A COMPARISON OF RANDOM FORESTS AND LINEAR STEPWISE REGRESSIONS TO MODEL AND MAP SOIL CARBON IN SOUTH-CENTRAL BRITISH COLUMBIA GRASSLANDS USING NORMALIZED DIFFERENCE VEGETATION INDEX BASED MODELS

By

HEATHER JENINE RICHARDSON

B.Sc., Queen's University 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN ENVIRONMENTAL SCIENCE
in the Faculty of Science

Thesis examining committee:

Lauchlan Fraser (PhD), Professor and Thesis Supervisor, Department of Natural Resource Sciences, Thompson Rivers University.

Wendy Gardner (PhD), Associate Professor and Committee Member, Department of Natural Resource Sciences, Thompson Rivers University.

David Hill (PhD), Associate Professor and Committee Member, Department of Geography and Environmental Studies, Thompson Rivers University.

Cameron Carlyle, External Examiner (PhD), Department of Agricultural, Food, and Nutritional Science, University of Alberta.

Summer 2015

Thompson Rivers University

Abstract

Industrialization, production and consumption of fossil fuels, and land use changes have resulted in increased concentrations of carbon dioxide (CO_2) and other greenhouse gases in the atmosphere causing changes in ecosystem structure and properties. Soil carbon (SC) sequestration, the process of storing CO_2 in the soil through crop residues and other organic solids, has been an area under much investigation as it relates to reducing atmospheric carbon (C) and mitigating climate change. Since grasslands predominately sequester C below ground through root growth and consequent soil-building processes, they have a high potential for long term C storage and therefore are of major importance for maintaining Earth's carbon cycle. Despite advances in SC determination in recent years, it remains a challenge to model and map SC across large regions. There are several factors, both anthropogenic and environmental, that influence C sequestration. Given this complex system, I have used Geographic Information Systems (GIS) data in conjunction with accurate field measurements to examine the mechanisms that affect SC storage in order to produce predictive SC maps for the southern interior grasslands of British Columbia (BC). Soil carbon prediction was based on the Normalized Difference Vegetation Index (NDVI), which has demonstrated high correlation with SC distribution in past studies. The relationship of SC and NDVI was evaluated on two scales using: i) the MOD 13Q1 (250 m/16 day resolution) NDVI data product from the Moderate Resolution Imaging Spectro-radiometer (MODIS) aboard the United States Terra satellite ($NDVI_{MODIS}$), and ii) a handheld Multispectral Radiometer (MSR16R, Cropscan Inc., 1 m resolution) device ($NDVI_{MSR}$). Other factors included in the model are: i) grazing, ii) climate data, iii) vegetation community zones, iv) soil classification and drainage, and v) topography. A traditional linear stepwise regression (SR) modelling approach was compared with random forest (RF) modelling, a recursive partitioning technique that employs randomized bagging and bootstrapping of samples. There was a strong relationship between NDVI derived from the MSR with SC in fenced systems (R^2=0.41), SOC in fenced systems (R^2=0.47), and SOC in grazed systems (R^2=0.34). When NDVI data derived from the MSR was used as model input, the percentage of explained variance was greater than for models which used NDVI derived from MODIS data (R^2 = 0.68 for SC in 2014 for fenced systems, modelled with SR based on NDVI data derived from MODIS ; R^2=0.77 for SC in 2014 for fenced systems, modelled with SR based on NDVI data derived from MSR). These results show the potential of increased model accuracy with higher resolution GIS data and the effectiveness of NDVI based models to predict SC and SOC. Significantly higher SC and SOC was recorded in 2014 as compared to 2013 (p=0.001 for SC and p=0.031 for SOC), demonstrating the potential for C sequestration in BC grasslands as a climate change mitigation tactic. Based on comparisons of R^2 and AIC values, SR produces models that explain more variance and are of better quality (R^2=0.49-0.77 and AIC = 0.30-0.13 for SR models in 2014; R^2=0.36-057 and AIC = 0.36-0.18). This project creates the groundwork for effective monitoring techniques of SC and SOC levels using GIS data in order to develop a carbon offset program for the ranching industry and can be used to help direct land management efforts to increase C sequestration in BC.

Keywords: carbon sequestration, climate change, soil carbon, random forest, stepwise regression, Normalized Difference Vegetation Index, predictive mapping

TABLE OF CONTENTS

Abstract ii

Table of Contents iv

Acknowledgments vi

List of figures vii

List of tables ix

List of abbreviations xi

Chapter 1 : Introduction 1

Defining Soil Carbon and Its Components 2

Advances in Soil Carbon Monitoring 3

Rationale and Research Aims 8

Works Cited 10

Chapter 2 : Modelling Soil Carbon in the Grasslands of British Columbia's Southern Interior 12

Introduction 12

Cattle Grazing and Soil Carbon 12

Environmental Factors and Soil Carbon 14

Decisions Trees 15

Objectives 18

Methods 20

Experimental Design and Study Sites 20

Field Methods 23

Biogeochemical Analysis 24

Data 26

Statistical Analysis 29

Modelling and Mapping 29
Results 32
Vegetation Indices 32
Grazed Vs Un-grazed areas 35
Depth 35
Time 36
Soil Carbon and Soil Organic Carbon Density 39
Stepwise Regressions and Random Forest Models 44
Discussion 52
Conclusion 53
Works Cited 55
Chapter 3 : General Conclusions 58
Limitations 58
Management implications and climate change 60
Future research directions 61
Conclusion 62
Works Cited 64
Appendix A: Site Codes and Locations 65
Appendix B: Script 67
Appendix C: Predictive maps 80
Appendix D: Grass spectral signatures 96
Appendix E: Supplemental Resources 97

Acknowledgments

I would like to thank my committee members: my supervisor Lauchlan Fraser for his constant guidance over the course of my project; David Hill for his patience and assistance, especially when teaching me to use R; and Wendy Gardner for reviews, suggestions, and advice.

Everyone at the Fraser Lab has contributed to a welcoming and educational working and learning environment. In particular, I am thankful for the encouragement, assistance with field work, and discussion from my lab mates and research assistants: Dan Denesiuk, Sabina Donnelly, Heath Garris, Kathy Baethke, Amanda Schmidt, Simon Oliver, Zoe Simon, and Matthew Coghill. Continuing support and positivity from my friends, family, and especially my parents Glenn and Shirley Richardson, has been invaluable throughout my education.

This research was funded by grants to Lauchlan Fraser from the Canada-British Columbia Ranching Task Force, Growing Forward 2, and BC Farm Adaptation Innovator Program. I was supported through a National Science and Engineering Research Council (NSERC) Industrial Partner Scholarship with the Grasslands Conservation Council of BC.

LIST OF FIGURES

Figure 1.1: Mineralization and transfer of organic matter in soil (Christensen, 1996). 3
Figure 1.2 Spectral signatures of soil from various texture classes and soil C content (Yang and Mouazen, 2012). Note the peaks at 1414, 1814, and 2208 wavelength of the sample with the highest soil carbon content. 5
Figure 1.3: What is Imaging Spectroscopy? (Modified from: Elowitz, 2014). 7
Figure 2.1 : Decision Tree layout (Han and Kamber, 2006). 16
Figure 2.2: Sample site locations of Range Reference Areas within 5 grassland regions of the South-Central Interior. 21
Figure 2.3: Examples of Range Reference Areas sites at a) Alkali Creek, Chilcotin Region; b) Lac du Bois, Thompson-Nicola Region; and c) Crump, Okanogan Region. 22
Figure 2.4: Sampling design at Range Reference Area exclosures. 23
Figure 2.5: GIS data layers for soil carbon modelling and mapping including: a)Mean Annual Precipitation (MAP), b) Mean Annual Temperature (MAT), c) Soil Type, d) Soil Drainage, e) Vegetation Community, f) Aspect, g) Slope, h) Elevation, and i) Normalized Difference Vegetation Index (NDVI) 28
Figure 2.6: Example of model Builder flowchart for pre-processing data and creating predicted soil carbon grids with the stepwise regressions for 2013 soil carbon model of fenced systems 31
Figure 2.7: Relationships between 2014 soil carbon (SC%) and soil organic carbon (SOC(g/kg)) versus normalized difference vegetation index (NDVI) derived from the multispectral radiometer. Plots a) and b) compare SC and SOC against NDVI derived from MODIS. Plots c) and d) compare SC and SOC against the MSR. 34
Figure 2.8: Comparison of a) Soil Carbon (SC%) and b) Soil Organic Carbon(SOC (g/kg)) over various depths (0-10cm, 10-20cm, and 20-30cm). Letters above boxes represent Post Hoc Tukey Test results where categories with different letters are significantly different. 36

Figure 2.9 Comparison of a) Soil Carbon (SC%) and b) Soil Organic Carbon (SOC (g/kg)) between 2013 and 2014. Letters above boxes represent Post Hoc Tukey Test results where categories with different letters are significantly different. 37

Figure 2.10: Soil Carbon and Soil Organic Carbon density by region: a) Grazed 2014, b) Fenced 2014, c) Grazed 2014, d) Fenced 2014, e) Grazed 2013.. Letters above the bars represent results from the Post Hoc Tukey Test where categories sharing the same letter are not significantly different.. 43

Figure 2.11 Predicted soil carbon (SC) and soil organic carbon (SOC) based on SR models: a) SOC Grazed 2014, b) SOC Grazed 2013, c) SOC Fenced 2014, d) SOC Fenced 2013, e) SC Grazed 2014, f) SC Grazed 2013, g) SC Fenced 2014, and h) SC Fenced 2013. 47

Figure 2.12 Predicted Soil Carbon (SC%) for fenced systems in 2013 based on Stepwise Regression. Soil carbon layer over-layed on elevation to show distribution in upper and lower grasslands. 48

Figure 2.13 Predicted soil carbon (SC) and soil organic carbon (SOC) based on RF models: a) SOC Grazed 2014, b) SOC Grazed 2013, c) SOC Fenced 2014, d) SOC Fenced 2013, e) SC Grazed 2014, f) SC Grazed 2013, g) SC Fenced 2014, and h) SC Fenced 2013. 51

LIST OF TABLES

Table 2.1: GIS for soil carbon modelling and mapping.. 27
Table 2.2: The relationship between green biomass (derived from various spectral indices) versus SC(%) and SOC(g/kg) for 0-10cm depth... 33
Table 2.3: Stepwise regression results of change in Soil Carbon and Soil Organic Carbon from 2013 to 2014 in Grazed (G) and Fenced (F) Systems. The coefficients (Est) and significance (Sig) of each variable in the model is displayed. Sig: significance codes of p-values represented by 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.. ' 1.. 38
Table 2.4: Stepwise regression for 2013 Soil Carbon and Soil Organic Carbon Density for grazed (G) and fenced (F) systems, and the change between the two (C). The coefficients (Est) and significance (Sig) of each variable in the model is displayed. Sig: significance codes of p-values represented by 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.. ' 1......................... 40
Table 2.5: Stepwise regression results for 2014 Soil Carbon and Soil Organic Carbon Density for grazed (G) and fenced (F) systems, and the change between the two (C). The coefficients (Est) and significance (Sig) of each variable in the model is displayed. Sig: significance codes of p-values represented by 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.. ' 1. 41
Table 2.6: Stepwise regression results for 2013 Soil Carbon and Soil Organic Carbon in grazed (G) and fenced (F) systems. The coefficients (Est) and significance (Sig) of each variable in the model is displayed. Sig: significance codes of p-values represented by 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.. ' 1. 45
Table 2.7: Stepwise regression for 2014 Soil Carbon and Soil Organic Carbon in grazed (G) and fenced (F) systems comparing Normalized Difference Vegetation Index derived from MODIS and the Multispectral Radiometer. The coefficients (Est) and significance (Sig) of each variable in the model is displayed. Sig: significance codes of p-values represented by 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.. ' 1. 46
Table 2.8: 2013 Random Forest results showing "%incMSE", the percent increase in Mean Square Error when variable is permutated. Variables with negative"%incMSE" values were removed from the model and therefore not displayed in the table............... 49

Table 2.9: 2014 Random Forest results showing "%incMSE", the percent increase in Mean Square Error when variable is permutated. Variables with negative"%incMSE" values were removed from the model and therefore not displayed in the table.............. 50

List of Abbreviations

BC	British Columbia
BD	Bulk Density
C	Carbon
CO_2	Carbon dioxide
DT	Decision Tree
GHG	Greenhouse Gases
LOI	Loss on Ignition
MAP	Mean annual precipitation
MAT	Mean annual temperature
NDVI	Normalized Difference Vegetation Index
MSE	Mean Square Error
MSR	Multispectral Radiometer
MODIS	Moderate-resolution Imaging Spectro-radiometer
RF	Random forest
RRA	Range reference area
SC	Soil carbon
SOC	Soil organic carbon
SOM	Soil organic matter
SR	Stepwise regression
VI	Vegetation index

Chapter 1 : INTRODUCTION

Due to the production and consumption of fossil fuels and land use changes, the concentration of carbon dioxide (CO_2) and other greenhouse gases (GHG) in the atmosphere have been on the rise since the Industrial Revolution. The increases in GHGs intensify the process of climate change and in turn cause changes in ecosystem structure and properties (Hansen, 2008). Grasslands may be affected by these intensified processes through drought and erosion, a decrease in biodiversity, and ecosystem degradation (Winslow et al., 2003). Land use changes have simultaneously resulted in the depletion of soil and soil carbon (SC) levels, releasing 50 to 100 GT of carbon (C) from soil into the atmosphere due to reduced plant root material and residues returned to the soil, increased decomposition from soil tillage, and increased soil erosion (Lal, 2009; Wall and Six, 2015). Depletion of SC stocks has created a SC deficit that represents an opportunity to store C in soil through an assortment of land management approaches. Improved land management may reverse this deficit through the opposite process of SC sequestration. SC sequestration, the process of storing C in the soil through crop residues and other organic solids, has been an area under much investigation as it relates to reducing atmospheric CO_2 and mitigating climate change. Soils with high C content are also associated with increased fertility, water retention, and vegetation (Schlesinger, 1999).

The ability of soil to store C is dependent on many environmental factors (e.g. climate and landscape) and management practices (e.g. grazing). Grasslands and open forests grazed by livestock accounts for approximately 40% of British Columbia's (BC) land base; hence, a large part of BC's total SC pool is potentially affected by range management (Wikeem et al., 1993). Since grasslands predominately sequester C below ground through root growth and consequent soil-building processes, they have a high potential for long term SC storage and therefore are of major importance for maintaining earth's C cycle (Parton et al., 1995). The process of C sequestration relies on respiration and photosynthesis, two basic processes of the C cycle. Carbon may also enter the soil in the form of roots, litter, harvest residues, and animal manure. These inputs also contribute to the SC sink and are stored as soil organic matter (SOM). In many areas, poor land use management can upset this process, thereby causing a net emission of C. Therefore, monitoring SC stocks is an important task to maintain grassland ecosystem function, support the cattle industry, and mitigate global climate change.

Defining Soil Carbon and Its Components

SC can be either organic or inorganic. Inorganic C consists of elemental C and carbonate materials such as calcite, dolomite, and gypsum (Lal, 2004). Soil organic carbon (SOC) includes plant, animals, and microbial residue in all stages of decomposition. Physically defined fractionations of SOC pools are delineated into two groups: the light fraction and the organo-mineral (Figure 1.1). The light fraction is not combined with mineral matter and has a high turnover rate. Once transformed by bacterial action, the majority of SOC is transformed and found in clay or silt sized organo-mineral complexes. Finally, a small portion of SOC is represented in microbial biomass, which mediates the transfer of SOC among inputs. The rates of transfer and transformations are influenced by biologically important factors including soil moisture and temperature.

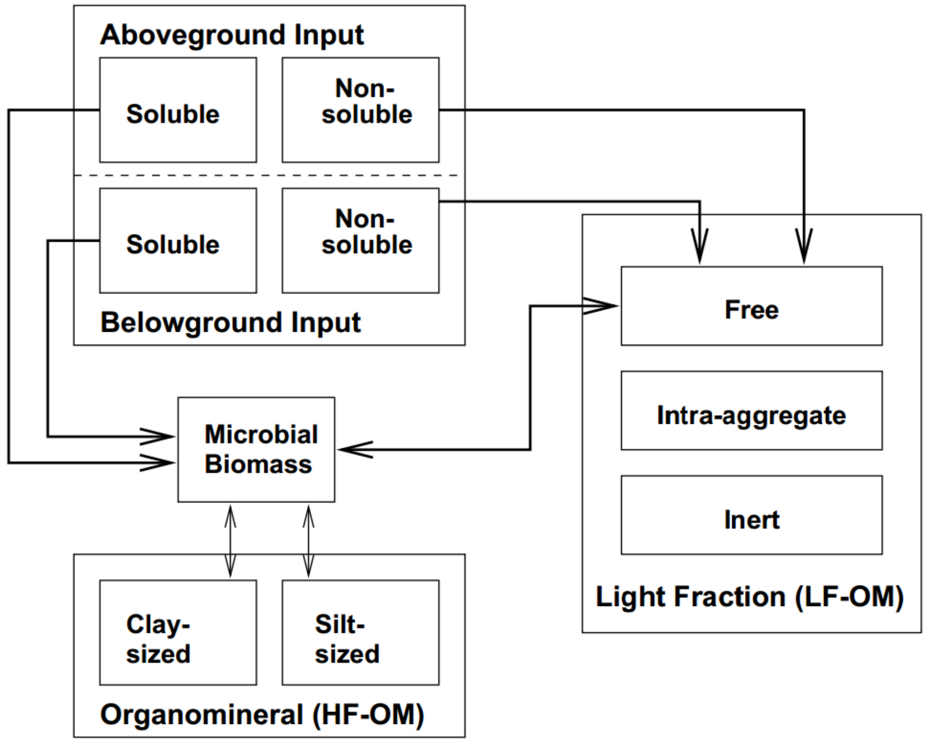

Figure 1.1: Mineralization and transfer of organic matter in soil (Christensen, 1996).

ADVANCES IN SOIL CARBON MONITORING

The traditional method of quantifying SC in the lab is the Walkey-Black (1934) method which uses a dry combustion technique of soil core samples. This method has notable limitations, being both time consuming and labor intensive (Gehl and Rice, 2007). Determination of total C by dry combustion, the measurement of CO_2 emitted from the oxidation of organic C and thermal decomposition of carbonate materials using an elemental analyzer (Nelson and Sommers, 1996), has become the predominant means of laboratory C analysis and was the method used in this project. However, some laboratories base C measurements on weight change rather the CO_2 emitted, presenting discrepancies in laboratory results from different areas (McCarty et al., 2002). In general, laboratory analysis of soil samples for C determination is too time consuming and costly for a constant monitoring system for SC over large spatial regions, such as the province of BC. Hence, ongoing investigations of remotely sensed (RS) monitoring systems for SC are critical.

