) \quad (19)$$

5. Lal (1979):

$$FC_W = 0.065 + 0.004(Clay\%) \quad (20)$$

Volumetric FCs (Eqs. 17-18) were converted to gravimetric by dividing volumetric FC by Db. FCs measured as weight fractions (Eqs. 16, 19-20) were converted to percentages by multiplying FC by 100. With this, model results are compared in Fig. 3.9.

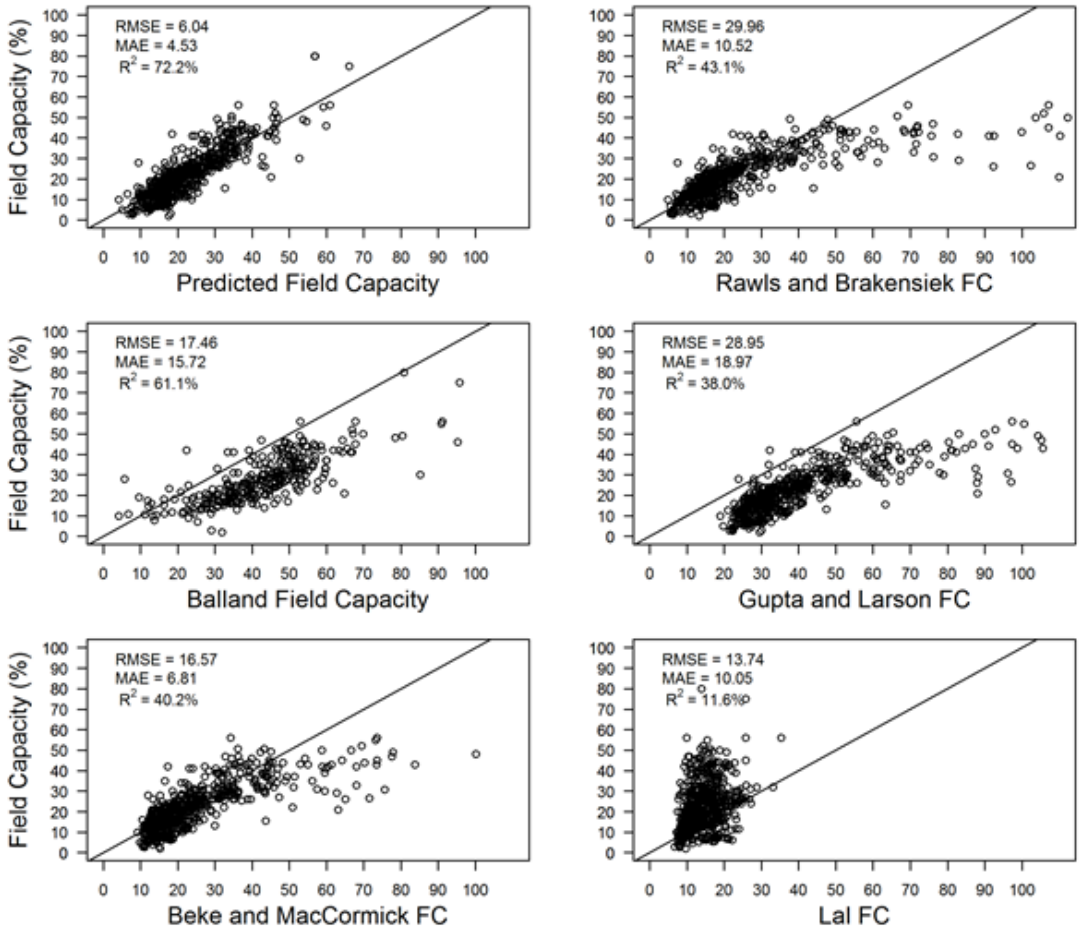

Figure 3.9. Comparison of actual vs. predicted FC (Eq. 15) (top left), and published PTFs, namely Rawls and Brakensiek (1982) (**top right**), Balland *et al*. (2008) (**mid left**), Lal (1979) (**mid right**), Beke and MacCormick (1985) (**bottom left**), and Oosterveld and Chang (1980) (**bottom right**) when compared to measured FC in database.

Permanent Wilting Point: Permanent wilting point (PWP, %) is the moisture content at the upper reaches of the pressure gradient (-1500kPa). At this pressure, moisture is held onto the soil particles so tightly via matric forces (adhesion) that it is no longer available for root uptake. As this occurs, plants reach a stage in which irreversible wilting occurs unless water is added to the soil (McBratney *et al.*, 2002; Weil and Brady, 2017). This limit is determined by applying a pressure of 1,500kPa on saturated soil

samples placed on a porous plate inside a pressure chamber, for 24 hours, and subsequently weight determining the remaining soil moisture content. The resulting PWP entries in the database were analyzed using the following formulation:

$$PWP_W = f(Intercept + SOM\% + Clay\% + Basal + \left(\frac{GLMA}{Basal}\right) + Residual + Colluvium), n = 566 \quad (21)$$

Where PWP_W is the moisture percentage at -1500kPa and $Residual + Colluvium$ is a binary field for that mode of deposition. Without reference to parent material, the formula becomes:

$$PWP_W = f(Intercept + SOM\% + Clay\%), n = 582 \quad (22)$$

The best-fitted results so obtained are listed in Table 3.10 and are plotted in Fig. 3.10.

Table 3.10. Comparison of best-fitted model results for the two gravimetric PWP PTFs, including intercept, regression coefficients, and associated t- and p-values.

Model	Variable	Estimate	t- value	p- value
All Variables (Eq. 21)	*Intercept*	-0.16404	-0.317	0.7511
	SOM%	1.3018	34.222	<0.0001
	Clay%	0.2197	13.446	<0.0001
	Basal	1.4570	3.603	0.0003
	GLMA/Basal	3.1493	3.572	0.0004
	Residual and Colluvium	6.3863	5.193	<0.0001
No Parent Material (PM) (Eq. 22)	*Intercept*	1.3726	4.672	<0.0001
	SOM%	1.3289	36.570	<0.0001
	Clay%	0.1972	14.878	<0.0001

The regression coefficients in Table 3.10 for SOM and $Clay$ confirm their role in retaining soil moisture against vegetation uptake at the PWP. The added contributions by soil parent material are of significance, likely due to the influence of parent material on

texture composition, but the PWP model with all variables produces a statistically insignificant intercept whereas excluding PM yields more significant estimates.

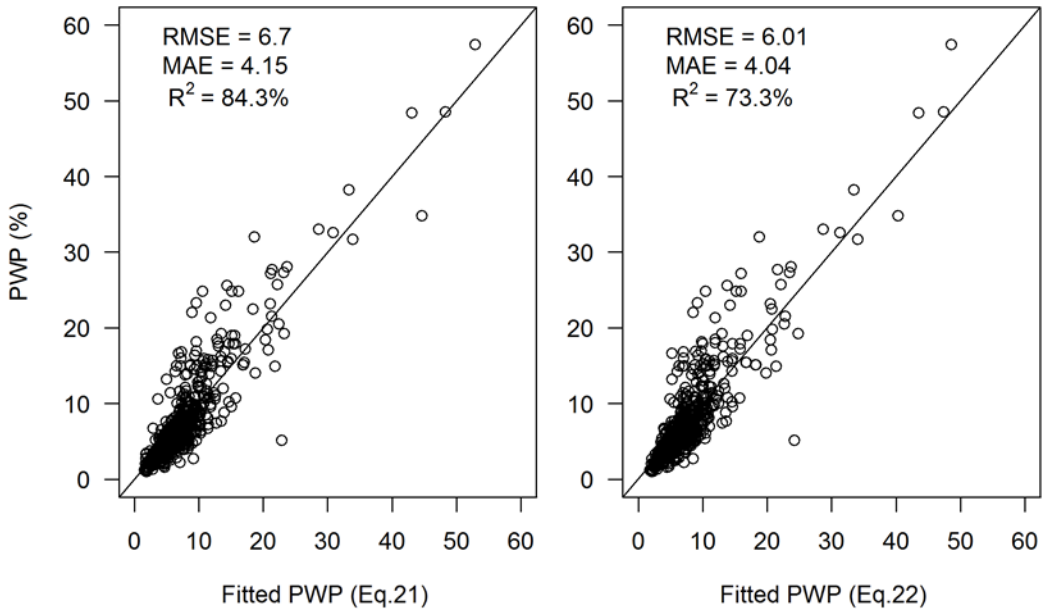

Figure 3.10. Comparison of PTF-derived PWP performance (actual vs. fitted) for all variables (**left**) and excluding parent material (**right**). Fitted values were compared to measured values to determine performance.

In terms of PWP literature comparisons, the results agree with those of Rawls and Brakensiek (1982), Beke and MacCormick (1985), and Balland *et al.* (2008) in that clay and SOM are the two dominant predictors. However, applying the corresponding equations to the PWP database entries produced wider actual versus predicted scatter plots with a general trend towards higher than actual PWP values and therefore lower R^2 values than noted in Fig. 3.10, as shown in Fig. 3.11. Arranging R^2s in descending order reveals:

Eq. 21 > Eq. 22 >> Balland *et al*. (2008) > Rawls and Brakensiek (1982) ≈ Beke and MacCormick (1985) >> Oosterveld and Chang (1980) > Lal (1979).

This sequence is somewhat determined by the combined inclusion of the SOM and clay variables. The advantage of the Balland et al. (2008) formulation is that it produces $0 < PWP_W < FC_W$ values only. The absence of $PWP < 0$ values is guaranteed by the other formulations except for the negative regression coefficient by soil depth in Oosterveld and Chang (1980). For Eq. 21, this can be enforced by setting the intercept equal to zero.

The plots in Fig. 3.11, other than the top left entry obtained via Eq. 21, are based on the following equations, and noting the volumetric to gravimetric $PWP_W = PWP_V/Db$ conversion.

1. Balland *et al.* (2008):

$$PWP_W = FC_W\left(1 - exp\left(\frac{-0.511(Clay_W\%) - 0.865(SOM_W\%)}{FC_W}\right)\right) \quad (23)$$

2. Beke and MacCormick (1985):

$$PWP_V = 4.41 + 0.66(Clay\%) + 1.07(SOM\% \cdot Db) \quad (24)$$

3. Rawls and Brakensiek (1982):

$$PWP_V = 0.026 + 0.005(Clay\%) + 0.0158(SOM\%) \quad (25)$$

4. Lal (1979):

$$PWP_V = 0.006 + 0.003(Clay\%) \quad (26)$$

5. Oosterveld and Chang (1980):

$$PWP_W = 4.035 + 0.299(Clay\%) - 0.034(Sand\%) - 0.016(Depth) \quad (27)$$

Where $Depth$ (cm) is the depth of the mid-point of the horizon.

Volumetric PWP results from of Beke and MacCormick and Rawls and Brakenseik were converted to gravimetric PWP as was FC. PWPs for Eqs. 24, 26, and 27 were converted to percentages by multiplying PWP by 100. Db was included in the analysis, but it was realized that this inclusion causes some PWP values to become less than zero, exceeding known thresholds for the moisture content. Thus, Db was excluded while modeling PWP. Once removed, PWP PTFs were compared to test conformance (Fig. 3.11).

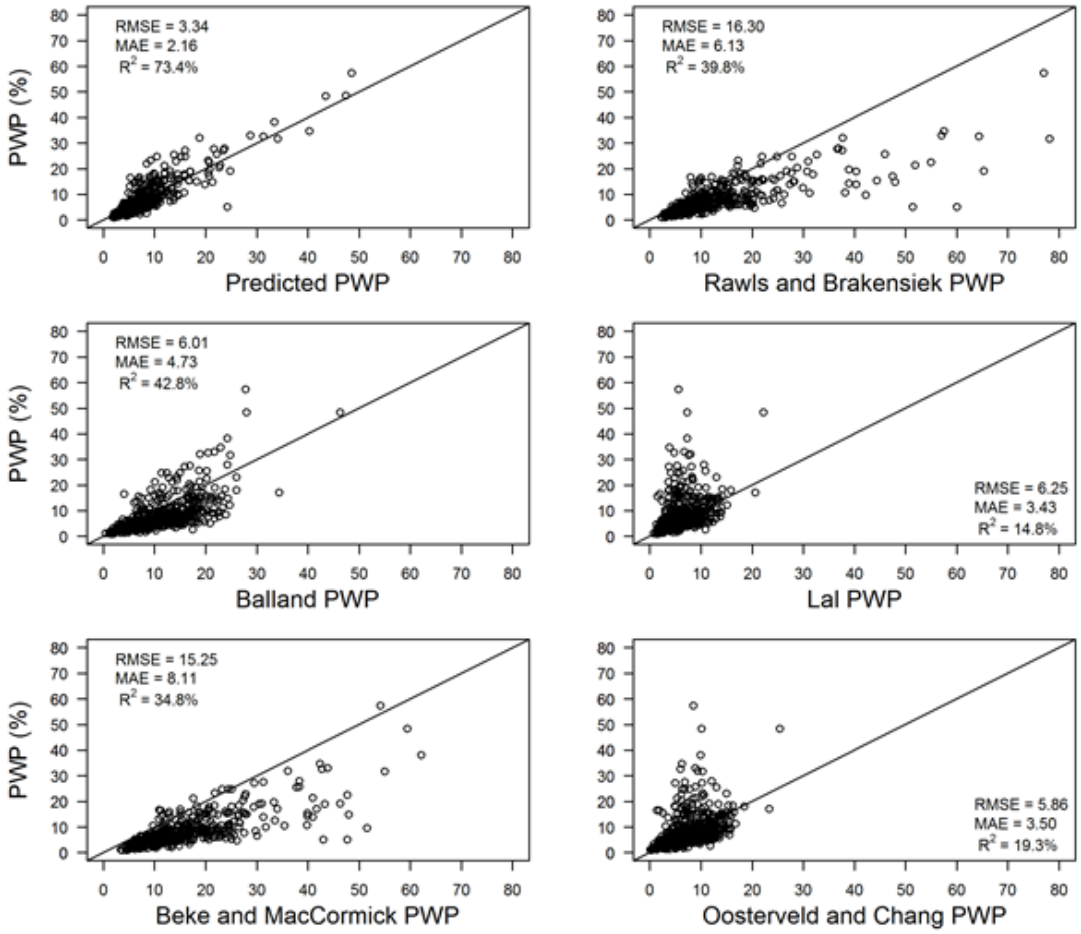

Figure 3.11. Comparison of actual vs. predicted PWP PTF (Eq. 22) (**top left**), and published PTFs, namely Rawls and Brakensiek (1982) (**top right**), Balland *et al*. (2008) (**mid left**), Lal (1979) (**mid right**), Beke and MacCormick (1985) (**bottom left**), and Oosterveld and Chang (1980) (**bottom right**) when compared to measured PWP values in database.

Cation Exchange Capacity: CEC, measured as meq/100g, is the total sum of negative charges within a soil and represents how many base cations can be held within that soil (Birkeland, 1999; Liao *et al.*, 2015; Olorunfemi *et al.*, 2016; Weil and Brady, 2017). CEC is recognized as the largest single indicator of soil fertility, nutrient holding capacity, and overall quality (Pitty, 1979; McBratney *et al.*, 2000). CEC strongly influences the ability

of the soil to buffer against acidity, retain nutrients, and supply nutrients to roots (Krogh *et al.*, 2000; Liao *et al.*, 2015).

Determination of CEC, like Db, is very costly and time consuming therefore it is common practice to employ PTFs to determine these values (Liao *et al.*, 2015). Traditionally, the literature states that CEC is a linear function of carbon (SOM) and clay percentages (Syers *et al.*, 1970; Brady, 1990; Meyer and Arp, 1994; Krogh *et al.*, 2000; Olorunfemi *et al.*, 2016). Building on this theory, CEC was tested against many soil properties resulting in Eq. 28:

$$CEC = f(Intercept + Clay_W + SOM_W + f + h), \mathrm{n} = 639 \quad (28)$$

CEC was also modeled comparing $Clay_W$ and SOM_W with and without intercepts:

$$CEC = f(Clay_W + SOM_W), n = 639 \quad (29)$$

$$CEC = f(Intercept + Clay_W + SOM_W), n = 639 \quad (30)$$

The best-fitted results obtained with these equations are listed in Table 3.11, with actual versus best-fitted values plotted in Fig. 3.12.

Table 3.11. Comparison of best-fitted model results for the three CEC PTFs, including intercept, regression coefficients, and associated t- and p-values.

Model	Variable	Estimate	t- value	p- value
All Variables (Eq. 28)	*Intercept*	3.7244	5.346	<0.0001
	$Clay_W$	22.3254	10.214	<0.0001
	SOM_W	207.5449	16.090	<0.0001
	f	3.3563	5.255	<0.0001
	h	13.5131	8.439	<0.0001
Clay + SOM (Eq. 29)	$Clay_W$	29.9340	21.57	<0.0001
	SOM_W	306.1330	47.46	<0.0001
Intercept + Clay + SOM (Eq. 30)	*Intercept*	3.4257	5.815	<0.0001
	$Clay_W$	17.9277	7.262	<0.0001
	SOM_W	292.1864	43.400	<0.0001

The results listed indicate that the significance of the CEC-determining variables decrease as follows:

$$SOM_W >> Clay_W >> \mathrm{h} > \mathrm{f}$$

Excluding the *f* and *h* binary variables would not affect the best-fitted results in a major way. Altogether, this leads to a best-fitted result with R^2 = 88.9%.

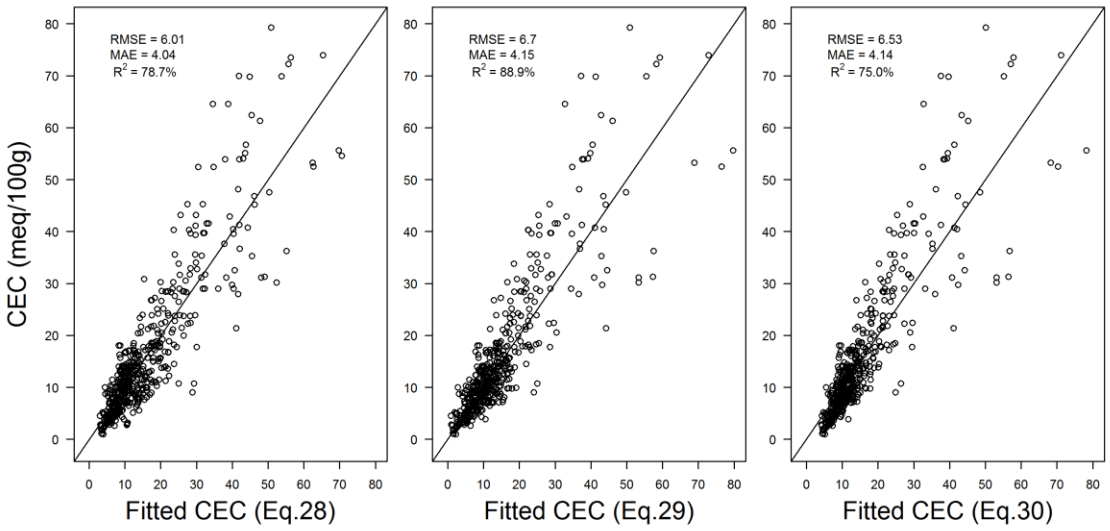

Figure 3.12. Comparison of PTF-derived CEC performance (actual vs, fitted) for all variables (**left**), just clay and SOM (no intercept) (**center**), and clay and SOM (with intercept) (**right**). Fitted values were compared to measured values to determine performance.

The modeled results for CEC agree with literature in that CEC is dominantly influenced by the amount of clay and SOM in a sample. This is because clay and SOM represent the colloidal fraction of the soil, with highest concentrations of negative electrostatic charges. These negative charges are what attract and hold base cations in the soil (Birkeland, 1999). Alternative equations were applied to the database to determine the feasibility of existing PTFs for determining CEC. These included:

1. Brady (1990):

$$CEC = 0.5(Clay\%) + 3.44(OC\%) \quad (31)$$

2. Liao *et al.* (2015):

$$CEC = -7.9 - 0.01(Sand\%) + 0.01(Silt\%) + 0.52(Clay\%) + 2.75(SOM\%) + 1.44(pH(CaCl_2) \quad (32)$$

3. Syers *et al.* (1970):

$$CEC = 1.68 + 3.33(OC\%) + 0.06(Clay\%) \quad (33)$$

4. Olorunfemi *et al.* (2016):

$$CEC = -13.93 + 2.645\big(pH(H_2O)\big) + 0.0446(Clay\%) + 2.267(OC\%) \quad (34)$$

5. Krogh *et al.* (2000):

$$CEC = 0.95 + 0.53(Clay\%) + 4.99(OC\%) \quad (35)$$

Plotting the actual CEC database entries versus the CEC model equations generated the scatter plots in Fig. 3.13. The goodness-of-fit sequence of these predictions is as follows:

Eq. 29 > Eq. 28 > Eq. 30 > Syers *et al.* (1970)> Krogh *et al.* (2000) > Liao *et al.* (2015) > Brady (1990) > Olorunfemi *et al.* (2016)

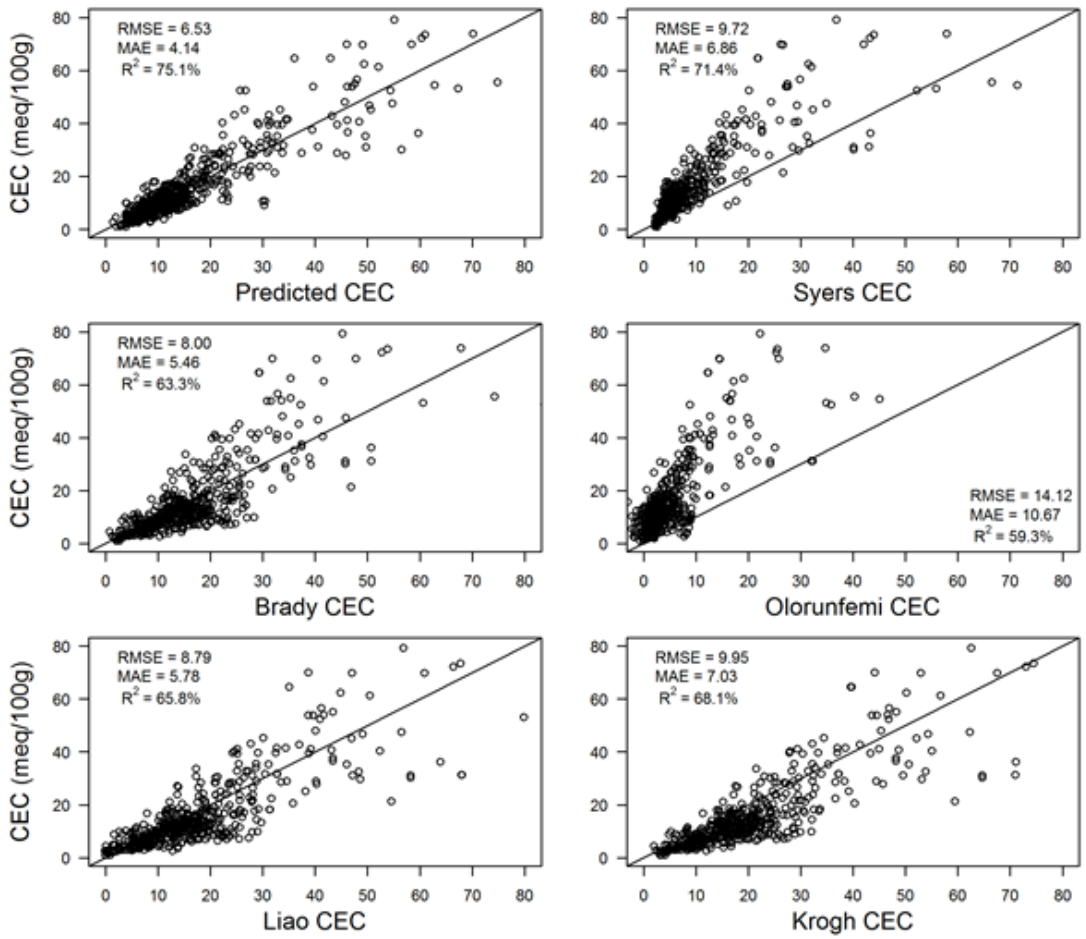

Figure 3.13. Comparison of actual vs. predicted CEC predictions (Eq. 30) (**top left**) with CEC predictions from Syers et al. (1970) (**top right**), Brady (1990) (**middle left**), Olorunfemi *et al*. (2016) (**middle right**), Liao *et al*. (2015) (**bottom left**), and Krogh *et al*. (2000) (**bottom right**) when compared to measured CEC values within database.

In this sequence, Brady (1990), Liao *et al.* (2015), and Krogh *et al.* (2000) predict higher than actual CEC values, while the opposite occurs with using the Syers *et al.* (1970) and Olorunfemi *et al.* (2016) models. These differences are, respectively, mainly due to high versus low regression coefficients for SOM, possibly related to temperate versus semi-tropical to tropical soil conditions. In addition, the clay coefficients also vary by a factor of about 10 as follows:

Krogh *et al.* (2000) ≈ Liao *et al.* (2015) ≈ Brady (1990) >> Syers *et al.* (1970) ≈ Olorunfemi *et al.* (2016)

These differences could be due to clay mineralogy, with CEC being low in soils or soil parent materials that developed under tropical conditions (e.g., Nigerian soils) or have very low clay content as in the sand chronosequence study of Syers *et al.* (1970). In addition, the negative intercepts and/or regression coefficients of the Liao *et al.* (2015) and Olorunfemi *et al.* (2016) models also lower overall CEC predictions.

Differences in CEC across grasslands, forested lands, and for the organic portion of forest soils (litter fermentation and humification layers combined) have been noted by Meyer and Arp (1994) as follows:

For the forest floor:

$$CEC\ (brodleaves) = -38 + 1.21(SOM\%) + 10.3pH, R^2 = 0.69 \quad (36)$$

$$CEC\ (conifers) = -31 + 0.58(SOM\%) + 12.1pH, R^2 = 0.58 \quad (37)$$

For mineral soil layers:

$$CEC\ (forest\ soils) = -7 + 0.29(Clay\%) + 1.4(SOM\%) + 1.4(pH), R^2 = 0.72 \quad (38)$$

$$CEC\ (wooded\ grasslands) = -6 + 0.31(Clay\%) + 2.24(SOM\%) + 1.0(pH), R^2 = 0.74 \quad (39)$$

$$CEC\ (grasslands) = -8.3 + 0.24(Clay\%) + 3.66\ (SOM\%) + 1.3(pH), R^2 = 0.79 \quad (40)$$

These equations imply that CEC increases with increasing soil organic matter decomposition and humification, i.e., the SOM% contributions are generally lower in the forest floor of conifers than of broadleaves, and, in mineral soils, increase from forests to grasslands. There is also a CEC contribution with increasing pH, and this would mainly be due to the increasing dissociation of the organically bound carboxylic acid into carboxylates with increasing pH, e.g.:

$$RCOOH + HCO_3^- \rightarrow RCOO^- + H_2O + CO_2 \quad (41)$$

Hence, data-derived models for predicting CEC cannot be generalized globally, since the mineral and organic contributions to CEC depend on local to regional differences in climate, vegetation type, and clay mineralogy. In this regard, note that the CEC models by Meyer et al. (1994) apply across Eastern Canada, and are therefore like Eqs. 28 to 30 in quantitative expression, and especially so in terms of their overall clay and SOM contributions to CEC.

3.5. CONCLUSIONS

This paper has outlined the development of PTFs for sand, silt, clay, CF, pH, SOM, Db, FC, PWP, and CEC with necessary data acquired through the amalgamation and harmonization of soil surveys, as outlined in Chapter 2. It was ensured that PTFs were developed from original data and that the results of one PTF were not utilized in the development of another. The results also adhere to the first principle of PTF development,

predicted property must be more difficult to measure than the predictor properties (McBratney *et al.*, 2002). These findings align with theoretical thresholds associated with each measured soil property, ensuring that FC and SOM do not exceed 100, the combination of sand, silt, and clay does not exceed 100, Db remains < 2.5, and FC > PWP > 0. PTFs were developed and utilized to fill gaps within the database pertaining to amalgamated soil surveys. With this, Table 3.12 represents the ability of these PTFs to fill in the gaps of the aspatial database:

Table 3.12. Overview of representation of soil properties within database and gaps filled via developed PTFs as outlined in this study. Note that SOM is standardized from SOM% and OC%, FC is standardized from FC_V and FC_W, and PWP standardized from PWP_V and PWP_W.

Property	Sample Size	% of Total	Equation	Number of Gaps Filled	% Filled	% of Total Filled
Sand	1306	52.45	Eq. 1	1128	45.30	97.75
Silt	1306	52.45	Eq. 1	1168	46.91	99.36
Clay	1306	52.45	Eq. 1	1128	45.30	97.75
CF	885	35.54	Eq. 2	1557	62.53	98.07
pH (H_2O)	1535	61.65	Eq. 3	183	7.35	61.72
pH ($CaCl_2$)	647	25.98	Eq. 4	1012	40.64	66.62
SOM	1202	48.27	Eq. 5	79	3.17	51.44
			Eq. 6	1282	51.49	99.76
Db	938	37.67	Eq. 8	0	0.00	37.67
			Eq. 9	0	0.00	37.67
			Eq. 10	332	13.33	51.00
FC	846	33.98	Eq. 13	149	5.98	39.96
			Eq. 14	161	6.47	40.45
			Eq. 15	504	20.24	54.22
PWP	738	29.64	Eq. 21	583	23.41	53.05
			Eq. 22	583	23.41	53.05
CEC	659	26.47	Eq. 28	596	23.94	50.41
			Eq. 29	596	23.94	50.41
			Eq. 30	596	23.94	50.41

The derived PTFs for Db, FC, PWP, and CEC were compared to published PTFs from different geographical locations (Eastern Canada, USA, Brazil, Africa, and Australia) to test

performance and the applicability of PTF transferability. As outlined, PTFs from this article performed better than published PTFs, but the fundamental underlying relationships between soil properties are similar between models. With this, it is apparent that SOM plays a dominant role in influencing Db, FC, PWP, and CEC with texture as being the second dominant factor (sand for Db and clay for FC, PWP, and CEC). Likely reasons for varying performance with changing geographic location is the variable influences that each soil forming factor play on soil formation. Soils in an area with recent glacial activity are more-likely influenced by surficial geology than those unaffected by glacial activity during the past ice age. Additionally, these unaffected areas are likely more influenced by other soil forming factors, namely topography and climate.

Some PTFs developed within this article, namely those incorporating parent material and soil horizon descriptors, differ from other PTFs due to the incorporation of these factors. These factors, although increasing model accuracy, limits the applicability of the model because the predictors may not be measured for a soil, or may differ from those in which the model was derived. Therefore, alternative models only using measured soil properties were also included. For all models developed, only the original, measured soil properties were used in model development and performance testing. Building models on modelled results will compound errors and likely lead to overfitted relationships. All models developed are based on simple and multiple linear regression analyses. Incorporating machine learning algorithms in the modeling process will likely result in higher predicting capabilities but doing so limits the ability to ensure results stay

within theoretical thresholds. Also, using simple models allows one to see the relationships between soil properties.

Soil surveys have long been recognized as useful sources of soil information for PTF development. Through soil survey amalgamation and data harmonization, soil properties can be predicted from a few more easily measured properties to a certain degree of confidence. Focusing on data amalgamation and harmonization improved model performances by increasing sample sizes and removing outliers.

CHAPTER 4 - FUSING DIGITAL ELEVATION MODELS TO IMPROVE HYDROLOGICAL INTERPRETATIONS

Shane Furze, Jae Ogilvie, Paul Arp

Faculty of Forestry and Environmental Management

University of New Brunswick, Fredericton, NB, Canada, E3B 6C2

Foreword:

The following chapter is an article published in the Journal of Geographic Information System. It was submitted on July 28th, 2017 and published September 17th, 2017.

Citation:

Furze, S., Ogilvie, J. and Arp, P.A. 2017. Fusing Digital Elevation Models to Improve Hydrological Interpretations. Journal of Geographic Information System, 9, 558-575. https://doi.org/10.4236/jgis.2017.95035

4.1. ABSTRACT

Improving the accuracy of digital elevation models is essential for reducing topographic derivation errors pertaining to, e.g., flow direction, basin borders, channel networks, depressions, flood forecasting, and soil drainage. This article demonstrates how a gain in this accuracy is improved through digital elevation model (DEM) fusion and using LiDAR-derived elevation layers for conformance testing and validation. This demonstration is done for the Province of New Brunswick (NB, Canada) as a case study, using five province-wide DEM sources (SRTM 90m; SRTM 30m; ASTER 10m; CDED 30m; NB-DEM 10m) and a five-stage process that guides the re-projection of these DEMs while

minimizing their elevational differences relative to LiDAR-captured bare-earth DEMs, through calibration and validation. This effort decreased the resulting non-LiDAR to LiDAR elevation differences by a factor of two, reduced the minimum distance conformance between the non-LiDAR and LiDAR-derived flow channels to ± 10 m at 8.5 times out of 10, and dropped the non-LiDAR wet-area percentages of false positives from 59 to 49 %, and of false negatives from 14 to 7 %. While these reductions are modest, they are nevertheless not only consistent with already existing hydrographic data layers informing about stream and wetland locations, they also extend these data layers across the province by comprehensively locating previously unmapped flow channels and wet areas.