Recently, several advanced, non-invasive methods have been utilized for SC research. Mid-infrared reflectance (MIR) and near-infrared reflectance (NIR) spectroscopy have each been assessed as a means to predict soil properties, including C content (Chang and Laird, 2001; McCarty et al., 2002). Reflectance spectroscopy provides a rapid and non-destructive method to indirectly determine SC based on diffusely reflected radiation of illuminated soil (Gehl and Rice, 2007). By comparing the spectral signature of soil samples with known SC contents, inferences can later be made about soils with similar properties. For example, regarding clay soils examined in a recent study, samples with high SC exhibited stronger absorption in the Vis-NIR spectra (Figure 1.2). The distribution of soils' reflectance over a spectrum of wavelengths creates identifiable characteristics – a spectral signature.

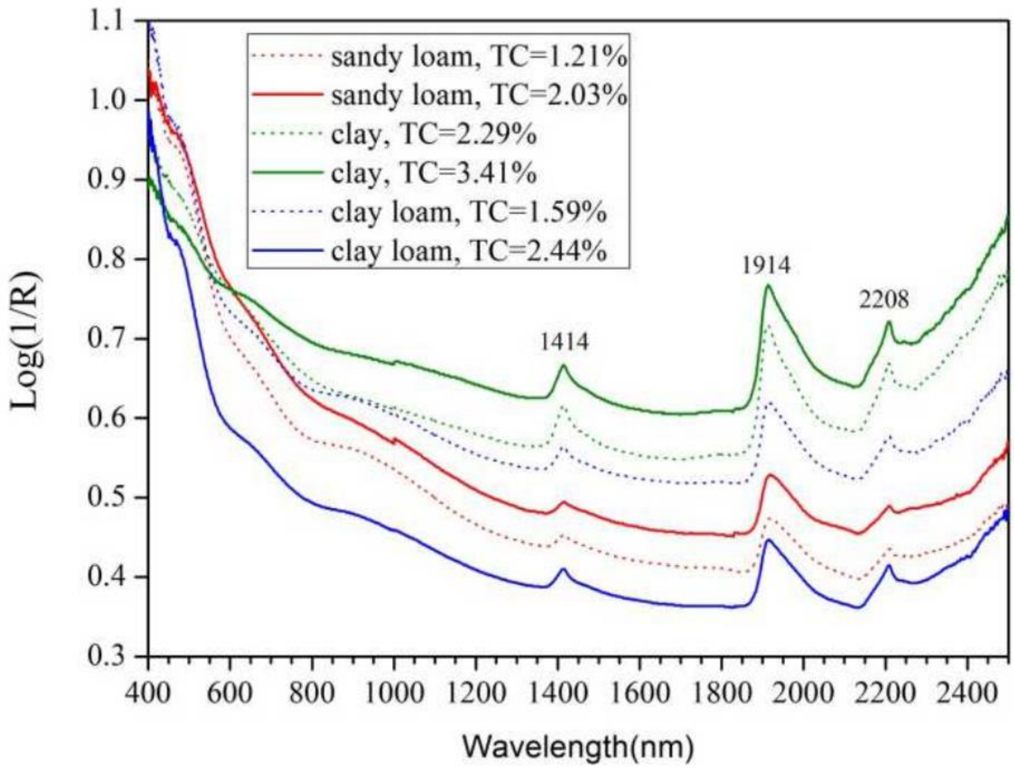

Figure 1.2 Spectral signatures of soil from various texture classes and soil C content (Yang and Mouazen, 2012). Note the peaks at 1414, 1814, and 2208 wavelength of the sample with the highest soil carbon content.

Constituents of organic matter each have unique absorptive or reflective properties due to stretching and bending vibrations of molecular bonds between elements (Gehl and Rice, 2007). Spectral signatures related to the various components of soil organic matter generally occur in the MIR (2.5–25 μm) range, although small overtones and combinations of fundamental vibrations occur in the NIR (0.7–2.5 μm) region (Shepherd and Walsh 2002). Field analysis of SC using spectral analysis minimizes soil disturbance while increasing expedience of analysis for C. Advanced spectral field methods of SC analysis should be capable of providing repetitive, successive measurements for evaluation at a finer spatial and temporal scale than previously feasible (Gehl and Rice, 2007).

Since different objects reflect radiation differently, the unique spectral reflectance curves (spectral signatures) of each object can be obtained, collected, and used for identification. This collection is called a spectral library. Using the correlation between SC and reflectance, a proximity based model in the spectral dimensions can be used to predict the SC content of an unknown sample. Using this concept, Bartholomeus et al. (2008) concludes it was possible to use spectral indices derived from laboratory measurements to predict SC in various soil types. However, a large variance within the spectral library in SC is required for the calibration of the prediction model, since extrapolation beyond the SC range in the training dataset results in large errors (Bartholomeus et al., 2008). Comparing SC results from laboratory analysis and field spectroscopy would assist the transition towards less invasive methods. Still, this method is limited in its ability to be applied to large spatial and/or temporal domains due to the time and costs associated with field analyses.

The theoretical basis for empirical-based vegetation indices is derived from examination of typical spectral reflectance signatures of leaves (Figure 1.3). The reflected energy in the visible spectrum is very low as a result of high absorption by photosynthetically active pigments, with maximum absorption values in the blue (470 nm) and red (670 nm) wavelengths. Nearly all of the near-infrared radiation (NIR) is scattered (reflected and transmitted) with very little absorption, in a manner dependent upon the structural properties of a canopy (LAI, leaf angle distribution, leaf morphology). As a result, the contrast between red and NIR responses is a sensitive measure of amount of vegetative land cover.

Remote sensing methods have also been used to predict SC by modelling methods focused on vegetation indices. Vegetation Indices (VIs) are combinations of surface reflectance at two or more wavelengths designed to highlight a particular property of vegetation, and they are commonly used as a surrogate for plant biomass. The distribution of SC has been proven to highly correspond with Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) (Zhang et al, 2012; Yang et al., 2008). Though other VIs exist, NDVI and EVI data products are freely available from satellite-borne instruments, such as the Moderate Resolution Imaging Spectro-radiometer (MODIS), and offer a helpful addition to SC prediction models.

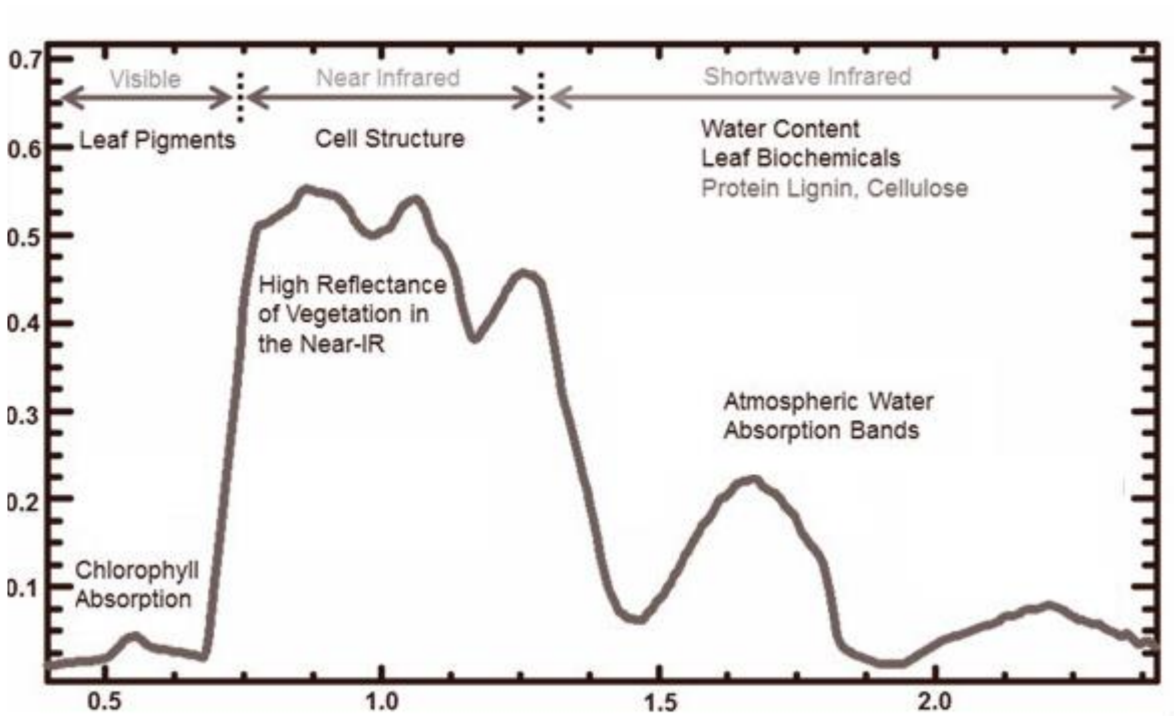

Figure 1.3: What is Imaging Spectroscopy? (Modified from: Elowitz, 2014).

The Normalized Difference Vegetation Index (NDVI) is a normalized transform of the near infrared (ρNIR) to red reflectance (ρRed) ratio (Equation 1). The formula for NDVI is:

$$\text{NDVI} = \frac{\rho NIR - \rho Red}{\rho NIR + \rho Red} \tag{1}$$

The Enhanced Vegetation Index (EVI) incorporates the atmospheric resistance concept, along with the removal of soil-brightness induced variations in VI (Equation 2). Additionally, EVI decouples the soil and atmospheric influences from the vegetation signal by including a feedback term for simultaneous correction. The formula for EVI is:

$$\text{EVI} = G * \frac{\rho NIR - \rho Red}{\rho NIR + C_1 * \rho Red - C_2 * \rho blue + L} \tag{2}$$

where ρ_x are the full or partially atmospheric-corrected (for Rayleigh scattering and ozone absorption) surface reflectance; L is the canopy background adjustment for correcting nonlinear, differential NIR and red radiant transfer through a canopy (L=1); C1 and C2 are the coefficients of the aerosol resistance term (which uses the blue band to correct for aerosol influences in the red band) (C1=6, C2=7.5,); and G is a gain or scaling factor (G=2.5).

EVI data has been used to estimate SOC storage in alpine grasslands in China and it was found that growing season EVI from MODIS datasets (500 m/16 day resolution) was strongly correlated with above ground biomass and SOC (Yang et al., 2008). A net ecosystem production (NEP) model using a piecewise regression tree approach was developed based on NDVI data from IKONOS (2m resolution), weather data sets, and NEP data from flux towers to produce a high accuracy result (r=0.88) (Zhang et al. 2012).

Despite advances in SC determination in recent years, it remains a challenge to model and monitor SC over large regions such as BC. There are several factors, both anthropogenic and environmental, that influence carbon sequestration. Given this complex system, the possibility of using RS applications in conjunction with accurate field measurements is a topic of much interest. Ideally, hybridization of both techniques will generate an updateable, efficient province-wide model.

Rationale and Research Aims

In order to promote the use of ranching techniques which have the greatest potential for carbon storage, rangeland carbon offsets should be viewed as an effective way for ranchers to be compensated for sustainable practices. Improving ranching management to optimize C sequestration and implementing a C offset program would address sustainable ranching practices. This improved system can also provide a unique revenue source for the ranching industry in BC. The implementation of such a program would be a remarkable advancement in the ranching industry in BC, improving economic sustainability, and offering the potential for increased C sequestration, improving environmental sustainability.

This thesis is part of the "Soil carbon sequestration in grasslands[1]" collaborative project in the Fraser Lab at Thompson Rivers University, which aims to investigate soil carbon storage potential in BC rangelands with respect to sustainable ranching practices. There are three streams of the project: i) grazing management and SC sequestration, ii) modelling SC in BC with respect to climatic, topographic, and vegetation differences, and iii) economical modelling of SC stocks for the ranching industry. Focusing on the second stream of the collaborative, my thesis will contribute the ecological background for economic modelling. Since we expect SC and SOC potential to vary throughout BC, it would be unfair to expect the same rates of SC sequestration from 2 different ranches. Before we assess how ranchers should be compensated for sustainable land management practices, we must know what SC stocks are expected at the undisturbed state of the land and what SC sequestration rates are possible.

While research has been conducted on the ability of pastures to sequester C and reduce GHG emissions (Franzluebbers, 2010), the proposed project will tackle the remaining challenges to develop protocols that include the implementation of sustainable range management and subsequent measurement and monitoring of carbon sequestration (Fynn et al., 2009). The first step to monitoring and improving sustainable ranching techniques is getting a better understanding of how various anthropogenic and environmental processes affect SC storage. Hence, the main objective of this research is to compare modelling techniques and identify optimal input variable combinations in order to most effectively map SC in BC grasslands.

[1] This project is a research initiative at the Fraser Lab, Thompson Rivers University, Kamloops, BC, Canada. See website for more information: https://grazingmgtandclimatechange.wordpress.com/

Works Cited

Bartholomeus, H. M., Schaepman, M. E., Kooistra, L., Stevens, A., Hoogmoed, W. B., and Spaargaren, O. S. P. (2008). Spectral reflectance based indices for soil organic carbon quantification. Geoderma, 145(1), 28-36.

Chang, C. W., Laird, D. A., Mausbach, M. J., and Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. Soil Science Society of America Journal, 65(2), 480-490.

Christensen, B. T. (1996). Matching measurable soil organic matter fractions with conceptual pools in simulation models of carbon turnover: revision of model structure. In Evaluation of soil organic matter models (pp. 143-159). Springer Berlin Heidelberg.

Franzluebbenrs, AJ. (2010). Achieving soil organic carbon sequestration with conservation agricultural systems in the southeastern United States. Soil Science Society of America Journal 74: 374-357.

Fynn, AJ, Alvarez P, Brown JR, George MR, Kustin C, Laca EA, Oldfield JT, Schohr T, Neely CL, and Wong CP. (2009). Soil carbon sequestration in U.S. rangelands: Issues paper for protocol development. Environmental Defense Fund, New York, NY, USA.

Gehl, R. J., and Rice, C. W. (2007). Emerging technologies for in situ measurement of soil carbon. Climatic Change, 80(1-2): 43-54.

Hansen, J. (2008). Target atmospheric CO_2: Where should humanity aim? The Open Atmospheric Science Journal 2: 217-231.

Lal, R. (2004). Soil carbon sequestration impact on global climate change and food security. Science 304, 1623-1627.

McCarty, G. W., Reeves, J. B., Reeves, V. B., Follett, R. F., and Kimble, J. M. (2002). Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Science SCiety of America Journal, 66(2), 640-646.

Nelson, D. W., Sommers, L. E., Sparks, D. L., Page, A. L., Helmke, P. A., Loeppert, R. H., ... and Sumner, M. E. (1996). Total carbon, organic carbon, and organic matter. Methods of soil analysis. Part 3-chemical methods., 961-1010.

Parton WJ, Scurlock JMO, Ojima DS, Schimel DS, Hall DO, Scopegram Group Members. (1995). Impact of climate change on grassland production and soil carbon worldwide. Global Change Biology 1: 13-22.

Schlesinger, William H. (1999) "Carbon sequestration in soils." Science 284.5423: 2095.

Shepherd, K. D., and Walsh, M. G. (2002). Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal, 66(3), 988-998.

Walkley, A., and Black, I. A. (1934). An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Science, 37(1), 29-38.

Wall DH and Six J. (2015). Give soils their due. Science 347: 695.

Wikeem, B.M. (1993). An overview of the forage resourve and beef production non Crown land in British Columbia. Canadian Journal of Animal Science. 73: 779-794.

Winslow CJ, Hunt R Jr, Piper SC. (2003). The influence of seasonal water availability on global C_3 versus C_4 grassland biomass and its implications for climate change research. Ecological Modeling. 163: 153-173.

Yang, Haiqing, and Abdul M. Mouazen. (2012) "Vis/Near-and Mid-Infrared Spectroscopy for Predicting Soil N and C at a Farm Scale." Infrared Spectroscopy-Life and Biomedical Sciences.185-210.

Zhang, W. et al. (2012). Ancillary information improves kriging on soil organic carbon data for a typical karst peak cluster depression landscape. J Sci Food Agric. 92(5):1094-102.

Chapter 2 : Modelling Soil Carbon in the Grasslands of British Columbia's Southern Interior

Introduction

Cattle Grazing and Soil Carbon

Despite the importance of rangelands for soil carbon (SC) storage, the impact of grazing on carbon (C) sequestration is not fully understood. Overgrazing can lead to poor range health and reduce the potential for rangelands to sequester C in the soil (Chapman and Lemaire, 1993). It is widely accepted that overgrazing is detrimental to plant communities due to grazing and trampling which may lead to a loss in species diversity, reduced vegetation biomass and density, and an increase in undesirable non-native invasive plants which thrive in disturbed ecosystems (Chapman and Lemaire, 1993). Grazing may also affect hydrology and soil properties such as increased soil erosion, reduced water infiltration and soil compaction, and lower soil quality and fertility (Schlesinger et al., 1990; Bremer, 2001).

On the other hand, recent research in the United States and in the southern interior of British Columbia (BC) suggests that moderate grazing may increase soil building processes and SC storage by increasing compensatory growth of forage grasses and turnover of plant roots, better facilitating soil development (Loeser et al., 2007; Schönbach et al., 2011). Light grazing may improve shoot turnover compared to fenced conditions (Schuman et al., 1999). Further, aboveground immobilization of C in standing dead plant material in fenced areas may lead to lower SC observed (Schuman et al., 1999).

A recent paper examining long term grazing effects of SC in upper grasslands of BC's Southern Interior determined that rough fescue above-ground and litter biomass were greater on fenced than grazed treatments, though this did not create differences in SC, which was similar on plots both with and without grazing (Krzic et al., 2014).

Another study carried out on bluebunch wheatgrass grasslands in the southern interior of BC (Evans et al., 2012) also reported that the long-term elimination of grazing did not lead to an increase of SC relative to the grazed pastures. In contrast, Schuman et al. (1999) found that 12 years of season-long cattle grazing at 0.67 and 2 AUM (animal unit months) per hectare led to 21 and 22% significantly higher total SC, respectively relative to non-grazed pasture. This increase in SC under grazing conditions was attributed to the increase in blue grama cover, a species which is known to develop a dense and continuous root mass in the upper soil layer and allocate more C and nutrients to roots than other species commonly found in the mixed-grass prairie. In a global review, Conant and Paustian (2002) found that of the studies they researched that showed increased soil organic matter (SOM) with higher grazing intensities, half of the sites contained blue grama grass. In their global review, it was also concluded that most C sequestration was located in areas that were lightly or moderately grazed, while only a small amount was located in strongly grazed grasslands (Conant and Paustian, 2002).

Evidently, previous studies have found both strong positive and negative grazing effects on SC. These contradicting results are explained by McSherry and Ritchie's (2013) conclusion that grazer effects on soil organic carbon (SOC) are highly context-specific and their causations interrelated. For instance, in their international study it was found that increasing grazing intensity increased SOC by 6-7% on C4-dominated and C4-C3 mixed grasslands but decreased SOC by an average 18% in C3-dominated grasslands (McSherry and Ritchie, 2013). Note that the native bunchgrasses in BC are C3 grasses (cool season grasses) while C4 grasses (warm season grasses) are less common and restricted to zeric habitats (Gayton, 2013).

My project compared the effect of long term grazing, in a similar fashion as Krzic et al. (2014), by sampling in grazed and fenced areas, separated by a permanent fenced exclosure (established for ~30 years, on average). Since there are over 60 sites included where exclosures have been in place for an upwards of 75 years, historical and current grazing practices are unknown at these locations. However, because of their distribution throughout BC, these sites encompass a variety of vegetation, soil, and climatic conditions.

Environmental Factors and Soil Carbon

Climate and Topography

The environmental variables that influence SC are often interconnected and relate to the productivity and stability of a landscape. Conant and Paustian (2002) modelled potential SC sequestration in overgrazed grassland ecosystems and established a positive linear relationship between potential SC sequestration and mean annual precipitation (MAP). The regression model predicted losses of SC with decreased grazing intensity in drier areas (MAP< 333 mm/ yr) but substantial sequestration in wetter areas; most (93%) C sequestration potential occurred in areas with MAP less than 1800 mm (Contant and Paustian, 2002).

Likewise, low-lying south facing slopes are typically drier therefore I expect aspect and elevation to be useful indicators as well. Since areas on steeper slopes may be more likely to experience erosion, I expect steep slopes to help indicate regions with low SC.

Soil Properties

Within BC's grasslands there was a diversity of soil types, encompassing several groups of the Canadian Soil Classification System which mark the differences in many soil characteristics including organic matter content, drainage, litter production, and soil texture. These characteristics influence SC directly and indirectly by affecting productivity, drainage, and stability. A study by Bhatti et al. (2003) has successfully used soil classifications to improve predictability of SC models. Specifically, the SC estimates were based on data from (i) analysis of pedon data from both the Boreal Forest Transect Case Study (BFTCS) area and from a national-scale soil profile database; (ii) the Canadian Soil Organic Carbon Database (CSOCD), which uses expert estimation based on soil characteristics; and (iii) model simulations with the Carbon

Budget Model of the Canadian Forest Sector (CBM-CFS2) (Bhatti et al., 2003). In McSherry and Ritchie's (2013) recent study, an increase in MAP of 600 mm resulted in a 24% decrease in 'grazer effects' on SOC for finer-textured soils, while the same increase in precipitation over sandy soils produced a 22% increase in 'grazer effects' on SOC.

Vegetation and Vegetation Indices

The abundance of organic C in the soil affects and is affected by plant production. Specifically, shoot/root allocations combined with vertical root distributions of different functional groups (e.g. grasses, shrubs, trees) have been found to affect the distribution of SOC with depth (Jobbágy and Jackson, 2000). Also, the presence of C4 grass and legume species was a key cause of greater soil C and N accumulation in both higher and lower diversity plant assemblages because legumes have unique access to N, and C4 grasses take up and use N efficiently, increasing below-ground biomass and thus soil C and N inputs (Fornara and Tilman, 2008). Past research has determined that the distribution of SC is related to Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), spectral indices compute from remote sensing (RS) data, which indicate the amount of green biomass present (Zhang et al., 2012; Yang et al., 2008).

DECISIONS TREES

To determine the appropriate set of input variables to be used to predict a certain characteristic, such as SC, regression analysis is commonly used. Though linear stepwise regressions (SRs) are classically used, decision trees (DT) are an alternative method for determining the interaction between variables and/or the independence of variables. For this research, I will compare SR with random forest (RF) models, a type of DT.

Decision tree induction is a supervised machine learning method that constructs a tree-based classifier based on a training dataset. In supervised learning, the input variable values (called attributes) are provided along with the observed response variable for each example in the training dataset. The DT classifier has a flowchart-like tree structure, where each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class (Figure 2.1). The topmost node in a tree is called the root node and identifies the most important input attribute while the nodes further down the tree are of lesser importance with each step down. Each terminal node contains a class label that is the expected outcome, based on the training dataset, of the unique combination of attribute values that define the path from the tree root to its leaf (Figure 2.1). To create a DT, a recursive partitioning method based on the information content of each input attribute in the dataset is used to produce the tree model. This method determines which of the input variable fields does the best job splitting the data. Then, it repeats the process for each sub-set until an end condition is reached. The splitting of the data is performed using an information-based metric to identify the splitting criterion that creates the most homogeneous data subsets following the split (Therneau and Atkinson, 2013).

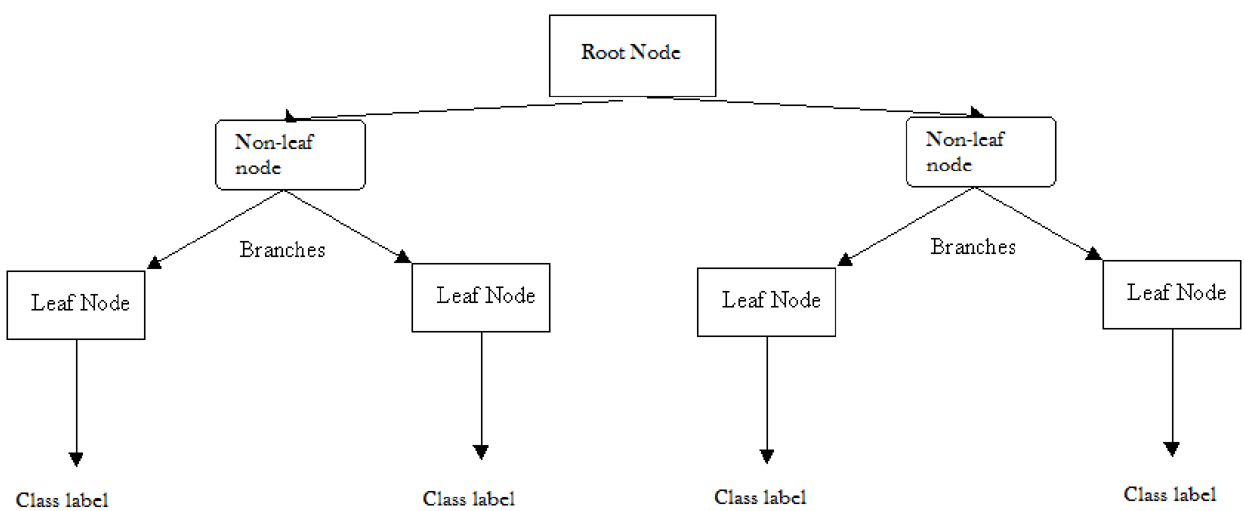

Figure 2.1 : Decision Tree layout (Han and Kamber, 2006).

Because of their natural graphical representation, DT models facilitate human understanding and interpretation via visual analysis (Therneau and Atkinson, 2013).

At the same time, DTs are fairly robust and typically perform well even with large data sets with different types of values (categorical, numerical, ranking) (Therneau and Atkinson, 2013). Furthermore, DT models are inherently non-linear and are robust to datasets that exhibit multicollinearity among the predictive variables (Therneau and Atkinson, 2013).

Random Forests

Random forest is an ensemble learning method for regression, that is based on the construction of many DTs (referred to as a 'forest'of DTs) during training. Random forest prediction is the expected value of the distribution of output values from trees within the forest. In the case of real-valued output, this value is calculated as the mean prediction of the individual trees. Although DT training may create models that are too specific to the training data and do not generalize well, a condition known as over-fitting, RFs correct for this tendency by bagging and bootstrapping the training data and by incorporating some randomness in selecting the attribute to split on. Each tree is built from a bootstrap sample of the original data set, which allows for robust error estimation with the remaining data, referred to as the 'Out-Of-Bag' (OOB) data. This is accomplished by predicting each example within the OOB data using a RF that was constructed from the bootstrap training samples. By aggregating the OOB predictions from the all trees within the RF, the mean square error of the prediction is then calculated (MSE_{OOB}) as:

$$MSE_{OOB} = n^{-1} \sum_{i=1}^{n} (z_i - {z_i}^{OOB})^2 \tag{3}$$

Where z_i is the ith OOB prediction and ${z_i}^{OOB}$ is the average of n OOB predictions for the ith observation, MSE_{OOB} is normalized as it depends on the unit of response variable and the percentage of explained variance (Var_{ex}) is calculated in Equation 4:

$$Var_{ex} = 1 - \frac{MSE_{OOB}}{Var_z} \tag{4}$$

Where Var_z is the total variance of the response variable. This is the goodness of fit.

The result of the RF is one single prediction which is the average of all the aggregations. One disadvantage of RF is that it is challenging to interpret a relationship between the input and response variables because so many DTs are produced in the forest, limiting the interpretation of the relationships between the response and then input variables. To explain these relationships, RF outputs an estimation of variable importance measured by the decrease in prediction accuracy before and after permuting a variable ('%incMSE').

OBJECTIVES

The main objective of this project is to evaluate the factors which influence SC using SR and RF modelling in order to subsequently map SC throughout BC's grasslands. We will compare various factors which influence SC values from 65 sites across BC's southern interior and have undergone total C determination by dry combustion using an automated elemental analyzer. Input factors evaluated in the model include: i) grazing; ii) climate zones based on historical temperature and precipitation data; iii) landscape variables including aspect, soil, and elevation; iv) Normalized Difference Vegetation Index imagery (MOD13Q1- 250m, 16day resolution); v) vegetation community zones; and vi) soil classifications within BC. The work conducted for this project will lay the basis for effective monitoring techniques of SC levels by using remote sensing (RS) techniques and explore the possibility for the implementation of a carbon-offset program for ranchers in BC. Specifically, the research questions to be covered in this chapter include:

i. What environmental and anthropogenic factors allow us to best predict SC?
ii. How is SC distributed across BC grasslands?
iii. What factors control sensitivity to grazing in regards to SC?
iv. What factors indicate high potential to store C with time?

v. How do SR models compare to RF models?

vi. How does increased resolution of NDVI data improve modelling?

METHODS

EXPERIMENTAL DESIGN AND STUDY SITES

Grasslands are a small but significant component of British Columbia's (BC) natural landscape. They are an important habitat for many wildlife species and support the ranching industry. Roughly 90% of BC's grasslands are grazed by domestic livestock, either through deeded private rangelands, grazing tenures on provincial crown land or grazing regimes on First Nations land (BC Grasslands Conservation Council, 2004). To capture the climatic, topographic, and vegetative differences throughout the grasslands of BC, 65 sites across the province were used to collect samples (Figure 2.2; see Appendix A for list of site locations, code names, and coordinates). In order to compare the effect of long term grazing, samples were taken at Range Reference Areas (RRAs) (Figure 2.3). RRAs are permanent fencing installations which are used to monitor the impact of livestock on BC rangelands and evaluate the accuracy of potential natural (climax) communities (PNC) estimates. These RRAs have been in place for 20-50 years and can therefore be used to identify the long term impacts of grazing; however, there is no information available on grazing intensity (e.g. stocking rates) or management practices at these sites. Hence, grazing is represented in 2 treatments -- grazed and fenced (fenced) (Figure 2.4). The sites cover a variety of local climates, plant communities, and physiographies. Among the sites, elevation ranges from 346 to1213 m and MAP ranges from 302mm/year to 538mm/year. Once analyzed, the SC and SOC values at these sites were used as training and testing data to construct and evaluate the SR and RF models.

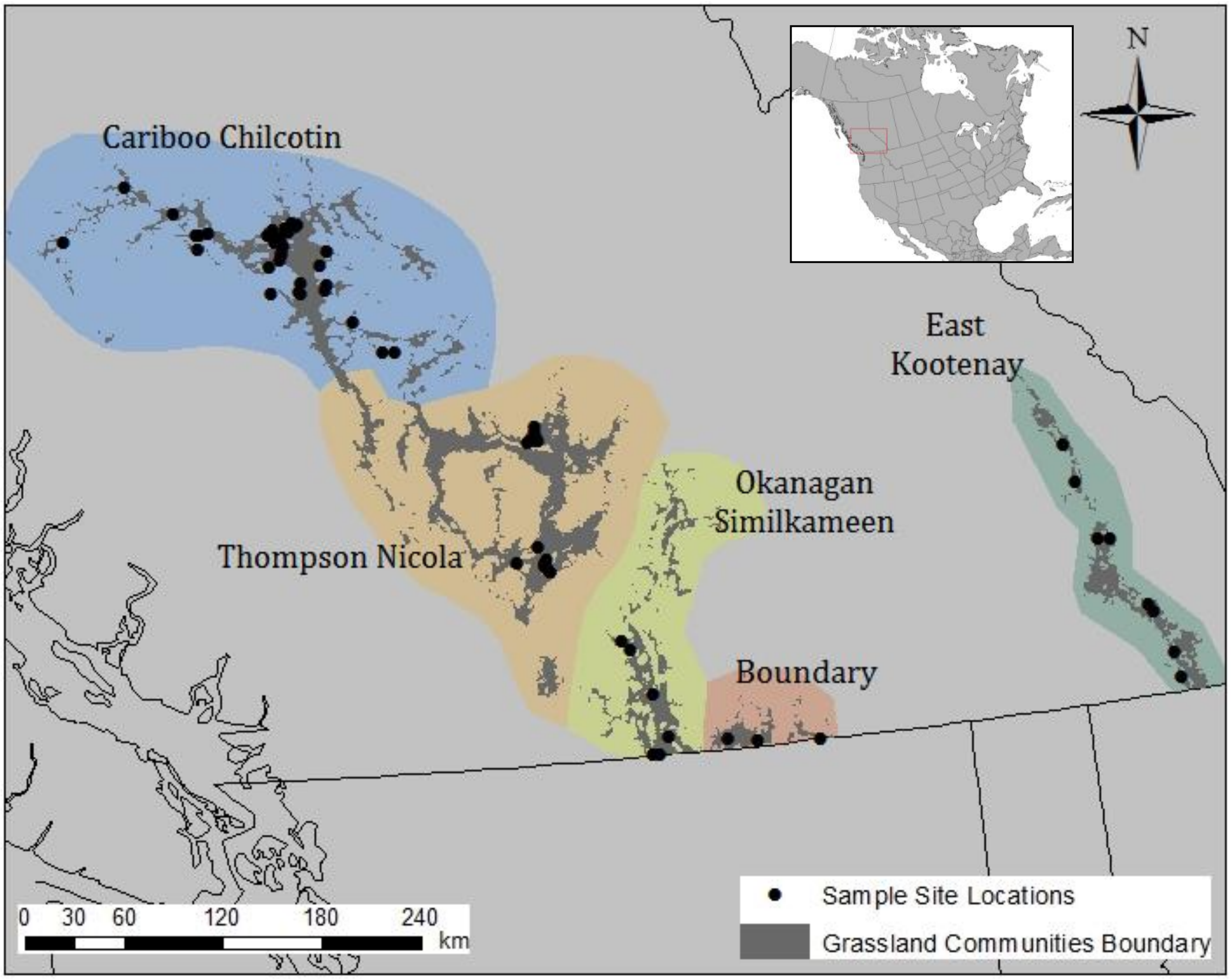

Figure 2.2: Sample site locations of Range Reference Areas within 5 grassland regions of the South-Central Interior.

a)

b)

c)

Figure 2.3: Examples of Range Reference Areas sites at a) Alkali Creek, Chilcotin Region; b) Lac du Bois, Thompson-Nicola Region; and c) Crump, Okanogan Region.

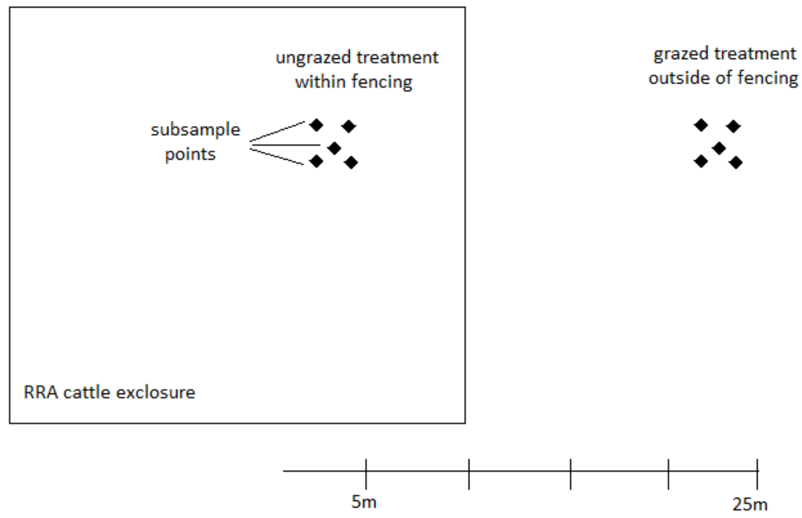

Figure 2.4: Sampling design at Range Reference Area exclosures.

Field Methods

At each RRA location, samples were collected at 2 sites: in grazed (outside exclosure) and non-grazed (inside fencing) areas (Figure 2.4). At each site, 5 30 cm deep holes were augured within a 5 m x 5 m plot. Using 2 of these holes, the removed soil was collected for bulk density (BD) analysis. Each of the 5 holes were used to collect soil C samples by scraping soil from the walls of the hole at each depth increment (0-10 cm, 10-20 cm, 20-30 cm). Vegetation analysis of cover class and dominant species was recorded within each 5 m x 5 m plot; however, this data was not used for modelling because the vegetation data collected was not strongly correlated to the SC or SOC and there was no equivalent available from a remotely sensed sourced therefore the data could not be used for mapping. Instead, vegetation community was derived from a GIS layer published by the *Grasslands Conservation Council of BC* (see 'Data' section below). Photos were taken and landscape variables were measured (slope with clinometer, elevation with GPS, aspect by sight). During the second field season, a DLC Multispectral Radiometer (*MSR16R, CROPSCAN Inc.)* was used to determine spectral reflectance (5 replicates per site).