Key words: DEM fusion, LiDAR-based calibration, hydrographic interpretations, stream network, wet-areas mapping.

4.2. INTRODUCTION

There is a growing demand for high-resolution datasets representing the Earth's surface as technology and applications of spatial modeling advance. Among these datasets are digital elevation models (DEMs), gridded datasets representing continuous elevation change spatially across landscapes at both local and global extents (Moore *et al.*, 1991; Yue *et al.*, 2007). In this regard, DEM elevation accuracies and resolutions vary based on differences in mode of elevation capture, processing, point-to-point interpolations, and timing of capture (Skidmore, 1989; Carrara *et al.*, 1997; Bater and Coops, 2009; Erdoğan, 2010; Smith, 2010). Thus, as DEMs are used for spatial visualization and modeling of topographic, geomorphologic, and hydrological properties (e.g., slopes,

soil erosion, basin borders, stream and river networks, and stream discharge), their ability to represent such topographic features vary in considerable detail (Wolock and Price, 1994; Thompson *et al.*, 2001; Shary *et al.*, 2002; Hengl, 2006; Wu *et al.*, 2008; Hopkinson *et al.*, 2009; Wilson, 2012).

In terms of acquisition and availability, DEMs of varying origins are becoming more and more freely accessible, with those utilized in this study listed in Table 4.1 (for details, see Carrara *et al.*, 1997; Carlisle, 2005; Chaplot *et al.*, 2006; Erdoğan, 2010; El-Sammany *et al.*, 2011; Florinsky, 2012). Of these, some provide global coverage, e.g., the Shuttle Radar Topography Mission (SRTM), and the Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER). Others provide national coverage, e.g., the Canadian Digital Elevation Model (CDED), and finally, provincial coverage, e.g., the New Brunswick DEM (NB DEM). The SRTM and ASTER DEMs are satellite-derived whereas the CDED was derived from elevation contours and spot heights (Center for Topgraphic Information Customer Support Group, 2000; Farr *et al.*, 2007; ASTER GDEM Validation Team, 2011). The NB DEM was derived through stereo-photogrammetry (Pegler, 1999; Service New Brunswick, 2001). Also becoming increasingly available are Light Detection and Ranging (LiDAR) DEMs, which, at 1m resolution, provide a convenient reference DEM with a vertical accuracy of ±15cm (Cunningham *et al.*, 2006). The accuracies of these DEMs vary greatly, depending on specifications of each DEM (Carrara *et al.*, 1997; Chaplot *et al.*, 2006) as outlined in Table 4.1. NB was selected as a case study due to the number of available DEMs and the amount of accessible LiDAR coverage.

Table 4.1. Overview of five open-sourced DEMs and one privately-acquired DEM utilized in this study. Note the variation in coverage, resolution, and vertical error. Also note that DSM_x represents digital surface models whereas DSM represents digital soil models.

	DEM					
	SRTM[1]	**SRTM[1]**	**ASTER[1]**	**CDED[2]**	**NB DEM[3]**	**LiDAR[4]**
Coverage	Global	Global	Global	National	Provincial	Site-Specific
Resolution (m)	90	30	30	22	10	1
Vertical Error (m)	±16	±7	±17	±6	±3	±0.15
Reference	(Sun *et al.*, 2003)	(Santillan and Makinano-Santillan, 2016)	(ASTER GDEM Validation Team, 2011)	(Happi Mangoua and Goïta, 2008)	(Pegler, 1999; Service New Brunswick, 2001)	(Cunningham *et al.*, 2006)
Data Type	DSM_x	DSM_x	DEM	DEM	DEM	DEM
Format	Integer	Integer	Integer	Integer	Floating Point	Floating Point
Acquisition Year	2000	2000	2000	2000	1999	Ongoing
Technology	Radar	Radar	Stereo Pairs	Contours, Spot Heights	Stereo Pairs	Laser
Original Spacing	3 arc second	1 arc second	1 arc second	0.75 arc second	75m	1m

1. http://www.snb.ca/gdam-igec/e/2900e_1c.asp
2. https://earthexplorer.usgs.gov/
3. http://geogratis.gc.ca/site/eng/extraction/
4. http://www.snb.ca/geonb1/e/DC/catalogue-E.asp

Elevation differences between LiDAR and other DEMs (hereby referred to as non-LiDAR DEMs) relate to data acquisition methodology which influence accuracy and resolution (Sun *et al.*, 2003; Cunningham *et al.*, 2006; Happi Mangoua and Goïta, 2008; ASTER GDEM Validation Team, 2011; Santillan and Makinano-Santillan, 2016). Non-LiDAR DEMs such as SRTM30 and STRM90 are based on surface reflections, while ASTER, CDED, NB-DEM attempt to capture bare-earth elevations. In this regard, the former refer to Digital Surface Models (DSM_xs), while the latter refer to Digital Elevation Models (DEMs).

Both are subsets of Digital Terrain Models (DTMs) (Wilson, 2012). LIDAR point cloud data can be used to generated DSM_xs as well as DEMs through determining the elevation differences between the first and last laser pulse returns, generally at a vertical accuracy of ±0.15m (Cunningham *et al.*, 2006).

All DEMs can be used to generate topographic interpretations pertaining to, e.g., flow directions, flow accumulation, stream channel networks, upslope basin areas and borders, location of depressions, and extent of areas subject to flooding and well-to-poor soil drainage. These interpretations are, however, influenced by DEM accuracy (Tarboton, 1997; Murphy *et al.*, 2007, 2008, 2009, 2011; Hobi and Ginzler, 2012; Mukherjee *et al.*, 2013). For example, 10 to 20 m resolution DEMs are preferable to determine seamless flow channel connectivities across roads, but lead to underestimating slope steepness especially along shorelines, road cuts and deeply incised stream valleys. High-resolution LiDAR DEMs can serve both purposes, by using them at their finest resolution to determine where roads potentially block channel flow and therefore need to be breached, or where the need to be re-sampled towards coarser resolutions to approximately connect flow channels across barriers without breaching. In contrast, attempts to increase the resolution of photogrammetrically-derived non-LiDAR DEMs through re-interpolation alone does not generate more information, but introduces DEM artifacts such as “ridging” (Mukherjee *et al.*, 2011).

Several approaches are available to reduce vertical and lateral DEM errors through fusion. These involve DEM re-projecting, re-sampling, re-interpolation, amalgamation,

and hydrological enforcement (e.g., Costantini *et al.* (2005); Karkee *et al.* (2008); Papasaika *et al.* (2008); Schindler *et al.* (2011); Robinson *et al.* (2014)). For example, the extent of "ridging" can be reduced through DEM editing by way of TIN Random Densification (Pegler, 1999) and/or Kalman spatial filtering (Wang, 1998; Albani and Klinkenberg, 2003). Luedeling *et al.* (2007) applied a DEM fill technique to fill SRTM voids with ASTER data. Fusing DEMs for the purpose of DEM-error reduction was reviewed by Schindler *et al.* (2011). Apart from DEM fusion, hydrological enforcement (Soille *et al.*, 2003; Callow *et al.*, 2007; Al-Muqdadi and Merkel, 2011; Zhang *et al.*, 2013; Tran *et al.*, 2014) is needed to guide DEM-based flow-direction and flow-accumulation algorithms towards already delineated streams, rivers, lakes and shorelines. This enforcement is done by (i) lowering the elevation of all stream and open-water referenced pixels to their immediate lowest neighborhood elevations, and (ii) ensuring that all streams and rivers drop monotonously in elevation towards their nearest shorelines. This enforcement is generally applied to non-LiDAR DEMs, but topographical adjustments are also needed for high-resolution LiDAR-DEMs to remove artifacts due to low laser pulse returns from open water surfaces such as lakes, rivers and shores.

The objective of this article was to generate a non-LiDAR DEM at 10m resolution with resulting hydrological data sets via DEM fusion with which the generated DEM has reduced elevation errors and is relatively free of surface artifacts in comparison to original non-LiDAR DEMs. This was done comprehensively for the Province of NB as a case study by way of:

1. non-LIDAR DEM fusion;

2. using bare-earth 1m LiDAR elevation data as reference DEM;

The extent of the DEM-based hydrological interpretation improvements was analyzed in terms of:

1. nearest-distance conformance testing of non-LiDAR versus LiDAR-derived flow channels;
2. evaluating false positives and false negatives within DEM-derived cartographic depth-to-water indices (DTW; (Murphy *et al.*, 2008, 2009)) for each DEM.

4.3. METHODOLOGY

Located in Eastern Canada and spanning an area of 7,282,014 ha, the Province of NB encompasses an array of diverse geomorphologies and landforms and has province-wide non-LiDAR DEM and hydrographic network coverages available from Service New Brunswick's GeoNB data catalogue as well as LiDAR coverages for parts of the province (Fig. 4.1, Table 4.1). The SRTM and ASTER DEMs were acquired through the United States Geological Survey Branch's EarthExplorer application. The CDED data were obtained from Natural Resources Canada.

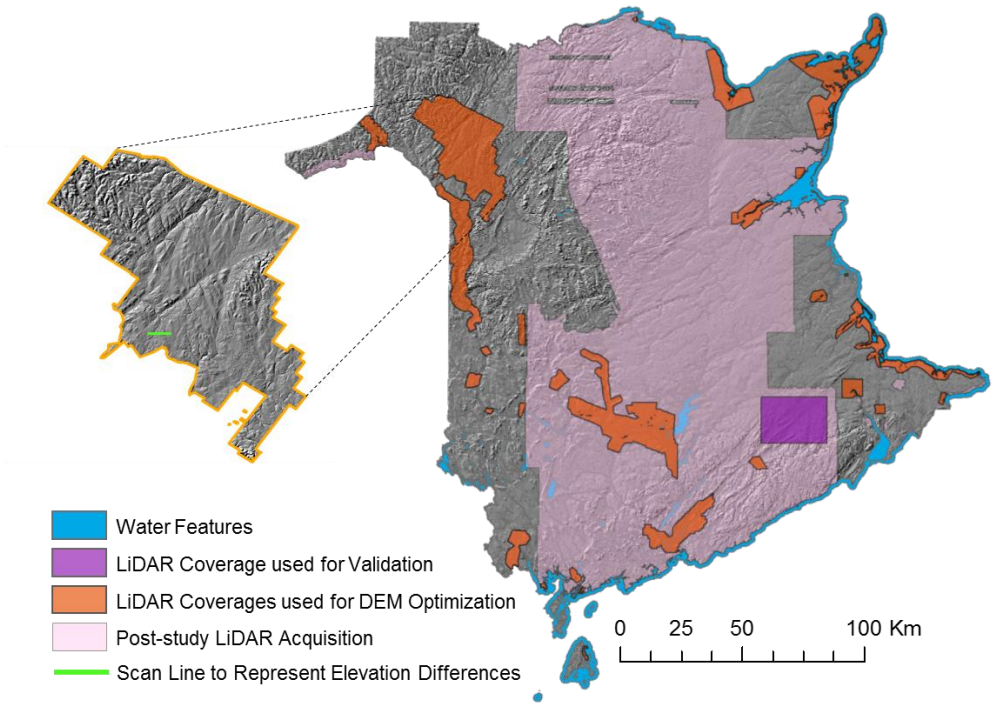

Figure 4.1. LiDAR DEM coverages (~11%) used for improving non-LiDAR DEMs across all of NB by way of a 5-stage process. Also shown: latest LiDAR-DEM acquisition for New Brunswick, LiDAR coverage used for optimization and process validation, and scan line (green) used to represent elevation differences associated with each non-LiDAR DEM. Background: hill-shaded NB-DEM. LiDAR-DEM source: http://www.snb.ca/geonb1/e/DC/DTM.asp.

A semi-automated five-stage process (Fig. 4.2) was developed (utilizing ArcGIS 10.1) as follows and tested in areas with available LiDAR coverage:

Stage 1. Homogeneity between the non-LiDAR DEMs in terms of the spatial reference and resolution was ensured through re-projection. All downloaded open-sourced DEM tiles were combined into seamless coverages and re-projected (CSRS NAD 1983 New Brunswick Stereographic) in three ways (bilinear, cubic, nearest) for three cell sizes (10m, default, 90m). These resolutions were chosen to align with the variability in resolution of available DEMs. Stage 1 was done to determine which projection matches the NB-DEM and LiDAR data the best.

Stage 2. The Stage 1 data layers, resampled at 8, 10, 12, 14, 16, 18, 20 m cell size, were randomly point sampled to generate a point shapefile containing n = 86,915 for each layer.

Stage 3. The points of the Stage 2 shapefile were re-interpolated using three interpolation methods (Inverse Distance Weighting, Kriging, Nearest Neighbor) and three cell sizes (5, 10, 20m).

Stage 4. The best-fitted non-LiDAR to LiDAR match was generated through

1. determining which of the Stage 3 generated data layers had the least non-LiDAR to LiDAR elevation differences, by DEM source (done through basic statistics and box plot comparisons); for this purpose, the bare-earth LiDAR-DEM, with the last-return points originally interpolated to 1 m, was re-sampled at 10 m resolution;

2. determining which combination and weighting of the best Stage-3 non-LiDAR DEMs should be used for the NB-DEM fusion process;

3. obtaining the best-fitted non-LiDAR combination through linear regression analysis, with the least-elevation generating DEM-layers including the fused NB-DEM layer as independent LiDAR-DEM predictor variables. This analysis was done based on a systematic selection of evenly-spaced points across the non-LiDAR DEMs within the LiDAR coverages, done at 1 point per 1 km^2, yielding 7,255 points for each layer. The output of the regression analysis included the best-fitted coefficients as well as their standard error of estimate, Student's t-value (best-

fitted regression coefficient estimate / stand error of estimate), and p-value (the probability that the best-fitted regression coefficient is significantly different from zero).

Stage 5. The best-fitted regression result (hereby referred to as "NBDEM-Optimized") was applied province-wide. With this, flow channel and associated DTW datasets were derived across the province. The results so produced were validated using a 40 by 40 km tile within a newly acquired LiDAR coverage (Fig. 4.1).

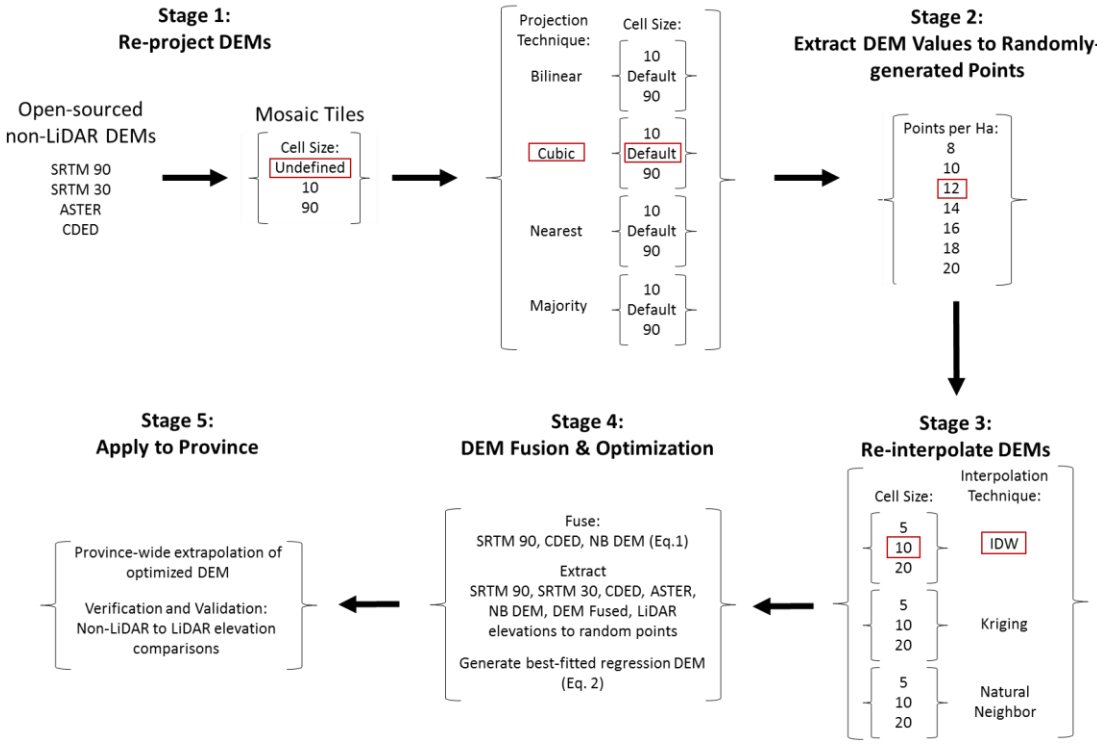

Figure 4.2. A 5-stage workflow developed for unifying both spatial reference and resolution and reducing non-LiDAR to LiDAR elevation differences in bare-earth elevation. The preferred technique for each stage is high-lighted in red.

The progressive non-LiDAR to LiDAR elevation difference reduction was achieved by selecting the best-fitting permutations from stages 1 and 2 based on layer-to-layer differences in terms of averages, standard deviations, minimum and maximum values, root mean square error (RMSE), and elevation difference percentages falling within the ± 2m and ± 4m LiDAR elevation ranges. Also obtained were percentile box plots for the elevation differences, by layer and by stage. The sequence of best choices so generated is high-lighted in Fig. 4.2. In detail, using default cell sizes and the cubic resampling technique for re-projection generally led to better results than using the bilinear, nearest, and majority re-projection methods (Stage 1). Similarly, randomly extracting re-projected DEM elevation points at 12 points per hectare (Stage 2), followed by IDW re-interpolation (Stage 3) at 10 instead of 5 or 20m resolution generally led to better re-interpolation results than using the kriging and natural neighbor methods.

Each of the non-LiDAR DEMs resulting from the 5-stage process and the LiDAR DEMs were used to derive continuous flow channel networks and associated cartographic depth-to-water (DTW) indices across New Brunswick. These processes involved using:

1. the D8 algorithm for deriving flow direction and flow accumulation datasets (Tarboton, 1997), with stream channels defined to have a minimum upslope flow accumulation area of 4 ha;

2. a least-cost algorithm to determine the extent of the least upward elevation change away from each flow channel and shore line (Murphy *et al.*, 2009).

The resulting DTW pattern generally reflects soil drainage patterns, with DTW ≤ 10cm very poorly drained, 10 < DTW ≤ 25cm poorly drained, 25 < DTW ≤ 50cm imperfectly drained, 50 < DTW ≤ 100cm moderately well drained, and DTW > 100cm well to excessively well drained. The distances between the nearest non-LiDAR to LiDAR-derived flow channels for each DEM were conformance tested by plotting their cumulative frequency distributions. The differences in each non-LiDAR- versus LiDAR-derived DTW (< 1m) coverages were evaluated by determining the area extent of false negatives and positives (predicted high DTW on dry site and predicted low DTW on wet site, respectively).

4.4. RESULTS

Elevation differences, by DEM Layer. The layer-by-layer non-LiDAR to LIDAR elevation differences are overlaid on the layer-corresponding hill-shaded DEMs in Fig. 4.3. Among these, the contour-derived CDED differences are smoothest, the NB-DEM differences are ridged, and the ASTER differences are the most variable. Across the LiDAR-DEM coverages, the differences tend to be largest on forested ridges and valleys (where steep elevation changes occur), and smallest on open areas. Negative SRTM and ASTER elevation differences relative to the LiDAR-DEM occur along steep slopes (valleys, shores, forest edges), undoubtedly due to the differences in resolution (1m versus 30 and 90 m). Owing to the general terrain and changing forest practices across New Brunswick, there are almost as many negative SRTM and ASTER to LiDAR-DEM differences as there are positive differences. As a result, their mean differences only have a small positive bias at

1.62 and 1.47m for ASTER and SRTM (90m) respectively. The extent of the elevation differences within and across the non-LiDAR generated data layers from the 5-stage process are summarized in Table 4.2. In comparison, the standard deviation and root mean square errors are largest for the ASTER DEM, with only 23.2% of its elevation range remaining within the ± 2m LiDAR elevation range, and least for the best-fitted NB-DEM Optimized regression result. For the latter, the remaining elevation differences with respect to the LiDAR DEM are within ± 2m at 70.1%, and within ± 4m at 92.9%.

Stage 1 to 4 non-LiDAR to LiDAR elevation difference reductions. The entries in Table 4.2 and the box plots in Fig. 4.4 concerning the non-LiDAR to LiDAR-DEM elevation differences display a narrowing towards the 0m difference. The fused and regression optimized DEMs for NB were obtained as follows:

$$NBDEM\text{-}Fused = 0.3\,NB\,DEM + 0.375\,SRTM90 + 0.325\,CDED \quad (1)$$

and

$$NBDEM - Optimized = a + b\,NBDEM\,Fused + c\,SRTM\,90 + d\,STRM\,30 + e\,CDED \quad (2)$$

(*a, b, c, d, and e* refer to intercept and regression coefficients). The weights for the NBDEM-Fused were chosen by systematically comparing different weight combinations to underlying LiDAR-derived elevation values. The best-fitted regression results for Eq. 2 are compiled in Table 4.3**.** In combination, the influence of stepwise non-LiDAR layer inclusion into the regression analysis generated the following sequence:

$$NBDEM\text{-}Fused >> SRTM30 > SRTM90 > CDED > > ASTER.$$

Hence, the SRTM, CDED and DEM Fused accounts for most of the non-LiDAR to LiDAR elevation difference reductions. Entering the selected CDED and SRTM elevations individually into the stepwise regression process is of further benefit, but only marginally so, but not so when entering the corresponding ASTER points. The main benefit of Eq. 1 stems from partially cancelling the elevation differences among NB-DEM, CDED, and SRTM while omitting qualitative errors associated with each, especially "ridging" embedded in the NB-DEM. In detail, the NBDEM-Fused and NBDEM-Optimization layers matched the LiDAR-DEM with the least bias, least minimum and maximum differences, and around 70% (i.e., 90th percentiles) of the elevation differences falling within the ±2 m LiDAR elevation range (Fig. 4.3).

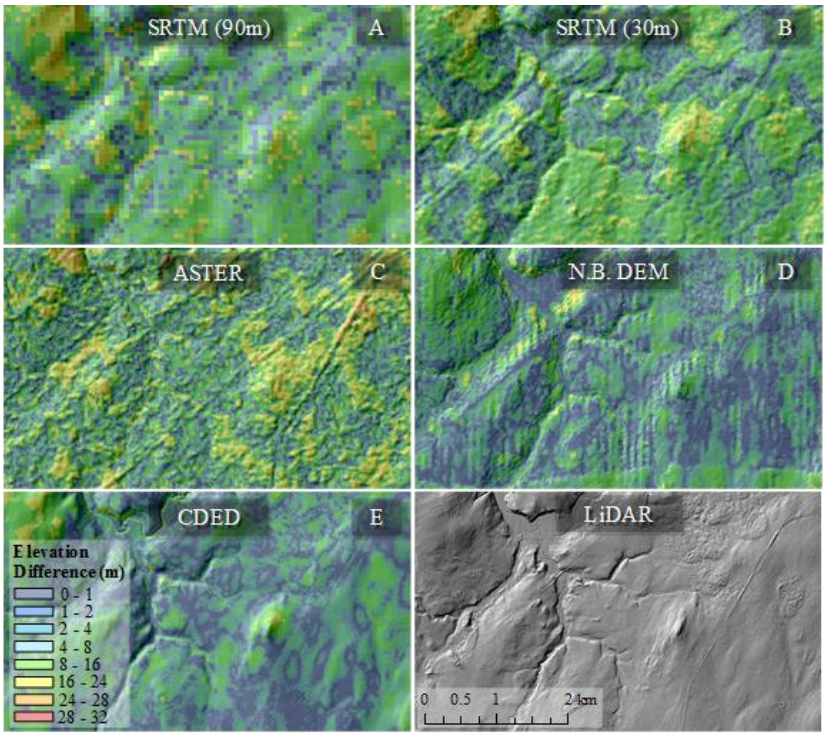

Figure 4.3. Example for non-LiDAR to LiDAR elevation differences (color-coded) draped over DEM-derived hill-shade for each non-LiDAR DEM, including hill-shaded bare-earth LiDAR DEM (**bottom right**).

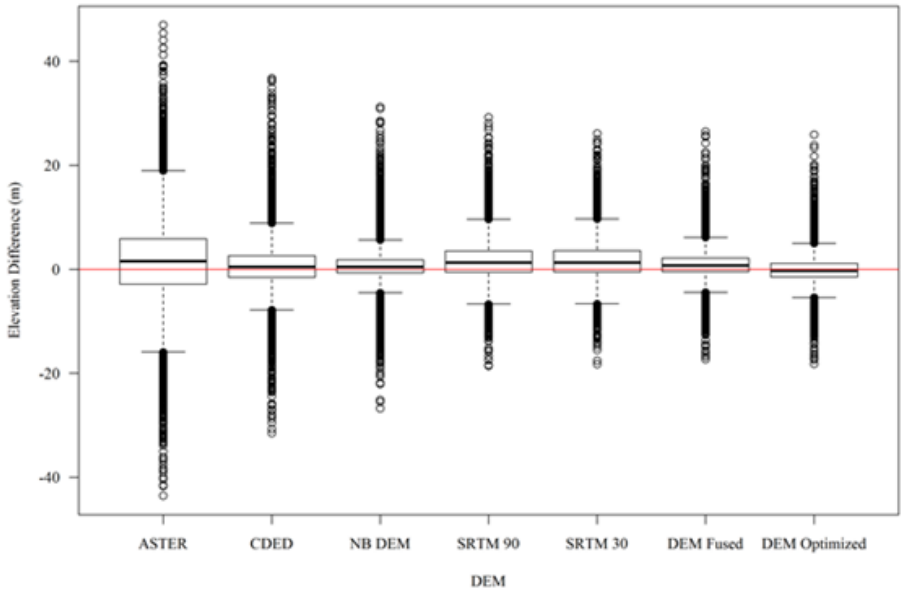

Figure 4.4. Box plots for the non-LiDAR to LiDAR elevation differences for SRTM90, SRTM30, CDED, NB-DEM, NBDEM-Fused, and NBDEM-Optimized across the DEM optimization extent, showing the 10th, 25th, 50th, 75th and 90th percentiles, the points above and below the 10th and 90th percentiles, and a reference line at 0m.

Table 4.2. Non-LiDAR and LiDAR DEM elevation difference comparison across the DEM optimization extent: statistical summary, in m, including increase in ≤ ±2 to ±4m elevation difference percentages.

DEM	Mean Difference (m)	Minimum Difference (m)	Maximum Difference (m)	Standard Deviation (m)	± 2m (m)	± 4m (m)	RMSE
ASTER	1.47	-47.51	46.78	6.78	23.21	44.53	6.93
SRTM (90)	1.62	-33.81	45.44	3.14	49.94	77.69	3.53
SRTM (30)	1.64	-16.91	28.66	3.04	49.89	77.84	3.46
CDED	0.58	-33.11	38.79	3.77	48.13	76.75	3.81
NB DEM	0.66	-24.06	30.70	2.76	66.88	87.83	2.84
DEM Fused	1.00	-18.10	27.13	2.42	68.31	88.89	2.62
DEM Optimized	-0.13	-18.09	21.81	2.31	70.11	92.85	2.32

Table 4.3. Stepwise regression results with LiDAR- DEM as dependent variable and non-LiDAR DEMs as predictor variables SRTM90, SRTM30m, CDED, and NBDEM-Fused (in m; R^2 values near 1), across DEM optimization extent.

Regression Models	Predictor Variables	Regression Coefficients	Std. Error	t-value	p-value
Model 1 RMSE = 2.26	Intercept	-0.308	0.120	-2.6	<0.0001
	CDED	-0.04	0.003	-13.3	<0.0001
	SRTM 30	0.702	0.012	58.5	<0.0001
	SRTM 90	-0.598	0.013	-46.0	<0.0001
	NB DEM-Fused	0.930	0.006	155.0	<0.0001
Model 2 RMSE = 2.27	Intercept	-0.228	0.011	-20.7	<0.0001
	SRTM 30	0.537	0.007	76.7	<0.0001
	SRTM 90	-0.500	0.008	-62.5	<0.0001
	NB DEM-Fused	-.957	0.004	239.3	<0.0001
Model 3 RMSE = 2.32	Intercept	-0.206	0.012	-17.2	<0.0001
	NB DEM-Fused	0.994	0.001	994.0	<0.0001

Table 4.4. Non-LiDAR and LiDAR DEM elevation difference comparison: statistical summary for validation site, in m, including percentage of elevation differences ≤ ±2m and ≤ ±4m.

DEM	Mean Difference (m)	Minimum Difference (m)	Maximum Difference (m)	Min-Max Range	Standard Deviation (m)	RMSE	± 2m (%)	± 4m (%)	RMSE
ASTER	4.88	-83.98	66.99	151.0	7.12	8.63	20.30	38.85	8.63
SRTM (90m)	2.45	-36.60	57.17	93.8	4.52	5.13	36.13	63.48	5.13
SRTM (30m)	2.42	-16.32	27.20	43.5	2.93	3.80	43.93	71.14	3.80
CDED	1.06	-29.43	27.69	57.1	5.65	5.74	30.65	55.41	5.74
NB DEM	1.03	-25.08	35.90	61.0	2.91	2.92	62.95	94.54	2.92
DEM Fused	1.31	-20.22	35.96	56.2	3.12	3.39	55.49	80.48	3.39
DEM Optimized	0.58	-18.07	29.64	47.4	2.84	2.90	60.49	95.43	2.90

Hydrological interpretations (flow channels, wet-areas), by DEM layer. Examples of the DEM–produced flow channel and DTW patterns are presented in Fig. 4.5, by data layer. As shown, the patterns obtained with the SRTM 90 and SRTM 30 DEMs are narrow, are widely pixelated and somewhat terraced across the ASTER DEM, and are ridged across NB-DEM. In contrast, the CDED, NBDEM-Fused, and NBDEM-Optimized and LiDAR DEM layers follow fairly consistent patterns.

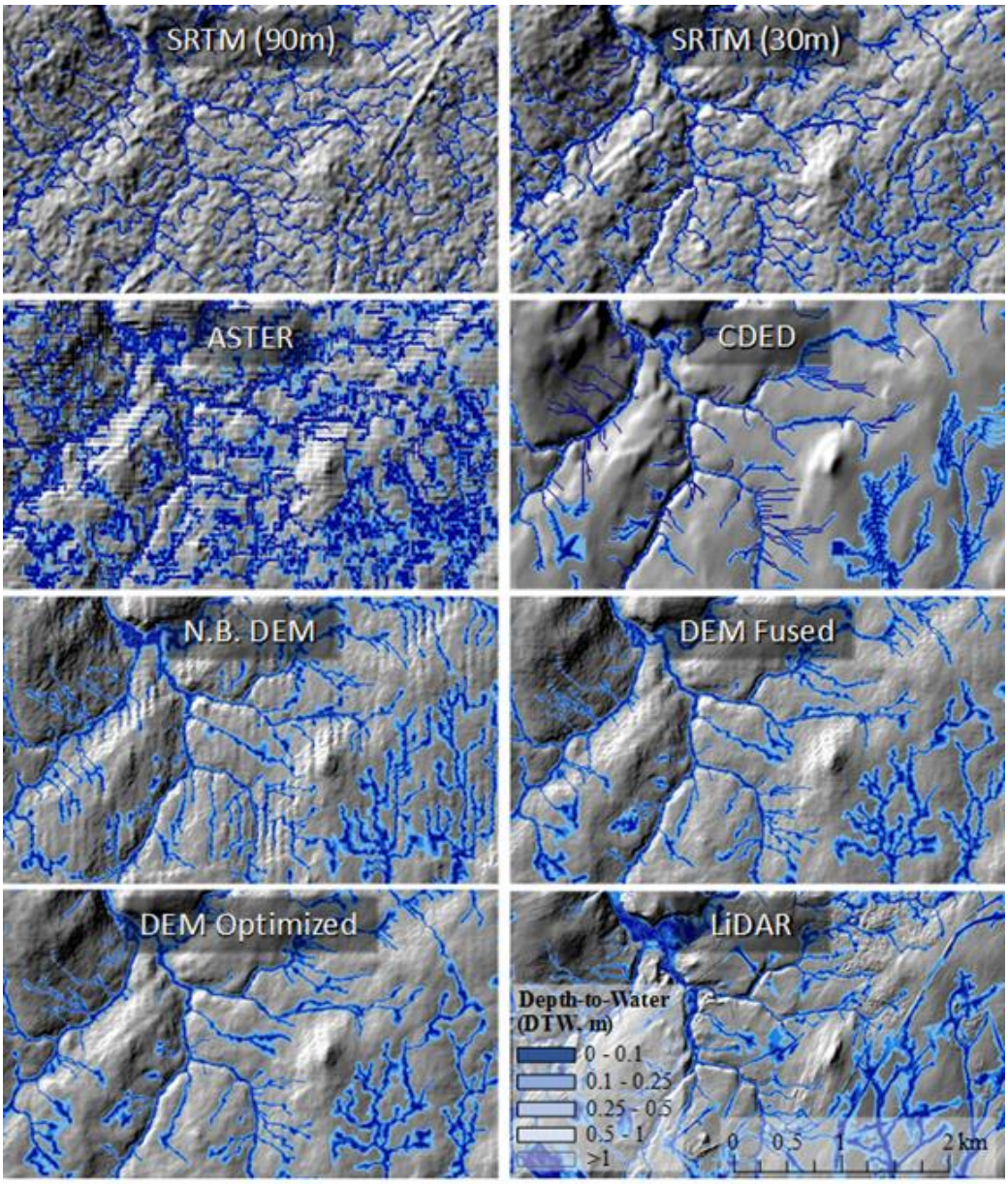

Figure 4.5. Comparison of flow channel and cartographic depth-to-water patterns overlaid on the hill-shaded non-LiDAR and LiDAR DEMs.

The 3,000m scan line in Fig. 4.6 (see Fig. 4.1 for location) provides an example of how the non-LiDAR to LiDAR elevation differences change across a forest management area in relation to a recent surface image and the corresponding hill-shaded LiDAR DEM. Also shown are: (i) the 0 to 1m cartographic depth-to-water index (DTW) pattern, shaded dark to light blue from, 0 to 1m, respectively, and (ii) the DTW and LiDAR-derived canopy

height profiles along the scan line. In detail, the LiDAR elevation profiles correspond closely with image- and hill-shaded DEM features, e.g., low tree height across recent cuts and wetlands, incised flow channel locations, elevated or incised roadbeds, and bare transmission corridors. The same, however, does not occur across the non-LiDAR elevation profiles. For these, the ASTER and CDED profiles have the widest excursions from the general elevation pattern. In contrast, the SRTM and NB-DEM profiles follow a similar pattern and reflect, in part, forest to non-forest tree-heights except for recent image-revealed forest cuts. The SRTM to LiDAR elevation differences in Fig. 4.6, however, only amount to about one-half to one third of the LiDAR-generated tree height maxima.

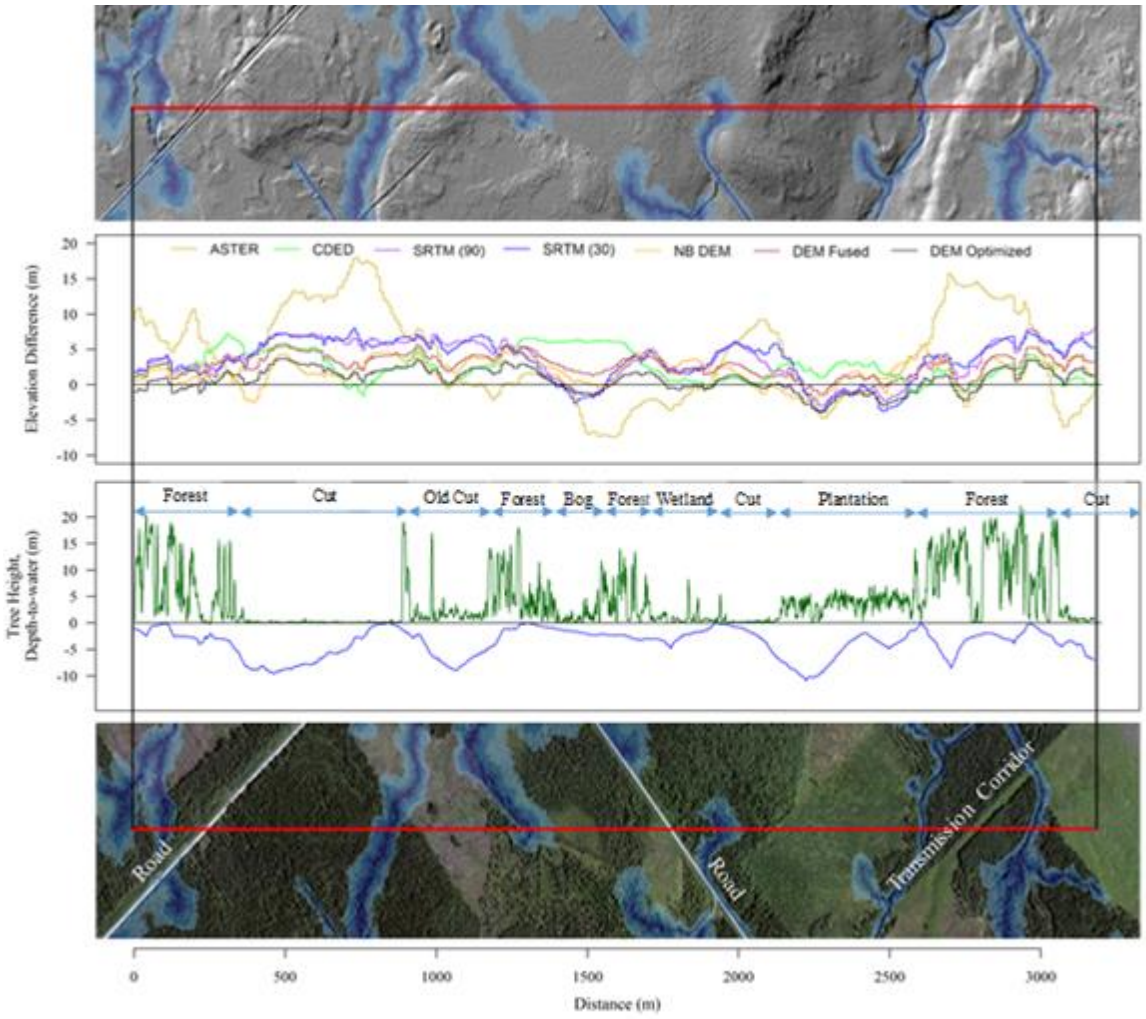

Figure 4.6. Non-LiDAR to LiDAR elevation differences along a 3,000m scan line (**Fig. 4.1**) related to the DTW pattern overlaid on recent surface image (**bottom**) and hill-shaded LiDAR DEM (**top**). Also shown, scan line for LiDAR-generated DTW and vegetation heights.

Validation results regarding flow-channel and wet-area mapping. The DEM validation results, completed for the tile outside Stage 3 and 4 processing areas (Fig. 4.1), are compiled in Tables 4.4 and 4.5. For this tile, all layers are upwardly biased, but least so for NBDEM-Optimized at 0.58m, with standard deviation and mean square root differences also being least at 2.84 and 2.90m, respectively. The ≤ ±2 and ±4m percentages of the elevation differences remained about the same overall across the DEM

optimization and validation extents but dropped somewhat from 70.1 to 60.5 % for ≤ ±2m and increased from 92.9 to 95.4% for ≤ ±4m (compare Table 4.2 with Table 4.4).

The conformance results for the nearest flow channel distances between the non-LiDAR and LiDAR-derived DEMs in Fig. 4.7 improved by layer as follows:

ASTER < NB-DEM < CDED < SRTM90 < SRTM30 < NBDEM-Fused < NBDEM-Optimized.

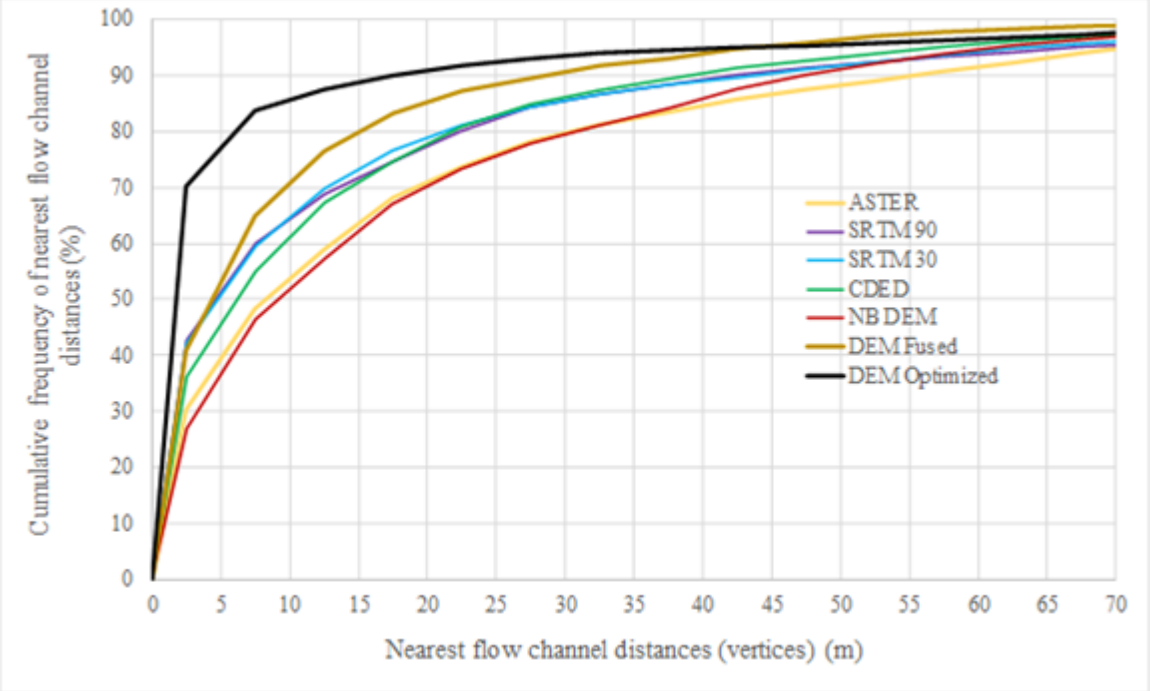

Figure 4.7. Cumulative frequency of nearest distances between the non-LiDAR- and LiDAR-derived flow channel networks, across validation extent.

There is a similar performance increase from ASTER to the NBDEM-Optimized layers in terms of false positive and false negative DTW < 1m reductions, with the optimized NB-DEM layer being closest to the corresponding 10 m re-sampled LiDAR DEM derivation (Table 4.5).

Table 4.5. Reduction in area (as percentage) of false positive and false negative DTW <1m relative to the 10m LiDAR-derived DTW pattern across validation extent.

DEM	Conformance (Wet)	False Positive	Conformance (Dry)	False Negative
ASTER	31.76	68.22	82.56	17.43
SRTM (90)	41.46	58.5	85.91	14.09
SRTM (30)	41.42	58.51	85.76	14.23
CDED	46.37	53.47	86.48	13.51
NB DEM	40.65	59.33	91.30	8.70
DEM Fused	49.55	50.42	92.29	7.71
DEM Optimized	50.50	49.50	92.69	7.31

In summary, the 5-stage process of combining the original NB DEM with the CDED and SRTM DEMs and subsequently calibrating the result with available LiDAR DEM coverages therefore not only improved but also extended the flow-channel delineation across New Brunswick in a systematic and comprehensive manner. This is illustrated in Fig. 4.8 by way of an example within validation tile.

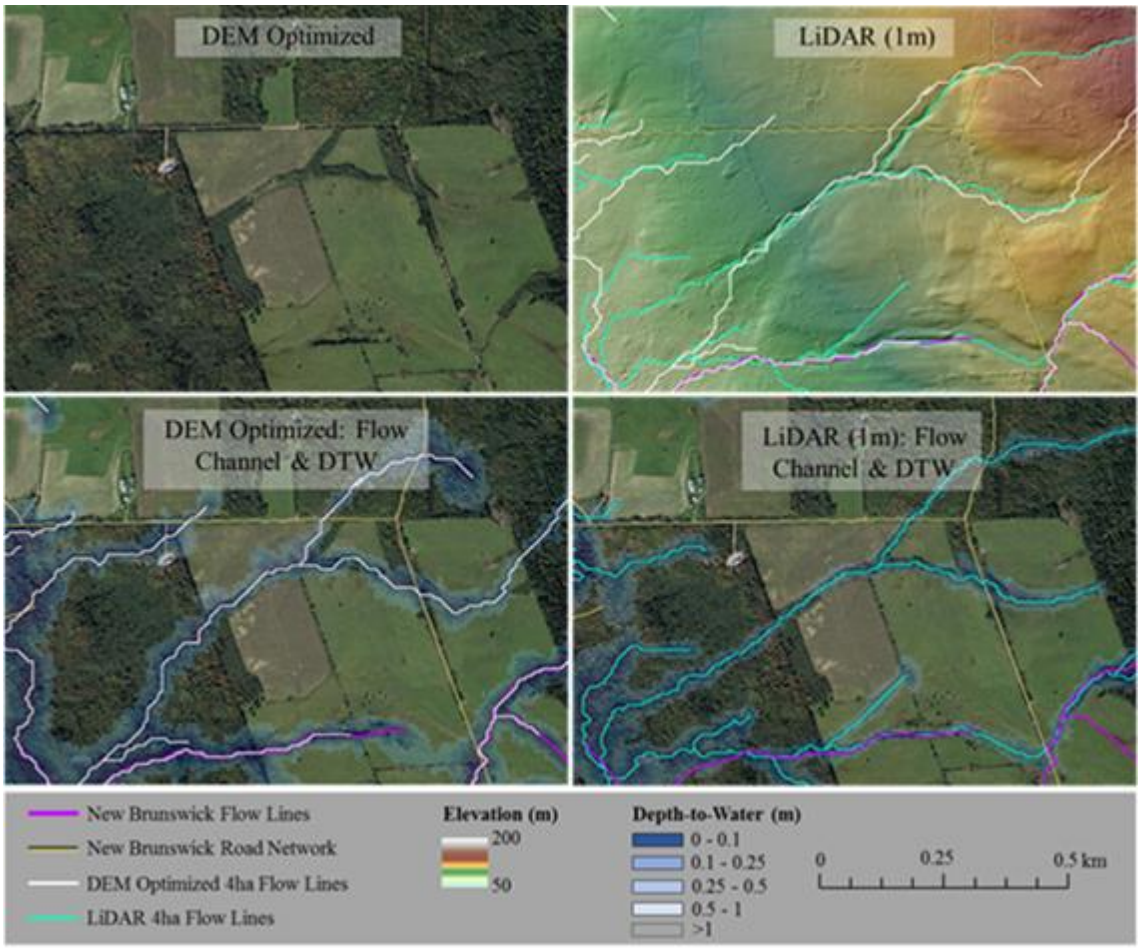

Figure 4.8. NBDEM-Optimized (**bottom left**) and LiDAR-DEM (**bottom right**) derived flow channel and DTW < 1m delineations versus the watercourses for the New Brunswick Hydrographic Network (purple). These delineations are overlain on aerial imagery and hill-shaded LiDAR-DEM (**top right only**). Correspondence between the DEM-derived flow channels and the riparian vegetation buffer is apparent. The New Brunswick Hydrographic Network is incomplete in this regard. Location: part of the validation tile in Fig. 4.1.

4.5. DISCUSSION

The above analysis demonstrates that the fusion of non-LiDAR DEMs can lead to systematic elevation difference reductions, which can be further enhanced through regression calibration with available LiDAR datasets. The DEM so optimized can then be used to generate province or region-wide hydrographic DEM interpretations that are similar to what can be derived from LiDAR-generated DEMs, at 10 m resolution at least.

The non-LiDAR-DEM error reductions by way of the 5-stage process appear to be small numerically at a 20% reduction of false positives and a 10% reduction of false negatives (Table 4.4). There is, however, a substantial 8 times out of 10 distance-to-flow-channel improvements from ±25 m (NB-DEM) to ±5m (NBDEM-Optimized) (Fig. 4.7). This, by itself, and in view of the illustrations in Fig.'s 4.5 - 4.7, has led to considerable improvements in comprehensively mapping flow channels and wet areas across NB. In turn, this has many potential positive effects on local and regional operations planning in, e.g., forest management (harvest block layout and access), transportation and trail routing, environmental impact assessments, land transactions, and emergency responses. In part, some of the province-wide benefits would accrue from improved DEM delineations and determinations of upslope basin areas. In turn, these areas are used to estimate, e.g., potential stream discharge rates and required culvert and bridge dimensions at any road-stream crossing by extreme weather events (e.g., 100mm or stream discharge per day, roughly equivalent to a 100-year storm event). In this regard, DEM error reductions are especially important across flat terrain, where minor elevations changes (artificial, natural, or man-made) can lead to substantial errors in determining storm water flow directions.

The optimized NB-DEM layer is also useful for checking province-wide wetland coverages. In general, already delineated wetlands fall into the DTW< 0.5m range (Murphy *et al.*, 2007). Extending the conformance checking between DEM-delineated and actual wetland border using the original NB-DEM layer amounted to ±80m, 8 times out of 10. Air-photo delineated wetland borders generally conform to distances within ±40m, 8

times out of 10. In contrast, LiDAR-DEM derived DTW = 0.5m contours conform to actual wetland borders within ±10m, 8 times out of 10 (details not shown). The optimized DEM, while somewhat less accurate than the LiDAR-DEM, can nevertheless be used to trace and survey-check wetland-to-wetland connections across the province.

It has been shown in literature that DEM fusion leads to DEM improvements in terms of vertical and lateral error reductions has been demonstrated repeatedly (e.g., Karkee *et al.* (2008), Papasaika *et al.* (2008), Al-Muqdadi and Merkel (2011), Zhang *et al.* (2013)), but the fusion processes pertaining to cell re-sizing, re-projection, re-sampling re-interpolation, DEM mix, weighting and noise filtering all vary. For example, Robinson *et al.* (2014) were able to produce "a nearly-global, void-free, multi-scale smoothed, 90m digital elevation model" called EarthEnv-DEM90 (http://geomorphometry.org/content/earthenv-dem90). Tran *et al.* (2014) fused ASTER with SRTM30 data through (i) DEM quality assessment and pre-processing, (ii) hydrologic DEM enforcement, (iii) void filling and projection shifting, (iv) DSM versus DTM bias elimination by landform, and (v) DEM de-noising. The 5-stage process in this study varies from Tran *et al.* (2014) by way of systematic cell size and DEM re-processing (re-projecting and re-interpolation), and using the regression process for bias removal. The reference DEMs also differ: using a DEM generated from the 1:10,000 topographic map 5m intervals and spot-height elevation data (i.e., similar to CDED) versus using 1m LiDAR DEM re-sampled at 10m.

In terms of applying the above approach to other areas with similar and/or different DEM including LiDAR DEM coverages, it is important to examine each re-projected, re-sampled and re-interpolated DEM with reference to artifacts and hydrographic correctness. In this regard, all open water surfaces need to be rendered flat, and streams and rivers need to drop monotonously through their surrounding terrain towards their receiving shores.

4.6. CONCLUSION

The systematic fusion of currently available DEM layers for all of NB not only led to considerable non-LiDAR-to LiDAR DEM elevation difference reductions, but also produced a closer and verifiable correspondence between the resulting flow-channels and wet-area derivations. In summary, the Stage 1 to 5 process as described in this article has shown that:

1. Non-LiDAR elevation differences relative to LiDAR DEMs can be reduced through careful analysis requiring re-projecting, re-sampling, and re-interpolation, followed by selective non-LiDAR DEM amalgamation.

2. The amalgamation process of the SRTM90, CDED and NB-DEM layers was effective in generating a much-improved non-LiDAR-DEM coverage across New Brunswick, resolving the issue of ‘ridging’ while slightly reducing vertical elevation error. Using the ASTER DEM as well did not improve the best-fitted fusion result so obtained.

3. The fusion process removed many layer-specific artifacts and large layer-to-layer elevation differences, while the non-LiDAR- to LiDAR-DEM regression process reduced the NBDEM-Optimized elevation bias to less than 1m.

4. While the results are specific to, and applied comprehensively across New Brunswick, similar non-LiDAR DEM improvements can be incurred elsewhere following the outlined procedure with open-sourced DEM coverages as long as precise elevation data is available, whether it be ground points of elevation data sets.

4.7. ACKNOWLEDGEMENTS

Financial support for this work was received from NSERC's AWARE Research Network, the Canadian Wood Fibre Research Centre, Natural Resources Canada, and from the Forest Watershed Research Center at UNB. We thank Bernie Connors (Service New Brunswick) for discussion and comments.

CHAPTER 5 – UTILIZING SOIL FORMING FUNDAMENTALS FOR HIGH-RESOLUTION SOIL PROPERTY MAPPING FOR NEW BRUNSWICK, CANADA

Shane Furze, Paul Arp

Faculty of Forestry and Environmental Management

University of New Brunswick, Fredericton, NB, Canada, E3B 6C2

5.1. ABSTRACT

This article presents an approach to modeling continua of soil property variation across New Brunswick, Canada by comparing field-collected soil profiles to underlying data sets representing soil forming factors, namely topography, climate, and surficial geology. Soil profiles were acquired and amalgamated from different sources into a spatial database (n=12,058) that was used for developing models for predicting the following soil properties: (i) drainage and horizons depths for A, B, and solum depth, and (ii) texture (sand, silt, and clay), coarse fragment (CF) content, soil organic matter (SOM) content, and bulk density (Db) at depth intervals of 0-15cm, 15-30cm, 30-60cm, 60-90cm, and on a solum basis. Procedures are outlined on how spatial data sets of topographic, climatic and geological features were updated and how variables were selected for statistical modeling using the machine learning algorithm Random Forest. Modeling results were evaluated on one third of the soil profile database by way of Kappa scores for soil drainage and RMSE and R^2 for remaining soil properties with validation performance ranging between 70-90%.

Key words: digital soil mapping, soil forming factors, multi scale, machine learning, digital elevation model, topography, surficial geology, climate.

5.2. INTRODUCTION

With traditional soil mapping techniques, high-resolution datasets of soil attributes either do not exist or are geographically limited to locally-defined extents. Instead, soils have been mapped as broad tesselated boundaries representing hierarchical, multi-component soil associations with each component representing a defined soil type. Each soil type is distinguished by soil drainage but not spatially represented within the association delineations (Pitty, 1979; Moore *et al.*, 1993; Zhu and Mackay, 2001; Heung *et al.*, 2014; Odgers *et al.*, 2014). This form of mapping represents a generalized simplification of soil property variation (Mora-Vallejo *et al.*, 2008) and contradicts continuous soil heterogeneity, both spatially across the landscape, and vertically with increasing depth (Pitty, 1979; Odgers *et al.*, 2011; Adhikari *et al.*, 2012). Thus, there is a growing interest toward digital soil mapping (DSM), modeling soil properties continuously via relationships with additional data sources representing soil forming factors (McBratney *et al.*, 2003; Smith *et al.*, 2006).

Numerous statistical models have been used for digital modeling of soil properties and drainage classes. The use of machine learning algorithms in DSM has been well-documented in literature with emphasis on regression trees, artificial neural networks, and support vector machines. Regression trees have been used in Henderson *et al.* (2005), Scull *et al.* (2005), Schmidt *et al.* (2008), and Behrens *et al.* (2010) for modeling soil

properties and in McKenzie and Ryan (1999) for predicting soil thickness. Regression trees have also been used for predicting soil types (Moran and Bui, 2002; Scull *et al.*, 2005; Schmidt *et al.*, 2008; Lemercier *et al.*, 2012). Alternatively, Odgers *et al.* (2014) applied classification trees for soil map disaggregation. Behrens *et al.* (2005), Zhao *et al.* (2010), and Zhao *et al.* (2013) have utilized artificial neural networks for predicting soil properties and soil drainage patterns. Additionally, Zhu (2000), Behrens *et al.* (2005), and Bodgahabadi *et al.* (2015) have used artificial neural networks for predicting soil classes.

This paper presents a framework for mapping individual soil properties spatially across the landscape at 10m resolution. This was completed by comparing a spatial database of amalgamated soil samples to underlying data sets representing soil forming factors, namely topography, surficial geology (mode of deposition, and lithology), and climate. Emphasis was placed on re-delineating soil forming factors prior to statistical modeling to mitigate the influence of error on predictions. Soil property mapping focused on soil texture (sand, silt, and clay), depth, CF content, SOM content, Db, horizon depths (A, B, and depth to C horizon (solum depth)), and soil drainage, all of which influence the medium for root growth and development.

5.3. METHODS

Study Extent: Spanning an area of 7.5 million ha, New Brunswick (NB) shares borders with the state of Maine (USA, west), Quebec (northwest) and Nova Scotia (southeast), the Nothumberland Straight (East), Chaleur Bay (Northeast), and the Bay of Fundy (South). As the northern range of the Appalachian Mountain range, NB ranges in

elevation from sea level along the coast to 811m above sea level in the north (Mount Carleton) (Happi Mangoua and Goïta, 2008) (Fig.5.1).

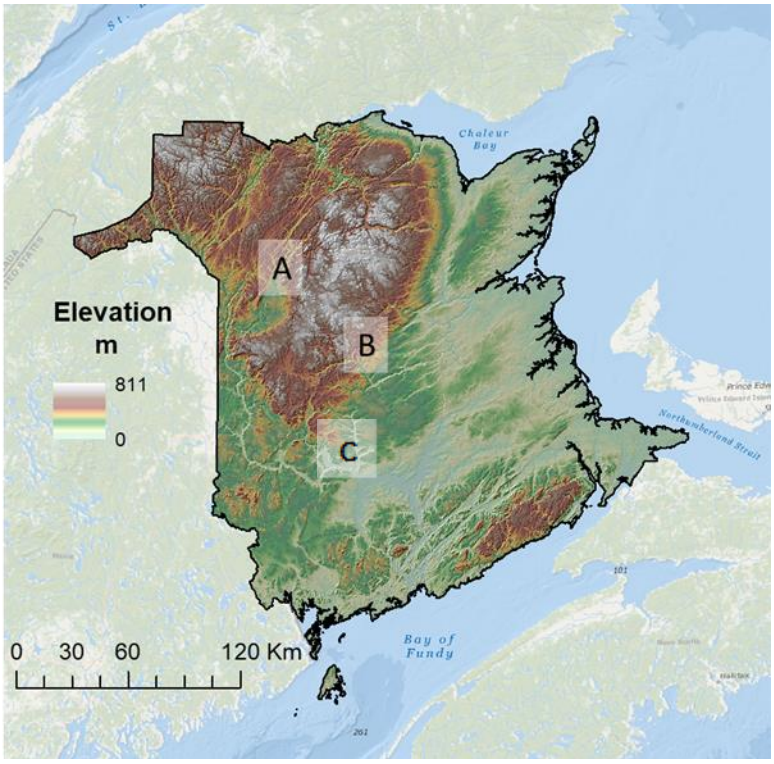

Figure 5.1. Outline of geographic location and elevation change of New Brunswick, Canada with surrounding provinces, state, and oceanic features. Background basemap retrieved from ESRI basemap portal within ArcGIS. "A" represents location used for representing DSM results and MRVBF comparison, "B" represents location used for LiDAR and curvature classification comparisons, and "C" represents location used for comparing surficial geology vector data sets.

This change in elevation results in varying climatic conditions from the coast to inland and higher elevations. Annual average rainfall ranges from 788mm in the northwest (higher elevations) to 895mm in the south and southeast. Annual average temperature follows a different trend with 0.4°C in the higher elevations (north and northwest) to 7.1°C in the south and southwest (Environment Canada, 2018).

Spatial Profile Database: Five datasets representing field-collected soil profiles for NB were acquired and amalgamated into a seamless spatial database (n = 12,058, Fig. 5.2). Table 5.1 outlines the datasets with sample size and source.

Table 5.1. Overview of acquired profiles (spatially-defined soil profiles) utilized in developing spatial database, including source and associated sample size.

Data Set	Source	Sample Size	% of Total
Permanent Sample Plots (PSPs)	NB Department of Natural Resources	2,594	21.5
Tri-Nations Sample Plots	Natural Resources Canada	123	1.0
National Pedon Database	Canadian Soil Information Service (CANSIS)	285	2.4
NB Soil Survey Plots	CANSIS	222	1.8
Till Geochemistry Site Cards	NB Department of Energy and Mines	8,834	73.3

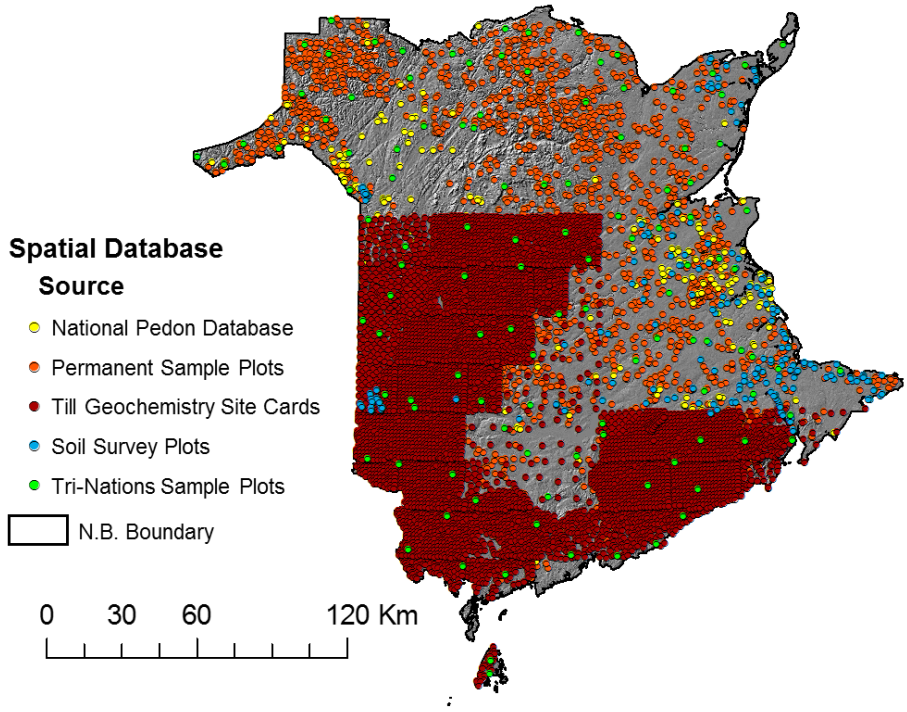

Figure 5.2. Spatial representation of profiles acquired for database with colours representing data sources, overlaying hill-shaded relief of NB.

Of these, only the till geochemistry site cards lack horizon-specific information for each sample, instead, properties for C horizon and general site attributes (drainage, landform,

etc.) were provided. Data harmonization followed the framework as outlined in Chapters 2 and 3. From this, it was ensured that correct classifications for texture and drainage were assigned to all samples and horizons. Finally, it was ensured that both spatial database and all spatial data sets aligned in terms of extent, resolution, and spatial reference (10m and NAD 1983 New Brunswick Stereographic).

Soil drainage, mineral horizon depths, texture, CF content, SOM content, and Db were chosen for analyses. Texture (sand, silt, and clay), CF, SOM, and Db were summarized by solum depth (A and B horizons) and depth intervals of 0-15, 15-30, 30-60, and 60-90cm (Table 5.2). Samples exceeding 90cm were not analyzed due to the infrequent occurrence of soil profiles reaching these depths.

Table 5.2. Summary of soil properties to be modeled via DSM including minimum, maximum, mean, and standard deviation, separated by depth intervals in which they are being modeled. Values for sand, silt, and clay retrieved from texture classes.

Property	Depth	Min	Max	Mean	Std. Dev.	# of profiles
Coarse Fragment Content (%)	0-15	0.00	100.00	30.95	21.84	2660
	15-30	0.00	100.00	36.12	23.29	2660
	30-60	0.00	100.00	42.32	25.54	2512
	60-90	0.00	100.00	35.17	28.74	711
	Solum	0.00	100.00	33.09	21.61	2660
Sand (%)	0-15	5.00	93.00	53.39	17.86	9064
	15-30	5.00	93.00	53.74	17.34	9093
	30-60	5.00	93.00	52.59	21.40	2580
	60-90	5.00	93.00	55.69	14.03	544
	Solum	5.00	93.00	53.46	17.17	9113
Silt (%)	0-15	4.00	90.00	30.76	13.73	9064
	15-30	4.00	90.00	30.09	12.98	9093
	30-60	4.00	90.00	28.80	16.02	2580
	60-90	4.00	65.00	26.39	7.21	544
	Solum	4.00	90.00	30.61	13.01	9139
Clay (%)	0-15	3.00	75.00	15.84	8.15	9064
	15-30	3.00	50.00	16.17	8.38	9093
	30-60	3.00	50.00	18.61	12.02	2580
	60-90	3.00	50.00	17.92	5.06	544
	Solum	3.00	75.00	15.93	8.19	9139
Organic Matter Content (%)	0-15	0.02	91.13	8.48	2.49	218
	15-30	0.00	93.57	4.13	1.44	335
	30-60	0.00	86.86	1.82	1.25	319
	60-90	0.00	81.18	1.85	1.00	131
	Solum	0.00	25.40	4.55	1.27	411
Bulk Density (g/cm^3)	0-15	0.10	2.25	1.05	0.17	227
	15-30	0.10	2.10	1.23	0.22	279
	30-60	0.10	2.10	1.50	0.26	280
	60-90	0.10	2.30	1.72	0.20	124
	Solum	0.10	2.05	1.17	0.23	360
A Horizon Depth (cm)	-	1.00	45.00	9.67	6.38	2889
B Horizon Depth (cm)	-	2.00	59.00	15.48	7.41	2899
Depth to C Horizon (cm)	-	4.00	250.00	42.06	14.96	6683
Drainage	-	-	-	-	-	9223

5.3a Spatial Data Quality Assessment

Spatial data sets representing soil forming factors of surficial geology, topography, and climate were updated as follows, yielding gridded data sets at 10m resolution:

Surficial Geology: Surficial geology data for NB were mapped directly via photo-interpretation with road cuts and bore holes, and indirectly via soil association maps, with the underlying similarity of tessellated delineations. Five vector data sets representing surficial geology were acquired and compared via a similarity model for re-delineating boundaries, including:

1. Soil Associations of NB (NB Department of Energy and Resource Development, 2012),
2. Forest Soil Associations of New Brunswick (NB Department of Energy and Resource Development, 2016),
3. Soil landscapes of Canada (Version 3.2) (Soil Landscapes of Canada Working Group, 2010),
4. NB Generalized Surficial Geology (NR8) (Rampton, 1984),
5. NB Granular Aggregate Database (New Brunswick Department of Energy and Resource Development, 2018).