BIOGEOCHEMICAL ANALYSIS

The soil samples were dried, sifted through a 2 mm sieve, and weighed on an analytical scale before analysis through an automated elemental analyzer (*CE-440 Elemental Analyzer, Exeter Analytical Inc.*) was used to determine C percent by thermal conductivity detection. Three of the five samples were run through the analyzer individually (not bulked). If the deviation between the 3 samples was too high, the additional 2 samples were run as well. SC% values from the elemental analyzer were then converted to carbon density:

$$\text{Carbon density } (g/cm^3) = \text{bulk density } (g/cm^3) \times \text{percent carbon } (\%) \tag{5}$$

To determine dry soil BD of all samples, Equation 6 was applied. :

$$\text{Bulk density 2013 } (\text{cm}^3) = \frac{\text{mass of dry soil (g)} - \text{mass of rocks (g)}}{\text{volume of core } (\text{cm}^3) - \text{volume of rocks} (\text{cm}^3)} \tag{6}$$

$$\text{Volume of rocks} = \text{mass of rocks (g)} \times \text{standard rock density } (g/cm^3) \tag{7}$$

Where rock mass was mass of total dry soil (g) subtracted by the mass of sieved dry soil (g) and the standard rock density was 2.65 g/cm^3 (Daly, 1966). To determine the mass of dry soil, all samples were dried in a constant temperature oven (*DKN818, Yamato*) and weighed with a top loading scale. To determine the mass of sieved dry soil the samples were sifted through a 2 mm sieve to remove rocks, and weighed again.

The Loss on Ignition (LOI) technique was used to determine the organic matter content in the soil samples. Following Wang et al. (2012), approximately 5 g of soil was placed in a weighed aluminum sample boat, heated at 105°C in a constant temperature oven (*DKN818, Yamato*) for 12 hours to remove soil moisture, then weighed with an analytical scale. Next, the soil was ignited in a programmable muffle furnace (*F26700, Barnstead International*) at 500°C for 5 hours, left in desiccator for 2 hours until room temperature, and weighed again. Soil organic matter (SOM) was calculated as the weight loss between 105°C and 375°C:

$$SOM_{LOI(gkg^{-1})} = \frac{Weight_{105C} - Weight_{500C}}{Weight_{105C}} \times 1000 \quad (8)$$

Using Wang et al.'s (2012) conversion factors, SOC may be calculated from SOM $_{LOI}$ using Equation 9:

$$SOM_{LOI} = \frac{SOM_{LOI} - 4.189}{1.792} \quad (9)$$

DATA

Data Sources

Several datasets were used to model and map SC and SOC (Table 2.1 and Figure 2.5). NDVI data derived from the Multi-Spectral Radiometer ($NDVI_{MSR}$) was used to model SC and SOC in order to demonstrate the increased modelling accuracy when using smaller scale spectral measurements; however this data could not be used for predictive mapping since $NDVI_{MSR}$ does not cover the grasslands province-wide. Models that included $NDVI_{MSR}$ data were not mapped because MSR data does not have province-wide coverage.

Pre-processing

The tiles which comprised the $NDVI_{MODIS}$ layer required significant pre-processing. The appropriate tiles were downloaded, projected in to the BC Albers Equal Area projection, and subsequently mosaicked into province-wide layers. A mosaic was produced for each 16 day composite, resulting in 69 layers from 2011-2013. To smooth out noise in NDVI data that is caused primarily by cloud contamination and atmospheric variability, layers were stacked sequentially and a pixel-by-pixel computation smoothed NDVI data over the 3 year time series using a Loess smoothing function. These functions were extremely time-consuming; for example, the smoothing function ran continuously for 3 weeks. Fortunately, to update models in the future as new $NDVI_{MODIS}$ data is released, these pre-processing steps can be easily reproduced with R code (Appendix B).

Table 2.1: GIS for soil carbon modelling and mapping.

Layer Name	Description (Year Created)	Format (Resolution)	Publisher
MAP	The average annual precipitation in millimetres for the period 1961 to 1990 (2005)	Raster (2.5 arc min)	Ministry of Forests and Range, Research Branch
MAT	The average temperature for the entire year in degrees Celsius for the period 1961 to 1990 (2005)	Raster (2.5 arc min)	Ministry of Forests and Range, Research Branch
Soil Type	Soil Development type derived from Soil Landscapes of Canada data (2008)	Vector	Agriculture and Agri-Food Canada
Soil Drainage	Describes the removal of water from the soil; derived from Soil Landscapes of Canada data (2008)	Vector	Agriculture and Agri-Food Canada
Vegetation Community	Vegetation community zones derived from Biogeoclimatic Ecosystem Classification zones (2004)	Vector	Grasslands Conservation Council of British Columbia
Aspect, Slope, Elevation	Topographic layers derived from gridded DEM created by the Terrain Resource Information Management program (2002)	Raster (1:20,000)	Base Mapping and Geomatic Services
$NDVI_{MODIS}$	Satellite derived Normalized Difference Vegetation Index data (16 day/ 250 m resolution) from Moderate-resolution Imaging Spectro-radiometer (MODIS) satellite, MOD13Q1 product (2012-2014)	Raster (250m)	USGS, MODIS Terra Land Processes Distributed Active Archive Center directory

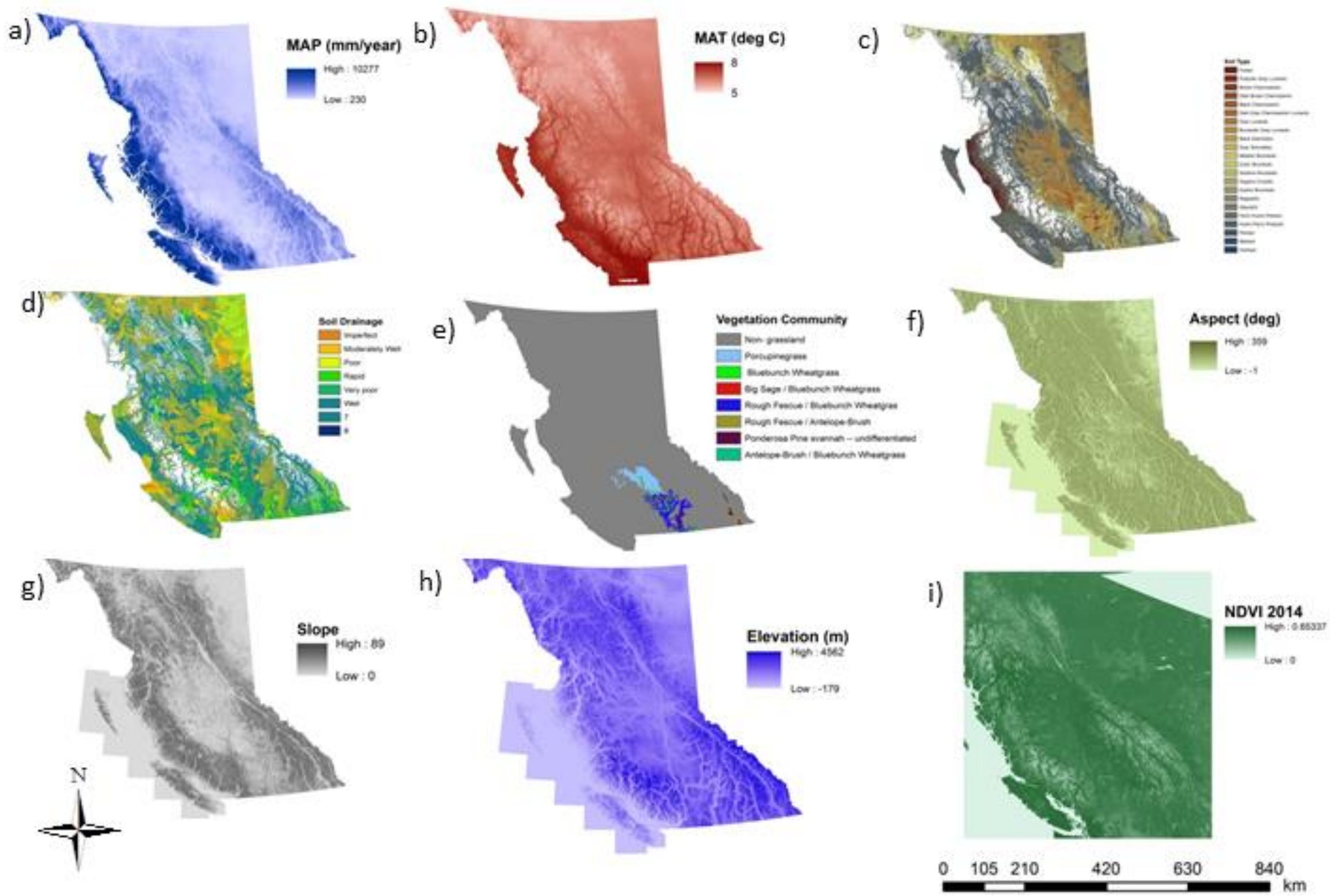

Figure 2.5: GIS data layers for soil carbon modelling and mapping including: a)Mean Annual Precipitation (MAP), b) Mean Annual Temperature (MAT), c) Soil Type, d) Soil Drainage, e) Vegetation Community, f) Aspect, g) Slope, h) Elevation, and i) Normalized Difference Vegetation Index (NDVI)

Note that given the 250 m pixel size, one MODIS pixel likely covers, at least partially, grazed and fenced areas at one location. In contrast, the MSR readings were taken separately for grazed and fenced areas.

All other layers were pre-processed using Model Builder in ArcGIS (Figure 2.6 shows Model Builder flowchart – in Modelling Section). Layers were re-projected if not already projected in the BC Albers Equal Area projection. Since the data was skewed, it was transformed using ln(n+1) to normalize it. The "+1" was used because some data points were originally zero and would create an error.

Statistical Analysis

Study design allowed for paired t-tests and one-way analysis of variance (ANOVA) tests using depth and region as factors to be tested on the grazed and fenced data from 2013 and 2014. No interactions were analyzed. Since data were ln(n+1) transformed, the ANOVA assumption of equality of variances was met. Post-hoc Tukey's HSD tests were performed on the data after ANOVAs if there were significant treatment effects. All analyses were performed using R (version 3.0.2) (R Development Core Team 2014) and the R package 'car' (Fox and Weisberg 2011).

Modelling and Mapping

Data was randomly divided into training and testing data. Two thirds of the sites were assigned as training data and one third was assigned as validation data. All factors affecting SC levels were evaluated simultaneously with the RF and SR, using the training data. The models were validated by predicting outcomes for the validation data-set and comparing with the observed data, to calculate a Mean Square Error (MSE) value. Goodness of fit was evaluated with adjusted R^2 for SR models and percent variance explained for RF models. The best SR models were selected automatically via forward-backward stepwise regression with the 'step.lm' function in R which selects for low Akaike Information Criterion (AIC) and high coefficient of determination (R^2) values. For the RF models, all variables were input into the initial model and subsequent models included only variables that with positive values from the sensitivity test, which quantifies the increase in MSE after the variable has been permutated. The final models were compared with the coefficient of determination (R^2), MSE, and AIC. To compare AIC between model types, the modified equation by Hastie et al (2001) was used to compute AIC manually:

$$\text{AIC} = MSE + s^2 \times \frac{d}{N} \tag{10}$$

Where s^2 is the squared sum of variance between the predicted and actual values of the test dataset *(N)* and *d* is the number of parameters. For SR, I is the number of variables in the output model plus 1 for variance and 1 for the intercept. For RF, *d* is calculated by:

$$d = average\ \#\ of\ times\ the\ variables\ are\ used + 1 \tag{11}$$

Where 1 is added for variance and the average number of times the variables are used in one tree of the forest was determined with the varUsed() function of the randomForest package (Breiman and Cutler, 2015) which calculates the amount of times each variable was used in the entire random forest. These values were summed and divided by the number of trees in the forest (501).

Since running province-wide calculations in R is extremely time consuming, the predictive maps created with SRs were generated with Model Builder in ArcGIS while the predictive maps created with RF models were generated in R using the RF predict function. With Model Builder, Raster Calculator was used to predict SC and SOC based on the stepwise regression equations (Figure 2.6). Finally, predictive maps were clipped to the extent of the BC Grasslands, as defined by a layer created by the Grasslands Conservation Council (2004).

Modelling and mapping was performed using ArcGIS (version 10.1) (ESRI 2012) and R (version 3.0.2) (R Development Core Team 2014) packages 'rgdal' (Bivand et al., 2015), 'raster' (Hijmans et al., 2015), 'randomForest' (Leo Breiman and Adele Cutler, 2015), and 'XML' (Lang et al., 2013). See Appendix B for script to manipulate MODIS data and create predictions based on RF models.

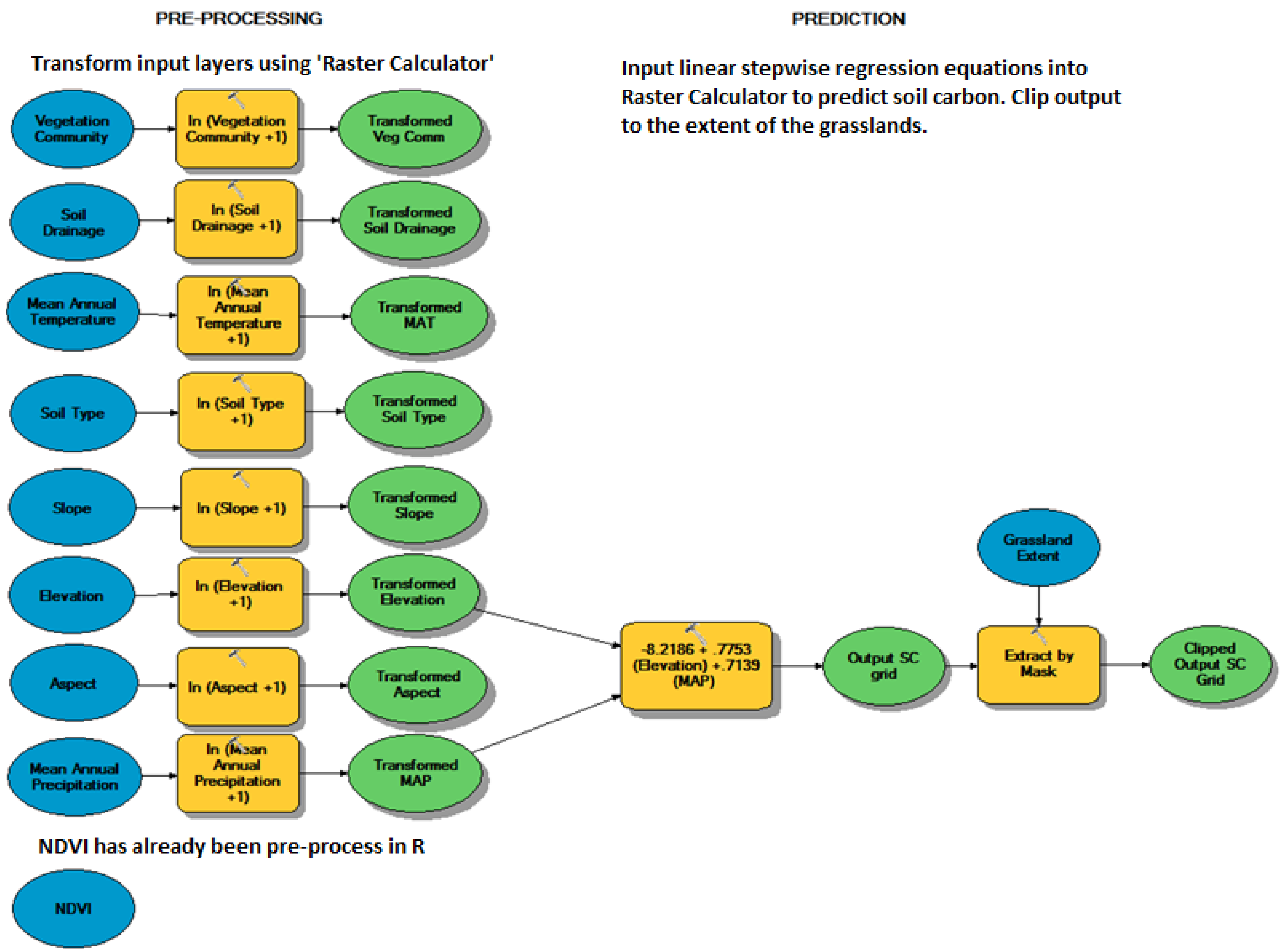

Figure 2.6: Example of model Builder flowchart for pre-processing data and creating predicted soil carbon grids with the stepwise regressions for 2013 soil carbon model of fenced systems

RESULTS

VEGETATION INDICES

Simple regressions were performed to compare various VIs versus SC and SOC results for grazed and fenced systems (Table 2.2 and Figure 2.7). MSR data was only available for 2014. MSR data for 3 sites (K004, O001, and O041) are missing due to poor weather conditions during sampling.

SC and SOC were most strongly correlated to NDVI derived from MSR data. Recall that the relatively large spatial resolution (250m) of the MODIS pixels result in a mixture of grazed (G) and fenced (F) treatments within a single pixel, and potentially land covers other than grassland. Therefore it is unsurprising that the MSR data has stronger correlations with SC and SOC. MODIS NDVI was more strongly correlated to SC and SOC than MODIS EVI, and therefore, it was selected as the model input to develop predictive carbon maps. This result is mildly surprising, since the EVI filters out signatures from background soil, and it was expected that the signature of bare soil would be significant given the low density of the grass canopy.

Table 2.2: The relationship between green biomass (derived from various spectral indices) versus SC(%) and SOC(g/kg) for 0-10cm depth.