The similarity model was applied as follows: mode of deposition and lithology information within each vector data set were extracted to a randomly-generated spatial point grid

with a spacing of 3 points per hectare. At each point the classifications from each vector data set were compared and uncertainty values assigned based on how many of the data sets provided the same classifications. If three or more were the same, then it was assumed that the geology at that site was known. If only half, or less, had the same information then additional research was conducted via geological reports (NB Department of Energy and Resource Development, 2018a). This occurred for 20% of the randomly-generated points.

With each point assigned classes, thiessen polygons were delineated and the boundaries of adjacent polygons dissolved if they shared the same class, completed for both mode of deposition and lithology. Using Moh's mineral hardness scale, mineral hardness values were assigned to each lithology class (Table 5.3). Additionally, primary lithology was used to group lithological types by dominant grain size (very fine to coarse), dominant rock type (i.e., sandstone, shale, granite), and dominant geology (igneous, sedimentary, metamorphic) (Fig. 5.3).

Table 5.3. Mineral hardness values assigned to each lithological class within spatial database based on Moh's mineral hardness scale.

Lithology	Hardness	Lithology	Hardness
Sandstone, gritstone, shale	6.75	Igneous, Slate	4.8
Mafic volcanics, gabbro, diorite	6.6	Calcareous sandstone and shale	4.5
Granite and metamorphic	6.5	Non-calcareous shale, some sandstone	4.5
Granite, some quartzite, sandstone	6.5	Sandstone and Conglomerate (calcareous if un-weathered)	4.5
Sandstone	6.5	Sandstone and conglomerate (some quartz)	4.5
Sandstone, some granite, quartzite, gneiss	6.5	weakly calcareous Sandstone and shale	4.5
Sandstone/ Undifferentiated	6.5	Metamorphosed non- to weakly-calcareous slate, quartzite, argillite, sandstone	4.2
Metamorphosed rhyolite, andesite, schist, slate, granite	6.4	Sandstone, conglomerate, shale	4.0
Granite, gneiss, basalt, felsite	6.3	Slate and argillite and sandstone and schist	4.0
Calcareous sandstone and quartzite, some argillite, shale	6.0	Weakly- to calcareous shale, slate, quartzite, some sandstone	4.0
Granite, quartzite, gneiss, argillite, volcanics, some sandstone	6.0	Sandstone, conglomerate, mudstone	3.8
Non- to weakly-calcareous sandstone, shale, quartzite	6.0	Sandstone, shale, mudstone	3.8
Non-calcareous sandstone and quartzite, some argillite, slate, shale, schist	6.0	Calcareous shale and slate with/without limestone, argillite	3.5
Sandstone/Sandstone and shale or siltstone	5.8	Slate	3.5
Sandstone/ clay from calcareous shale	5.5	Calcareous shale, slate, quartzite, argillite	3.3
Strongly metamorphosed slate, quartzite, volcanics	5.3	Calcareous Shale	3.0
Highly calcareous shale, quartzite, argillite, sandstone	5.0	Weakly calcareous shale, mudstone	2.75
Quartz, Schist, Igneous	5	Clay	2.0

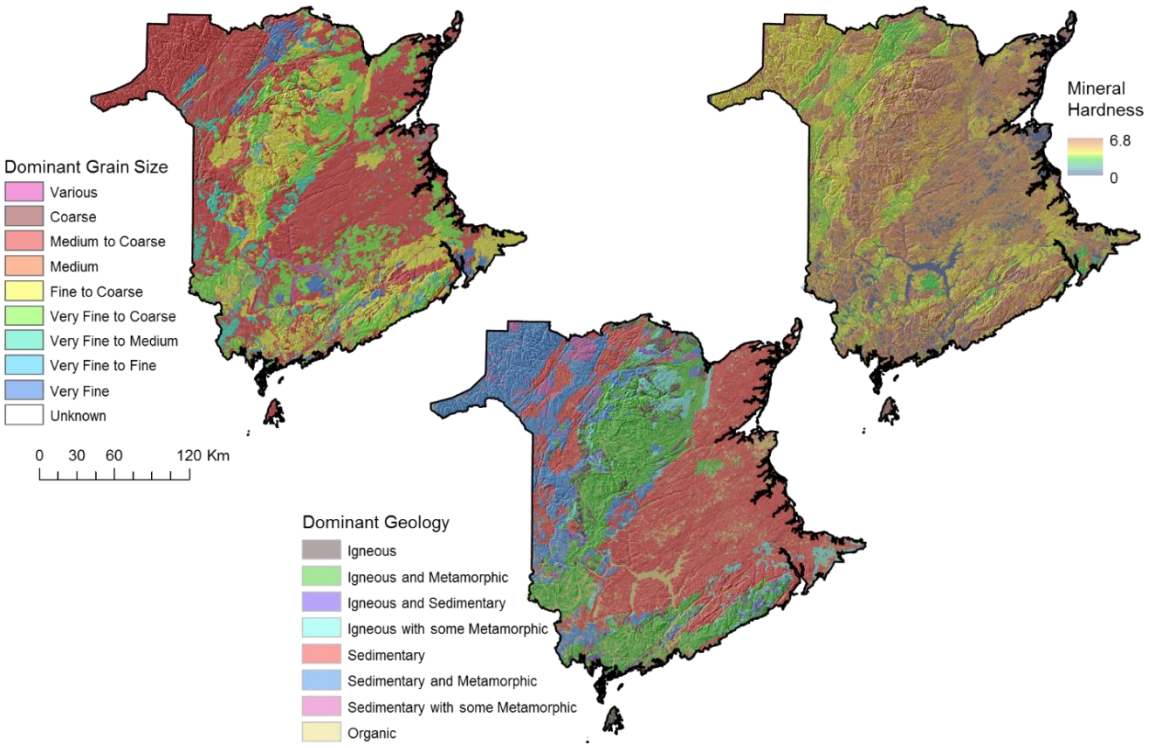

Figure 5.3. Visual examples of variability in dominant grain size, mineral hardness, and geological type for NB, derived from updated lithological representations via the similarity model.

Topography: Both primary and secondary topographic derivatives were derived from a 10m DEM for NB. This was completed in both ArcGIS and SAGA GIS software (Conrad *et al.*, 2018; Environmental Systems Research Institute, 2018) utilizing the DEM derived in Chapter 4. Prior to calculating the topographic derivatives, the DEM was hydro-conditioned to ensure correct flow patterns from ridge to valley (Saunders and Maidment, 1996; Murphy *et al.*, 2008). Next, a suite of derivatives were calculated with emphasis on those representing soil-landscape relationships, namely those influencing the rate, dissipation, and accumulation of nutrients, water, and minerals across the landscape. Also included were derivatives that represent atmospheric exposure (position

on landscape, open and flat areas). Appendix III outlines both topographic and hydrologic derivatives incorporated into the modeling procedure.

Climate: To develop continuous maps of climatic variables of interest, meteorological stations within NB and Ontario, Canada (n = 151) were spatially plotted. Next, annual average values for temperature (annual average and annual daily maximum, °C) and precipitation (annual average rainfall, mm) calculated for the past 29 years (1981 – 2010) were assigned to each weather station (retrieved from Environment Canada (2018)). The measured climatic properties were compared to DEM-derived topographic derivatives (those believed to influence environmental exposure, i.e., topographic openness, elevation, topographic position, aspect, slopes, and horizontal distance to coast) and spatial data sets representing variations of latitude and longitude, for NB and Ontario, via linear regression analyses. Regression results were applied spatially at 10m resolution resulting in gridded data sets representing climatic variation.

DEM Resolution and Multi-Scale Analysis: Since landscape processes are scale dependent, some of the above DEM derivatives were calculated multiple times at different resolutions due to the artificial presence of many small-scale mounds and pits within the 10 m DEM (Fig. 5.4). This was in part resolved through resampling the DEM at 30 m resolution. Some of the resulting derivatives were smoothed further based on a determining their average values within their 5 x 5 cell neighborhoods, followed by subsequent re-interpolation at 10 m. This procedure was applied to all topographic data sets representing curvatures and the results utilized in following analyses.

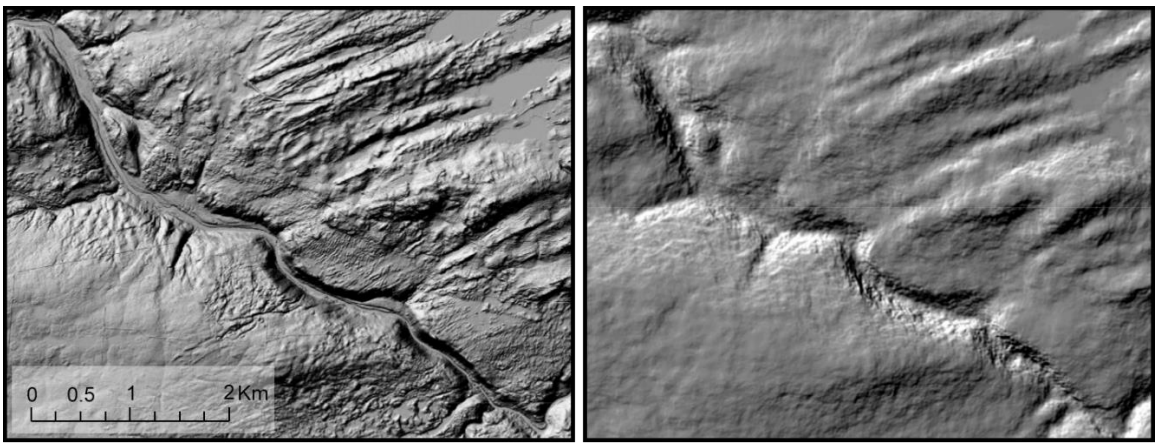

Figure 5.4. Comparison of hill-shaded relief for NB fused DEM (10m, **right**) and LiDAR-Derived DEM (1m, **left**) representing the omission of topographic variation and the presence of pits and peaks, particularly on hill sides within the fused DEM.

5.3b Statistical Analyses

Variable Correlation Analysis: Spatial gridded data sets pertaining to topography, hydrology, climate, and surficial geology (hereby titled "covariates") were compared via Pearson correlation analysis (Table 5.4) to determine the extent of multicollinearity. Correlations ≥ 70% were considered significant to ensure correlation and to help aid in variable reduction. For each covariate, counts were calculated to determine the number or correlated covariates. Those with the highest counts were retained while their correlated counterparts were removed. This process was conducted iteratively until all remaining covariates were un-correlated. The first covariates assessed were those calculated at different neighborhood sizes, including topographic position index (200, 300, 500, 700m search radius) and terrain surface texture, terrain surface convexity, terrain surface concavity, and terrain ruggedness index (each calculated at 60m, 200m, 500m search radii). The search radii applied were selected based on qualitative inspection of the resulting data set and how it corresponds with changing hillslope. Overall, this technique resulted in a variable reduction of 50%, from 56 to 30 covariates.

Table 5.4. Overview of Pearson correlation results on covariates, with results of recursive covariate removal depending on correlation values. Covariates considered correlated if absolute correlation values were at least 70%.

Covariates Retained	Code	Correlated Covariates	Correlation %
Slope (Degrees)	SLP	Slope Length	86
		Real Surface Area	88
		Topographic Negative Openness	-85
		Topographic Positive Openness	-86
		Terrain Surface Texture (220m search)	-70
		Terrain Ruggedness Index (110m Search)	88
Topographic Position Index (110m Search)	TPI	Downslope Curvature	81
		General Curvature	72
		Local Downslope Curvature	71
		Local Curvature	75
		Upslope Curvature	77
		Local Upslope Curvature	73
Daily Maximum High Temperature	Admax	Annual Average Rainfall (mm)	89
		Annual Average Temperature (°C)	100
		Digital Elevation Model	-82
		Latitude	-87
Curvature Classification	Cur_Cls	Catchment Slope	100
		Horizontal Distance to Stream	84
		SAGA Wetness Index	100
Catchment Area	CA	Specific Catchment Area	100
		Modified Catchment Area	86
Tangential Curvature	Tan_Curve	Cross-sectional Curvature	97
Terrain Surface Convexity	Ts_Conv	Terrain Surface Concavity	-90
Planar Convexity	Plan_Conv	Profile Convexity	100
Longitude	Long_10	Horizontal Distance to Coast	-73
Landform	Land	-	-
Lithology	Lithology	-	-
Grain Size	Grain_Size	-	-
Rock Type	Rock_Type	-	-
Geology Type	Geo_Type	-	-
Aspect	ASP	-	-
Diurnal Anistropic Heating	Di_Heat	-	-
Downslope Distance Gradient	Down_Grade	-	-
Flow Accumulation	FA	-	-
Flowline Curvature	Flw_Curve	-	-
Mineral Hardness	Hardness	-	-
Longitudinal Curvature	Lng_Curve	-	-
Maximum Curvature	Max_Curve	-	-
Minimum Curvature	Min_Curve	-	-
Planar Curvature	Plan_Curve	-	-
Profile Curvature	Prof_Curve	-	-
Total Curvature	Tot_Curve	-	-
Cartographic Depth-to-water Index	DTW	-	-
Multi-resolution Index of Valley Bottom Flatness	MRVBF	-	-
Multi-resolution Index of Ridgetop Flatness	MRRTF	-	-
Valley Depth	Val_Depth	-	-

Each soil profile within the database was then compared to underlying covariates via the Random Forest machine learning algorithm (Breiman, 2001) in the 'caret' package for R (Kuhn, 2017; R Core Team, 2018). In this, a randomly-selected 30% of the database was set aside as a validation dataset to assess all modeling results. During modeling, the training data was partitioned so that 70% of each drainage class was randomly sampled for training while the remaining 30% was used for model validation. With Random Forest, instead of assigning a specific number of predictors to test at each node, a tuning grid was developed which tests a matrix of predictor variables (ranging from 3 to 24 at intervals of 3) at each node within the decision trees of the random forest model. Model training was repeated with k-fold cross validation using 20 replicates. Each random forest run was completed with a tree size of 10 (number of regression trees to develop and test).

Model performance was assessed via two methods, RMSE and R^2 for continuous predictions (depth, texture, CF, SOM, and Db), and confusion matrix with Kappa index for categorical predictions (drainage class). Model performance was determined by comparing predictions within training data to the separate validation data. Variable importance for the 5 dominant predictors of each model are provided in Appendix IV.

5.4. RESULTS

Soil Drainage: Soil drainage is represented as 7 classes, ranging from very poor to excessively drained, as outlined by Working Group on Soil Survey Data (1982). The rate at which water passes through soils is a function of topographic position, soil physical

properties, and geomorphology (Wilson and Gallant, n.d.; Gallant and Hutchinson, 1997; Zhu and Mackay, 2001; Taylor *et al.*, 2013).

The distribution of soil drainage classes within the database is biased toward soils on moderately well- to well- drained sites (Table 5.5) due to the different objectives of each project in which the soil profile data were collected. Drainage classes were grouped into broader classes to increase sample size of each group for better representation, omitting transitions between classes.

Table 5.5. Representation of technique applied to amalgamate individual soil drainage classes (at the profile level) into groups representing the dominant drainage classes. This result was necessary to increase the sample sizes of the under-sampled classes to improve model predictions.

Original Classification		**New Classification**	
Class	Count	Class	Count
VP	33	VP	34
VP-P	1		
P	226	P	229
P-I	3		
I	982	I	983
I-MW	1		
MW	2450	MW	2455
MW-W	5		
W	4890	W	4895
W-R	5		
R	648	R	658
R-EX	9		
EX	1		

Soil drainage model results yielded a Kappa score of 79.8% with the confusion matrix outlined in Table 5.6. The imperfectly-, moderately well-, and well-drained classes had the greatest variations in predictions, but still performed well in predicting the proper class. Overall, model performance and visual inspection do coincide with theoretical

expectations of changes in soil drainage in relation to changes in topography for most of the landscape, low lying areas have poorer drainage than steeper slopes and ridgetops. Qualitatively, performance decreases in northern portions of the province where variations in topography are more frequent and intense (Fig. 5.5). This is likely due to the lack of samples on more poorly-drained soils within both the training and validation data as well as the influence of climate and longitude on predicting drainage. Additionally, coarser DEM resolution (10m) results in the inability to capture small depressional and low-lying areas where drainage is likely to be classified as poor or very poor. Also, coarser DEM resolution results in small and frequent variations in elevation (pits and peaks) over short distances. This results in abruptly changing drainage classes over short distances, and at times, one cell may be allocated a drainage class whereas all adjacent cells have the same, but different drainage class from the initial cell.

Table 5.6. Confusion matrix results of soil drainage class predictions via the Random Forest Model.

Prediction	**Reference**					
	VP	P	I	MW	W	R
VP	6	0	2	0	0	0
P	0	38	4	3	1	1
I	2	2	184	9	13	2
MW	1	5	10	428	48	8
W	1	8	31	64	1030	21
R	0	0	5	4	15	115

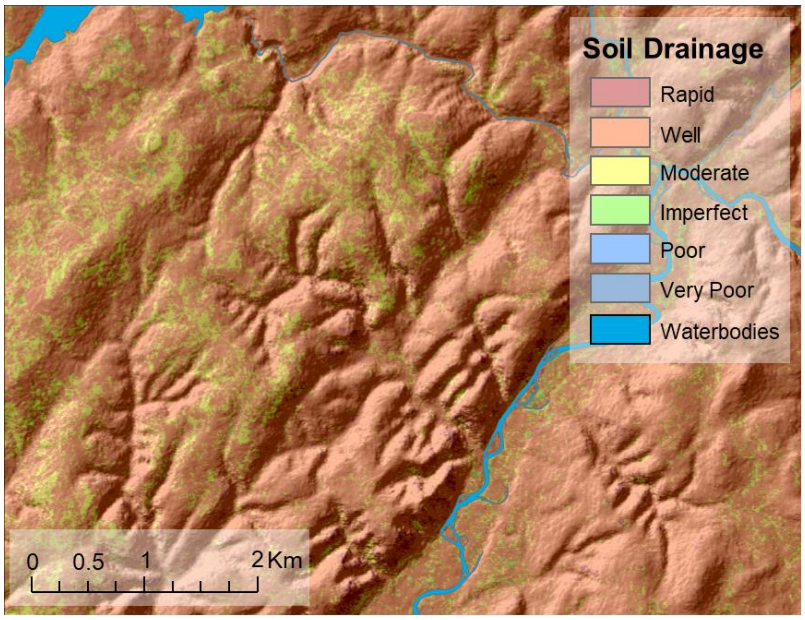

Figure 5.5. Visual example of soil drainage model result located in Northwestern NB. Note the tendency to predict well- and moderately well- drained classes.

There are few articles on predicting soil drainage classes via DSM, with most focusing on predicting soil hydraulic properties. Campling *et al.* (2002) utilized logistic modeling to map the similar soil drainage classes as outlined in this study, but at 25m resolution within Southeastern Nigeria. Their results performed well at predicting hydric vs. non-hydric soils (99%), but model performance decreased (65%) when predicting imperfect- vs. moderately well-drained soil classes. Bell *et al.* (1994) determined the probability of drainage classes in Pennsylvania, USA, via multivariate discriminant analysis. Their results agreed with published soil drainage maps at 67%. Murphy *et al.* (2011) compared the cartographic depth-to-water index (DTW) and topographic wetness index (TWI) from LiDAR data for mapping soil drainage classes for NB, with results explaining 64% and 25% of the variability in drainage classes, respectively. Their study only focused on three drainage classes, very poor, poor, and imperfectly-well-drained.

Finally, Zhao *et al.* (2018) utilized artificial neural networks to predict soil drainage classes for New Brunswick. Their results yielded a correct prediction for 52% of the data, but their study did not mention what resolution or source of the DEM used.

Depth: Each mineral horizon sampled was provided an upper and lower depth at which they occur (i.e., 0-6cm). With this, profiles were summarized by depth to C horizon (solum), and average depth of each horizon (i.e., A and B horizon depth). Additionally, soil properties of each profile were averaged by depth classes representing 0-15cm, 15-30cm, 30-60cm, 60-90cm, 90-150cm, and >150cm (Fig. 5.6). Doing so allowed for soil properties to be summarized by these classes to not only represent how soil properties vary by depth, but also how they compare from site to site on specific depth intervals. Additionally, the influence of topography on soil properties varies with increasing depth, thus, better predictions of soil may result from separating properties into depth intervals (Florinsky *et al.*, 2002). Analyses were not conducted on samples exceeding 90cm (labelled as '150' in Fig. 5.6) due to the infrequent occurrence of soil profiles reaching these depths. Fig. 5.7 represents the Random Forest modeled depths for A and B horizon thickness, and depth to C horizon (solum depth). Alternatively, spline functions could be used to incorporate the influence of depth on soil properties, but it becomes difficult when developing spatial predictions of soil property variation.

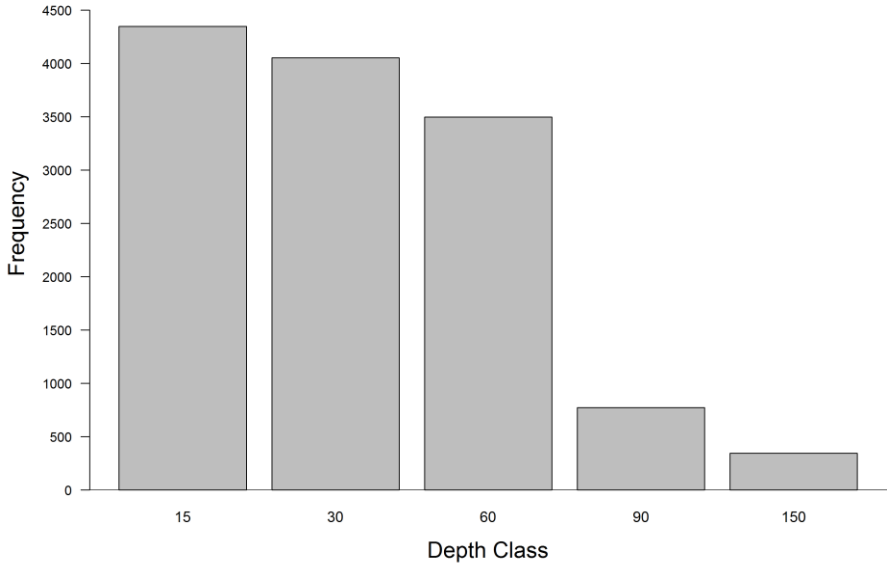

Figure 5.6. Frequency distribution of defined depth classes within spatial database for profiles in which soil horizons were provided (23%). Depths exceeding 90cm (labeled as '150') were not analyzed but shown to represent the lack of profiles reaching these depths.

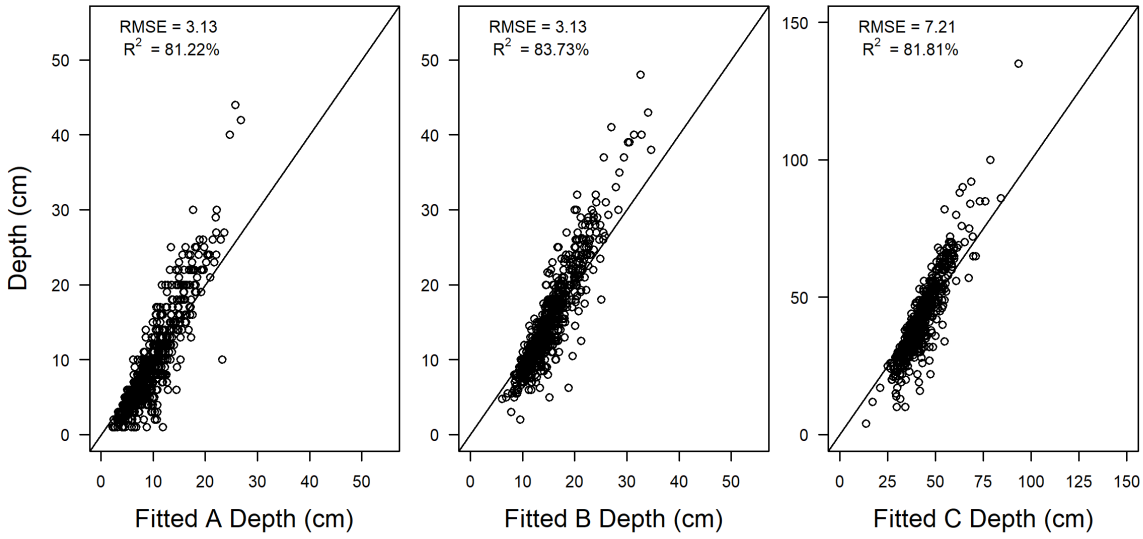

Figure 5.7. Comparison of performance plots (measured vs. fitted) for A-horizon depth (**left**), B-horizon depth (**center**), and solum depth (**right**) predictions.

Modeling A horizon, B horizon, and solum depth to underlying covariates yielded strong relationships with the B horizon model explain the greatest variation. In

comparison to these findings, Florinsky *et al.* (2002) were able to explain 20% of the variation in solum thickness for an agricultural area in the Canadian prairies. Hengl *et al.* (2004a) compared different techniques to correct DEM errors with which these were used to predict solum thickness in Eastern Croatia. Their best performing model explained 40% of the variation in solum thickness. Taylor *et al.* (2013) compared imagery with topographic derivatives, both ranging from 10 and 90m resolution for mapping soil depth. Using a multi-scale approach, they were able to predict 91% of the variation in soil depth in the Peyne Watershed, Southern France. Gessler *et al.* (1995) developed models for predicting both solum depth and A-horizon depth from curvatures and the compound topographic index with which they explained 63% and 68% of the variation in A-horizon and solum depths, respectively. Similarly, Moore *et al.* (1993) utilized the compound topographic index to explain 55% of the variation in A-horizon depth.

Texture: Texture classes were converted to average proportions of sand, silt, and clay (70.4% of database, Table 5.7) and values summarized by depth class intervals. The frequencies of the resulting sand, silt, and clay percentages are plotted in Fig. 5.8. The resulting best-fitted sand, silt, and clay percentage emulations with varying depth intervals are shown in Fig. 5.9, capturing > 85 % of the percentage values for each class.

Table 5.7. Distribution of texture classes sampled on a horizon basis within database with average proportions of sand, silt, and clay within each class, by increasing sand content.

Texture Class	Abbreviation	% Sand	% Silt	% Clay	Sample Size	% of Total
Silty Clay	SiC	5	49	46	32	0.18
Silt	Si	5	90	5	5	0.03
Silty Clay Loam	SiCL	10	56	34	67	0.37
Heavy Clay	HC	12	13	75	2	0.01
Silt Loam	SiL	20	65	15	1847	10.22
Clay	C	25	25	50	396	2.19
Clay Loam	CL	32	34	34	1375	7.61
Clay Loam – Loam	CL-L	35	39	40	1	0.01
Loam	L	41	40	19	5295	29.29
Silt Loam – Sandy Loam	SiL – SL	45	50	5	1	0.01
Sandy Clay	SC	53	4	43	5	0.03
Loam – Sandy Loam	L - SL	53	36	11	6	0.03
Sandy Clay Loam	SCL	60	13	27	1270	7.02
Sandy Clay Loam – Sandy Loam	SCL – SL	65	15	20	5	0.03
Sandy Loam	SL	68	22	10	5851	32.36
Sandy Loam – Loamy Sand	SL – LS	77	14	9	7	0.04
Loamy Sand	LS	80	13	7	1512	8.36
Loamy Sand – Sand	LS – S	88	7	5	3	0.02
Sand	S	93	4	3	400	2.21

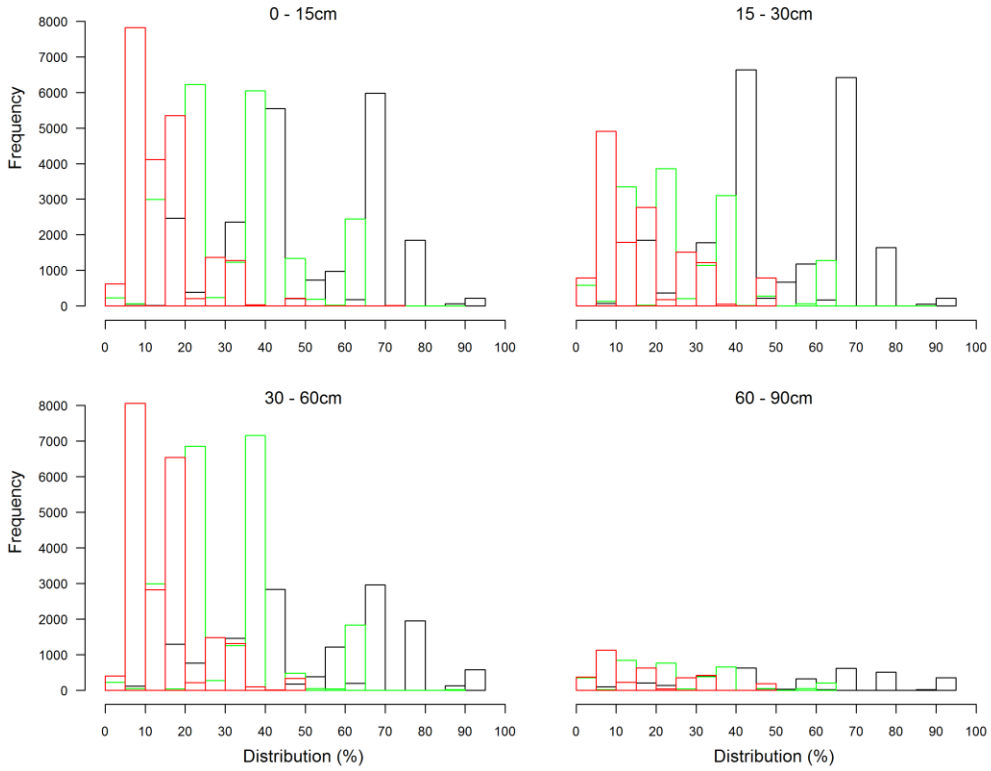

Figure 5.8. Comparison of sand (**black**), silt (**green**), and clay (**red**) composition for changing depth intervals from 0-15cm (**top left**), 15-30cm (**top right**), 30-60cm (**bottom left**), and 60-90cm (**bottom right**). Note the reduced sample sizes of the finer-textured soils (silts and clays) even with increasing depth.