	Year	Grazed/ Fenced	EVI_{MODIS} F (p)	EVI_{MODIS} R^2	$NDVI_{MODIS}$ F (p)	$NDVI_{MODIS}$ R^2	$NDVI_{MSR}$ F (p)	$NDVI_{MSR}$ R^2
SC	2012	G	20.52 (<0.001)	0.28	15.71 (<0.001)	0.23	N/A	
		F	15.44 (<0.001)	0.21	16.61 (<0.001)	0.23	N/A	
	2013	G	20.70 (<0.001)	0.26	14.43 (<0.001)	0.2	N/A	
		F	8.99 (3.95E-03)	0.13	8.83 (4.26E-03)	0.13	N/A	
	2014	G	19.80 (<0.001)	0.03	23.31 (<0.001)	0.35	27.96 (2.29E-06)	0.18
		F	25.45 (<0.001)	0.31	31.12 (<0.001)	0.35	38.19 (8.77E-08)	0.41
SOC	2012	G	9.24 (3.67E-03)	0.15	8.14 (6.18E-03)	0.13	N/A	
		F	11.18 (1.48E-03)	0.17	10.85 (1.72E-03)	0.16	N/A	
	2013	G	14.70 (<0.001)	0.2	14.88 (<0.001)	0.2	N/A	
		F	10.01 (2.48E-03)	0.15	11.06 (1.53E-03)	0.16	N/A	
	2014	G	10.20 (2.29E-03)	0.15	11.84 (1.09E-03)	0.17	27.95 (2.30E-06)	0.34
		F	19.91 (<0.001)	0.26	20.28 (<0.001)	0.26	47.69 (5.86E-09)	0.47

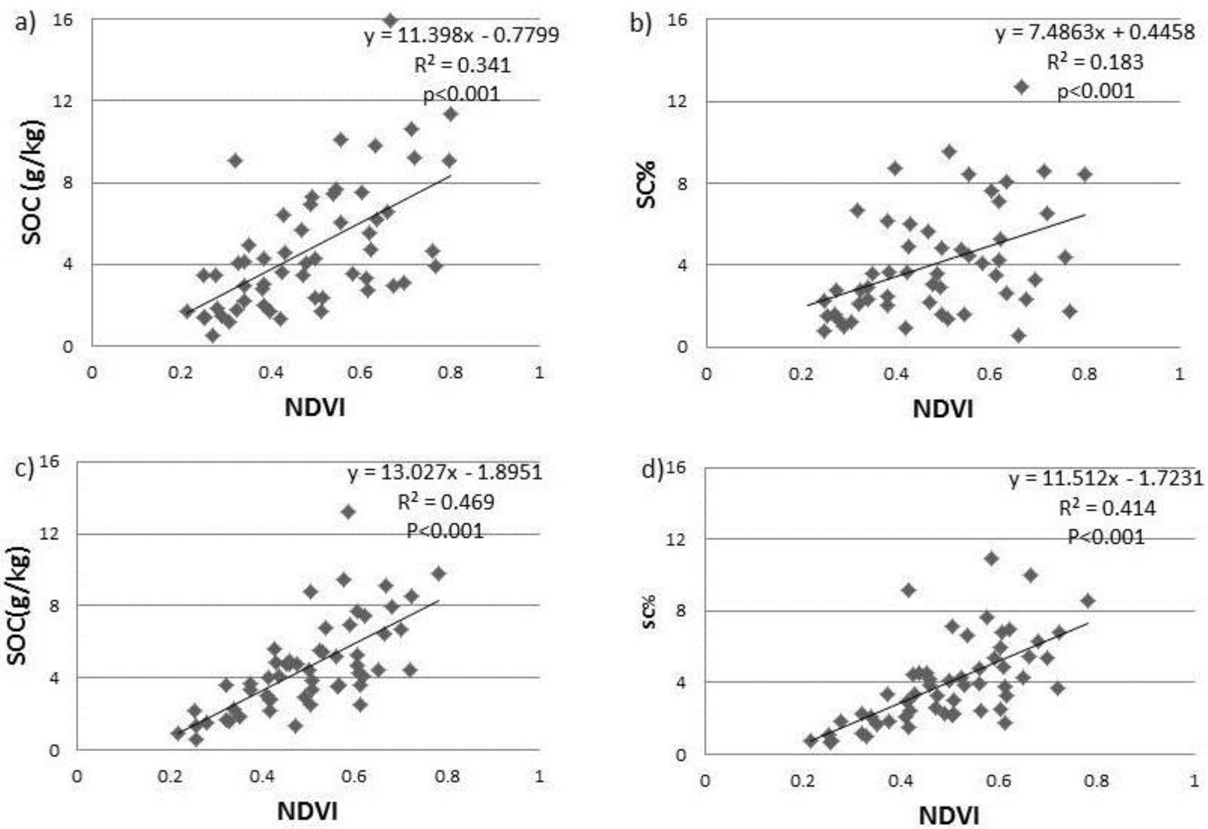

Figure 2.7: Relationships between 2014 soil carbon (SC%) and soil organic carbon (SOC(g/kg)) versus normalized difference vegetation index (NDVI) derived from the multispectral radiometer. Plots a) and b) compare SC and SOC against NDVI derived from MODIS. Plots c) and d) compare SC and SOC against the MSR.

In 2014, 3 notable outliers exist within MODIS dataset at the sampled locations (See Appendix A for list of sites, code names, and locations). They were not removed because they represent the inherent error caused by remote sensing imagery at this scale. These cases occur where sites were too close to non-grassland features to capture NDVI properly (i.e., the pixels contained mixed land cover types):

- N010 Quilchena is located between an agricultural field and a steep hill
- C010 Morrison Meadows is close to a dried pond

- C030 Bald Mt Holding exists in a small grassland clearing within a forest stand which is <100m away

Grazed Vs Un-grazed areas

T-tests showed that there is no significant differences between grazed areas versus fenced areas with respect to SOC (p=0.190) or SC (p=0.614). Factorial ANOVA results showed that there were no significant interactions between grazing and elevation/MAP/NDVI with respect to SC or SOC in 2013 or 2014.

Depth

ANOVAs were performed to compare SC and SOC in 2013 against depth as an input variable. The null hypotheses were that there were no differences for SC and SOC when compared by depth. The hypothesis was rejected at a 5% level for SC (F= 5.577, p=0.004) and SOC (F = 7.216, p=0.001). Specifically, a post hoc Tukey test has determined that there is significantly greater soil C in 0-10 cm than 10-20 cm (p=0.038) and 0-10 cm than 20-30 cm (p=0.006) and significantly greater SOC in 0-10 cm than 10-20 cm (p=0.016) and 0-10cm than 20-30 cm (p=0.001). Note that for this analysis, sites that did not reach all depth increments were excluded. Figure 2.8 shows the distribution of SC and SOC by depth at 0-10cm, 10-20cm, and 20-30cm. Results from the Post Hoc Tukey Test show which categories are significantly different.

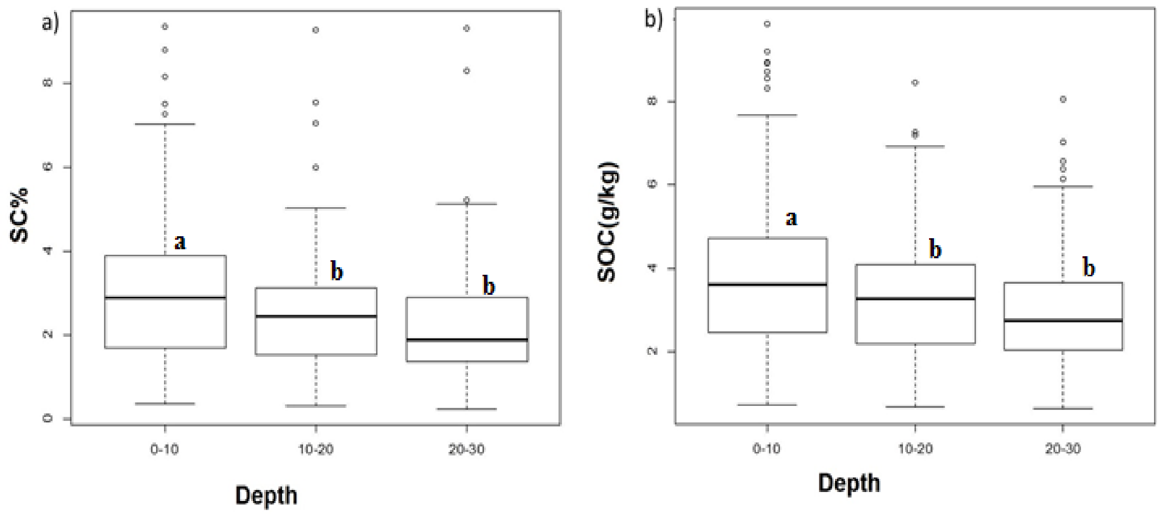

Figure 2.8: Comparison of a) Soil Carbon (SC%) and b) Soil Organic Carbon(SOC (g/kg)) over various depths (0-10cm, 10-20cm, and 20-30cm). Letters above boxes represent Post Hoc Tukey Test results where categories with different letters are significantly different.

TIME

To compare the change in SC and SOC from 2013 to 2014, a paired two-sample t-test was performed. Soil was only sampled to 10 cm in 2014; therefore, only 0-10 cm samples were used to compare the change in soil C over time because deeper soil stores less carbon and is less impacted by grazing. For the t-test, the null hypotheses were that the 2013 SC and SOC were greater or equal to the 2014 SC and SOC. These hypotheses were rejected at a 5% level ($p=0.001$ for SC and $p=0.031$ for SOC). Figure 2.9 shows the distribution of SC and SOC by year. Results from the Post Hoc Tukey Test show which categories are significantly different.

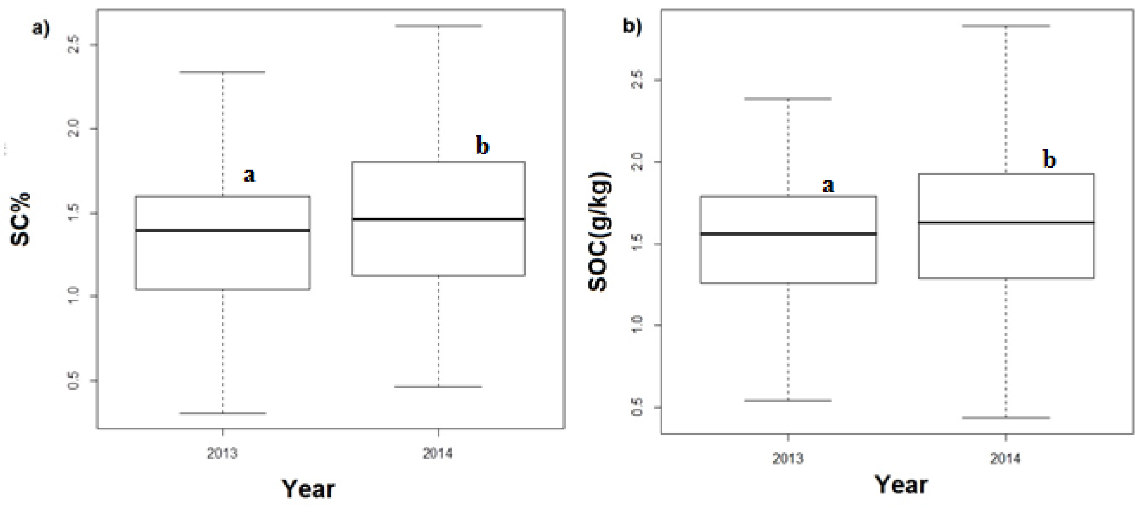

Figure 2.9 Comparison of a) Soil Carbon (SC%) and b) Soil Organic Carbon (SOC (g/kg)) between 2013 and 2014. Letters above boxes represent Post Hoc Tukey Test results where categories with different letters are significantly different.

Table 2.3 displays the SR results for change in SC and SOC from 2013 to 2014. Elevation was the only input variable that was consistently a significant variable.

Table 2.3: Stepwise regression results of change in Soil Carbon and Soil Organic Carbon from 2013 to 2014 in Grazed (G) and Fenced (F) Systems. The coefficients (Est) and significance (Sig) of each variable in the model is displayed. Sig: significance codes of p-values represented by 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.. ' 1.

	Soil Carbon				Soil Organic Carbon			
	G		F		G		F	
	Est	Sig	Est	Sig	Est	Sig	Est	Sig
Intercept	-3.57	*	-3.02	.	6.17		-5.31	*
Elevation	0.01	**	0.01	*	-0.01	**	0.01	*
Aspect								
Slope								
MAT					-0.01	.		
MAP					-0.01	.	0.01	
NDVI_MODIS					9.77			
Soil Type								
Soil Drainage					0.94	**	-0.36	
Vegetation Community								
Adjusted R^2	0.12		0.09		0.24		0.17	

Soil Carbon and Soil Organic Carbon Density

SRs were used to model SC and SOC density for grazed (G) and fenced (F) systems, and to display the change between the two systems (grazed-fenced)(C) (Tables 2.4 and 2.5). The results from 2013 do not reveal any significant variables for predicting SC and SOC density. In 2014, elevation, slope, aspect, and vegetation community were significant variables.

Table 2.4: Stepwise regression for 2013 Soil Carbon and Soil Organic Carbon Density for grazed (G) and fenced (F) systems, and the change between the two (C). The coefficients (Est) and significance (Sig) of each variable in the model is displayed. Sig: significance codes of p-values represented by 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.. ' 1..

	Soil Carbon						Soil Organic Carbon					
	G		F		C		G		F		C	
	Est	Sig	Est	Sig	Est	Sig	Est	Sig	Est	Sig	Est	Sig
Intercept	5.57	*	0.73		1.31		1.04	***	5.81		-5.24	
Elevation												
Aspect					-0.01				-0.01		0.01	.
Slope							0.10	.				
MAT									-1.84	*	1.71	*
MAP												
$NDVI_{MODIS}$									16.95	*	-16.51	
Soil Type			-0.04		-0.06							
Soil Drainage			0.15	.					0.86		-0.72	*
Veg Community					-0.11				-0.60		0.69	.
Adj R^2	0.07		0.07		0.17		0.05		0.18		0.23	

Table 2.5: Stepwise regression results for 2014 Soil Carbon and Soil Organic Carbon Density for grazed (G) and fenced (F) systems, and the change between the two (C). The coefficients (Est) and significance (Sig) of each variable in the model is displayed. Sig: significance codes of p-values represented by 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.. ' 1.

	Soil Carbon						Soil Organic Carbon					
	G		F		C		G		F		C	
	Est	Sig	Est	Sig	Est	Sig	Est	Sig	Est	Sig	Est	Sig
Intercept	-4.32	*	-2.74	*	-1.74		-2.45		-35.52	**	22.10	**
Elevation	0.00	*	0.00	*			0.00	**	0.03	***	-0.02	***
Aspect	0.01	*	0.00	.			0.00		0.04	*	-0.04	*
Slope	-0.24	*	-0.17	*			-0.29	*	-1.57	*	1.31	.
MAT					0.37	.						
MAP												
$NDVI_{MODIS}$							-3.87					
Soil Type												
Soil Drainage	0.34	.	0.29	*			0.28		1.64			
Veg Community	0.30	*	0.21	*			0.37	*	2.03	*	-1.41	.
Adj R^2	0.20		0.20		0.04		0.19		0.33		0.29	

SC and SOC Density by Region

ANOVAs were performed to compare Soil Carbon Density (SCD) and Soil Organic Carbon Density (SOCD) against Region as an input variable. The null hypothesis were that there were no differences for SCD and SOCD for grazed (G) and fenced (F) systems, and the change between G and F (C) for 2013 and 2014 when compared by region. The hypothesis was rejected at a 5% level for SOCD for grazed systems in 2014 ($F=9.3$, $p=0.000$), SOCD for fenced systems in 2014 ($F=6.538$, 0.001), SCD for grazed systems in 2014 ($F=10.45$, $p=0.000$), SCD for fenced systems in 2014 ($F=12.78$, $p=0.000$), and SOCD for grazed systems in 2013 ($F=5.05$, $p=0.003$). Significant results ($p<0.05$) were plotted in Figure 2.10 which shows the distribution of SC and SOC density by region. Results from the Post Hoc Tukey Test show which categories are significantly different.

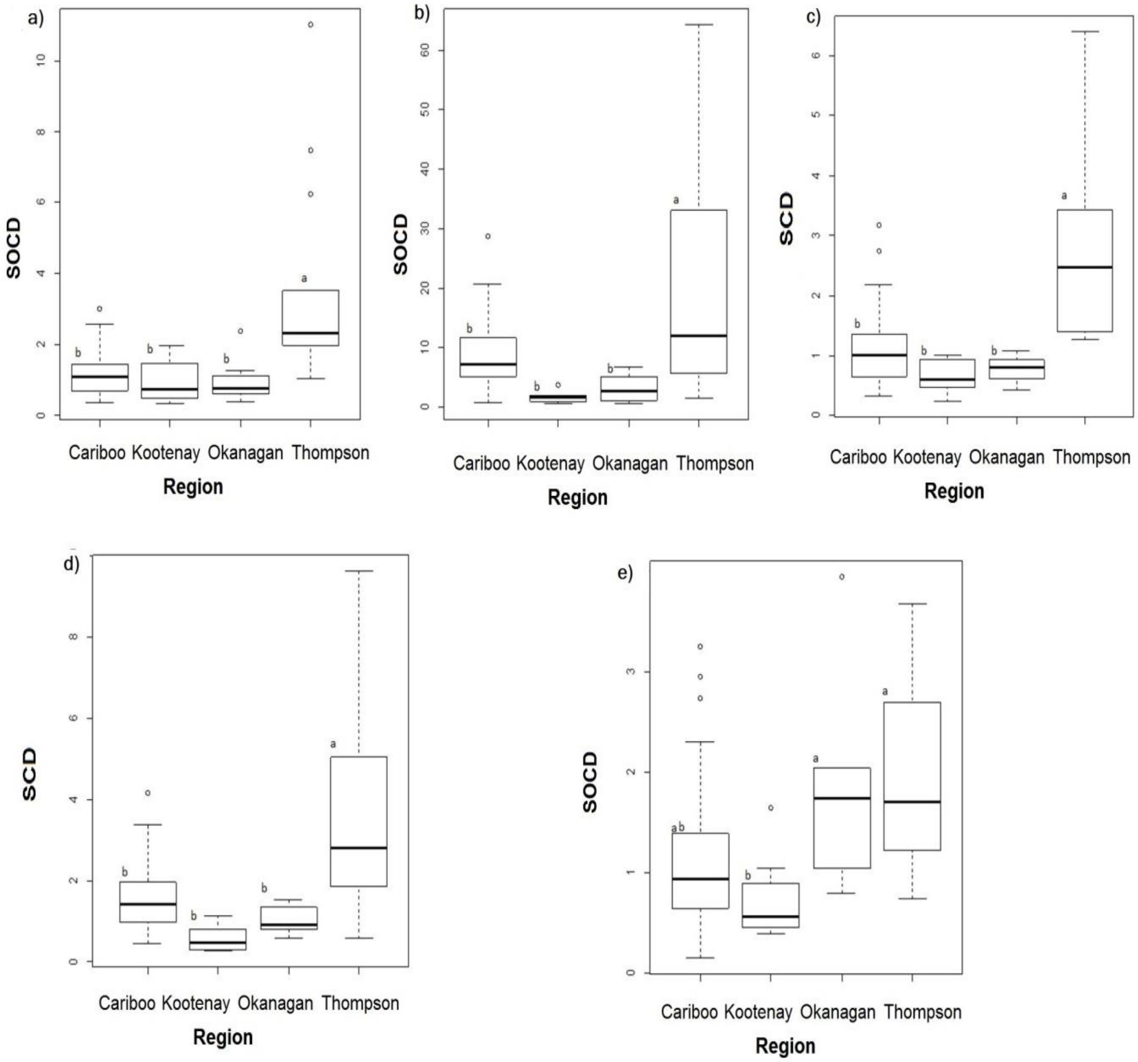

Figure 2.10: Soil Carbon and Soil Organic Carbon density by region: a) Grazed 2014, b) Fenced 2014, c) Grazed 2014, d) Fenced 2014, e) Grazed 2013.. Letters above the bars represent results from the Post Hoc Tukey Test where categories sharing the same letter are not significantly different.