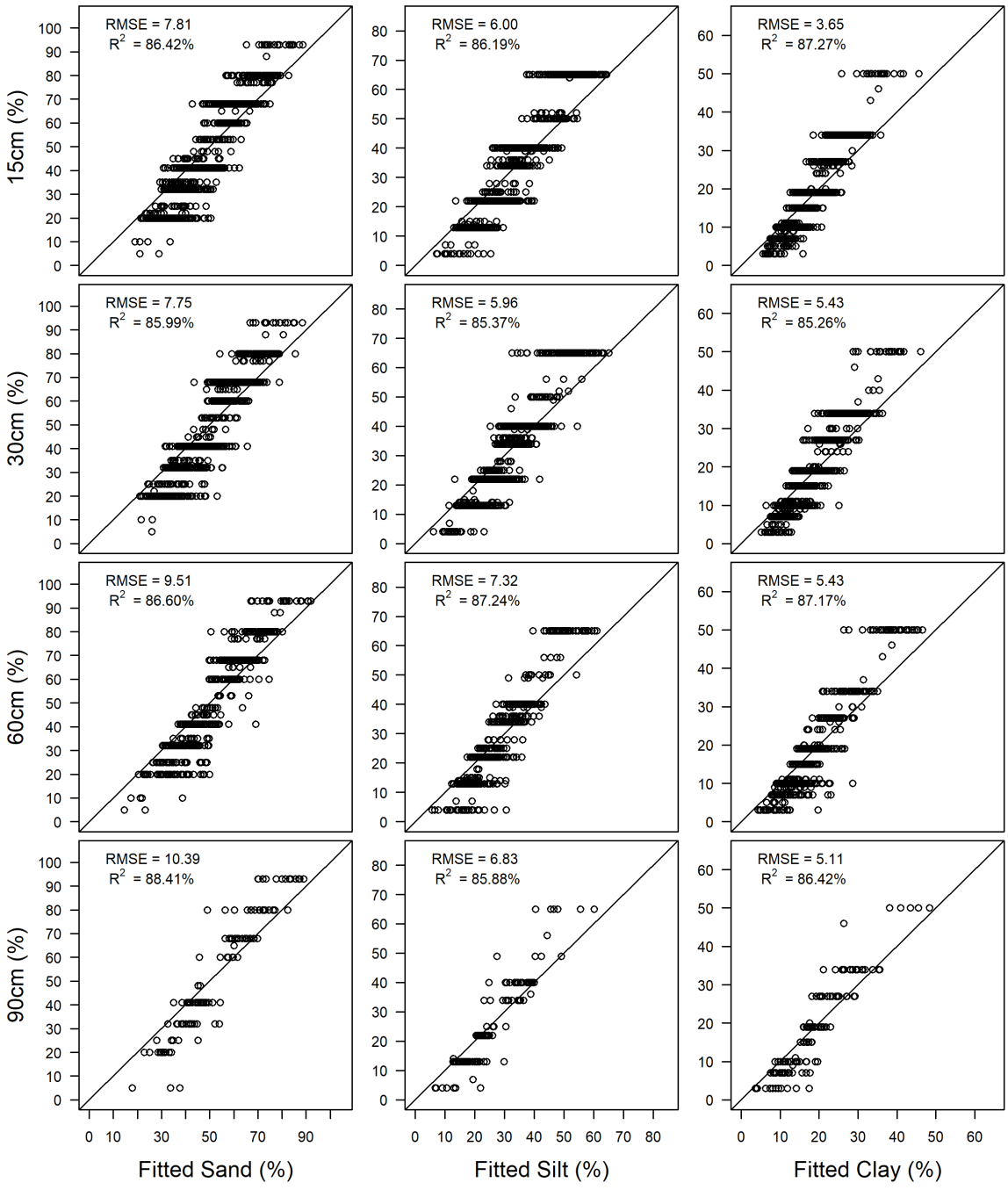

Figure 5.9. Comparison of performance plots (measured vs. fitted) for sand (**left**), silt (**center**) and clay (**right**) models at defined depth intervals of 0-15cm (**top row**), 15-30cm (**second row**), 30-60cm (**third row**), and 60-90cm (**bottom row**). Proportions of sand, silt, and clay assigned from texture class.

From this, it is apparent that model performance increases with increasing depth due to increasing influence of parent material and topography on texture. Near the surface

texture is influenced by environmental factors in the form of water and acidity, altering the texture composition of the soil.

In comparison, Odeh *et al.* (1994) were able to explain 51% of the variation in subsoil clay content using a few topographic derivatives. Their study focused on an area with minimal change in lithology and at coarse resolutions (500m to 1km). Adhikari *et al.*, (2012) predicted silt, clay, and sand (fine and coarse) both spatially and vertically, using depth intervals, for Denmark utilizing the cubist model. They found that the ability to model soil texture decreased with increasing depth (best performance was 54% of the variation explained for clay at 0-5cm). Zhao *et al.* (2013a) utilized artificial neural networks to predict clay content then used the results to predict soil nutrient regimes for Nova Scotia, Canada. This study utilized a 10m DEM and soil information from existing soil polygons. Again with the use of artificial neural networks, Zhao *et al.* (2018) modeled sand and clay contents within the Black Brook watershed in Northwestern NB, Canada. This study also utilized existing soil association polygons and lookup tables for soil information to include in the modeling. Akumu *et al.* (2015) predicted soil texture within the clay belt of Ontario using a fuzzy logic approach with a 10m LiDAR-derived DEM. They predicted six textural classes at 79% accuracy. At a broader scale, Mansuy *et al.* (2014) applied k-nearest neighbors to predict soil texture within the top 15cm of forest soils at 250m resolution across Canada. With this, they were able to explain up to 43% of the variation in topsoil clay content, with weaker results for sand and silt (20 and 13%, respectively).

Coarse Fragment Content: Reviewing the changing CF content with depth, it became apparent that soil profiles exceeding 60cm typically coincide with deep parent materials lacking CF's (i.e., deep alluvial floodplains, Fig. 5.10). The distributions also show that CF content increases with increasing depth from 0 to 60cm. Actual vs. fitted model results for CF content at defined depth intervals are depicted in Fig. 5.11.

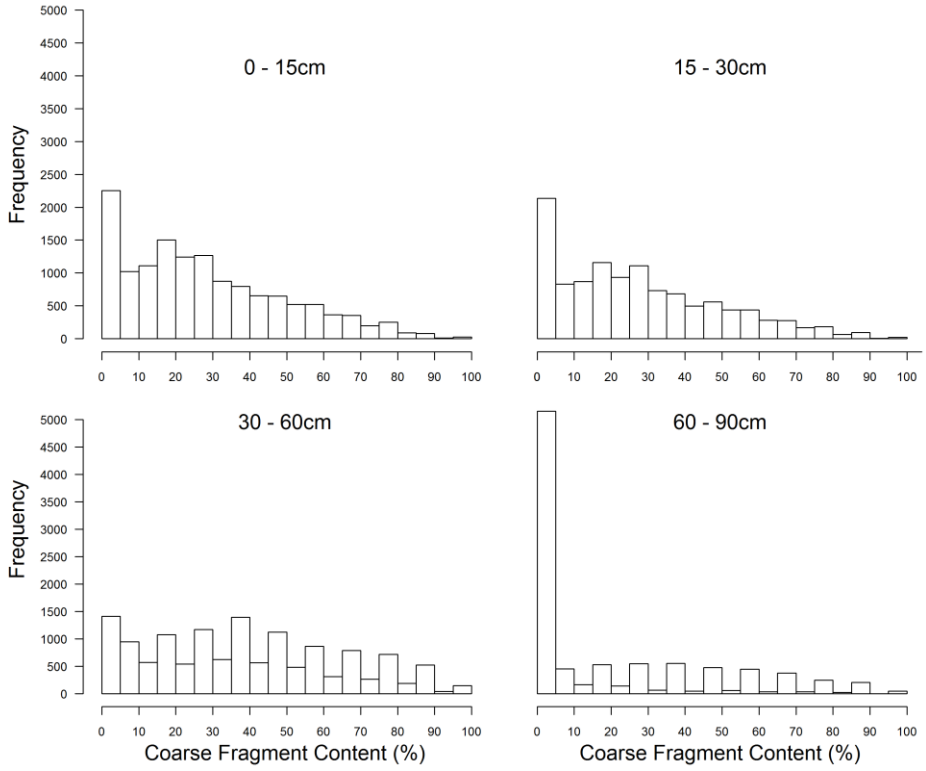

Figure 5.10. Frequency distributions for coarse fragment content at defined depth intervals of 0-15cm (**top left**), 15-30cm (**top right**), 30-60cm (**bottom left**), and 60-90cm (**bottom right**).

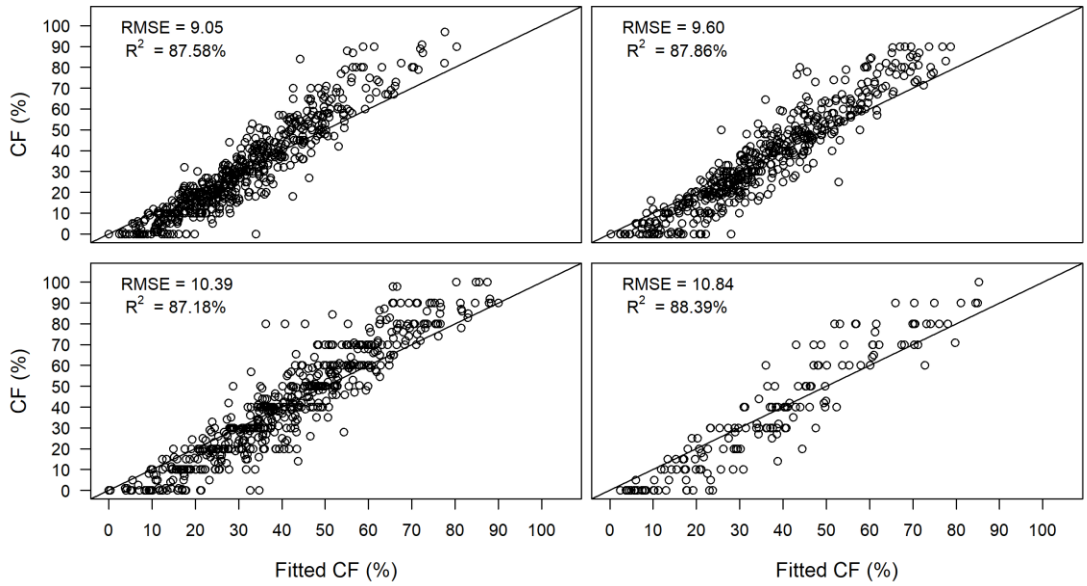

Figure 5.11. Comparison of performance plots (measured vs. fitted) for CF models at defined depth intervals of 0-15cm (**top left**), 15-30cm (**top right**), 30-60cm (**bottom left**), and 60-90cm (**bottom right**).

Limited research has been conducted on DSM for CF content. Mosleh *et al.* (2016) attempted to model CF content along with additional physical and chemical properties within topsoil within the Chaharmahal-Va-Bakhtiari province of Iran, but they were unable to retrieve any significant relationship. Vaysse and Lagacherie (2015) tested different DSM techniques for mapping a suite of soil properties across Europe at 100m resolution at different soil depth intervals. The best RMSE achieved for CF% was 8.37 at a depth interval of 15-30cm with an R^2 of 16%.

Soil Organic Matter Content: In comparison to the other soil properties, SOM data were sparse. The actual versus best-fitted emulation results using the log (ln) transformed SOM are shown in Fig. 5.12 by soil depth class, accounting for about 80% variations within the 0-15 and 15-30 cm soil layers, but only about 30 to 13 % for the 30 -60 and 60-90 cm data layers, for which the organic matter accumulations are also decidedly lower, as to

be expected. The literature reviewed explained 28-67% of the variation in SOM content and 69-80% of the variation in C content (Table 5.8).

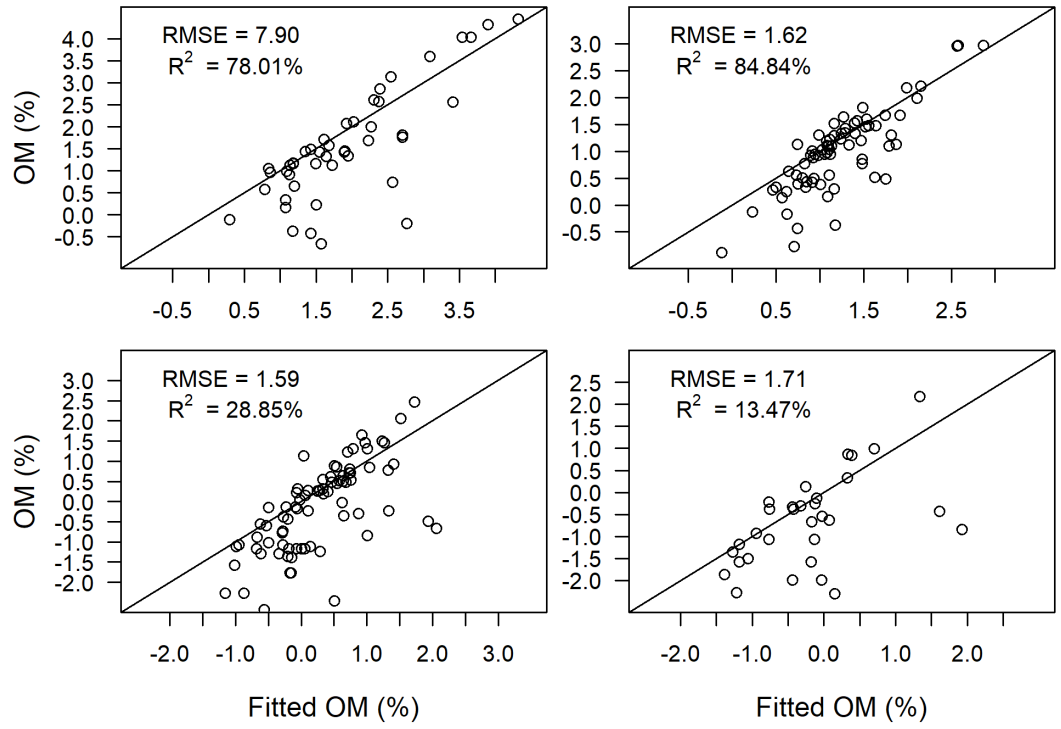

Figure 5.12. Comparison of log-transformed performance plots (measured vs. fitted) for SOM models at defined depth intervals of 0-15cm (**top left**), 15-30cm (**top right**), 30-60cm (**bottom left**), and 60-90cm (**bottom right**).

Table 5.8. Overview of literature in which DSM was applied for modeling both carbon and SOM content at different resolutions and varying geographic extents, with accuracy represented as percentage of variation explained.

Source	Location	Resolution (m)	Property (SOM, C)	Accuracy (% var. Explained)
Florinsky *et al.* (2002)	Manitoba, Canada	15	SOM	37
Hengl *et al.* (2004b)	Croatia	100	SOM	53.3 (OK), 66.5 (RK)
Gessler *et al.* (2000)	California, USA	1	C	80
Bui *et al.* (2006)	Australia	250	C	69
Thompson *et al.* (2006)	Kentucky, USA	10	SOM	28
Zushi (2006)	Japan	10	SOM	50.4

Bulk Density: The frequency distribution of the profile-based Db data is plotted in Fig. 5.13, indicating a shift towards higher densities with increasing depth, as to be expected for (i) a generally glaciated landscape with extensive areas subject to glacial regolith compaction, and (ii) for the general decreasing in soil-aggregating SOM content with increasing soil depth. The actual versus best-fitted results (Fig 5.13) account for about 85% of the of the profile reported data.

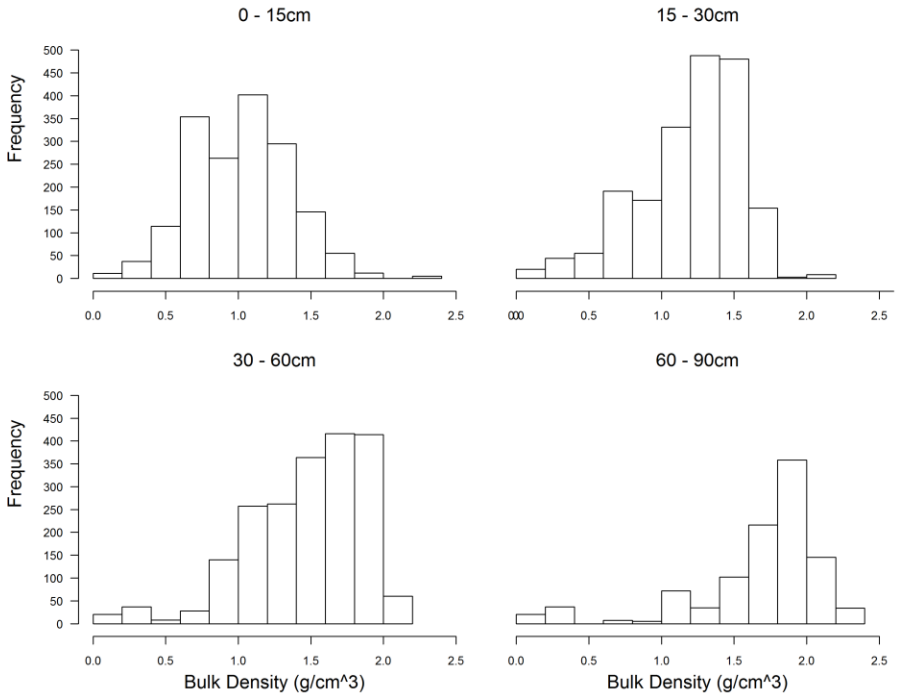

Figure 5.13. Frequency distributions for Db at defined depth intervals of 0-15cm (**top left**), 15-30cm (**top right**), 30-60cm (**bottom left**), and 60-90cm (**bottom right**). Note how bulk density increases with increasing depth.

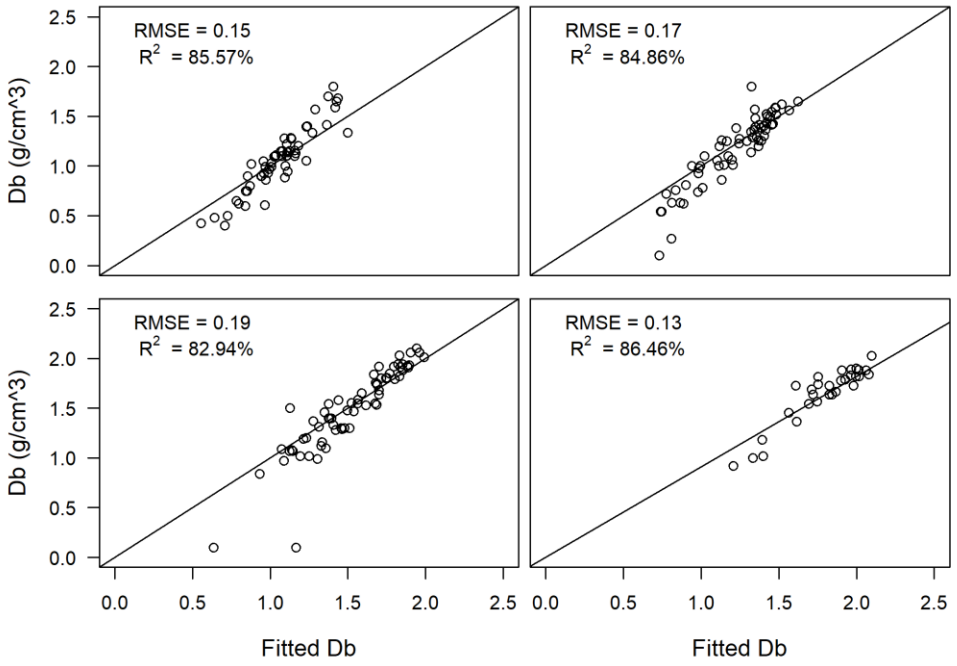

Figure 5.14. Comparison of performance plots (measured vs. fitted) for Db models at defined depth intervals of 0-15cm (**top left**), 15-30cm (**top right**), 30-60cm (**bottom left**), and 60-90cm (**bottom right**).

Only few studies have applied DSM techniques for modeling Db of which Ballabio *et al.* (2016) modeled topsoil sand, silt, and clay and used this to predict Db over the extent of the European Union at 100m resolution. Mansuy *et al.* (2014) predicted 17% of the variation in Db at 250m resolution for the Canadian forested land base.

Solum-based model results for actual vs. fitted predictions of sand, silt, clay, CF, SOM, and Db are represented in Fig. 5.15 with resulting spatial predictions depicted in Fig. 5.16.

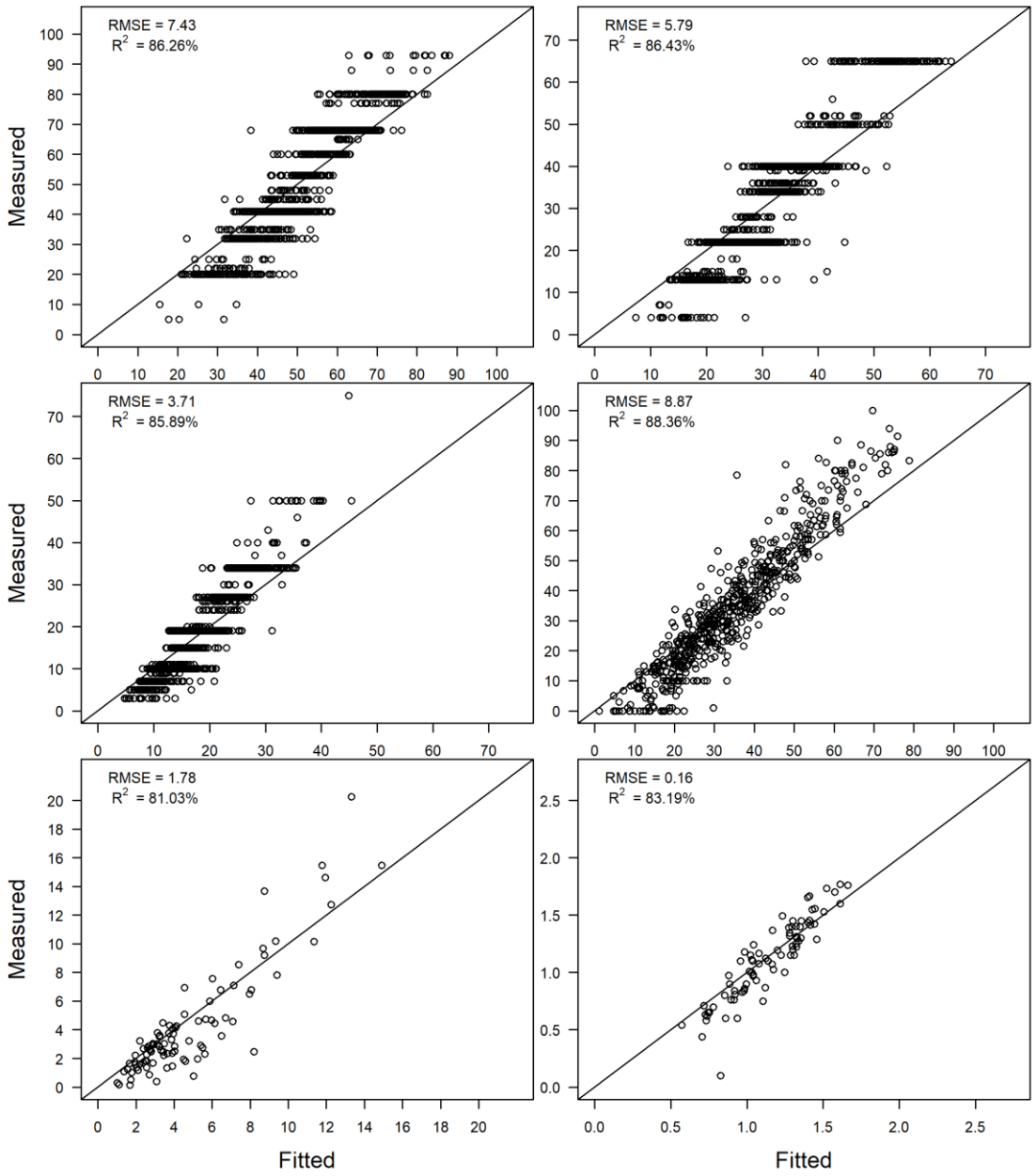

Figure 5.15. Comparison of performance plots (measured vs. fitted) for sand (**top left**), silt (**top right**), clay (**mid left**), CF (**mid right**), SOM (**bottom left**), and Db (**bottom right**) measured on the solum basis.

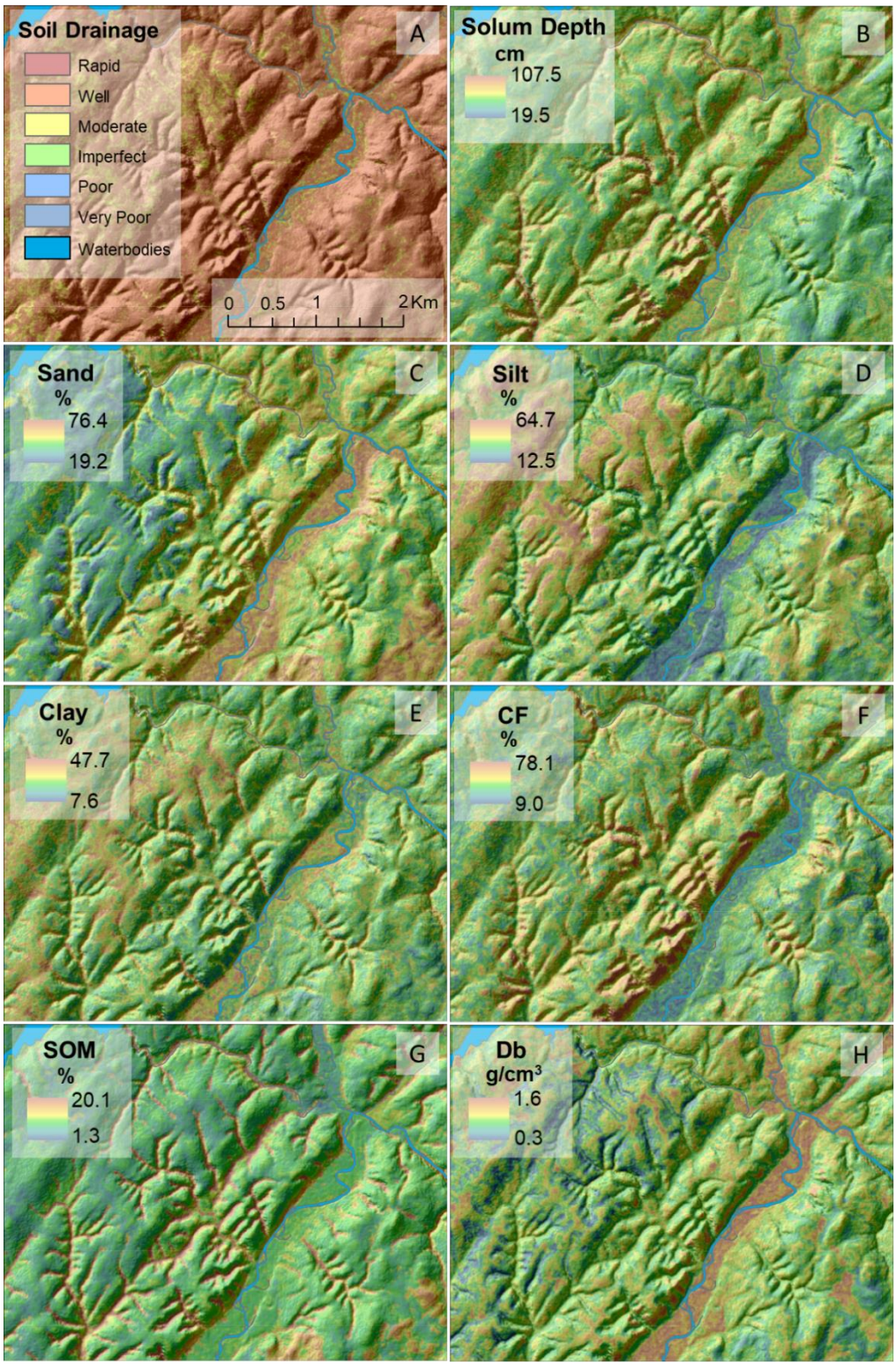

Figure 5.16. Visual representation of model predictions of drainage (**A**), solum depth (**B**), and sand (**C**), silt (**D**), clay (**E**), CF (**F**), SOM (**G**), and Db (**H**) measured on the solum basis within the study extent outlined in Fig. 5.1.

With respect to texture and CF content, there is a bias to over-predict low values, and to under-predict high values by about 10%. The RMSE values range from ± 3.7 (clay) to ± 8.7 (CFs). Actual data for SOM and Db were sparse. The associated RMSE values amounted to 1.8% and 0.12 g cm^{-3}, respectively. For the bottom A-, B- and solum depths (Fig. 5.7), high-value under-predictions amounted to 10 to 20 cm, while low-value over-predictions amount to 5 to 10 cm. Overall RMSE values amounted to ± 3.1, ± 3.1 and ± 7.2 cm, respectively.

5.5 DISCUSSION

The DSM results outlined above represent a limitation to modeling in which model performance is dependent on input data and performance cannot be assessed by quantitative measures only. Having limited data on the poor, very poorly-, and rapidly-drained sites results in a biased drainage model which seldom predicts those drainage classes. Additionally, model performance is limited when small sample sizes are utilized. Of 12,058 samples, Db at 60-90cm was only measured 124 times whereas texture was measured 544 times. Both small samples are a result of infrequent occurrence of profiles reaching these depths, but it is likely that these limited samples do not capture the full variability of soil-forming factors as they change throughout NB. Thus, model performance will be biased toward the sites with soil forming factors similar to those in which the samples were measured.

Soil samples were acquired and amalgamated from different sources increasing the likelihood of inconsistencies between data sets. It is assumed that soil and site

properties were measured in the same manner, but, it is likely that some measurements were done so incorrectly, namely those measured in the field such as drainage classification and field-based texture and CF assessments.

Surficial Geology (Parent Material): Photogrammetrically-derived spatial datasets representing surficial geology tend to be crude in nature, omitting small-scale variations in both mode of deposition and lithology. This is particularly true when dealing with heavily-glaciated landscapes. Invariably, surficial geology maps with sufficient detail are generally unavailable at provincial extents. The rationale behind using the soil polygons as predictors in surficial geology delineations is that these vector data sets are based on diverse expert knowledge and on soil-landscape relationships, although not fully conforming with topographic changes in the landscape. Additionally, as depicted in Fig. 5.17, there are many inconsistencies between data sets when delineating the same landform types, i.e., ablation till, within the same geographic location.

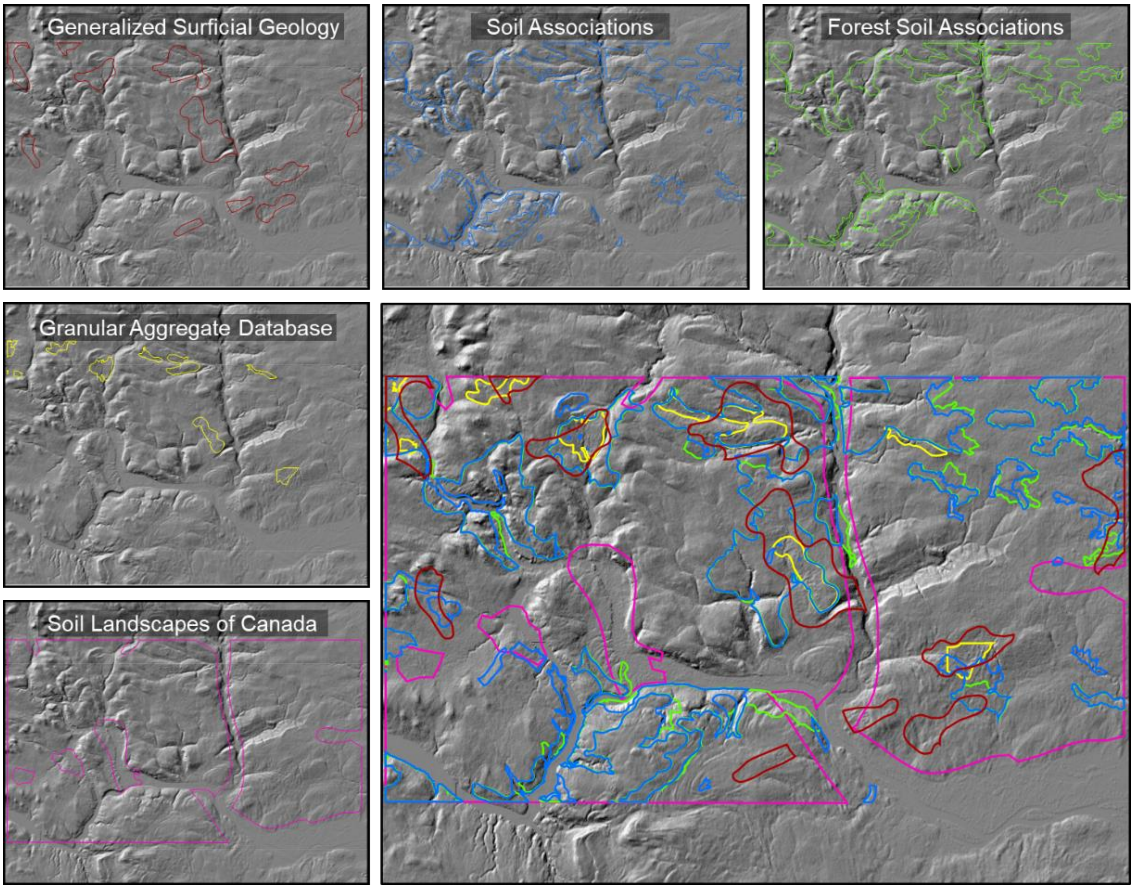

Figure 5.17. Visual comparison of ablation till delineations from different vector data sets representing surficial geology within NB with all data sets overlain (**bottom right**). As seen, many inconsistencies occur between delineations for extent outlined as "C" in Fig. 5.1.

Therefore, combining soil delineations with original surficial geology maps via similarity modeling may yield better representations of variations in landforms (mode of deposition) and lithologies.

Topography: Prior to topographic development, the DEM was hydro-conditioned to ensure hydrologically-correct flow patterns, particularly in low-lying and flat areas (Saunders and Maidment, David, 1996; Murphy *et al.*, 2008). This process is necessary prior to developing hydrological derivatives from any DEM. Once hydrologically-

corrected, the hydrologic derivatives were developed while topographic derivatives were produced from the original DEM.