STEPWISE REGRESSIONS AND RANDOM FOREST MODELS

Tables 2.6 and 2.7 show the SR results for SC and SOC for grazed and fenced systems in 2013 and 2014, respectively. The SR indicated that growing season average of NDVI, Elevation, and MAP were the most useful factors in predicting SC and SOC in 2013 and 2014 (Tables 2.8 and 2.9). Figure 2.11 displays predicted SC and SOC for 2013 and 2014 in grazed and fenced systems. Figure 2.12 displays the predicted SC values across BC grasslands based on the SR results for 2013 fenced systems (See Appendix C for larger maps). Notice the distribution of higher SC values in upper grasslands, at high elevations which are associated with more moisture and vegetation. In 2013, an interesting pattern shows that MAP is a significant in fenced systems but not in grazed systems. In 2014, higher R^2 and lower MSE and AIC values indicate that SR created better models when NDVI was derived from the MSR.

Table 2.6: Stepwise regression results for 2013 Soil Carbon and Soil Organic Carbon in grazed (G) and fenced (F) systems. The coefficients (Est) and significance (Sig) of each variable in the model is displayed. Sig: significance codes of p-values represented by 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.. ' 1.

	Soil Carbon				Soil Organic Carbon			
	G		F		G		F	
	Est	Sig	Est	Sig	Est	Sig	Est	Sig
Intercept	-4.06	*	-8.22	**	-2.23		-8.80	***
Elevation	0.52	..	0.78	**	0.38		0.80	***
Aspect								
Slope								
MAT								
MAP			0.71	**			0.82	*
$NDVI_{MODIS}$	1.89	..			1.77			
Soil Type					-0.27			
Soil Drainage	0.52	.			0.47			
Veg Community								
R^2	0.41		0.35		0.48		0.46	
MSE	0.08		0.08		0.08		0.05	
AIC	0.11		0.11		0.13		0.08	

Table 2.7: Stepwise regression for 2014 Soil Carbon and Soil Organic Carbon in grazed (G) and fenced (F) systems comparing Normalized Difference Vegetation Index derived from MODIS and the Multispectral Radiometer. The coefficients (Est) and significance (Sig) of each variable in the model is displayed. Sig: significance codes of p-values represented by 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.. ' 1.

	Model inputs include $NDVI_{MODIS}$								Model inputs include $NDVI_{MSR}$							
	Soil Carbon				Soil Organic Carbon				Soil Carbon				Soil Organic Carbon			
	G		F		G		F		G		F		G		F	
Variables	Est	Sig	Est	Sig	Est	Sig	Est	Sig	Est	Sig	Est	Sig	Est	Sig	Est	Sig
Intercept	-15.86	***	-13.50	***	-8.60	***	-13.96	***	-13.05	***	-8.67	**	-6.09	*	-11.67	***
Elevation	1.31	***	1.11	***	1.55	***	1.50	***	0.67	.	0.42	..	1.07	**	0.77	**
Aspect									0.13	..						
Slope					-0.28	*							-0.35	*		
MAT	0.75	..	0.50	..												
MAP	1.02	*	0.99	**			0.89	*	1.28	*	0.96	**			1.15	***
$NDVI_{MODIS}$	2.05	.	1.74	*					--	--	--	--	--	--	--	--
$NDVI_{MSR}$	--	--	--	--	--	--	--	--	2.96	***	3.33	***	2.23	**	2.86	***
Soil Type									0.44	*						
Soil Drainage											0.28	..				
Veg Community			-0.14	..					-0.37	**	-0.30	**				
R^2	0.56		0.68		0.49		0.59		0.70		0.77		0.62		0.73	
MSE	0.11		0.13		0.16		0.13		0.16		0.09		0.15		0.08	
AIC	0.29		0.23		0.23		0.18		0.30		0.16		0.22		0.13	

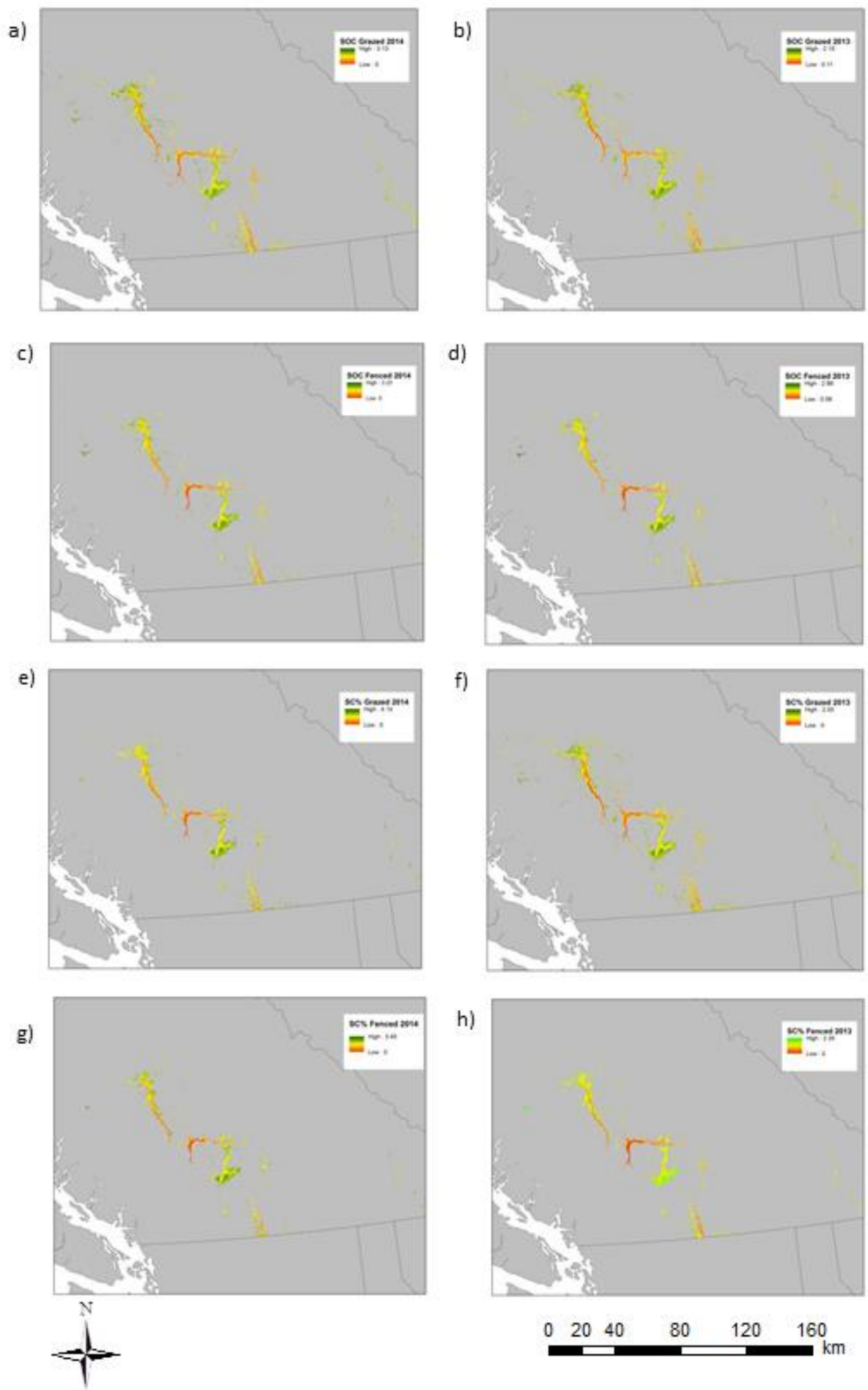

Figure 2.11: Predicted soil carbon (SC) and soil organic carbon (SOC) based on SR models: a) SOC Grazed 2014, b) SOC Grazed 2013, c) SOC Fenced 2014, d) SOC Fenced 2013, e) SC Grazed 2014, f) SC Grazed 2013, g) SC Fenced 2014, and h) SC Fenced 2013.

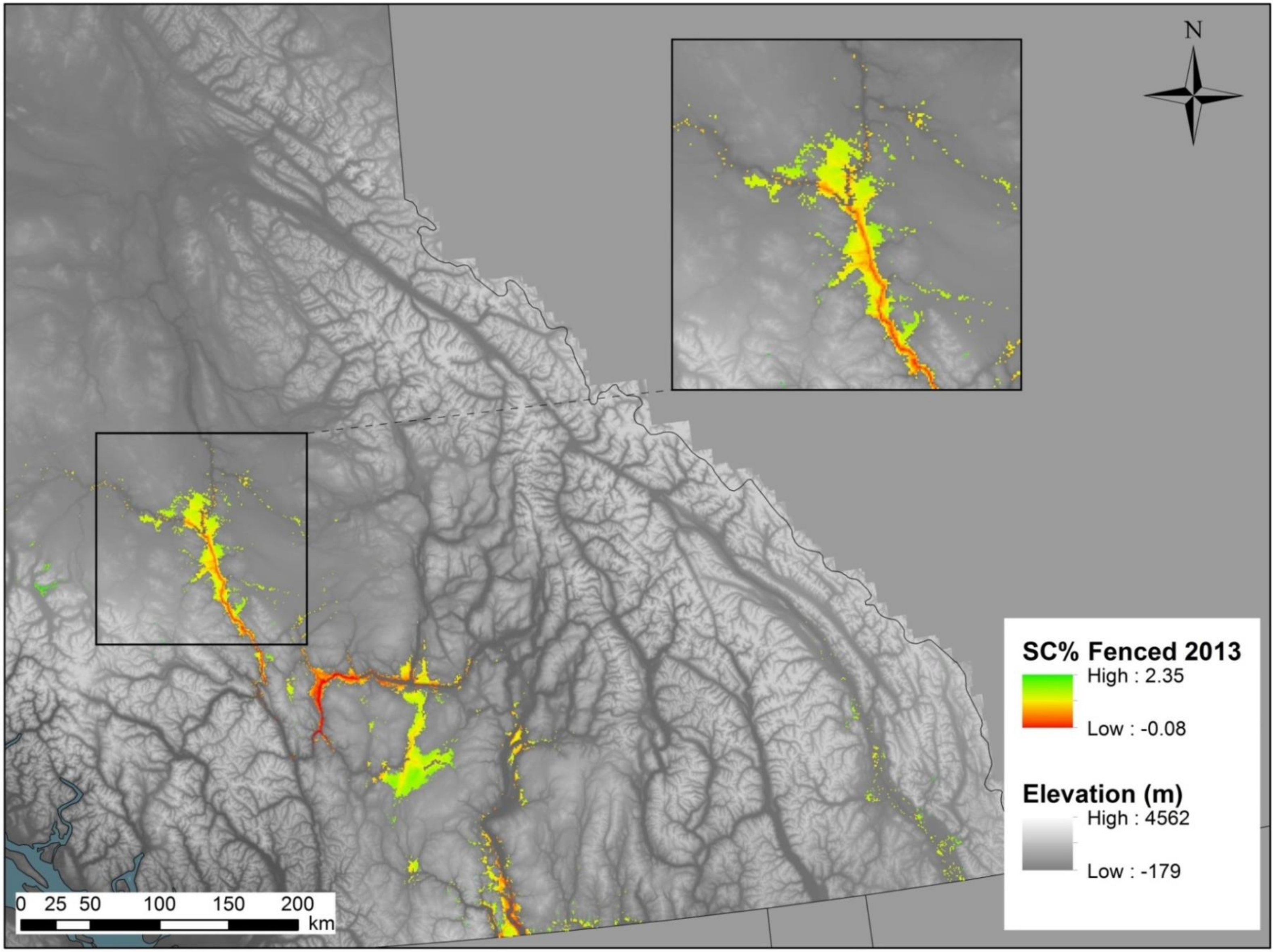

Figure 2.12: Predicted Soil Carbon (SC%) for fenced systems in 2013 based on Stepwise Regression. Soil carbon layer over-layed on elevation to show distribution in upper and lower grasslands.

Based on comparisons of R^2 and AIC values, SR produces models that explain more variance and are of better quality (R^2=0.49-0.77 and AIC = 0.30-0.13 for SR models in 2014; R^2=0.36-057 and AIC = 0.36-0.18) (Tables 2.6-2.9). In 2013, all SR models have higher R^2 and lower AIC values when compared to their RF counter-parts (Tables 2.6-2.9). In 2014, 6 of the 8 SR models have higher R^2 and lower AIC values in comparison to the RF models (Tables 2.6-2.9). Consistent with SR results, 2014 RF results indicated that the input variables Elevation, MAP, and NDVI were important in SC and SOC prediction. In 2013, RF showed soil drainage to be a more important variable than MAP or NDVI (Table 2.8) Figure 2.13 shows the predicted SC and SOC for 2013 and 2014 based on RF models. Visually, SR and RF models produced similar patterns when mapped.

Table 2.8: 2013 Random Forest results showing "%incMSE", the percent increase in Mean Square Error when variable is permutated. Variables with negative"%incMSE" values were removed from the model and therefore not displayed in the table.

	Soil Carbon		Soil Organic Carbon	
	G	F	G	F
Elevation	10.61	11.00	6.25	8.62
Aspect	1.43	1.30	2.93	1.14
Slope			2.18	2.46
MAT				
MAP		1.75	0.12	5.07
$NDVI_{MODIS}$	5.47		7.31	4.81
Soil Type			2.17	1.01
Soil Drainage	8.84	3.98	6.21	4.82
Veg Community	4.96	3.44	1.88	2.35
R^2	.28	.16	.24	.22
MSE	0.07	0.09	0.07	0.07
AIC	0.79	1.09	1.13	1.25

Table 2.9: 2014 Random Forest results showing "%incMSE", the percent increase in Mean Square Error when variable is permutated. Variables with negative"%incMSE" values were removed from the model and therefore not displayed in the table.

	Remote Sensing				RS and MSR			
	Soil Carbon		Soil Organic Carbon		Soil Carbon		Soil Organic Carbon	
	G	F	G	F	G	F	G	F
Elevation	15.13	16.24	18.95	14.33	19.55	18.57	25.89	19.26
Aspect								
Slope	4.72	7.59	9.89	6.13		4.06	9.17	6.33
MAT								0.77
MAP	7.55	8.79	0.07	7.47	7.95			2.62
$NDVI_{MODIS}$	6.08	12.88	5.06	6.39	--	--	--	--
$NDVI_{MSR}$	--	--	--	--	7.33	5.83	7.96	15.74
Soil Type								
Soil Drainage	0.69	1.14		0.42				3.16
Veg Community	4.06	6.22	3.68	3.50	4.56	7.55		3.15
R^2	0.38	0.53	0.36	0.38	0.38	0.44	0.51	0.57
MSE	0.11	0.13	0.15	0.12	0.07	0.10	0.10	0.08
AIC	0.27	0.30	0.36	0.24	0.18	0.23	0.23	0.19

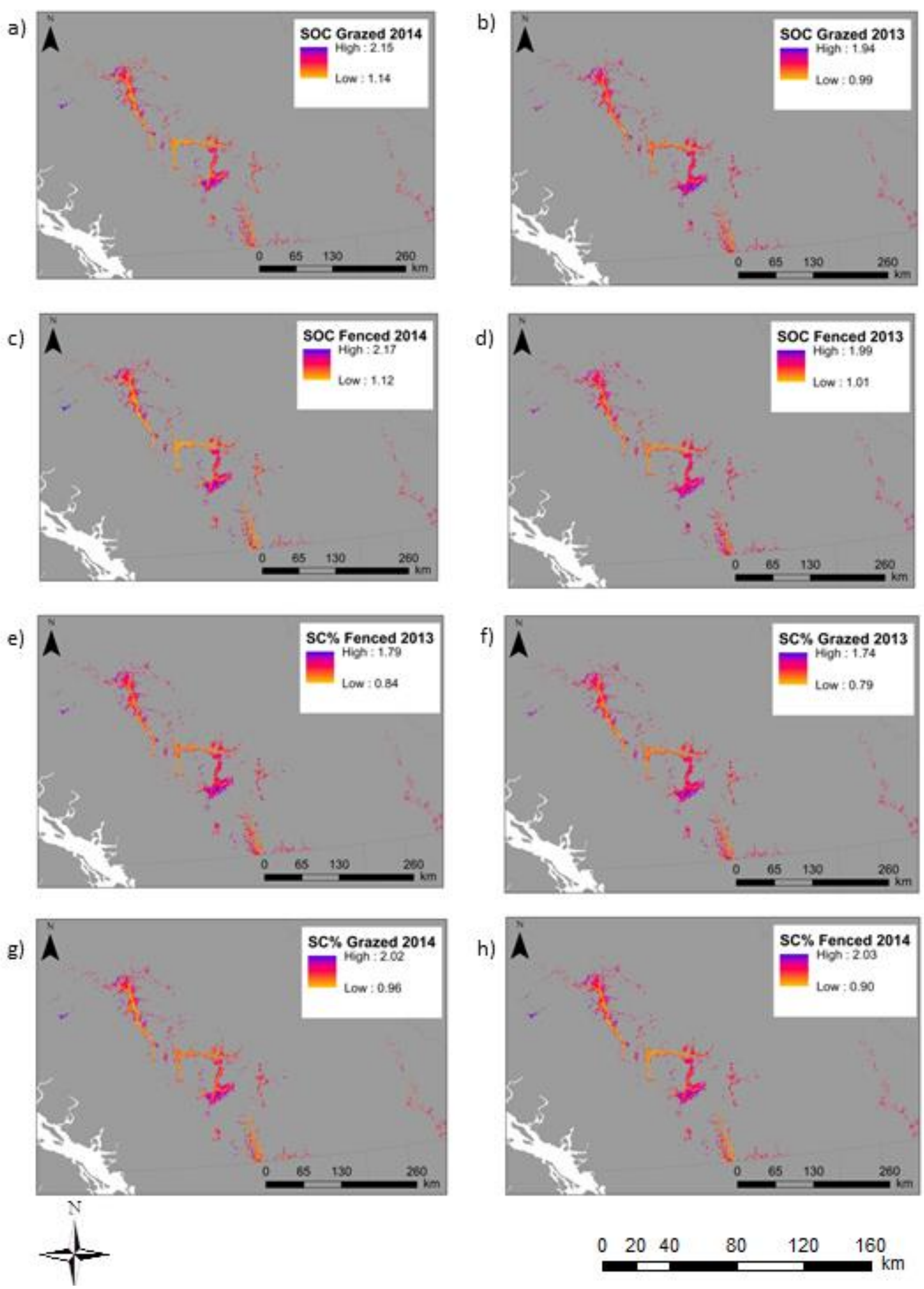

Figure 2.13: Predicted soil carbon (SC) and soil organic carbon (SOC) based on RF models: a) SOC Grazed 2014, b) SOC Grazed 2013, c) SOC Fenced 2014, d) SOC Fenced 2013, e) SC Grazed 2014, f) SC Grazed 2013, g) SC Fenced 2014, and h) SC Fenced 2013.