With the fused DEM from Chapter 4, the influence of coarser resolutions and interpolation is evident in the form of abrupt changes in elevation at small scales ('pits and peaks') (Fig. 5.4). From this, computed topographic derivatives recognize these variations as changes in hillslope position, resulting in noise which is then inputted into statistical modeling. Higher resolution DEMs, namely LiDAR, may yield better predictions in soil property variation due to its ability to adequately represent varying topographic positions at high resolution and the scale at which soil properties vary is finer than the DEM resolution of 10m. Adversely, due to its high resolution, topographic derivatives from LiDAR DEMs under-represent the full hillslope from ridge to valley due to the broad extent at which these occur. Thus, it is likely that topographic derivatives from LiDAR would require a smoothing technique and/or larger search neighborhood to fully capture the hillslope.

As a case study, the curvature classification, utilized as one of 30 covariates for DSM, was developed under varying DEM resolutions and neighborhood sizes. Retrieved from Birkeland (1999) and originally from Dikau (1988), curvature classification distinguishes different forms of a hillslope, ranging from 0-8 depending on the concavity and convexity in both planar and profile form (Fig. 5.18).

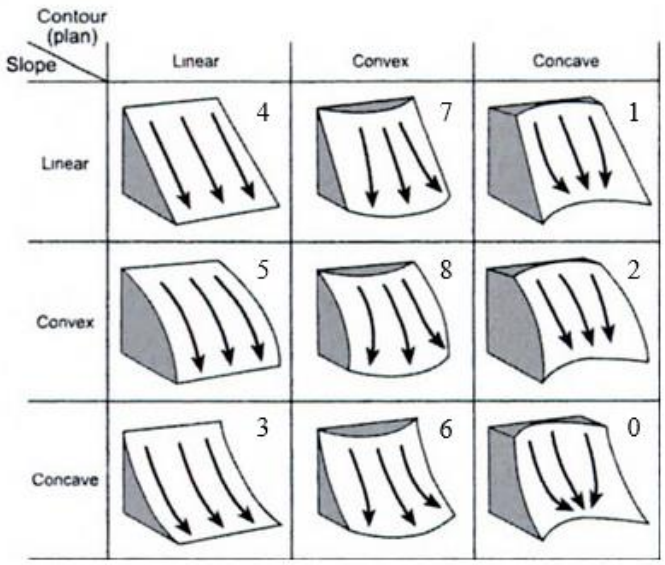

Figure 5.18. Visual representation of the 8 curvature classifications derived from "Curvature Classification" script in SAGA (Conrad *et al.*, 2018), with the value of each class shown. Classes and figure retrieved from Dikau (1988) and Birkeland (1999) respectively.

At 10m resolution, the curvature classification shows much variation over a locally-defined spatial extent (Fig. 5.19). Smoothing the curvature classification grid using a 5 x 5m rectangular neighborhood (search window) yields different classifications at a slightly broader extent. Alternatively, aggregating the DEM from 10 to 30m and re-calculating the curvature classification, both on the original DEM at 30m and smoothed DEM, shows how the curvature classification begins to conform with changing topographic features at the hillslope extent. This shows that topographic derivatives from one DEM resolution and at one neighborhood size may not be suitable for DSM, instead, there is a need for a multi-scale approach in which both DEM resolution and neighborhood size vary depending on the derivative being produced.

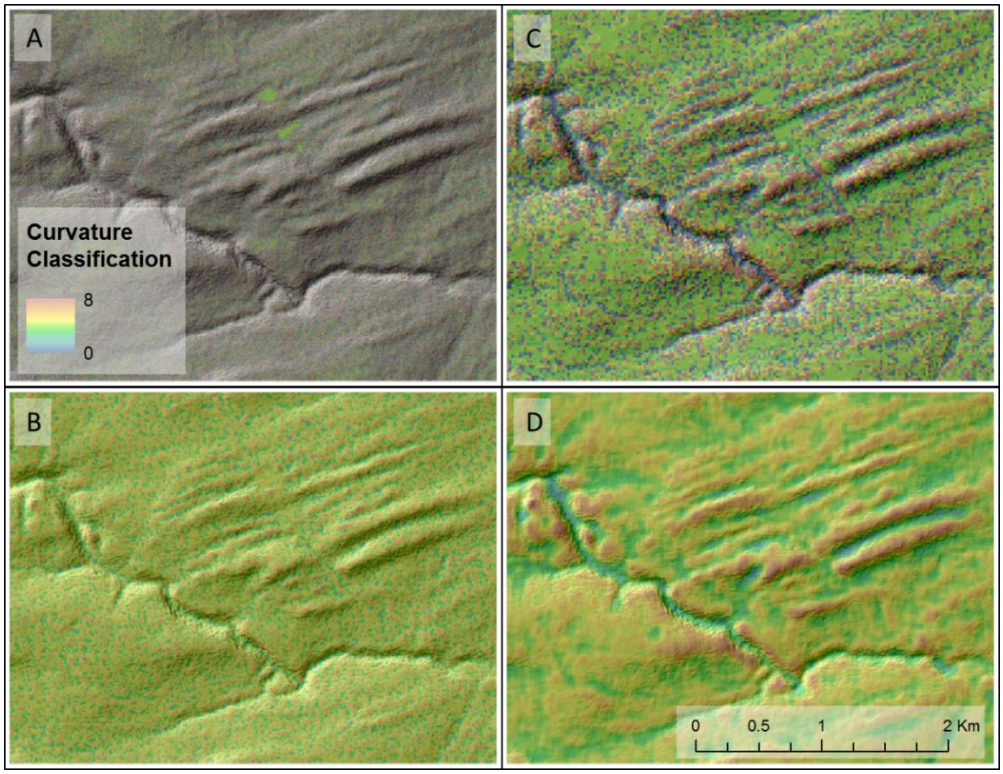

Figure 5.19. Comparison of curvature classifications derived from varying DEM resolutions and neighborhoods of original 10m DEM (**A**), 10m DEM smoothed via 5 x 5 cell window (**B**), 10m DEM aggregated to 50m (**C**), and 50m DEM smoothed via 5 x 5 cell search window (**D**) for study extent "B" outlined in Fig. 5.1. Note how the ability to delineate hillslope positions increases with increasing cell size from 10m to 50m smoothed.

Alternatively, the DEM could be smoothed, by either one or multiple neighborhood sizes prior to developing topographic derivatives. Although feasible, this approach may lose representation of subtle variations in topography which explain soil property variation.

Varying resolutions and neighborhood sizes influences the depictions of hillslope positions and the scale at which these features are delineated. This approach can be applied to any topographic derivative, depending on derivative, allowing for a multi-scale approach to DSM. This is necessary because the influence of hillslope on soil properties occurs over multiple scales. Adversely, some existing algorithms which apply a multi-scale

approach, namely “Multi Resolution Index of Valley Bottom Flatness” (MRVBF) and its counterpart, “Multi-Resolution Index of Ridgetop Flatness” (MRRTF), result in different values for the same location depending on the scale in which they are applied. For example, MRVBF was calculated for a 600,000ha test area in Northern NB (Fig. 5.20) and compared to MRVBF that was calculated for the entire province and extracted to the test area. It is apparent that the resulting data sets provide different values for the same site. In terms of DSM, this difference could result in incorrect, or missed, relationships.

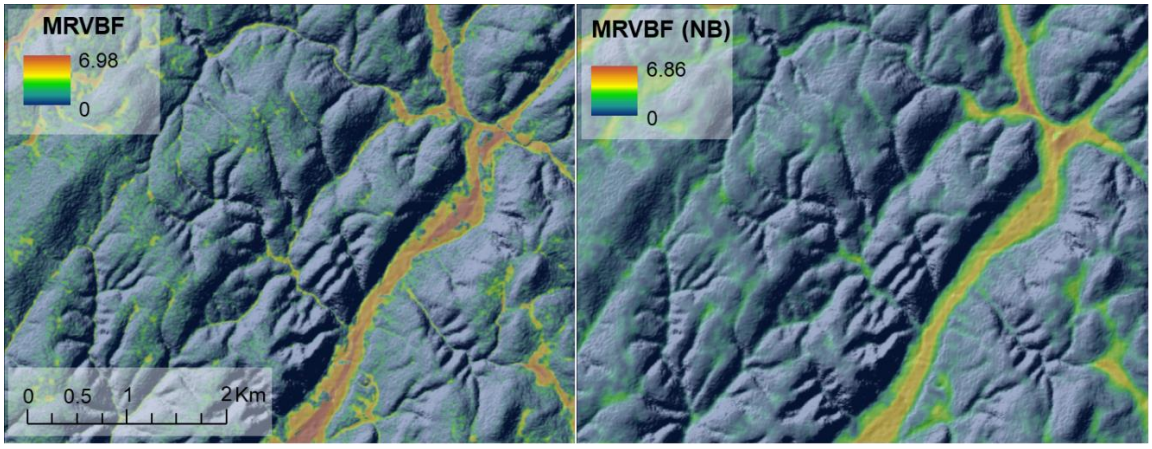

Figure 5.20. Comparison of Multi Resolution Index of Valley Bottom Flatness (MRVBF) derived at two different scales, one for a 600,000ha test site in Northern NB outlined in Fig. 5.1. (**left**) and one at full provincial extent (7.2million ha) (**right**), both overlaying hill-shaded relief of 10m DEM. The results show how the same algorithm developed different results for the same area, depending on the extent in which it is produced.

The multi-scale approach was assessed by Smith *et al.* (2006) for use in digital soil surveys, and Behrens *et al.* (2010) and Zhu *et al.* (2008) for use in DSM. All studies have concluded that there is an optimal threshold for different topographic derivatives to capture the spatial variation in soil properties. With this, it is possible to develop the same topographic derivative at one resolution, but at multiple neighborhoods, and utilize these

in combination for DSM since the influence in which topography affects soil property variation changes with scale.

Climate: Although influential to soil formation and nutrient retention, climatic attributes are typically not included in DSM studies likely due to their coarse resolution in comparison to high-resolution variability of soil properties. Instead, as mentioned, topography is used as a surrogate for climate. Climate attributes were incorporated into this study by comparing climate 'normals' (from 1989-2010) to underlying elevation, latitude, and longitude, allowing for the development of models and the prediction of temperature and precipitation variables at 10m resolution, as outlined in Table 5.9.

Table 5.9. Model results for annual average temperature, daily maximum high temperature, and annual average rainfall from comparing climate normal from 21 years for 150 weather stations to elevation and geography.

Predicted	Variables	Estimate	t- value	p- value	Adj. R^2 (%)
Annual Average Temperature °C	Intercept	48.2867	41.043	<0.0001	92.1
	Elevation	-0.0048	-9.621	<0.0001	
	Latitude	-1.0059	-44.591	<0.0001	
	Longitude	-0.0537	-6.457	<0.0001	
Daily Maximum Temperature °C	Intercept	46.7065	29.134	<0.0001	81.8
	Elevation	-0.0044	-6.491	<0.0001	
	Latitude	-0.8397	-27.315	<0.0001	
	Longitude	-0.0415	-3.659	0.0003	
Annual Average Rainfall mm	Intercept	2762.5580	22.910	<0.0001	63.4
	Latitude	-27.5780	-11.250	<0.0001	
	Longitude	9.5540	13.270	<0.0001	

Weather stations for Ontario were included in the analysis to help increase sample size (n=39 for NB) and to help quantify the influence of geography on varying climatic patterns.

Variable Selection: Incorporating all 60 data sets (48 topographic, 3 climatic, 5 geological) into spatial analyses for modeling soil properties at a provincial extent would be both computationally- and time- intensive. Appendix I outlines all variables developed in this study, and from this, it is apparent that some topographic derivatives are very similar, if not the same, to others. The differentiating factor being the model used to calculate the derivative.

As an alternative to the correlation technique, all derivatives could be incorporated into the modeling phase and have the training procedure remove insignificant variables. This can be completed via assessment of variable inflation factors (VIFs), utilizing variable importance plots (depending on modeling approach), or using principle components analysis. Principle components analysis (PCA) can be an effective technique, but limitations arise in which the technique groups input predictors into components, thus, it is challenging to apply the model result spatially afterwards.

Prediction Results: Model results (variable importance) for A, B, and solum depth, drainage, and sand, silt, clay, CF, SOM, and Db on solum basis are shown in Table 5.10.

Table 5.10. Random forest model results (variable importance for 5 most dominant predictors) for horizon depths (A, B, and solum), drainage, and sand, silt, clay, CF, SOM, and DB on a solum basis. Results for each analysis shown for 5 dominant predictors using abbreviations outlined in Table. 5.4.

A Depth		B Depth		Solum Depth		Drainage	
Covariate	**Importance**	**Covariate**	**Importance**	**Covariate**	**Importance**	**Covariate**	**Importance**
Admax	100.00	CA	100.00	Admax	100.00	SLP	100.00
FA	91.10	MRVBF	96.65	Cur_Cls	93.37	Max_Curve	87.95
TPI	90.22	Cur_Cls	87.31	CA	76.53	Admax	78.61
Geo_Type	85.66	Admax	77.84	Tan_Curve	69.86	TPI	77.35
Max_Curve	82.13	TPI	75.99	Di_Heat	68.43	Down_Grade	71.32
Sand		**Silt**		**Clay**		**CF**	
Covariate	**Importance**	**Covariate**	**Importance**	**Covariate**	**Importance**	**Covariate**	**Importance**
Admax	100.00	Admax	100.00	Wetland	100.00	Val_Depth	100.00
Rock_Type	69.31	Lng_Curve	66.70	Rock_Type	80.94	MRVBF	85.45
Prof_Curve	58.77	Geo_Type	55.59	Cur_Cls	68.09	Admax	71.98
Land	48.10	SLP	54.72	Crs_Curve	63.06	Rock_Type	61.93
Lithology	41.50	Prof_Curve	54.21	DTW	62.17	MRRTF	60.48
SOM		**Db**					
Covariate	**Importance**	**Covariate**	**Importance**				
Geo_Type	100.00	Admax	100.00				
Val_Depth	91.71	Crs_Curve	92.05				
Rock_Type	83.54	SLP	91.73				
Max_Curve	52.80	Ts_Conv	91.11				
Crs_Curve	73.40	MRVBF	80.15				

Covariates produced within this study, both climatic and geological, are dominant predictors for all soil properties. It is accepted that geological covariates influence the physical properties of the soil including texture and CF content. These covariates also strongly influence SOM within the solum and this is likely due to the influence these have on texture composition, and in turn, how texture influence SOM retention. As expected, daily maximum high temperature ("Admax") was a dominant predictor for all properties except clay and SOM. This is likely due to the influence of environmental factors on soils properties within the solum, with stronger environmental influence within topsoil and this influence decreasing with depth. Although decomposition and mineralization of SOM is controlled by climate, both SOM and clay are the two particles frequently transported with moving water through the soil.

In addition to climatic and geolgical covariates, topographic covariates, namely those representing curvature, appear as dominant predictors in all models except CF content. This is expected as curvature influences the rate of movement, accumulation, and dissipation of nutrients, minerals, and water across the landscape. For modeling soil depth, these results are in agreement with other literature (Moore *et al.*, 1993; Boer *et al.*, 1996; Florinsky *et al.*, 2002; Hengl *et al.*, 2004a; Ziadat, 2010; Taylor *et al.*, 2013) in that surface curvature strongly influences soil depth. In addition, Mansuy *et al.* (2014) had shown that curvature, in combination with other topographic derivatives, has proven useful for measuring soil texture, Db, and SOM content across Canada. Also, Creed *et al.* (2002), Pennock (2003), and Terra *et al.* (2004) discovered that surface curvature were useful in predicting SOC contents. Finally, Romano and Palladino (2002), Mulder *et al.* (2011), and Taylor *et al.* (2013) discovered the curvature helped explain water redistribution, water table depth, and water retention, respectively.

5.6. CONCLUSIONS

This study presented an approach to multi-scale digital soil mapping of select soil properties using New Brunswick, Canada as a case study. Input predictor variables represented the dominant soil-forming factors of topography, climate, and surficial geology with the number of variables reduced via correlation analyses. The soil modeling results present access to soil information previously unavailable with traditional soil mapping techniques as confirmed in Fig. 5.21 with provincial representation shown in Fig. 5.22.

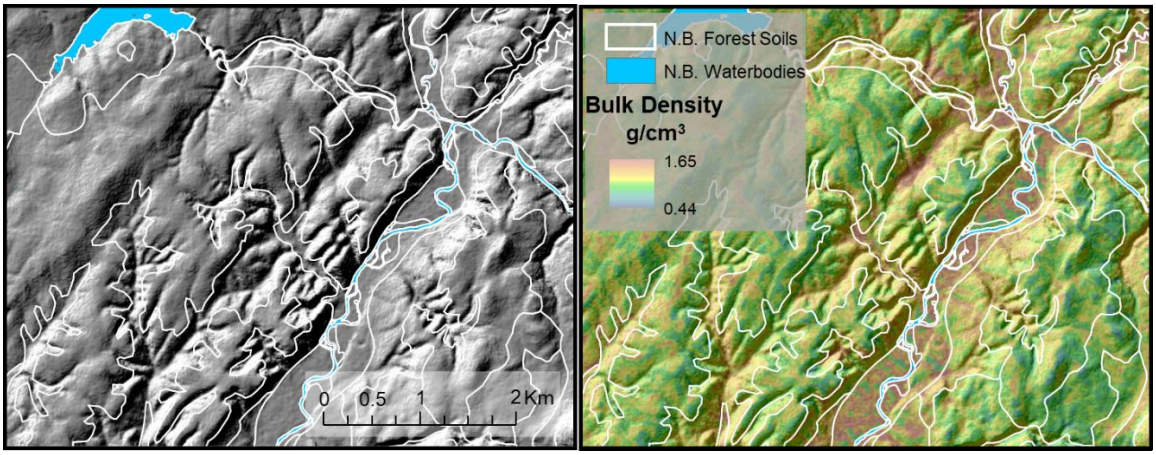

Figure 5.21. Comparison of conventional soil association mapping (**left**) and digital soil mapping (**right**) using bulk density measured on the solum basis for study extent outlined in Fig. 5.1. Note the variability of soil properties now assessable within confines of conventional, polygon-based soil maps.

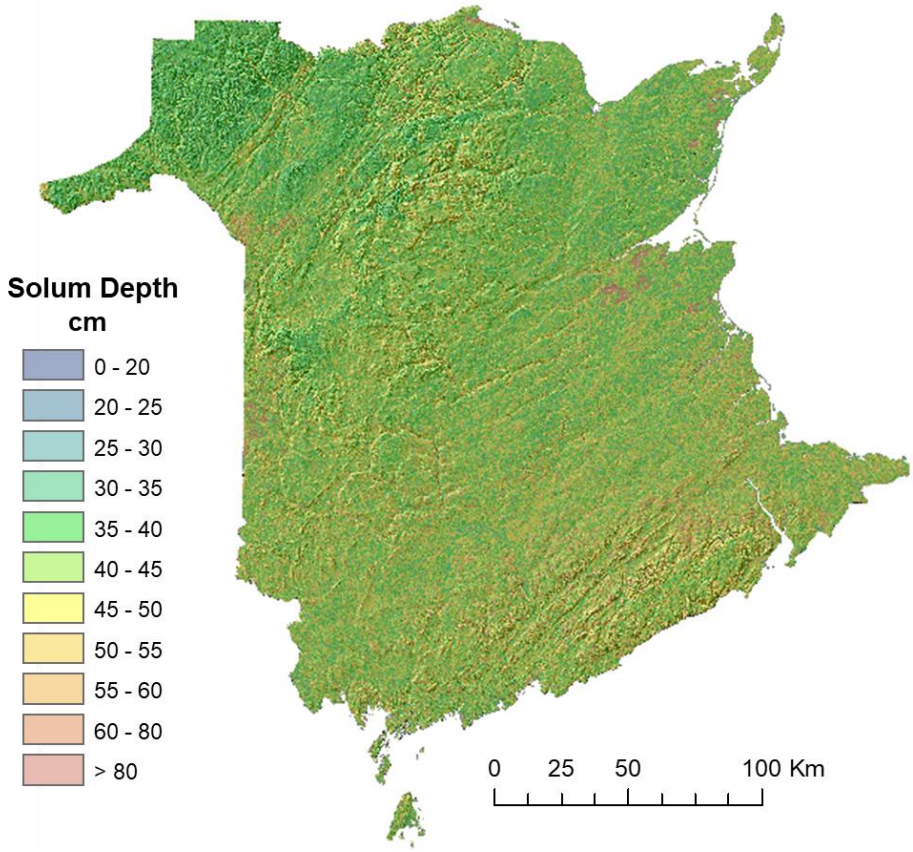

Figure 5.22. Visual example of DSM results representing changing solum depth at 10m resolution for NB, Canada.

The projected soil property maps are in reasonable agreement with the 30% of the amalgamated soil profile database used for model validation. As such, soils on ridges

are seen to be thinner and coarser than soils in floodplains. Soil Dbs also vary as expected by being denser on ridges tops and in areas with high sand content, and lower in areas with high SOM content. Texture is sandier in the upper reaches of floodplains while silt content appears to be higher in higher than in lower lying areas. These trends, however, vary by landform. To confirm this, additional testing will be required, especially at the practical level of field operations, to the effect that the soil properties maps generated by the above approach lead to quantitative improvements in soil and land management. This testing should also involve generating profile data for landforms and drainage conditions that are under-represented.

The precision of a DEM significantly influences the ability to model soil properties, and DEM derivatives require calculation at different resolutions and at different neighborhood search windows. The curvature classification example shows how noise in the DEM skews the output of topographic derivatives. This results in unrealistic relationships between underlying topographic features and soil properties. The same is true for the extent in which these derivatives are produced. Producing topographic derivatives at a small geographic extent shows much different results than the same derivative produced at a much larger extent. The example of MRBFV and MRRTF are provided. Thus, there is a need for a multi-scale approach to DSM in which topographic derivatives are developed at multiple scales. This holds true due to the influence in which topography plays on soil property variation. This is not static but varies with scale. Thus, better representing these relationships at different scales will likely yield better predictive models.

Amalgamated soils information from different sources resulted in a meaningful, harmonized database for application in DSM. With this database, soil properties were able to be modeled at the provincial extent. These individual data sets had not been used for application in DSM prior to this study. Much additional work can be done by further broadening this mapping efforts, e.g., by expanding this work to other soil properties such as pH, CEC, base saturation, and exchangeable base cations. In part, this can be done using the PTFs described in Chapter 3 and evaluating the same with corresponding soil profile-captured data. Also of interest is to determine the extent to which the above approach can be further improved by using province-wide LiDAR DEM coverages generated at 1m resolution.

With growing complexity and availability of data sets representing soil-forming factors, non-parametric DSM techniques are necessary when trying to model soil property variation. The relationship between soil property and underlying soil-forming factors is seldom linear in nature. Thus, advanced statistical models are necessary in trying to understand these relationships. Although linear modeling can explain some variation, machine learning algorithms are likely to capture more of the variation. Adversely, model transferability becomes more difficult with increasing complexity of statistical models and software.

With updated soil property maps, digital soil assessments can be conducted for use in many fields, i.e., soil erosion modeling, habitat suitability mapping, floodplain delineations, forest and crop productivity assessments, and carbon stock assessments.

CHAPTER 6 – SUMMARY, CONTRIBUTIONS, AND RECOMMENDATIONS

6.1. DISSERTATION SUMMARY

This dissertation presented an approach to digital soil mapping of soil physical and chemical properties from numerous data sources for NB, Canada. Soil properties of interest were those of which influence rooting, namely soil drainage, texture, CF content, Db, SOM content, CEC, and water retention at FC and PWP. This was accomplished by:

1. Amalgamating and harmonizing county-based soil surveys into a complete aspatial database focusing on general site characteristics (drainage, stoniness, rockiness, vegetation, and parent material) and soil profile descriptions with lab-measured soil properties on a soil associate basis (Chapter 2),

2. Utilizing the aspatial database to develop PTFs for soil properties with changing depth, namely texture, CF content, Db, SOM content, CEC, and water retention at FC and PWP (Chapter 3),

3. Developing a new provincial DEM at 10m resolution via fusion of open-sourced DEMs (ASTER, SRTM at 30 and 90m resolution, CDED, and original NB DEM) with LiDAR-derived DEM calibration (Chapter 4),

4. Developing a spatial database of soil profiles, compiled from different sources, and re-delineating data sets representing soil forming factors, namely topographic derivatives (from DEM in Chapter 4), surficial geology, and climate data sets, and

5. Comparing soil properties within database to underlying representations of topography, surficial geology and climate. The results allowed for the modeling and mapping of soil properties across the landscape at 10m resolution. PTFs (Chapter 3) can be applied to the spatial soil property maps to predict additional soil properties of CEC and moisture retention at FC and PWP.

6.2. ORIGINAL CONTRIBUTIONS

This dissertation established a framework for digitally mapping soil properties, which influence root development, across the landscape at 10m resolution. Original contributions to this framework involved:

1. Developing an aspatial database from soil surveys with unified terminology and harmonized attributes (Chapter 2),

2. Developing robust PTFs relating soil attributes to one another and to soil parent material while comparing these to published equations as a performance measure (Chapter 3),

3. Creating a semi-automated procedure for fusing open-sourced DEMs across provincial boundaries, resulting in an improved DEM for NB at 10m resolution with substantially decreased vertical errors (Chapter 4),

4. Utilizing the fused DEM from (3) to develop a suite of topographic and hydrological derivatives for NB (Chapter 5),

5. Updating NB-wide geological data sets pertaining to landforms and lithology (mineral hardness, grain size, and rock type) per landform,

6. Generating continuous province-wide data sets for climate 'normals' (annual average temperature, annual average rainfall, and annual average daily maximum high temperature) at 10m resolution by comparing 150 weather stations from NB and Ontario to underlying DEM and changes in latitude and longitude (Chapter 5),

7. Producing a spatial database representing soil profiles within NB with original data retrieved from numerous government agencies (Chapter 5), with 60 data columns to include information about dominant topographic, geologic and climatic soil forming influenced at each spatial location, and

8. Developing data-validated NB-wide spatial models for predicting soil drainage and soil properties (horizon depths, and sand, silt, clay, CF, SOM, and Db at multiple depth intervals (Chapter 5).

6.3. RECOMMENDATIONS

Although this project was completed for NB, similar approaches can be applied to any geographical location depending on access, and quality, of available data. The underlying principles of the outlined framework are applicable elsewhere. The idea was to predict soil properties by way of soil forming factors based on the assumption that relationships between soil properties and underlying soil-forming factors would remain the same, although the combined influence of topography, geology and climate would vary. The lessons learned, and limitations realized in doing this led to the following recommendations for further study:

1. There is a need for a systematic and fully standardized surveying approach to soil sampling and analysis. Developing this dissertation revealed many difficulties associated with compiling soil data and related data layers from various sources. Apart from being labour intensive, a substantial portion of the collected data was not found to be useful for several reasons:
    - too vague,
    - too much missing data,
    - data entries being categorical instead of numerical,
    - analytical protocols differ from report to report,
    - geographic positions unavailable,
    - over–representation of certain soil types, while under-representation of others.

2. There is also a need to address organic soils. This, however, could not be completed because only 3% of the 12,500 samples within the spatial database occurred on very poorly- or poorly-drained sites with bogs, marshes, and swamps excluded as well. Adding wet-area soil data to the database and ensuring equal data representation by drainage class would greatly improve the general applicability of the DSM approach as proposed in Chapter 5.

3. Additionally, there is a need for using LiDAR-derived DEMs for DSM, because LiDAR-generated DEMs reflect the elevation of the terrain at ground level. As indicated in Chapter 4, fusing the non-LiDAR DEMs improved the vertical accuracy of the fused DEM from about ± 20 m to ± 2 m 8 times after LiDAR DEM calibration. This advance generated much-improved DEM-derived hydrographic layers in terms of their 10 m alignments with existing flow channels, river, lake, shoreline and wetland delineations.

4. In the same way, one can expect that LiDAR-based landform delineations will also become more accurate than what is currently available by locations, extent and form. The most easily recognized landforms refer to upland ridges, steep slopes, floodplains, eskers, and drumlins. The delineation of moraines and other localized glacio-fluvial deposits will require the development of specialized landform recognition algorithms, likely in the form of machine-based learning procedures like the approach taken in Chapter 5.

5. With the consistent increases in DEM resolutions, thought must be put into choosing an appropriate resolution for the scope of the project. Ideally, DSM developments should be evaluated based on a multi-scale and multi-window approach in which DEMs of different resolutions, each with topographic derivatives developed at different window sizes, be assessed in a DSM framework. This is because the relationships between soil forming factors and soil property variability change with both scale and study extent. Thus, this needs to be adjusted for when modeling soil properties as depicted in Chapter 5.

6. There needs to be further assessment on the choice of statistical model for DSM because this study only applied one of many machine learning approaches. Alternatively, geostatistical techniques could be applied. A comparison of how different models compare using the data in this project would be very useful.

7. Further improvements and expansions can be made in terms of assessing which PTFs are most accurate for capturing more difficult to measure soil properties. The reliability of existing PTFs also needs to be assessed for each new area/ region under investigation, as demonstrated in Chapter 3.

8. More needs to be done on delineating features across boundaries, whether provincial or soil survey boundaries. (i) end abruptly at adjacent borders in a non-seamless fashion, and (ii) following differing soil terminology conventions. These

differences need to be resolved prior to cross-border DMS applications. Natural processes are continuous in nature and do not abide by these borders. Thus, meaningful relationships between spatially-derived datasets and natural processes may be missed due to the lack of continuity in spatial datasets across borders.

9. Finally, continuous improvements can be made in updating the borders of landform- and lithology-defined soil associations by correctly placing them in the context of topographically-correct digital terrain models. Once delineated, soil associations can then be disaggregated into individual soil types by changing drainage and soil property patterns, i.e., wet vs. dry, shallow vs. deep, fine-textured vs. coarse, and high vs. low CF content. Doing so allows for quick determinations of overall soil and site conditions for land management practices.

6.4. CONCLUSION

The chapters in this dissertation present the requirements to shift from conventional soil mapping techniques to a DSM framework by spatially determining how soil properties vary with local and regional variations in topography, climate, surficial geology, and other soil properties. The development of this framework proceeded from the compilation of aspatial and spatial soil information and the re-delineation of spatial data sets deemed essential for modeling and mapping soil properties. The results outlined have feasible application in land use management, namely forestry due to the variability of soil properties within stands, which, was otherwise omitted from conventional soil

mapping approaches. The information generated and documented will be helpful for allowing further scientific studies and continuing DSM improvements.

LITERATURE CITED

Adhikari, K., Kheir, R.B., Greve, M.B., Bøcher, P.K., Malone, B.P., Minasny, B., McBratney, A.B., Greve, M.H. 2012. High-Resolution 3-D Mapping of Soil Texture in Denmark. *Soil Sci. Soc. Am. J.* 77: 1–17.

Aksoy, E., Özsoy, G., Dirim, M.S. 2009. Soil mapping approach in GIS using landsat satellite imagery and DEM data. *African J. Agric. ….* 4: 1295–1302.

Akumu, C.E., Johnson, J.A., Etheridge, D., Uhlig, P., Woods, M., Pitt, D.G., McMurray, S. 2015. GIS-fuzzy logic based approach in modeling soil texture: Using parts of the Clay Belt and Hornepayne region in Ontario Canada as a case study. *Geoderma*. 239–240: 13–24.

Al-Muqdadi, S.W., Merkel, B.J. 2011. Automated Watershed Evaluation of Flat Terrain. *J. Water Resour. Prot.* 3: 892–903.

Albani, M., Klinkenberg, B. 2003. A Spatial Filter for the Removal of Striping Artifacts in Digital Elevation Models. *Photogramm. Eng. Remote Sens.* 69: 755–765.

Anderson, D.W. 1988. The Effect of Parent Material and Soil Development on Nutrient Cycling in Temperate Ecosystems. *Biogeochemistry*. 5: 71–97.

ASTER GDEM Validation Team 2011. ASTER Global Digital Elevation Model Version 2-Summary of Validation Results.

Azlan, A., Aweng, E.R., Ibrahim, C.O., Noorhaidah, A. 2013. Correlation between Soil Organic Matter, Total Organic Matter and Water Content with Climate and Depths of Soil at Different Land use in Kelantan, Malaysia. *J. Appl. Sci. Environ. Manag.* 16: 353–358.

Ballabio, C., Panagos, P., Monatanarella, L. 2016. Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*. 261: 110–123.

Balland, V., Pollacco, J.A.P., Arp, P.A. 2008. Modeling soil hydraulic properties for a wide range of soil conditions. *Ecol. Modell.* 219: 300–316.

Barka, I., Vladovic, J., Máli, Š. 2011. Landform Classification and Its Application in Predictive Mapping of Soil and Forest Units. *GIS Ostrava*. 11.

Bater, C.W., Coops, N.C. 2009. Evaluating error associated with lidar-derived DEM interpolation. *Comput. Geosci.* 35: 289–300.

Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.D., Goldschmitt, M. 2005. Digital soil mapping using artificial neural networks. *J. Plant Nutr. Soil Sci.* 168: 21–33.

Behrens, T., Zhu, A.X., Schmidt, K., Scholten, T. 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*. 155: 175–185.

Beke, G.J., MacCormick, M.I. 1985. Predicting Volumetric Water Retentions for Subsoil Materials from Colchester County, Nova Scotia. *Can. J. Soil Sci.* 65: 233–236.

Bell, J.C., Cunningham, R.L., Havens, M.W. 1994. Soil Drainage Class Probability Mapping using a Soil-landscape Model. *Soil Sci. Soc. Am. J.* 58: 464–470.

Benites, V.M., Machado, P.L.O.A., Fidalgo, E.C.C., Coelho, M.R., Madari, B.E. 2007. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. *Geoderma*. 139: 90–97.

Birkeland, P. 1999. Soils and Geomorphology, 3rd ed. Oxford University Press, Oxford, England, UK.

Bodgahabadi, M.B., Marta□Nez-Casasnovas, J.A., Salehi, M.H., Mohammadi, J., Borujeni, I.E., Toomanian, N., Gandomkar, A. 2015. Digital soil mapping using artificial neural networks and terrain-related attributes. *Pedosphere*. 25: 580–591.

Boehner, J., Koethe, R., Conrad, O., Gross, J., Ringeler, A., Selige, T. 2002. Soil Regionalisation by Means of Terrain Analysis and Process Parameterisation, in: Micheli, E., Nachtergaele, F., Montanarella, L. (Eds.), Soil Classification 2001. European Soil Bureau, Luxembourg. pp. 213–222.

Boehner, J., Selige, T. 2006. Spatial prediction of soil attributes using terrain analysis and climate regionalisation, in: Boehner, J., McCloy, K.R., Strobl, J. (Eds.), SAGA - Analyses and Modelling Applications. Goettinger Geographische Abhandlungen, Goettingen. pp. 13–28.

Boer, M., Del Barrio, G., Puigdefábregas, J. 1996. Mapping soil depth classes in dry Mediterranean areas using terrain attributes derived from a digital elevation model. *Geoderma*. 72: 99–118.

Bouma, J. 1989. Using Soil Survey Data for Qualitative Land Evaluation. *Adv. Soil Sci.* 9: 177–213.

Brady, N.C. 1990. The Nature and Properties of Soils, Tenth. ed. Macmillan Co., New York, New York, USA.

Breiman, L. 2001. Random forests. *Mach. Learn.* 45: 5–32.

Bui, E.N., Henderson, B.L., Viergever, K. 2006. Knowledge discovery from models of soil properties developed through data mining. *Ecol. Modell.* 191: 431–446.

Callow, J.N., Van Niel, K.P., Boggs, G.S. 2007. How does modifying a DEM to reflect known

hydrology affect subsequent terrain analysis? *J. Hydrol.* 332: 30–39.

Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J. 2013. A methodology for digital soil mapping in poorly-accessible areas. *Geoderma*. 192: 341–353.

Campling, P., Gobin, A., Feyen, J. 2002. Logistic modeling to spatially predict the probability of soil drainage classes. *Soil Sci. Soc. Am. J.* 66: 1390–1401.

Carlisle, B. 2005. Modelling the Spatial Distribution of DEM Error. *Trans. GIS*. 9: 521–540.

Carrara, A., Bitelli, G., Carla, R. 1997. Comparison of techniques for generating digital terrain models from contour lines. *Int. J. Geogr. Inf. Sci.* 11: 451–473.

Carré, F., McBratney, A.B., Mayr, T., Montanarella, L. 2007. Digital soil assessments: Beyond DSM. *Geoderma*. 142: 69–79.

Center for Topgraphic Information Customer Support Group 2000. Canadian Digital Elevation Data Standards and Specifications. Centre for Topographic Information Customer Support Group, NRCan, Quebec, Canada. 2000. Sherbrooke, Quebec.

Chaplot, V., Darboux, F., Bourennane, H., Leguédois, S., Silvera, N., Phachomphon, K. 2006. Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density. *Geomorphology*. 77: 126–141.

Chaudhari, P.R., Ahire, D. V, Ahire, V.D., Chkravarty, M., Maity, S. 2013. Soil Bulk Density as related to Soil Texture , Organic Matter Content and available total Nutrients of Coimbatore Soil. *Int. J. Sci. Resaerch Publ.* 3: 1–8.

Colpitts, M.C., Fahmy, S.H., Macdougall, J.E., Ng, T.T.M., Mclnnis, B.G., Zelazny, V.F. 1995. Forest Soils of New Brunswick. Agriculture and Agri-Food Canada Research Branch, Natural Resources Canada Canadian Forest Service.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Bohner, J. 2018. System for Automated Geoscientific Analyses (SAGA).

Costantini, M., Malvarosa, F., Minati, F., Zappitelli, E., Seifert, F.M. 2005. A Data Fusion Technique for Mosaicking of Different Sources Digital Elevation Models, in: Fringe 2005 Workshop. European Space Agency, Frascati, Italy. p. 5.

Creed, I.F., Trick, C.G., Band, L.E., Morrison, I.K. 2002. Characterizing the Spatial Pattern of Soil Carbon and Nitrogen Pools in the Turkey Lakes Watershed: a Comparison of Regression Techniques. *Water, Air Soil Pollut.* 2: 81–102.

Cunningham, D., Grebby, S., Tansey, K., Gosar, A., Kastelic, V. 2006. Application of airborne LiDAR to mapping seismogenic faults in forested mountainous terrain,

southeastern Alps, Slovenia. *Geophys. Res. Lett.* 33: 1–5.

Desmet, P.J.J., Govers, G. 1996. A GIS Procedure for Automatically Calculating the USLE LS Factor on Topographically Complex Landscape Units. *J. Soil Water Conserv.* 51: 427–433.

Dikau, R. 1988. Entwurf einer geomorphographisch-analytischen Systematik von Reliefeinheiten, 5th ed. Heidelberger Geographische Bausteine, Heidelberg.

Dobos, E., Micheli, E., Baumgardner, M.F., Biehl, L., Helt, T. 2000. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma*. 97: 367–391.

Dobos, E., Montanarella, L., Negre, T., Micheli, E. 2001. A regional scale soil mapping approach using integrated AVHRR and DEM data. *J. Appl. Geod.* 3: 1–13.

Ecosystem Classification Working Group 2007. Our Landscape Heritage: The Story of Ecological Land Classification in New Brunswick, 2nd ed. New Brunswick Department of Natural Resources, Fredericton, New Brunswick.

El-Sammany, M., Abou El-Magd, I.H., Hermas, E.-S.A. 2011. Creating a Digital Elevation Model (DEM) from SPOT 4 Satellite Stereo-Pair Images for Wadi Watiier - Sinai Peninsula, Egypt. *Nile Basin Water Sci. Eng. J.* 4: 49–59.

Environment Canada 2018. Canadian Climate Normals [WWW Document]. URL http://climate.weather.gc.ca/climate_normals/index_e.html (accessed 2.8.18).

Environmental Systems Research Institute 2018. ArcGIS Deskop.

Erdoğan, S. 2010. Modelling the spatial distribution of DEM error with geographically weighted regression: An experimental study. *Comput. Geosci.* 36: 34–43.

Fahmy, S.H., Hann, S.W.R., Jiao, Y. 2010. Soils of New Brunswick: The Second Approximation. Agriculture and Agri-Food Canada. Fredericton, New Brunswick.

Fairfield, J., Leymarie, P. 1991. Drainage networks from grid digital elevation models. *Water Resour. Res.* 27: 709–717.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D. 2007. The Shuttle Radar Topography Mission. Pasadena, California.

Florinsky, I.. 2012. Digital Terrain Analysis in Soil Science and Geology, 1st ed. Elsevier. Elsevier, The Boulevard, Langford Lane, Kidlington, Oxford, United Kingdom.

Florinsky, I.., Eilers, R.., Manning, G.., Fuller, L.. 2002. Prediction of soil properties by digital terrain modelling. *Environ. Model. Softw.* 17: 295–311.

Freeman, G.T. 1991. Calculating catchment area with divergent flow based on a regular grid. *Comput. Geosci.* 17: 413–422.

Fullen, M.A., Jankauskas, B., Jankauskienė, G., Booth, C.A., Slepetiene, A. 2007. Inter-relationships between soil texture and soil organic matter content in eroded Eutric Albeluvisols in Lithuania. *Zemes Ukio Moksl.* 14: 9–18.

Gallant, J.C., Dowling, T.I. 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* 39: n/a–n/a.

Gallant, J.C., Hutchinson, M.F. 1997. Scale dependence in terrain analysis. *Math. Comput. Simul.* 43: 313–321.

Gerrard, A.J. 1981. Soils and Landforms. An Integration of Geomorphology and Pedology. George Allen & Unwin Ltd., London, United Kingdom.

Gessler, P.E., Chadwick, O.A., Chamran, F., Althouse, L., Holmes, K. 2000. Modeling Soil-Landscape and Ecosystem Properties using Terrain Attributes. *Soil Sci. Soc. Am. J.* 64:.

Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J. 1995. Soil-landscape modelling and spatial prediction of soil attributes. *Int. J. Geogr. Inf. Syst.* 9: 421–432.

Grimm, R., Behrens, T. 2010. Uncertainty analysis of sample locations within digital soil mapping approaches. *Geoderma*. 155: 154–163.

Grohmann, C.H., Smith, M.J., Riccomini, C. 2011. Multiscale analysis of topographic surface roughness in the Midland Valley, Scotland. *IEEE Trans. Geosci. Remote Sens.* 49: 1200–1213.

Group, M.S.W. 1981. A soil mapping system for Canada: Revised. Ottawa, Ontario, Canada.

Guertin, R.K., Schuppli, P.A., Sheldrick, B.H., Wires, K.C., Zebchuk, W.D. 1984. Analytical Methods Manual 1984. Ottawa, Ontario, Canada.

Guisan, A., Weiss, S.B., Weiss, A.D. 1999. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol.* 143: 107–122.

Gupta, S.C., Larson, W.E. 1979. Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density. *Water Resour. Res.* 15: 1633–1635.

Happi Mangoua, F., Goïta, K. 2008. A comparison between Canadian Digital Elevation Data

(CDED) and SRTM Data of Mount Carleton in New Brunswick (Canada). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XXXVI: 1423–1430.

Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., Schröder, B. 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma*. 185–186: 37–47.

Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P. 2005. Australia-wide predictions of soil properties using decision trees. *Geoderma*. 124: 383–398.

Hengl, T., Gruber, S., Shrestha, D.P. 2004a. Reduction of errors in digital terrain parameters used in soil-landscape modelling. *Int. J. Appl. Earth Obs. Geoinf.* 5: 97–112.

Hengl, T., Hannes, I.R. (Eds.) 2009. Geomorphometry: Concepts, Software, Applications. Elsevier.

Hengl, T., Heuvelink, G.B.M., Stein, A. 2004b. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*. 120: 75–93.

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.

Heuscher, S. a, Brandt, C.C., Jardine, P.M. 2005. Using Soil Physical and Chemical Properties to Estimate Bulk Density. *Soil Sci. Soc. Am. J*. 69: 1–7.

Hjerdt, K.N., McDonnel, J.J., Seibert, J., Rodhe, A. 2004. A new topographic index to quantify downslope controls on local drainage. *Water Resour. Res.* 40: 1–6.

Hobi, M.L., Ginzler, C. 2012. Accuracy assessment of digital surface models based on WorldView-2 and ADS80 stereo remote sensing data. *Sensors (Basel).* 12: 6347–68.

Hopkinson, C., Hayashi, M., Peddle, D. 2009. Comparing alpine watershed attributes from LiDAR , Photogrammetric , and Contour-based Digital Elevation Models. *Hydrol. Process.* 23: 451–463.

Hugget, R.J. 2007. Fundamentals of Geomorphology. Advances in neonatal care : official journal of the National Association of Neonatal Nurses.

Irvin, B.J., Ventura, S.J., Slater, B.S. 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. *Geoderma*. 77: 137–154.

Iwahashi, J., Pike, R.J. 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology*. 86: 409–440.

Jenny, H. 1941. Factors of Soil Formation: a System of Quantitative Pedology. McGraw-Hill, New York, New York, USA.

Jutras, M., Arp, P.A. 2010. Determining Hydraulic Conductivity from Soil Characteristics with Applications for Modelling Stream Discharge in Forest Catchments. *Hydraluic Cnductivity- Issues, Determ. Appl.* .

Karkee, M., Steward, B.L., Aziz, S.A. 2008. Improving quality of public domain digital elevation models through data fusion. *Biosyst. Eng.* 101: 293–305.

Kerry, R., Goovaerts, P., Rawlins, B.G., Marchant, B.P. 2012. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma*. 170: 347–358.

Keys, K. 2007. Forest Soil Types of Nova Scotia- Identification, Description, and Interpretation. Nova Scotia Department of Natural Resources Manual FOR 2007-2. Datmouth, Nova Scotia.

Kinnell, P.I. 2005. Alternative Approaches for Determining the USLE-M Slope Length Factor for Grid Cells. *Soil Sci. Soc. Am. J.* 69: 674–680.

Klingebiel, A.A., Horvath, E.H., Reybold, W.U., Moore, D.G., Fosnight, E.A., Loveland, T.R. 1988. A Guide for the use of Digital Elevation Model Data for Making Soil Surveys. U.S. Department of Interior, U.S. Geological Survey, U.S. Department of Agriculture. Sioux Falls, South Dakota.

Krogh, L., Breuning-Madsen, H., Greve, M.H. 2000. Cation-Exchange Capacity Pedotransfer Functions for Danish Soils. *Acta Agric. Scand. Sect. B-Soil Plant Sci.* 50: 1–12.

Kuhn, M. 2017. Caret: Classification and Regression Training.

Lal, R. 1979. Physical properties and moisture retention characteristics of some nigerian soils. *Geoderma*. 21: 209–223.

Langmaid, K.K., MacMillan, J.K., Losier, J.G. 1980. Soils of Madawaska County, New Brunswick. Fredericton, New Brunswick.

Lemercier, B., Lacoste, M., Loum, M., Walter, C. 2012. Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach. *Geoderma*. 171–172: 75–84.

Liao, K., Xu, S., Zhu, Q. 2015. Development of ensemble pedotransfer functions for cation exchange capacity of soils of Qingdao in China. *Soil Use Manag.* 31: 483–490.

Luedeling, E., Siebert, S., Buerkert, A. 2007. Filling the voids in the SRTM elevation model

- A TIN-based delta surface approach. *ISPRS J. Photogramm. Remote Sens.* 62: 283–294.

MacMillan, R. a., Pettapiece, W.W., Brierley, J. a. 2005. An expert system for allocating soils to landforms through the application of soil survey tacit knowledge. *Can. J. Soil Sci.* 85: 103–112.

Mansuy, N., Thiffault, E., Paré, D., Bernier, P., Guindon, L., Villemaire, P., Poirier, V., Beaudoin, A. 2014. Digital mapping of soil properties in Canadian managed forests at 250m of resolution using the k-nearest neighbor method. *Geoderma*. 235–236: 59–73.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B. 2003. On digital soil mapping. *Geoderma*. 117: 3–52.

McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W. 2002. From pedotransfer functions to soil inference systems. *Geoderma*. 109: 41–73.

McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A.A., Dunbar, M.S., Shatar, T.M. 2000. An overview of pedometric techniques for use in soil survey. *Geoderma*. 97: 293–327.

McKeague, J.A. 1978. Manual on soil sampling and methods of analysis.

McKeague, J.A., Stobbe, P.C. 1978. History of Soil Survey in Canada 1914- 1975, 11th ed. Research Branch, Canada Department of Agriculture, Ottawa, Ontario, Canada.

McKenzie, N.J., Ryan, P.J. 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma*. 89: 67–94.

Meyer, W.L., Arp, P. a. 1994. Exchangeable cations and cation exchange capacity of forest soil samples: Effect s of drying, storage, and horizon. *Can. J. Soil Sci.* 421–429.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A. 1993. Soil Attribute Prediction Using Terrain Analysis. *Soil Sci. Soc. Am. J.* 57: 443–452.

Moore, I.D., Grayson, R.B., Ladson, A.R. 1991. Digital Terrain Modeling: A Review of Hydrological, Geomorphological, and Biological Applications. *Hydrol. Process.* 5: 3–30.

Mora-Vallejo, A., Claessens, L., Stoorvogel, J., Heuvelink, G.B.M. 2008. Small scale digital soil mapping in Southeastern Kenya. *Catena*. 76: 44–53.

Moran, C.J., Bui, E.N. 2002. Spatial data mining for enhanced soil map modelling. *Int. J. Geogr. Inf. Sci.* 16: 533–549.

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. *Environ. Monit. Assess.* 188: 1–13.

Mukherjee, S., Garg, R.D., Mukherjee, S. 2011. Effect of systematic error on DEM and its derived attributes: A case study on Dehradum area using Cartosat-1 stereo data. *Indian J. Landsc. Syst. Ecol. Stud.* 34: 45–58.

Mukherjee, S., Joshi, P.K., Mukherjee, S., Ghosh, A., Garg, R.D., Mukhopadhyay, A. 2013. Evaluation of vertical accuracy of open source Digital Elevation Model (DEM). *Int. J. Appl. Earth Obs. Geoinf.* 21: 205–217.

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R. 2011. The use of remote sensing in soil and terrain mapping - A review. *Geoderma*. 162: 1–19.

Murphy, P.N.C., Ogilvie, J., Arp, P. 2009. Topographic modelling of soil moisture conditions: a comparison and verification of two models. *Eur. J. Soil Sci.* 60: 94–109.

Murphy, P.N.C., Ogilvie, J., Connor, K., Arp, P.A. 2007. Mapping wetlands: A comparison of two different approaches for New Brunswick, Canada. *Wetlands*. 27: 846–854.

Murphy, P.N.C., Ogilvie, J., Meng, F., Arp, P.A. 2008. Stream Network Modelling using Lidar and Phtogrammetric Digital Elevation Models: a Comparison and Field Verification. *Hydrol. Process.* 22: 1747–1754.

Murphy, P.N.C., Ogilvie, J., Meng, F., White, B., Bhatti, J.S., Arp, P.A. 2011. Modelling and mapping topographic variations in forest soils at high resolution: A case study. *Ecol. Modell.* 222: 2314–2332.

Murphy, P.N.C.P.N.C., Ogilvie, J., Meng, F.R., White, B., Bhatti, J.S.J.S., Arp, P.A.P.A. 2011. Modelling and mapping topographic variations in forest soils at high resolution: A case study. *Ecol. Modell.* 222: 2314–2332.

N.B. Department of Energy and Resource Development 2012. Soil Associations of New Brunswick.

N.B. Department of Energy and Resource Development 2016. Forest Soil Associations of New Brunswick.

N.B. Department of Energy and Resource Development 2018. N.B. Mineral Report of Work Search Query.

Nanko, K., Ugawa, S., Hashimoto, S., Imaya, A., Kobayashi, M., Sakai, H., Ishizuka, S., Miura, S., Tanaka, N., Takahashi, M., Kaneko, S. 2014. A pedotransfer function for estimating bulk density of forest soil in Japan affected by volcanic ash. *Geoderma*. 213: 36–45.

New Brunswick Department of Energy and Resource Development 2018. New Brunswick Granular Aggregate Database [WWW Document].

Nobre, A.D., Cuartas, L.A., Hodnett, M., Renno, C.D., Rodrigues, G., Silveira, A., Waterloo, M., Saleska, S. 2011. Height Above the Nearest Drainage - a hydrologically relevant new terrain model. *J. Hydrol.* 404: 13–29.

O'Callaghan, J.F., Mark, D.M. 1984. The extraction of drainage networks from digital elevation data. *Comput. Vision, Graph. Image Process.* 28: 323–344.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J. 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*. 63: 197–214.

Odgers, N.P., McBratney, A.B., Minasny, B. 2011. Bottom-up digital soil mapping. II. Soil series classes. *Geoderma*. 163: 30–37.

Odgers, N.P., Sun, W., McBratney, A.B., Minasny, B., Clifford, D. 2014. Disaggregating and harmonising soil map units through resampled classification trees. *Geoderma*. 214–215: 91–100.

Olaya, V. 2006. Basic Land-Surface Parameters, in: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry: Concepts, Software, Applications. Developments in Soil Science. Elsevier. pp. 141–169.

Olorunfemi, I., Fasinmirin, J., Ojo, A. 2016. Modeling cation exchange capacity and soil water holding capacity from basic soil properties. *Eurasian J. Soil Sci.* 5: 266.

Oosterveld, M., Chang, C. 1980. Empirical relations between laboratory determinations of soil texture and moisture retention. *Can. Agric. engeneering*. 21: 149–151.

Ostovari, Y., Asgari, K., Cornelis, W. 2015. Performance Evaluation of Pedotransfer Functions to Predict Field Capacity and Permanent Wilting Point Using UNSODA and HYPRES Datasets. *Arid L. Res. Manag.* 29: 383–398.

Papasaika, H., Poli, D., Baltsavias, E. 2008. A framework for the fusion of digital elevation models. *Int Arch Photogramm Remote Sens*. 37: 811–818.

Park, S.., McSweeney, K., Lowery, B. 2001. Identification of the spatial distribution of soils using a process-based terrain characterization. *Geoderma*. 103: 249–272.

Parr, J.F., Hornick, S.B., Papendick, R.I. 1992. Soil Quality: The Foundation of a Sustainable Agriculture.

Pegler, K.H. 1999. An examination of alternative compensation methods for the removal of the Ridging Effect from Digital Terrain Model Data Files. Thesis Diss. University of

New Brunswick.

Pennock, D.J. 2003. Multi-site assessment of cultivation-induced soil change using revised landform segmentation procedures. *Can. J. Soil Sci.* 83: 565–580.

Petry, F.E., Robinson, V.B., Cobb, M.A. (Eds.) 2005. Fuzzy Modeling with Spatial Information for Geographic Problems. Springer-Verlag Berlin, Heidelberg.

Pitty, A.F. 1979. Geography and Soil Properties. Methuen, London.

Poggio, L., Gimona, A., Brewer, M.J. 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. *Geoderma*. 209–210: 1–14.

Pollacco, J.A.P. 2008. A generally applicable pedotransfer function that estimates field capacity and permanent wilting point from soil texture and bulk density. *Can. J. Soil Sci.* 88: 761–774.

Pribyl, D.W. 2010. A critical review of the conventional SOC to SOM conversion factor. *Geoderma*. 156: 75–83.

Prima, O.D.A., Echigo, A., Yokoyama, R., Yoshida, T. 2006. Supervised landform classification of Northeast Honshu from DEM-derived thematic maps. *Geomorphology*. 78: 373–386.

Pronk, A.G., Ruitenberg, A.A. 1991. Applications of surficial mapping to forest management in New Brunswick. *Atl. Geol.* 27: 209–212.

Quinn, P.F., Beven, K.J., Chevallier, P., Planchon, O. 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrol. Process.* 5: 59–79.

R Core Team 2018. R: A language and environment for statistical computing.

Rampton, V.N. 1984. Generalized Surficial Geology Map of New Brunswick.

Rawls, W.J., Brakensiek, D.L. 1982. Estimating soil water retention from soil properties. *J. Irrig. Drain. Div.* 108: 166–171.

Riley, S.J., DeGloria, S.D., Elliot, R. 1999. A Terrain Ruggedness Index that Qauntifies Topographic Heterogeneity. Intermt. J. Sci.

Robinson, N., Regetz, J., Guralnick, R.P. 2014. EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *ISPRS J. Photogramm. Remote Sens.* 87: 57–67.

Román Dobarco, M., Orton, T.G., Arrouays, D., Lemercier, B., Paroissien, J.B., Walter, C., Saby, N.P.A. 2016. Prediction of soil texture using descriptive statistics and area-to-point kriging in Region Centre (France). *Geoderma Reg.* 7: 279–298.

Romano, N., Palladino, M. 2002. Prediction of soil water retention using soil physical data and terrain attributes. *J. Hydrol.* 265: 56–75.

Salter, P.J., Williams, J.B. 1965a. The influence of texture on the moisture characteristics of soils. I: A critical comparison of techniques for determining the available-water capacity and moisture characteristic curve of a soil. *J. Soil Sci.* 16: 310–317.

Salter, P.J., Williams, J.B. 1965b. The influence of texture on the moisture characteristics of soils II. Available-water capacity and moisture release characteristics. *J. Soil Sci.* 16: 310–317.

Santillan, J.R., Makinano-Santillan, M. 2016. Vertical accuracy assessment of 30-M resolution ALOS, ASTER, and SRTM global DEMS over Northeastern Mindanao, Philippines. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* XLI-B4: 149–156.

Saunders, W.K., Maidment, David, R. 1996. a Gis Assessment of Nonpoint Source Pollution in the San Antonio-Nueces Coastal Basin. Online. Austin, Texas.

Schindler, K., Papasaika-hanusch, H., Schütz, S., Baltsavias, E. 2011. Improving Wide-Area DEMs Through Data Fusion – Chances and Limits.

Schmidt, K., Behrens, T., Scholten, T. 2008. Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. *Geoderma*. 146: 138–146.

Scull, P., Franklin, J., Chadwick, O. a. 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecol. Modell.* 181: 1–15.

Service New Brunswick 2001. User Guide to the Digital Topographic Data Base 1998 (DTDB98) of New Brunswick. New Brunswick.

Simbahan, G.C., Dobermann, A., Goovaerts, P., Ping, J., Haddix, M.L. 2006. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. *Geoderma*. 132: 471–489.

Simonson, R.W. 1959. Outline of a Generalized Theory of Soil Genesis. *Soil Sci. Soc. Am. J.* 23: 152–156.

Skidmore, A.K. 1989. A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. *Int. J. Geogr. Inf. Syst.* 3: 323–334.

Smith, M.J. 2010. Digital elevation models for research: UK datasets, copyright and

derived products, in: Flemming, C., Marsh, S.H., Giles, J.R.A. (Eds.), Elevation Models for Geoscience. Geological Society of London, London. pp. 129–133.

Smith, M.P., Zhu, A.-X., Burt, J.E., Stiles, C. 2006. The effects of DEM resolution and neighborhood size on digital soil survey. *Geoderma*. 137: 58–69.

Soil Classification Working Group 1998. The Canadian System of Soil Classification, 3rd ed. National Research Council of Canada, Ottawa, Ontario, Canada.

Soil Landscapes of Canada Working Group 2010. Soil Landscapes of Canada Version 3.2.

Soille, P., Vogt, J., Colombo, R. 2003. Carving and Adaptive Drainage Enforcement of Grid Digital Elevation Models. *Water Resour. Res.* 39: 13.

Southorn, N. 2003. The role of soil quality criteria in assessing farm performace, in: 14th International Farm Management Congress. Perth.

Sun, G., Ranson, K.., Kharuk, V.., Kovacs, K. 2003. Validation of surface height from shuttle radar topography mission using shuttle laser altimeter. *Remote Sens. Environ.* 88: 401–411.

Syers, J.K., Campbell, A.S., Walker, T.W. 1970. Contribution of organic carbon and clay to cation exchange capacity in a chronosequence of sandy soils. *Plant Soil*. 33: 104–112.

Tarboton, G. 1997. A New Method for the Determination of Flow Directions and Upslope Areas in Grid Digital Elevation Models. *Water Resour. Res.* 33: 309–319.

Taylor, J. a., Jacob, F., Galleguillos, M., Prévot, L., Guix, N., Lagacherie, P. 2013. The utility of remotely-sensed vegetative and terrain covariates at different spatial resolutions in modelling soil and watertable depth (for digital soil mapping). *Geoderma*. 193–194: 83–93.

Terra, J.A., Shaw, J.N., Reeves, D.W., Raper, R.L., van Santen, E., Mask, P.L. 2004. Soil Carbon Relationships with Terrain Attributes, Electrical Conductivity, and Soil Survey in a Coastal Plain Landscape. *Soil Sci.* 169: 819–831.

Thompson, J. a., Pena-Yewtukhiw, E.M., Grove, J.H. 2006. Soil–landscape modeling across a physiographic region: Topographic patterns and model transportability. *Geoderma*. 133: 57–70.

Tran, T.A., Raghavan, V., Masumoto, S., Vinayaraj, P., Yonezawa, G. 2014. A geomorphology-based approach for digital elevation model fusion - Case study in Danang city, Vietnam. *Earth Surf. Dyn.* 2: 403–417.

Valladares, G.S., Hott, M.C. 2008. The Use of GIS and Digital Elevation Model in Digital Soil Mapping – A Case Study from Sao Paulo , Brazil, in: Hartemink, A.E. (Ed.), Digital Soil

Mapping with Limited Data. Springer Science and Business Media B.V., Campinas, Brazil. pp. 349–356.

Vaysse, K., Lagacherie, P. 2015. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Reg.* 4: 20–30.

Wang, P. 1998. Applying two dimensional kalman filtering for digital terrain modelling, in: Fritsch, D., Englich, M., Sester, M. (Eds.), ISPRS Commission IV Symposium on GIS. Stuttgart, Germany. pp. 649–656.

Weil, R.R., Brady, N.C. 2017. The Nature and Properties of Soils, Fifteenth. ed. Pearson Education, United States.

Whitbread, A.M. 1992. Soil Organic Matter: Its Fractionation and Role in Soil Structure 124–130.

Wilson, J.P. 2012. Digital terrain modeling. *Geomorphology*. 137: 107–121.

Wilson, J.P., Gallant, J.C. 2000. Primary Topographic Attributes, in: Wilson, J.P., Gallant, J.C. (Eds.), Terrain Analysis: Principles and Applications. John Wiley & Sons Inc. pp. 51–85.

Wilson, J.P., Gallant, J.C. n.d. Terrain Analysis- Principles and Applications. John Wiley & Sons Inc.

Wolock, D.M., Price, C. V. 1994. Effects of digital elevation model map scale and data resolution on a topography-based watershed model. *Water Resour. Res.* 30: 3041–3052.

Working Group on Soil Survey Data 1982. The Canada Soil Information System Manual for describing soils in the field 1982 Revised. Ottawa, Ontario, Canada.

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. *J. Hydrol.* 251: 123–150.

Wu, S., Li, J., Huang, G.H. 2008. A study on DEM-derived primary topographic attributes for hydrologic applications: Sensitivity to elevation data resolution. *Appl. Geogr.* 28: 210–223.

Wu, W., Fan, Y., Wang, Z., Liu, H. 2008. Assessing effects of digital elevation model resolutions on soil–landscape correlations in a hilly area. *Agric. Ecosyst. Environ.* 126: 209–216.

Yokoyama, R., Shirasawa, M. 2002. Visualizing topography by openness: A new

application of image processing to digital elevation models. *Photogramm. Eng. Remote Sensing*. 68: 251–266.

Yue, T.-X., Du, Z.-P., Song, D.-J., Gong, Y. 2007. A new method of surface modeling and its application to DEM construction. *Geomorphology*. 91: 161–172.

Zevenbergen, L.W., Thorne, C.R. 1987. Quantitative analysis of land surface topography. *Earth Surf. Process. Landforms*. 12: 47–56.

Zhang, H., Huang, G.H., Wang, D. 2013. Establishment of channel networks in a digital elevation model of the prairie region through hydrological correction and geomorphological assessment. *Can. Water Resour. J.* 38: 12–23.

Zhao, Z., Ashraf, M.I., Keys, K.S., Meng, F. 2013a. Prediction of soil nutrient regime based on a model of DEM-generated clay content for the province of Nova Scotia, Canada. *Can. J. Soil Sci.* .

Zhao, Z., Chow, T.L., Rees, H.W., Yang, Q., Xing, Z., Meng, F.R. 2009. Predict soil texture distributions using an artificial neural network model. *Comput. Electron. Agric.* 65: 36–48.

Zhao, Z., MacLean, D.A., Bourque, C.P.-A., Swift, E., Meng, F.-R. 2013b. Generation of Soil Drainage Equations from an Artificial Neural Network-analysis Approach. *Can. J. Soil Sci.* 93: 329–342.

Zhao, Z., Meng, F.-R., Yang, Q., Zhu, H. 2018. Using Artificial Neural Networks to Produce High-Resolution Soil Property Maps, in: Advanced Applications for Artificial Neural Networks. INTECH.

Zhao, Z., Yang, Q., Benoy, G., Chow, T.L., Xing, Z., Rees, H.W., Meng, F.-R. 2010. Using artificial neural network models to produce soil organic carbon content distribution maps across landscapes. *Can. J. Soil Sci.* 90: 75–87.

Zhu, A.-X. 2000. Mapping soil landscape as spatial continua: The neural network approach. *Water Resour. Res.* 36: 663–677.

Zhu, A.-X., Burt, J.E., Smith, M., Wang, R., Gao, J. 2008. The Impact of Neighbourhood Size on Terrain Derivatives and Digital Soil Mapping, in: Zhou, Q., Lees, B., Tang, G. -a. (Eds.), Advances in Digital Terrain Analysis. Springer. p. 462.

Zhu, A.-X., Mackay, S.D. 2001. Effects of spatial detail of soil information on watershed modeling. *J. Hydrol.* 248: 54–77.