DISCUSSION

What environmental and anthropogenic factors allow us to best predict SC? And how is SC distributed across BC grasslands?

From the SR and RF results, the factors that best predicted SC and SOC values include elevation, MAP, and NDVI, with elevation being the most significant (Tables 2.7-2.10). Accordingly, when SC and SOC were mapped to show distribution, SC and SOC values were greatest at high elevation areas where there is more moisture and vegetation (Figure 2.11 and 2.12). In terms of vertical distribution, there are significantly greater SC and SOC stocks at 0-10 cm, compared to 10-20 cm or 20-30 cm (Figure 1.1). Since the soil profile depths across BC are unknown, no estimates of SC and SOC per area were produced. Spatial mapping was confined to 0-10cm soil depth. Understanding SC and SOC distribution in BC grasslands is a necessary input for economic models of carbon stocks and land use management. Before the ranching industry can be incorporated successfully into carbon offset programs in BC, we must know the SC and SOC potential of the land.

What factors control sensitivity to grazing in regards to SC?

SR and RF results revealed that precipitation, soil drainage, and slope differentially impact grazed versus fenced areas (Tables 2.6, 2.9, and 2.7). 2013 SR indicated MAP significantly ($p<0.01$) positively impacts SC and SOC distribution in fenced areas but not grazed areas (Table 2.6). 2014 RF indicated soil drainage controlled SOC in fenced areas but not grazed areas (Table 2.9). 2014 SR indicated that slope negatively impacts the distribution of SC and SOC in grazed areas but not fenced areas (Table 2.7). Therefore, steep areas may be more sensitive to grazing; however, further testing is needed to confirm results since this result is not consistent between years and model types.

What factors indicate high potential to store C with time?

Elevation was the only variable consistently significant as a input for SC and SOC change from 2013 to 2014 (Table 2.4). Since the greatest SC and SOC stocks exist at the highest elevations, it is logical that these areas also have the greatest potential to increase C storage over time.

How do SR models compare to RF models?

By comparing R^2 and AIC values between SR and RF models, I found that SR generally produced models that explained more of the variance and were less complex (Figures 2.7-2.10). As previously mentioned, one disadvantage of RF is that it is challenging to interpret the relationship between the input and response variables because so many DTs are produced when creating a RF model; this limits the interpretation of the relationships between the response and then input variables. For the application of predicting soil carbon in BC grasslands with the purpose of developing a carbon offset program and aiding land management, the ability of a model to be easily interpreted is important. In my comparison of SR and RF, SR is the better tool for my purpose in terms of predictability, complexity, and the ability to be interpreted.

How does increased resolution of NDVI data improve modelling?

In 2014 models, when $NDVI_{MSR}$ data was substituted for $NDVI_{MODIS}$, more variance was explained and lower MSE values were generally produced (Tables 2.6 and 2.9). Since one MODIS tile covers a 250m^2 area and likely encompasses grazed and fenced treatments at one site, SC and SOC were more strongly correlated to $NDVI_{MSR}$ than $NDVI_{MODIS}$ (Table 2.2). Therefore, it is logical that models using $NDVI_{MSR}$ data created better predictions and explained more variance. Since NDVI data is readily updated over time, improved modelling accuracy with MSR data demonstrates the potential for efficient and continuous SC and SOC monitoring with NDVI-based models using high resolution data.

Conclusion

To mitigate the effects of climate change, the idea of reducing atmospheric CO_2 by sequestration C into the terrestrial ecosystems is an area of much interest. Studies examining other regions have shown rehabilitation of rangelands to effectively prevent

SC emissions in other regions (Dean et al., 2012); however, it is expensive to alter land management practices. The development of a C offset program for ranchers may be incentive to employ (and continue to employ) sustainable practices, compensating ranchers for providing the valuable ecosystem service of C sequestration.

Past research has found both strong positive and negative grazing effects on SC, (Chapman and Lemaire, 1993; Schlesinger et al., 1990; Bremer, 2001; Loeser et al., 2007; Schönbach et al., 2011; Schuman et al., 1999). McSherry and Ritchie (2013) conclude that grazer effects on SC are highly context-specific and their causations interrelated. Due to the contradicting impacts on C sequestration, it is difficult to quantify the effect of land management (grazing). Despite this controversy, studies have shown a potential for additional C storage. Thompson et al. (2008) modeled potential rate of C sequestration by three ecosystem types over the next 100 years, concluding that agricultural lands store 0.21 GtC/year, reforestation store 0.31 GtC/year and pasture lands store 0.15 GtC/year. Conant et al. (2001) estimated that grassland ecosystems under different management scenarios would be able to sequester 0.54 MgC/year per hectare, illustrating the potential for BC grasslands to store C in order to mitigate the effects of GHGs in the atmosphere.

Chevrolet's deal with North Dakota ranchers further demonstrates the potential for carbon crediting in partnership with the ranching industry (USDS, 2014). Chevrolet's purchase was undertaken as part of their commitment to reduce eight million tons of CO_2 from being emitted (USDS, 2014). Landowners voluntarily place their land under a perpetual easement but keep their rights for grazing and growing hay. The carbon storage benefits are quantified, verified by a third party (United States Department of Agriculture), and registered as carbon credits available for purchase (by Chevrolet) (USDS, 2014). Despite this step forward, carbon offsets for ranchers are not yet available in BC. My thesis discusses the climatic and topographic restraints of SC distribution in BC's grasslands, a necessary backbone for developing a functioning economic model for C offsets and ranching.

WORKS CITED

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

Bremer DJ et al., (2001). Evapotranspiration in a Prairie Ecosystem: Effects of grazing by cattle. Agronomy Journal 93: 338-348.

Chapman, D. F., and G. Lemaire. (1993). Morphogenic and structural determinants of plant regrowth after defoliation, in Grasslands for Our World, edited by M. J. Baker, pp. 55–64, SIR, Wellington, New Zealand,

Conant, R. T., and Paustian, K. (2002). Potential soil carbon sequestration in overgrazed grassland ecosystems. Global Biogeochemical Cycles, 16(4), 1143.

Dean, C. et al. (2012). Carbon management of commercial rangelands in Australia: Major pools and fluxes. Agriculture, Ecosystems and Environment 148: 44– 64.

Daly, R. A., Manger, G. E., and Clark, S. P. (1966). Section 4: Density of rocks. Geological Society of America Memoirs, 97, 19-26.

Evans, C. R. W., Krzic, M., Broersma, K. and Thompson, D. J. (2012). Long-term grazing effects on grassland soil properties in southern British Columbia. Can. J. Soil Sci. 92: 685693.

Fornara, D. A. and Tilman, D. (2008), Plant functional composition influences rates of soil carbon and nitrogen accumulation. Journal of Ecology, 96: 314–322. doi: 10.1111/j.1365-2745.2007.01345.

USDS. (2014). "USDA and Partners Complete First-of-Its-Kind Sale of Carbon Credits from Working Ranch Grasslands". USDS Office of Communications. November 17, 2014. http://www.usda.gov/wps/portal/usda/usdamediafb?contentid=2014/11/0253.xmlandprintable=trueandcontentidonly=true

Gayton, D. (2013). British Columbia's grassland resources and climate change. Journal of Ecosystems and Management 14(2):1–16. Published by FORREX Forum for Research and Extension in Natural Resources.

Grasslands Conservation Council Report. (2009). An Ecological area assessment for Lac du Bois Grasslands, Kamloops, BC. 35p.

Han, J., and Kamber, M. (2006). Data Mining: Concepts and Techniques. San Francisco: Morgan Kauffman.

Henderson, D.C., (2000). Carbon storage in grazed prairie grasslands of Alberta. MS thesis, University of Alberta, Edmonton, AB, Canada.

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

Krzic, M., Newman, R. F., Lamagna, S. F., Bradfield, G., and Wallace, B. M. (2014). Long-term grazing effects on rough fescue grassland soils in southern British Columbia. Canadian Journal of Soil Science, 94, 1-9.

Loeser, M. R., Sisk, T. D., and Crews, T. E. (2007). Impact of grazing intensity during drought in an Arizona grassland. Conservation Biology, 21(1), 87-97.

McSherry, M. E. and Ritchie, M. E. (2013), Effects of grazing on grassland soil carbon: a global review. Global Change Biology, 19: 1347–1357.

Sapozhnikova, A. (2012). “The Effects of Precipitation and Clipping on Carbon Sequestration in Temperate Grasslands”. Unpublished masters dissertation. Thompson Rivers University, Kamloops, BC, Canada.

Schlesinger, William H. (1999). Carbon sequestration in soils. Science 284(5423), 2095.

Schönbach, P., Wan, H., Gierus, M., Bai, Y., Müller, K., Lin, L. and Taube, F. (2011). Grassland responses to grazing: effects of grazing intensity and management system in an Inner Mongolian steppe ecosystem. Plant and Soil, 340(1-2), 103-115.

Schuman, G. E., Reeder, J. D., Manley, J. T., Hart, R. H. and Manley, W. A. (1999). Impact of grazing management on the carbon and nitrogen balance of a mixed-grass rangeland. Ecological Applications. 9: 6571.

Therneau, T., Atkinson, E. (2013). An Introduction to Recursive Partitioning Using RPART Routines: Technical Report. Mayo Clinic Division of Biostatistics.

Thompson, A.M., Izaurralde, R.C., Smith, S., Clarke, L.E. (2008). Integrated estimates of global terrestrial carbon sequestration. Global Environmental Change 18: 192-203

Walkley, A., and Black, I. A. (1934). An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Science, 37(1), 29-38.

Wikeem, B.M. (1993). An overview of the forage resource and beef production non Crown land in British Columbia. Canadian Journal of Animal Science. 73: 779-794.

Winslow CJ, Hunt R Jr, Piper SC. (2003). The influence of seasonal water availability on global C_3 versus C_4 grassland biomass and its implications for climate change research. Ecological Modeling 163: 153-173.

Zhang, W. et al. (2012). Ancillary information improves kriging on soil organic carbon data for a typical karst peak cluster depression landscape. Journal of the Science of Food and Agriculture. 92(5):1094-102.

Chapter 3 : GENERAL CONCLUSIONS

Given the high coefficient of determination (R^2)values which range between 0.36 and 0.77 for soil carbon (SC) and soil organic carbon (SOC) models in 2014, this research has demonstrated the effectiveness of Normalized Difference Vegetation Index (NDVI) based models to predict SC and SOC. Since NDVI imagery derived from the MODIS satellite is updated every 16days at a 250m resolution and NDVI is highly correlated to SC and SOC, these models represent the framework for a monitoring system for SC and SOC in BC grasslands. The results will help facilitate the usage of a carbon credit program for sustainable ranching in British Columbia (BC).

LIMITATIONS

Sampling

The soil samples collected reached 30 cm, only encompassing portions of some soil profiles. Since the soil profile depths across BC are unknown and the samples collected did not capture the entire profile, no estimates of SC and SOC per area were produced. Distribution maps were only created for 0-10cm soil depth because we could not reach 30cm at many sites. Using an auger to drill a 20cm wide core was our best option for reaching as deep as possible without creating a soil pit; however, this method proved less effective in rocky terrain.

Another obstacle in regards to sampling was the clumped distribution of the Range Reference Areas (RRA) and therefore of the sampled locations; this limited the accuracy of SC and SOC prediction in grassland areas that are far from sampled RRA sites. As suggested by Henderson (2004), a collaborative effort is needed for greater coverage and model accuracy:

> “Definitive answers lay in a combination of similarly controlled experimental sites, with replicated grazing regimes, that use a single

> sampling and reporting protocol. Coordination amongst researchers in the design, execution and analysis of long-term grazing experiments will yield both accurate and precise data for meta-analysis and regional models of grazing impacts on ecosystem processes and patterns. Such coordination is necessary to provide data sets useful for developing general theory on the ecological impacts of grazing, and management prescriptions for carbon sequestration.” (Henerderon, 2004)

Understanding the Impact of Grazing

Since there was no information about grazing intensity at my study sites and no significant difference between SC or SOC in grazed versus fenced areas, the impact of grazing is difficult to interpret from the results. Several other studies carried out on rough fescue grasslands in Alberta (Johnston et al., 1971), on bluebunch wheatgrass grasslands in the southern interior of British Columbia (Evans et al., 2012), and on rough fescue grasslands in British Columbia (Krzic et al., 2013) also reported that the long-term elimination of grazing did not lead to an increase of SC when compared to grazed pastures. It is difficult to detect trends in SC difference between grazed and fenced areas for a number of reasons. First, SC and SOC are more variable between sites than between grazing treatments (fenced or grazed samples at one site). Second, the effect of grazing may be positive or negative, and these opposite effects cancel out when compare overall trends. Therefore, we must examine the differences in the models of grazed and fenced systems in order to identify in what contexts grazing increases and decreases C storage. For example, grazing may reduce SC and SOC levels in steep areas (Table 2.7).

The best units in which to report SC and SOC

There are several ways to report SC and SOC: (1) concentration as percent carbon or g kg-1, (2) carbon mass per either soil volume or area per soil depth increment, or (3) carbon mass per equivalent soil mass. The latter 2 units allow spatial scaling up of results, but carbon mass per volume may be misleading since bulk density (BD) is also influenced by grazing treatment. For example, a compact, heavy sample of soil from the uppermost 10 cm of soil may contain more C by mass than a loose sample of soil from the same depth, even though both may have the same % carbon; however, to conclude that grazing increases SOC would be incorrect because there were no corrections for treatment differences in BD. Therefore, future studies should report soil elemental concentrations on an equivalent mass basis to provide a quantitative measure independent of treatment differences in BD, as suggested by Ellert and Bettany (1995) and Carter et al. (1993).

MANAGEMENT IMPLICATIONS AND CLIMATE CHANGE

Due to their sensitivity and susceptibility to degradation, lower elevation grasslands should only be grazed with extreme caution (McCulloch, 2013). For all grassland types, grazing strategies that best maintain grassland ecosystem function should be promoted (Maestre et al., 2012) in order to prevent degradation.

Many types of grassland are already experiencing changing climate regimes that will continue to change in the future; however, for various reasons a lag phase exists in the vegetation response to the mismatch (Gayton, 2012). Wang et al. (2012) expects a new ecological climate zone, hotter and drier than anywhere currently found in British Columbia, will likely emerge in the South Okanagan/Similkameen and the hottest parts of the Thompson River valley (Gayton, 2012). This means that areas that are already sensitive to grazing with respect to plant community and C storage, will only become more sensitive.

As grasslands are affected by changing precipitation and temperature patterns, grazing systems will have to evolve to suit the plant communities that grow in the new climate regime. This will, in turn, impact carbon sequestration systems. Therefore, as climate change induced increases of seasonal temperature and decreased continue to affect the grasslands, continued monitoring of plant community change and their corresponding impact on carbon storage is necessary.

FUTURE RESEARCH DIRECTIONS

Vegetation Classification Layer

Currently, another project in the Fraser Lab is comparing the differential amounts of fine root growth by various grassland species on the effect of on SC content. Since fine roots represents a significant means for grassland ecosystems to store carbon, it would be helpful to know how strongly fine root biomass and SC content are related and if so, the species which have the greatest root mass and potentially contribute the most SC. These results may improve the model if multi or hyperspectral satellites can be spectrally unmixed to identify certain species with high SC storage potential. For instance, a larger spectral library of grassland species could be developed to add to a trial study conduction during the summer of 2014 (APPENDIX C). More simply, weights could be added to the existing Vegetation Community layer developed by the GCC to indicate species which contribute more to SC.

Hyperspectral Satellites to Quantify SC

Technologies have been developing to quantify SC rapidly using hyperspectral satellites over much larger scales. For instance, a study by Gomez et al. (2008) has compared measurements in the field with an AgriSpec portable spectrometer (350–2500 nm) and remotely from the Hyperion hyperspectral sensor onboard satellite (400–2500 nm). The spectral resolution did not change the accuracy of the model regardless of the size of SC ranges (between 0.54 and 1%, between 1.08 and 5.1%, or between 0.54 and 5.1%) or number of soil samples (56, 72 or 146) used in the prediction models (Gomez et al., 2008). These results demonstrate the potential for the use of hyperspectral remote sensing for predictions of soil organic carbon. Gomez et al. (2008) suggests the use of Environmental Mapping and Analysis Program (EnMAP) satellite for future projects. It has an onboard hyperspectral sensor which will provide high-spectral resolution observations over the wavelength range from 420 to 2450 nm (Stuffler et al., 2007). The spatial ground sampling distance will be 30 m and the Signal to Noise Ratio of EnMAP should be better than that of Hyperion. Future products will record bio-physical, bio-chemical and geo-chemical variables on a global basis.

The model could also be tweaked for smaller scale projects. To monitor specific ranch/ rangelands, a MSR could be mounted to a tractor in order to obtain a greater coverage of spectral data for SC analysis with an efficient and non-destructive technique. This approach may be appropriate for pilot studies with specific ranchers during the start-up of the provincial monitoring project. Eventually, all in situ methods can be phased out, leaving a highly efficient and low cost monitoring system in place.

Development of the Grassland Carbon Profit (GCPF) framework

The Grassland Carbon Profit model (Sapozhnikova, 2012) has been developed to represent the profit potential of the ecological service, carbon sequestration. Profit potential describes the ability of a given location to generate grassland C, based on the economics and biology of the location. It represents the net profit that could be obtained by selling the entire potential C. The results of my research will be used to update an economical model that better represents the biological factors.