Ziadat, F.M. 2010. Prediction of Soil Depth from Digital Terrain Data by Integrating Statistical and Visual Approaches. *Pedosphere*. 20: 361–367.

Zushi, K. 2006. Spatial distribution of soil carbon and nitrogen storage and forest productivity in a watershed planted to Japanese cedar (Cryptomeria japonica D. Don). *J. For. Res.* 11: 351–358.

APPENDICES

APPENDIX I. OVERVIEW OF NEW BRUNSWICK SOIL ASSOCIATIONS

Landform	Lithology	Soil Association	Soil Associate		
			W-R	I-MW	VP-P
Residual	Weakly calcareous Sandstone and shale	Undine	Undine	-	-
	Sandstone, gritstone, shale	Riley Brook	Riley Brook	-	-
		Cornhill	Cornhill	-	-
	Granite, gneiss, basalt, felsite	Big Bald Mountain	Big Bald Mountain	-	-
Residual & Colluvium	Strongly metamorphosed slate, quartzite, volcanics	Serpentine	Serpentine	Jenkins	Adder
Ablation	Sandstone and Conglomerate (calcareous if un-weathered)	Anagance	Anagance	Dusinane	Dusinane
		Jeffries Corner	Jeffries Corner	-	-
		Aulac	Aulac	Tidnish	Tidnish
	Granite, quartzite, gneiss, argillite, volcanics, some sandstone	Irving	Irving	Goodfellow	Halls Brook
		Juniper	Juniper	Jummet Brook	McKiel
	Calcareous shale and slate with/without limestone, argillite	Caribou	Caribou	Carlingford	Washburn
		Thibault	Thibault	Guercheville	Lauzier
	Sandstone and quartzite, some argillite, slate, shale, schist	Monquart	Monquart	-	-
	Sandstone	Big Hole	Big Hole	Beaver Lake	-
		Fair Isle	Fair Isle	Black Brook	-
		Sunbury	Sunbury	Hoyt	Cork
	Highly calcareous shale, quartzite, argillite, sandstone	Jardine	Jardine	Nickel Mill	Five Fingers
	Sandstone, shale, mudstone	Becaguimec	Becaguimec	Snyder	Snyder
	Metamorphosed rhyolite, andesite, schist, slate, granite	Jacquet River	Jacquet River	-	-
Ablation/ Residual	Calcareous sandstone and shale	Harquail	Harquail	-	-
	Calcareous Shale	Erb Settlement	Erb Settlement	-	-
	Metamorphosed non- to weakly-calcareous slate, quartzite, argillite, sandstone	Glassville	Glassville	Temiscouata	Foreston
	Metamorphosed rhyolite, andesite, schist, slate, granite	Lomond	Lomond	-	-
	Sandstone and conglomerate (calcareous if unweathered)	Parleeville	Parleeville	Midland	Midland
		Queenville	Queenville	Deed	Deed

	Sandstone and quartzite, some argillite, slate, shale, schist	Quisbis	Quisbis	Dube	Big Spring
	Sandstone, gritstone, shale	Tobique	Tobique	-	-
	Slate and argillite and sandstone and schist	Boston Brook	Boston Brook	Skin Gulch	Yellow Brook
Alluvium	Undifferentiated	Interval	Interval	Waasis	East Canaan
		Sussex	Sussex	Hampton	-
	-	Bottomland	Bottomland	-	-
Ancient Alluvium	Undifferentiated	Flemming	Flemming	Martial	Kelly
Alluvium & Glaciofluvial	Calcareous shale, slate, quartzite, argillite	Maliseet	Maliseet	Wapske	Wapske
	Strongly metamorphosed slate, quartzite, volcanics	Benedict	Benedict	-	-
Basal	Calcareous sandstone and quartzite, some argillite, shale	Siegas	Siegas	Salmon	Bourgoin
	Calcareous shale	Kedgwick	Kedgwick	-	-
		Saltspring	Saltspring	Byrns	Byrns
	Granite, gneiss, basalt, felsite	Parry	Parry	Midway	Midway
	Granite, quartzite, gneiss, argillite, volcanics, some sandstone	Rogersville	Rogersville	Acadieville	Rosaireville
		Tuadook	Tuadook	Redstone	Lewis
	Granite, some quartzite, sandstone	Pinder	Pinder	Coronary	McAdam
		Catamaran	Catamaram	-	-
	Sandstone and conglomerate (calcareous if un-weathered)	Wakefield	Wakefield	-	-
	Sandstone and quartzite, some argillite, slate, shale, schist	Holmesville	Holmesville	Johnville	Poitras
		Violette		Violette	Violette
	Sandstone, conglomerate, mudstone	Parsons Brook	Parsons Brook	-	-
	Sandstone, conglomerate, shale	Petitcodiac	Petitcodiac	Kings	Kings
		Salisbury	Salisbury	Harewood	Hicksville
		Salem	Salem	-	-
	Sandstone, shale, mudstone	Dorchester	Dorchester	-	-
		Knightville	Knightville	Byrns	Byrns
		Tracy	Tracy	Wirral	Rooth
	Strongly metamorphosed slate, quartzite, volcanics	Long Lake	Long Lake	Blue Mountain	Colter Mountain
	weakly calcareous Sandstone and shale	Shemogue	Shemogue	-	-
		Stony Brook	Stony Brook (Queens)	Blackville	Cambridge (Kings)
		Tormentine	Tormentine	Tidnish	Tidnish

	Weakly calcareous shale, mudstone	Kingsclear	Kingsclear	Plaster Rock	Nackawic
	Weakly- to calcareous shale, slate, quartzite, some sandstone	Carleton	Carleton	Canterbury	Canterbury
		Green Road	Green Road	-	-
		Lower Ridge	Lower Ridge	-	-
	Mafic volcanics, gabbro, diorite	Kingston	Kingston	Deed	Deed
		Tetagouche	Tetagouche	-	-
	Metamorphosed rhyolite, andesite, schist, slate, granite	Popple Depot	Popple Depot	-	-
Basal/ Residual	Metamorphosed non- to weakly-calcareous slate, quartzite, argillite, sandstone	Green River	Green River	-	-
Colluvium & Water re-worked till	Granite, gneiss, basalt, felsite	Clearwater	Clearwater	Ogilvie Lake	Yellow Lake
	Metamorphosed non- to weakly-calcareous slate, quartzite, argillite, sandstone	McGee	McGee	Nason	Trafton
	Non- to weakly-calcareous sandstone, shale, quartzite	Victoria	Victoria	McCluskey	Cote
	Strongly metamorphosed slate, quartzite, volcanics	Britt Brook	Britt Brook	Babbit Brook	Portage Lake
Glaciofluvial	Granite	Gagetown	Gagetown	Geary	Penobsquis
		Island Lake	Island Lake	Pennfield	Pennfield
	Metamorphosed non- to weakly-calcareous slate, quartzite, argillite, sandstone	Grand Falls	Grand Falls	Sirois	Cyr
	Sandstone	Kennebecasis	Kennebecasis	Quispamsis	Nevers Road
		Lord and Foy	Lord and Foy	-	-
	Sandstone and conglomerate (some quartz)	Gulquac	Gulquac	-	-
	Sandstone, some granite, quartzite, gneiss	Guimond River	Guimond River	St. Oliver	St. Theodule
	Weakly- to calcareous shale, slate, quartzite, some sandstone	Muniac	Muniac	Ennishore	Cyr
	-	Bransfield	Bransfield	-	-
	-	Chockpish	Chockpish	-	-
Glaciofluvial & Marine	Sandstone	Aldouane	Aldouane	Marquant	
		Baie du Vin	Baie du Vin	Napan	Fontaine
		Kouchibouguac	Kouchibouguac	Potters Mill	Vautour
		Caissie	Caissie	Robichaud	

		Richibucto	Richibucto	Cap Limiere	Nevers Road
		Riverbank	Riverbank	Oromocto	Nevers Road
Glaciolacustrine	Undifferentiated/ weakly calcareous clay	Bellefleur	Bellefleur	-	-
Glaciomarine	Calcareous sandstone and shale	Tracadie	Tracadie	Bouleau	Sheila
	Non-calcareous shale, some sandstone	Mount Hope	Mount Hope	Boland	Cambridge
	Sandstone	Babineau	Babineau	-	-
		Escuminac	Escuminac	Baie St. Anne	-
		Galloway	Galloway	Smelt Brook	Briggs Brook
Glaciomarine/ Basal	Sandstone/ clay from calcareous shale	Upper Caraquet	Upper Caraquet	Little Shippegan	Shediac
	Sandstone/Sandstone and shale or siltstone	Barrieau	Barrieau	Cote D'or	Shediac
		Buctouche	Buctouche	Michaud	Neguac
		Bretagneville	Bretagneville	-	-
		St. Charles	St. Charles	-	-
Glaciomarine/ Marine	Sandstone/ clay from calcareous shale	Caraquet	Caraquet	Middle Caraquet	Neguac
Lacustrine	Clay	Fundy	-	Fundy	Canobie
Marine/ Basal	Undifferentiated	Research Station	Research Station	Baker Brook	-
Marine	Undifferentiated	Belledune	Belledune	-	-
		Blackland	Blackland	-	-
		Acadia	Acadia	Acadia	Acadia
Ablation/ Basal	Sandstone/Sandstone and shale or siltstone	Reece	Reece	Chipman	Pangburn
		Harcourt	Harcourt	Coal Branch	Grangeville
Organic	Undifferentiated	Lavilette	-	-	Lavilette
		Acadie Siding	-	-	Acadie Siding
		Bog	-	-	Bog
		Fen	-	-	Fen
		SDTWp	-	-	SDTWp
		Legaceville	-	-	Legaceville
		Chelmsford	-	-	Chelmsford
		St. Quentin	-	-	St. Quentin

APPENDIX II. OVERVIEW OF SPATIAL REPRESENTATION OF SOIL ASSOCIATIONS WITH OVERALL PROFILE REPRESENTATION FOR EACH SOIL ASSOCIATE WITHIN ASPATIAL DATABASE.

Soil Association	% Spatial Coverage	Soil Associates	# of Profiles	% Representation
Acadia	3.83	Acadia (w)	0	0.36
		Acadia (I)	2	
		Acadia (P)	0	
Aldouane	-	Aldouane	3	0.71
		Marquant	1	
Anagance	-	Anagance	2	0.36
		Dusinane (I)	0	
		Dusinane (P)	0	
Aulac	-	Aulac	0	-
		Tidnish (I)	0	
		Tidnish (P)	0	
Babineau	-	Babineau	0	-
Baie du Vin	-	Baie du Vin	5	1.07
		Napan	0	
		Fontaine	1	
Barrieau	1.31	Barrieau	6	2.50
		Cote D'or	5	
		Shediac	3	
Becaguimec	0.18	Becaguimec	0	-
		Snyder (I)	0	
		Snyder (P)	0	
Belledune	-	Belledune	0	-
Bellefleur	-	Bellefleur	3	1.07
		Rob	1	
		St. Amand	2	
Benedict	-	Benedict	1	0.18
Big Bald Mountain	0.67	Big Bald Mountain	0	-
Big Hole	-	Big Hole	3	0.71
		Beaver Lake (I)	1	
Blackland	-	Blackland	0	-
Boston Brook	-	Boston Brook	7	1.96
		Skin Gulch	3	
		Yellow Brook	1	
Bottomland	-	Bottomland	1	0.18
Bransfield	-	Bransfield	1	0.18
Bretagneville	-	Bretagneville	1	0.18
Britt Brook	3.28	Britt Brook	2	0.71
		Babbit Brook	1	
		Portage Lake	1	
Buctouche	1.31	Buctouche	3	0.54
		Michaud	0	
		Neguac	0	

Caissie	-	Caissie	0	-
		Robichaud	0	
Caraquet	-	Caraquet	0	0.36
		Middle Caraquet	2	
		Neguac	0	
Caribou	2.74	Caribou	11	3.75
		Carlingford	6	
		Washburn	4	
Carleton	3.33	Carleton	1	1.07
		Canterbury (I)	3	
		Canterbury (P)	2	
Catamaran	1.62	Catamaran	0	-
Chockpish	-	Chockpish	1	0.18
Clearwater	-	Clearwater	3	
		Ogilvie Lake	1	
		Yellow Lake	1	
Cornhill	0.33	Cornhill	1	0.18
Dorchester	-	Dorchester	1	0.18
Erb Settlement	0.12	Erb Settlement	1	0.18
Escuminac	-	Escuminac	0	-
		Baie St. Anne	0	
Fair Isle	0.88	Fair Isle	1	0.36
		Black Brook	1	
Flemming	-	Flemming	3	1.79
		Martial	4	
		Kelly	3	
Fundy	-	Fundy (I)	2	0.54
		Canobie (P)	1	
Gagetown	1.17	Gagetown	8	1.79
		Geary	1	
		Penobsquis	1	
Galloway	-	Galloway	3	0.71
		Smelt Brook	1	
		Briggs Brook	0	
Glassville	2.68	Glassville	9	2.14
		Temiscouata	1	
		Foreston	2	
Grand Falls	1.00	Grand Falls	6	1.61
		Sirois	2	
		Cyr	1	
Green River	-	Green River	1	0.18
Green Road	-	Green Road	2	0.36
Guimond River	-	Guimond River	3	0.54
		St. Oliver	0	
		St. Theodule	0	
Gulquac	-	Gulquac	2	0.36
Harcourt	7.33	Harcourt	5	2.68
		Coal Branch	6	
		Grangeville	4	

Harquail	-	Harquail	3	0.54
Holmesville	4.47	Holmesville	18	7.50
		Johnville	12	
		Poitras	12	
Interval	0.55	Interval	4	1.25
		Waasis	2	
		East Canaan	1	
Irving	1.67	Irving	1	0.54
		Goodfellow	1	
		Halls Brook	1	
Island Lake	-	Island Lake	1	0.18
		Pennfield (I)	0	
		Pennfield (P)	0	
Jacquet River	1.37	Jacquet River	0	-
Jardine	-	Jardine	3	1.43
		Nickel Mill	2	
		Five Fingers	3	
Jeffries Corner	-	Jeffries Corner	1	0.18
Juniper	3.37	Juniper	3	1.07
		Jummet Brook	1	
		McKiel	2	
Kedgwick	1.28	Kedgwick	4	0.71
Kennebecasis	0.33	Kennebecasis	2	0.36
		Quispamsis	0	
		Nevers Road	0	
Kingsclear	-	Kingsclear	4	1.61
		Plaster Rock	2	
		Nackawic	3	
Kingston	0.87	Kingston	0	-
		Deed (I)	0	
		Deed (P)	0	
Knightville	-	Knightville	1	0.36
		Byrns (I)	1	
		Byrns (P)	0	
Kouchibouguac	-	Kouchibouguac	2	0.54
		Potters Mill	0	
		Vautour	1	
Lomond	2.30	Lomond	2	0.36
Long Lake	4.63	Long Lake	2	0.71
		Blue Mountain	1	
		Colter Mountain	1	
Lord and Foy	-	Lord and Foy	1	0.18
Lower Ridge	-	Lower Ridge	1	0.18
Maliseet	-	Maliseet	6	1.43
		Wapske (I)	1	
		Wapske (P)	1	
McGee	4.64	McGee	6	1.96
		Nason	3	
		Trafton	2	

Monquart	-	Monquart	8	1.43
Mount Hope	-	Mount Hope	5	1.79
		Boland	1	
		Cambridge	4	
Muniac	0.40	Muniac	4	1.07
		Ennishore	2	
		Cyr	0	
Organic	11.38	Acadie Siding	3	6.43
		Lavilette	6	
		Muck	4	
		St. Quentin	5	
		Legaceville	2	
		Bog	6	
		Fen	3	
		SDTWp	5	
Parleeville	0.08	Parleeville	7	1.61
		Midland (I)	1	
		Midland (P)	1	
Parry	2.14	Parry	5	1.07
		Midway (I)	1	
		Midway (P)	0	
Parsons Brook	-	Parsons Brook	4	0.71
Petitcodiac	-	Petitcodiac	0	0.18
		Kings (I)	1	
		Kings (P)	0	
Pinder	0.53	Pinder	1	0.36
		Coronary	1	
		McAdam	0	
Popple Depot	2.71	Popple Depot	0	-
Queenville	-	Queenville	2	0.36
		Deed (I)	0	
		Deed (P)	0	
Quisbis	-	Quisbis	4	1.07
		Dube	1	
		Big Spring	1	
Reece	7.18	Reece	6	1.96
		Chipman	3	
		Pangburn	2	
Research Station	-	Research Station	0	-
		Baker Brook (I)	0	
Richibucto	-	Richibucto	9	1.79
		Cap Limiere	0	
		Nevers Road	1	
Riley Brook	-	Riley Brook	1	0.18
Riverbank	2.05	Riverbank	9	1.79
		Oromocto	1	
		Nevers Road	0	
Rogersville	0.54	Rogersville	1	0.54
		Acadieville	1	

		Rosaireville	1	
Salem	-	Salem	2	0.36
Salisbury	2.30	Salisbury	6	2.14
		Harewood	3	
		Hicksville	3	
Saltspring	0.13	Saltspring	1	0.71
		Byrns (I)	2	
		Byrns (P)	1	
Serpentine	0.56	Serpentine	2	0.71
		Jenkins	1	
		Adder	1	
Shemogue	-	Shemogue	2	0.36
Siegas	0.60	Siegas	3	2.14
		Salmon	6	
		Bourgoin	3	
St. Charles	-	St. Charles	1	0.18
Stony Brook	6.38	Stony Brook (Queens)	11	4.11
		Blackville	6	
		Cambridge (Kings)	6	
Sunbury	3.85	Sunbury	7	1.43
		Hoyt	1	
		Cork	0	
Sussex	-	Sussex	2	0.54
		Hampton	1	
Tetagouche	0.60	Tetagouche	0	-
Thibault	2.95	Thibault	10	2.50
		Guercheville	2	
		Lauzier	2	
Tobique	-	Tobqiue	2	0.36
Tormentine	-	Tormentine	7	2.86
		Tidnish (I)	5	
		Tidnish (P)	4	
Tracadie	0.47	Tracadie	1	0.89
		Bouleau	1	
		Sheila	3	
Tracy	0.74	Tracy	4	1.96
		Wirral	5	
		Rooth	2	
Tuadook	1.96	Tuadook	2	0.71
		Redstone	1	
		Lewis	1	
Undine	0.24	Undine	6	1.07
Upper Caraquet	-	Upper Caraquet	0	0.18
		Little Shippegan	0	
		Shediac	1	
Victoria	1.99	Victoria	6	2.14
		McCluskey	3	
		Cote	3	

Violette	-	Violette (I)	2	0.71
		Violette (P)	2	
Wakefield	-	Wakefield	1	0.18

APPENDIX III. COVARIATES ASSESSED FOR USE IN DSM WITH DESCRIPTION, REFERENCE, AND SOFTWARE UTILIZED.

Derivative	Definition	Reference	Software Used
Catchment Area	Upslope area contributing discharge into any given cell, calculated from flow accumulation	Freeman (1991), Tarboton (1997)	ArcGIS
Flow Accumulation	Total number of cells flowing into any given cell	O'Callaghan and Mark (1984), Fairfield and Leymarie (1991), Freeman (1991), Quinn *et al.* (1991)	ArcGIS
Horizontal Distance to Stream	Euclidean distance to nearest stream	This study	ArcGIS
Vertical Distance to Stream	Vertical distance from nearest stream based on increasing elevation	O'Callaghan and Mark (1984), Nobre *et al.* (2011)	ArcGIS
Cartographic Depth-to-Water Index	Index of depth-to-water table for any given cell	Murphy *et al.* (2007, 2009)	ArcGIS
Distance to Coast	Euclidean distance to nearest coastal shore	This study	ArcGIS
Distance to Wetland	Euclidean distance to nearest mapped wetland feature	This study	ArcGIS
Slope (degrees)	Rate of elevation change between adjacent cells (in °)	Zevenbergen and Thorne (1987), Wilson and Gallant (2000)	SAGA GIS
Slope (%)	Rate of elevation change between adjacent cells (as %)	Zevenbergen and Thorne (1987), Wilson and Gallant (2000)	SAGA GIS
Aspect (degrees)	Cardinal direction of slope	Zevenbergen and Thorne (1987), Wilson and Gallant (2000)	SAGA GIS
Annual Average Rainfall	Annual average rainfall (mm) retrieved from weather stations, interpolated from elevation, latitude and longitude	This study	ArcGIS
Annual Average Temperature	Annual average temperature (°C) retrieved from weather stations, interpolated from elevation, latitude and longitude	This study	ArcGIS
Average Daily Maximum High Temperature	Annual average daily maximum high temperature (°C) retrieved from weather stations, interpolated from elevation, latitude and longitude	This study	ArcGIS
Latitude	Gridded representation of changing latitude from north to south (decimal degree format)	This study	ArcGIS

Longitude	Gridded representation of changing longitude from west to east (decimal degree format)	This study	ArcGIS
Landform	Gridded representation of parent material (surficial geology) mode of deposition	This study	ArcGIS
Lithology	Gridded representation of parent material (surficial geology lithology)	This study	ArcGIS
Mineral Hardness	Gridded representation of the hardness of lithologic types based on Moh's hardness scale	This study	ArcGIS
Grain Size	Gridded representation of changing mineral sizes from very fine to coarse based on lithology	This study	ArcGIS
Rock Type	Gridded representation of changing rock types based on lithology	This study	ArcGIS
Geology Type	Gridded representation of changing geological types (sedimentary, igneous, metamorphic)	This study	ArcGIS
Elevation	Representation of elevation change across landscape, height above sea level	Moore *et al.* (1991), Yue *et al.* (2007)	ArcGIS
Diurnal Anistropic Heating	Influence of topography on diurnal heat balance	Hengl and Hannes (2009)	SAGA GIS
Real Surface Area	Calculation of 'real' cell area	Grohmann *et al.* (2011)	SAGA GIS
Catchment Slope	Slope of catchment, derived as intermediate with SAGA Wetness Index	Boehner *et al.* (2002), Boehner and Selige (2006)	SAGA GIS
Modified Catchment Area	Adjustment to catchment area calculation to correct for flow in low-lying flat areas.	Boehner *et al.* (2002), Boehner and Selige (2006)	SAGA GIS
Specific catchment area	Catchment area divided by cell width	Freeman (1991)	ArcGIS
Topographic Position Index	Position on hillslope in relation to adjacent cells	Guisan *et al.* (1999), Wilson and Gallant (2000)	SAGA GIS
Multi-resolution index of valley bottom flatness (MRVBF)	Delineation of valley bottom flatness from surrounding hillslopes	Gallant and Dowling (2003)	SAGA GIS
Multi-resolution index of ridgetop flatness (MRRTF)	Delineation of ridgetops flatness from surrounding hillslopes, intermediate of MRVBF	Gallant and Dowling (2003)	SAGA GIS
Terrain Surface Convexity	Determines percentage of upward and convex cells within a defined radius	Iwahashi and Pike (2007)	SAGA GIS
Terrain Surface Concavity	Determines percentage of upward and concave cells within a defined	Iwahashi and Pike (2007)	SAGA GIS

	radius, same algorithm as terrain surface convexity		
Terrain Surface Texture	Determines the variability in frequency and intensity of pits and peaks within a defined radius	Iwahashi and Pike (2007)	SAGA GIS
Topographic Positive Openness	Represents landscape exposure to atmosphere	Yokoyama and Shirasawa (2002), Prima *et al.* (2006)	SAGA GIS
Topographic Negative Openness	Represents landscape enclosure/ protection from atmosphere	Yokoyama and Shirasawa (2002), Prima *et al.* (2006)	SAGA GIS
Curvature Classification	Representation of surface curvature/ 9 geometric forms of hillslopes	Dikau (1988), Birkeland (1999), Hugget (2007)	SAGA GIS
Maximum Curvature	Calculates the maximum slope of the slope on a defined search radius (secondary derivative)	Zevenbergen and Thorne (1987), Irvin *et al.* (1997), Petry *et al.* (2005), Olaya (2006)	SAGA GIS
Minimum Curvature	Calculates the minimum slope of the slope on a defined search radius (secondary derivative)	Zevenbergen and Thorne (1987), Irvin *et al.* (1997), Petry *et al.* (2005), Olaya (2006)	SAGA GIS
Upslope Curvature	Average local curvature of upslope contributing area for any given cell	Freeman (1991)	SAGA GIS
Downslope Curvature	Like profile curvature	Freeman (1991)	SAGA GIS
General curvature	Combination of planar and profile curvature	Zevenbergen and Thorne (1987)	SAGA GIS
Local Downslope Curvature	Same as local curvature but only looks at downslope cells	Freeman (1991)	SAGA GIS
Local Upslope Curvature	Same as local curvature but only looks at upslope contributing cells	Freeman (1991)	SAGA GIS
Local Curvature	Calculates the total gradient to neighboring cells	Freeman (1991)	SAGA GIS
Profile Curvature	Curvature parallel to slope	Zevenbergen and Thorne (1987)	SAGA GIS
Planar Curvature	Curvature perpendicular to slope	Zevenbergen and Thorne (1987)	SAGA GIS
Longitudinal Curvature	Like profile curvature	Zevenbergen and Thorne (1987)	SAGA GIS
Cross-sectional Curvature	Like planar curvature	Zevenbergen and Thorne (1987)	SAGA GIS
Tangential Curvature	Determines areas of convex and concave flows, calculated from planar and profile curvature	Wilson and Gallant (2000)	SAGA GIS
Flow Line Curvature	Like profile curvature	Freeman (1991)	SAGA GIS
Total Curvature	Curvature of the surface	Wilson and Gallant (2000)	SAGA GIS
Downslope Distance Gradient	Measures impact of local slope on hydraulic gradient	Hjerdt *et al.* (2004)	SAGA GIS

SAGA Wetness Index	Computes a modified topographic wetness index	Boehner *et al.* (2002), Boehner and Selige (2006)	SAGA GIS
Slope Length and Steepness	Calculates slope length factor, typically used in Revised Universal Soil Loss Equation (RUSLE)	Desmet and Govers (1996), Kinnell (2005), Boehner and Selige (2006)	SAGA GIS
Terrain Ruggedness Index	Total change in elevation within a defined radius of any given cell	Riley *et al.* (1999)	SAGA GIS
Valley Depth	Like inverse of vertical distance to channel network	-	SAGA GIS

APPENDIX IV. RANDOM FOREST-DERIVED VARIABLE IMPORTANCE FOR 5 DOMINANT PREDICTORS FROM MODEL RESULTS FOR SAND, SILT, CLAY, CF, SOM, AND DB AT DEFINED DEPTH INTERVALS AND BY SOLUM DEPTH.

Property	Depth									
	0-15cm		15-30cm		30-60cm		60-90cm		Solum	
	Var.	Imp.	Var.	Imp.	Var.	Imp.	Var.	Imp.	Var.	Imp.
Sand	Rock_Type	100	Rock_Type	100.00	SLP	100	Val_Depth	100	Admax	100.00
	Admax	77.17	Admax	97.33	Rock_Type	98.06	Land	71.57	Rock_Type	69.31
	Min_Curve	69.06	Lithology	91.13	Max_Curve	91.79	Max_Curve	71.16	Prof_Curve	58.77
	Land	68.51	Max_Curve	70.01	Admax	88.69	Cur_Cls	58.76	Land	48.10
	Lithology	66.08	Di_Heat	66.42	MRRTF	82.15	Geo_Type	52.75	Lithology	41.50
Silt	Admax	100.00	Admax	100.00	Lithology	100.00	Val_Depth	100.00	Admax	100.00
	DTW	53.83	Min_Curve	64.04	Admax	89.94	Geo_Type	97.79	Lng_Curve	66.70
	Rock_Type	45.40	Max_Curve	63.73	Grain_Size	76.29	Rock_Type	75.33	Geo_Type	55.59
	CA	44.36	SLP	63.65	DTW	73.75	Dwn_Grade	78.06	SLP	54.72
	Cur_Cls	34.10	Lng_Curve	55.34	SLP	64.35	Wetland	72.12	Prof_Curve	54.21
Clay	Wetland	100.00	Rock_Type	100.00	TPI	100.00	Admax	100.00	Wetland	100.00
	MRRTF	74.72	Admax	94.72	Val_Depth	94.58	Max_Curve	80.41	Rock_Type	80.94
	Lng_Curve	48.95	MRVBF	78.65	Rock_Type	71.45	Val_Depth	76.17	Cur_Cls	68.09
	MRVBF	44.12	Lithology	70.69	Max_Curve	70.24	Cur_Cls	66.26	Crs_Curve	63.06
	Rock_Type	40.94	Geo_Type	57.36	Admax	67.23	Wetland	62.79	DTW	62.17
CF	Admax	100.00	SLP	100.00	Admax	100.00	Admax	100.00	Val_Depth	100.00
	MRRTF	73.29	Admax	71.57	Tan_Curve	52.38	Land	76.40	MRVBF	85.45
	Cur_Cls	50.56	Dwn_Grade	67.77	Prof_Curve	47.84	Min_Cuve	74.49	Admax	71.98
	MRVBF	43.71	Max_Curve	63.97	Max_Curve	45.40	Geo_Type	67.07	Rock_Type	61.93
	Rock_Type	39.04	ASP	62.09	DTW	42.84	DTW	66.61	MRRTF	60.48
OM	Wetland	100.00	Land	100.00	Admax	100.00	Plan_Curve	100.00	Geo_Type	100.00
	Dwn_Grade	94.37	CA	98.38	TPI	53.46	Admax	95.22	Val_Depth	91.71
	Lng_Curve	86.35	Lng_Curve	92.63	ASP	53.26	ASP	91.18	Rock_Type	83.54
	Land	83.18	Rock_Type	92.09	Land	53.19	Crs_Curve	8694	Max_Curve	82.80
	Crs_Curve	77.21	FA	90.95	Wetland	53.02	Tan_Curve	86.28	Crs_Curve	73.40
Db	MRVBF	100.00	Lithology	100.00	Admax	100.00	Lithology	100.00	Admax	100.00
	MRRTF	75.72	Max_Curve	94.27	SLP	57.71	ASP	54.59	Crs_Curve	92.05
	Prof_Curve	72.07	Rock_Type	91.55	Cur_Cls	51.95	DTW	54.29	SLP	91.73
	Val_Depth	72.01	Wetland	87.03	CA	45.34	Land	53.67	Ts_Conv	91.11
	Lng_curve	71.17	DTW	77.10	ASP	45.02	Cur_Cls	49.55	MRVBF	80.15
Others	A Depth		B Depth		Solum Depth		Drainage			
	Admax	100.00	CA	100.00	Admax	100.00	SLP	100.00		
	FA	91.10	MRVBF	96.65	Cur_Cls	93.37	Max_Curve	87.95		
	TPI	90.22	Cur_Cls	87.31	CA	76.53	Admax	78.61		
	Geo_Type	85.66	Admax	77.84	Tan_Curve	69.86	TPI	77.35		
	Max_Curve	82.13	TPI	75.99	Di_Heat	68.43	Dwn_Grade	71.32		

VITAE

Candidate's full name: Shane Robert Furze

Universities Attended:

University of New Brunswick 2013-2018
Doctor of Philosophy in Forestry
(Candidate)

University of New Brunswick 2011-2013
Master of Environmental Management

St. Francis Xavier University 2006-2010
Bachelor of Science in Environmental Science
in Biology

Publications:

Furze, S., Castonguay M., Ogilvie J., Nasr, M., Cormier, P., Gagnon, R., Adams, G., Arp, P.A. 2017. Assessing Soil-Related Black Spruce and White Spruce Plantation Productivity. *Open Journal of Forestry*, 7, 209-227.

Furze, S., Ogilvie, J., Arp, P.A. 2017. Fusing Digital Elevation Models to Improve Hydrological Interpretations. *Journal of Geographic Information System*, 9, 558-575.

Furze, S., Arp, P.A. 2018. Amalgamation and Harmonization of Soil Survey Reports into a Multi-purpose Database: An Example. *Open Journal of Soil Science* (submitted).

Furze, S., Arp, P.A. 2018. From Soil Surveys to Pedotransfer Function Development and Performance Assessment. *Open Journal of Soil Science* (submitted).

Conference Presentations:

Soil Erosion and Sedimentation Estimates for the Northwest Miramichi Watershed, New Brunswick, Canada. ESRI Atlantic User Conference, Halifax, NS, November, 2013.

Removing Digital Elevation Model (DEM) Artifacts through a Weighted Average Optimization. ESRI Atlantic User Conference, Fredericton, NB, October, 2014.

Improving Available DEMs for Enhanced Applications (Poster). Canadian Symposium on Remote Sensing, St. Johns, Newfoundland, June, 2015

Enhancing Digital Elevation Models for Improved Soils Mapping. Canadian Symposium on Soil Science, Montreal, Quebec, July, 2015.

Enhancing Digital Elevation Models (DEMs) for Improved Soils Mapping. ESRI Atlantic User Conference, Halifax, NS, October, 2015.

Enhancing Digital Elevation Models (DEMs) for Improved Soils Mapping (Poster). ESRI International User Conference, San Diego, California, July, 2016.