CONCLUSION

Modelling and mapping SC and SOC in BC grasslands is an important step towards making a carbon offset program for ranching in BC a reality. Though there are several limitations associated with modelling SC and SOC over BC, working with the other members of the “Soil carbon sequestration in grasslands” project will help address these limitations. For example, the ‘Grazing management[2]’ stream of the project compares Management-intensive Grazing versus traditional grazing management at 7 ranches in BC. This research focuses on a smaller spatial scale but is based on a strong understanding of the grazing management practices. The data from both projects will be used in conjunction to form an economic model (the third stream of the “Soil carbon sequestration in grasslands” project).

---

[2] This research is being conducted by Dan Denesiuk (MSc Candidate at Thompson Rivers University, Kamloops, BC, Canada). For more information visit: https://grazingmgtandclimatechange.wordpress.com/research/management-intensive-grazing/

Collectively, the "Soil carbon sequestration in grasslands" project aims to quantify SC stocks in BC grasslands, determine the impact of grazing on SC sequestration, and assign monetary values to SC stocks. Ultimately, this research is working to validate that C sequestration in rangelands should be considered a viable climate change mitigation strategy and incorporated into CO_2 emissions abatement policy.

WORKS CITED

Carter, M. R. (Ed.). (1993). Soil sampling and methods of analysis. CRC Press.

Ellert, B. H., and Bettany, J. R. (1995). Calculation of organic matter and nutrients stored in soils under contrasting management regimes. Canadian Journal of Soil Science, 75(4), 529-538.

Evans, C. R. W., Krzic, M., Broersma, K. and Thompson, D. J. (2012). Long-term grazing effects on grassland soil properties in southern British Columbia. Can. J. Soil Sci. 92: 685693.

Gayton, D. (2013). British Columbia's grassland resources and climate change. Journal of Ecosystems and Management 14(2):1–16. Published by FORREX Forum for Research and Extension in Natural Resources.

Gomez, C., Rossel, R. A. V., and McBratney, A. B. (2008). Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. Geoderma, 146(3), 403-411.

Hastie, T. Tibshirani, R., and Friedman, J. (2001). The Elements of Statistical Learning: Data Mining, Inference and Prediction. New York: Springer.

Henderson, D.C., (2000). Carbon storage in grazed prairie grasslands of Alberta. MS thesis, University of Alberta, Edmonton, AB, Canada

Johnston, A., J.F. Dormaar, and S. Smoliak. (1971). Long-term grazing effects on fescue grassland soils. J. Range Manage.24:185-188.

Krzic, M., Newman, R. F., Lamagna, S. F., Bradfield, G., and Wallace, B. M. (2014). Long-term grazing effects on rough fescue grassland soils in southern British Columbia. Canadian Journal of Soil Science, 94, 1-9.

Maestre, F.T., Quero, J.L., Gotelli, N.J., Escudero, A., Ochoa, V., Delgado-Baquerizo, M., et al. (2012) Plant species richness and ecosystem multifunctionality in global drylands. Science, 335, 214–8.

McCulloch, J. (2013). Effects of changing precipitation patterns and clipping on the shrub-steppe grassland plant communities of the southern interior of British Columbia. Kamloops, BC, Canada.

Sapozhnikova, A. (2012). The Effects of Precipitation and Clipping on Carbon Sequestration in Temperate Grasslands. Kamloops, BC, Canada.

Stuffler, T., Kaufmann, C., Hofer, S., Förster, K. P., Schreier, G., Mueller, A., and Haydn, R. (2007). The EnMAP hyperspectral imager—an advanced optical payload for future applications in Earth observation programmes. Acta Astronautica, 61(1), 115-120.

Appendix A: Site Codes and Locations

I created site codes based on the region the sites are located in: boundary (B), Cariboo-Chilcotin (C), Kootenay (K), Okanagan (O), Thompson- Nicola (T). Site names were developed by the RRA program at the Ministry if Forests and Range. Coordinates were recorded with a GPS unit on site (Geographic Coordinate System, NAD83, in decimal degrees).

Site_ID	Name	Latitude	Longitude
B001	Johnstone Creek	49.0532	-119.048
B002	Overton-Moody	49.00883	-118.284
B042	Murray Gulch	49.03262	-118.792
C001	Wild Goose Lake	51.44033	-121.951
C002	Cow Camp	51.27158	-121.604
C003	Little White Lake	51.27887	-121.709
C004	Cottonwood Corrals	51.6101	-122.405
C005	Big Flat	51.62623	-122.406
C006	Cultus Lake	51.671	-122.393
C007	Alex Lake	51.61645	-122.659
C008	Cow Lake	51.76267	-122.658
C010	Morrison Meadow	52.40975	-125.15
C011	Polywog Lake	51.93745	-124.465
C012	Villa	52.07172	-123.49
C013	Punti Lake	52.21577	-123.92
C014	Stone Pasture Lower	51.95923	-123.189
C015	Haines Lak	51.95227	-123.263
C017	Tsuh Lake	51.87788	-123.283
C018	Snake Pit	51.98355	-122.415
C019	Loran C	51.98532	-122.396
C023	Big Sage Farwell	51.8245	-122.546
C024	Needlegrass Farwell	51.83223	-122.548
C025	Mile 35	51.87353	-122.528
C026	Thaddeus Lake	51.93742	-122.669
C027	Dog Lake	51.94458	-122.633
C028	Wineglass	51.8984	-122.609
C029	Bald Mountain Big B	51.96017	-122.617
C030	Bald Mountain Holding	51.92605	-122.589

C031	N Long Lake	51.91193	-122.557
C032	Toosey	51.94998	-122.496
C033	Cotton Lake	51.95127	-122.479
C034	Alkali Creek	51.83153	-122.144
C035	Joes Lake	51.75465	-122.214
C036	Sting and Vert	51.6527	-122.166
C037	Vert Lake	51.62205	-122.19
K001	Gold Creek	49.08118	-115.24
K002	Bronze Lake	49.44472	-115.393
K003	Bull River	49.49163	-115.43
K004	Skookumchuck	49.88255	-115.764
K005	Rushmere Rd	50.4119	-115.956
K043	Premier Ridge	49.87078	-115.672
K044	Buck	49.2133	-115.265
K045	Sun Lakes	50.20442	-115.895
N001	Drum	50.09323	-120.674
N002	Minnie W	50.0313	-120.403
N003	Minnie E	50.03133	-120.4
N004	Summit N	50.06355	-120.429
N005	Summit S	50.06345	-120.429
N006	Hamilton Fork	50.08662	-120.451
N007	Goose Lake	50.10312	-120.427
N008	Stipa Rich	50.06648	-120.447
N009	Stipa Nel	50.0784	-120.449
N010	Quilchena	50.16768	-120.491
O001	Crump	49.63168	-119.856
O002	McLellan	49.32847	-119.628
O003	Hayes Lease	49.0953	-119.526
O004	Chopaka	49.01167	-119.676
O005	East Chopaka	49.00997	-119.61
O041	Roddy Flats	49.5834	-119.782
T001	CDA LG 2	50.73817	-120.427
T002	CDA Lower Grazing	50.73877	-120.433
T014	CDA-M	50.76518	-120.434
T016	LDB Pond	50.78687	-120.449
T017	Frolek	50.81632	-120.439
T038	West Mara	50.74488	-120.496
T040	Tranquille 1981	50.73293	-120.517

APPENDIX B: SCRIPT

The script below is broken down into 5 sections:

1) Download MODIS data
2) Mosaic MODIS data
3) Process MODIS data and extract data from site locations
4) Loess smooth MODIS data and create raster of growing season averages for each year
5) Use Random Forest to Generate Soil Carbon and Soil Organic Carbon Predictions

```
############################################################
############################################################

## 1) Download MODIS data
## Created by David Hill Sept. 27, 2013

############################################################

# SPECIFY PREFERENCES
WD ='F:/MScThesis/IndStd/Rproj/'
WD
HDFDIR= paste(WD,'hdf/',sep='')    # LOCAL DIRECTORY TO HOLD HDF FILES
HDFDIR
YEARS = 2011:2014              # APPLICABLE YEARS AS VECTOR
# MODIS tile codes
TILES = c('h09v03', 'h09v04',
          'h10v02', 'h10v03', 'h10v04',
          'h11v02', 'h11v03',
          'h12v02', 'h12,v03')
PROD <- "MOD13Q1"              # PRODUCT IDENTIFIER
# Web address of product
PRODURL<- "http://e4ftl01.cr.usgs.gov/MOLT/MOD13Q1.005/"

############################################################

# load XML library
# The XML library permits us to parse XML documents
# HTML is a XML-like language
# install.packages("XML")
library(XML)

############################################################

# GET AVAILABLE DATES FOR APPLICABLE YEARS
DATES=NULL
```

```
for( i in 1:length(YEARS) ){
	html.tree=htmlTreeParse(PRODURL, useInternalNodes=TRUE)
	nodes <- getNodeSet(html.tree, "//a[@href]")
	links <- sapply(nodes, function(x) x <- xmlAttrs(x)[[1]])
	PAT = paste(YEARS[i], '.*.*/', sep='')
	tmp<-grep(x=links, pattern=PAT, value=TRUE)
	DATES=c(DATES,tmp)
}

DATES
##########################################################

#  GET APPLICABLE HDF FILES
NAMES=NULL
for( i in 1:length(DATES) ){
	######################################
	# get filenames
	URL=paste(PRODURL,DATES[i],sep='')
	html.tree=htmlTreeParse(URL, useInternalNodes=TRUE)
	nodes <- getNodeSet(html.tree, "//a[@href]")
	links <- sapply(nodes, function(x) x <- xmlAttrs(x)[[1]])
	TMP=NULL # define TMP OUTSIDE OF LOOP FOR VARIABLE SCOPING
	for(j in 1:length(TILES) ){
		PATTERN=paste(PROD,'.*.',TILES[j],'.*.hdf', sep='')
		TMP=c(TMP,grep(x=links, pattern=PATTERN, value=TRUE))

	}
	# strip XML hangers-on
    TMP=grep(x=TMP,pattern="*.xml",value=TRUE, invert=TRUE)
TMP


NAMES
	######################################
	# DOWNLOAD FILES
	for( j in 1:length(TMP) ){
		Resource=paste( PRODURL, DATES[i],TMP[j], sep='')
		Destination=paste(HDFDIR, TMP[j], sep='')
		print(sprintf('Putting %s in location %s', Resource, Destination ) )
		download.file(Resource, Destination, mode='wb')
	}
}

write.table( x=NAMES, file= 'F:/MScThesis/IndStd/Rproj/hdffilelist.txt',
row.names=FALSE, col.names=FALSE)
```

```
############################################################
############################################################

## 2) Mosaic MODIS files
##  Created by David Hill Sept. 27, 2013

############################################################

# SPECIFY PREFERENCES
# working directory
WD ='F:/MScThesis/IndStd/Rproj/'
# LOCAL DIRECTORY TO HOLD HDF FILES
HDFDIR=paste(WD,'hdf/',sep='')
# LOCAL DIRECTORY HOLDING MRT
MRTDIR = paste(WD,'MRT/MRT_Win/bin/',sep='')
# LOCAL DIRECTORY TO HOLD MOSAIC
MOSDIR = paste(WD,'mosaic/',sep='')
# FILE HOLDING TILE FILE NAMES
TFILE = paste(WD,'hdffilelist.txt', sep='')

############################################################

# load rgdal library
# rgdal provides bindings to Frank Warmerdam's
# Geospatial Data Abstraction Library (GDAL) (>= 1.6.3)
# and access to projection/transformation operations
# from the PROJ.4 library.
install.packages('rgdal')
library(rgdal)

############################################################

# read names
NAMES = read.table(TFILE, header=FALSE, as.is=TRUE)
NAMES

############################################################

# mosaic the blocks
for( T in 1:length( NAMES[,1] ) ){
#T=1 # for debug
	# Filename base for mosaic files
	MOSname=paste(MOSDIR,NAMES[T,1], sep='')
	moslist = file(paste(MOSname, ".list.prm", sep=""), open="wt")
	write(paste(HDFDIR, NAMES[T,2], sep=""), moslist)
	for( i in 3:length(NAMES[T,]) ){
```

```
            write(paste( HDFDIR,NAMES[T,i], sep=""), moslist)
        }
        close(moslist)
        # generate temporary mosaic:
        COMMAND=paste(MRTDIR, 'mrtmosaic.exe -i ', MOSname, '.list.prm -s "1 0 0 0 0
0 0 0 0 0 0" -o ', MOSname, '.mosaic.hdf', sep="")
        shell( cmd=COMMAND)

############################################################
        # See Appendix A of MRT documentation for
        # format of parameter file
        # resample to epsg=3005
        # This is BC Albers Equal Area Conic
        parfile = file(paste(MOSname, ".proj", ".prm", sep=""), open="wt")
        write(paste('INPUT_FILENAME = ', MOSname,'.mosaic.hdf', sep=""), parfile)
        write('  ', parfile, append=TRUE)
        write('SPECTRAL_SUBSET = ( 1 )', parfile, append=TRUE)
        write('  ', parfile, append=TRUE)
        write('SPATIAL_SUBSET_TYPE = OUTPUT_PROJ_COORDS', parfile,
append=TRUE)
        write('  ', parfile, append=TRUE)
        write('SPATIAL_SUBSET_UL_CORNER = ( 637278.0 1701350.0 )', parfile,
append=TRUE)
        write('SPATIAL_SUBSET_LR_CORNER = ( 1907278.0 335350.0 )', parfile,
append=TRUE)
        write('  ', parfile, append=TRUE)
        write(paste('OUTPUT_FILENAME = ', MOSname, '.mosaic.tif', sep=""), parfile,
append=TRUE)
        write('  ', parfile, append=TRUE)
        write('RESAMPLING_TYPE = NEAREST_NEIGHBOR', parfile, append=TRUE)
        write('  ', parfile, append=TRUE)
        write('OUTPUT_PROJECTION_TYPE = AEA', parfile, append=TRUE)
        write('  ', parfile, append=TRUE)
        write('OUTPUT_PROJECTION_PARAMETERS = ( ', parfile, append=TRUE)
        write(' 0.0 0.0 50.0', parfile, append=TRUE)
        write(' 58.5 -126.0 45.0', parfile, append=TRUE)
        write(' 1000000.0 0.0 0.0', parfile, append=TRUE)
        write(' 0.0 0.0 0.0', parfile, append=TRUE)
        write(' 0.0 0.0 0.0 )', parfile, append=TRUE)
        write('  ', parfile, append=TRUE)
        write('DATUM = NAD83', parfile, append=TRUE)
        write('  ', parfile, append=TRUE)
        write('OUTPUT_PIXEL_SIZE = 250', parfile, append=TRUE)
        write('  ', parfile, append=TRUE)
        close(parfile)
```

```
    # Run resampler
        COMMAND=paste(MRTDIR, 'resample -p ', MOSname, ".proj.prm", sep="")
        shell(cmd=COMMAND)
        GDALinfo(paste(MOSname,".mosaic.250m_16_days_NDVI.tif", sep='') )

}

############################################################
############################################################

##  3) Process rasters to create time-series of values from mosaics
##Created by David Hill

############################################################

WD = 'F:/MScThesis/IndStd/Rproj/'
MOSDIR = paste(WD,'mosaic/', sep= '')

############################################################

# load raster library
# raster provides raster data processing support
install.packages('raster')
library(raster)

#############################################################
load rgdal library
# rgdal provides bindings to Frank Warmerdam's
# Geospatial Data Abstraction Library (GDAL) (>= 1.6.3)
# and access to projection/transformation operations
# from the PROJ.4 library.
install.packages('rgdal')
library(rgdal)

#file holding site locations
LOCATIONS = paste (WD,'ndvisitelocations.txt', sep='')
LATLONG = read.table(LOCATIONS, header=FALSE, as.is=TRUE)
LATLONG
ALLNDVIplot = NULL
for( j in 1:length( LATLONG[,1] ) ){
#j=2 # for debug # this and dont run above
        TargetLat = LATLONG [j,2]
        #TargetLat = 50.67611
        TargetLon = LATLONG [j,3]
        #TargetLon = -120.3408
```

```
	# FILE HOLDING TILE FILE NAMES
	TFILE = paste(WD,'hdffilelist.txt', sep='')
	TFILE


	################################
	# Reproject Target Lat/Long to BC Albers
	ptsBCAlb<-project(cbind(TargetLon,TargetLat),   "+init=epsg:3005")
	ptsBCAlb


	#########################################
	# read names
	NAMES = read.table(TFILE, header=FALSE, as.is=TRUE)
	NAMES


	########################################
	# Process each image and extract EVI
	NDVI=NULL
	for( i in 1:length( NAMES[,1] ) ){
	 RFILE=paste(MOSDIR, NAMES[i,1], '.mosaic.250m_16_days_NDVI.tif',sep='')
		MOSAIC<-raster( RFILE )
		idxCol=colFromX(MOSAIC,ptsBCAlb[1])
		idxRow=rowFromY(MOSAIC,ptsBCAlb[2])
		datestamp = as.character(NAMES[i,1])
		year = as.numeric( substr(datestamp,1,4) )
		month = as.numeric( substr(datestamp,6,7) )
		day = as.numeric( substr(datestamp,9,10) )

		#jdn is julian day
		jdn = 
as.numeric(as.Date(sprintf("%d/%d/%d",month,day,year),format="%m/%d/%Y"))+2
440588
		#  NDVI stored in raster as integer, we need to multiply by
		#0.0001 to convert to actual NDVI value.  See scale factor on page 10 in
		#  MODIS MOD13 product documentation (same as EVi scale factor)
		dataRow = c( year, month, day, MOSAIC[idxRow,idxCol]*0.0001 )
		NDVI=rbind(NDVI,c( year, month, day, jdn, 
MOSAIC[idxRow,idxCol]*0.0001 ) )
	}


############################################################
```

```
############################################################

## 4) Generate data-frame of all NDVI, use LOESS function to smooth data across the 2011-2014 time span at each pixel, calculate average NDVI for each year's growing season, populate raster with NDVI values for each year
##Created by Heather Richardson

#############################################################

#set preferences and install packages
install.packages('rgdal')
library(rgdal)
install.packages('raster')
library(raster)
memory.limit() #increase memory or data frame cant be created

#############################################################

# create data frame of all NDVI data to I can loess data across #the 2011-2014 time span at each pixel
#load all rasters of mosaicked NDVI at each time period
#use values function to list data cell by cell (row major)
r<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2011.09.30.mosaic.250m_16_days_NDVI.tif"))
s<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2011.10.16.mosaic.250m_16_days_NDVI.tif"))
t<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2011.11.01.mosaic.250m_16_days_NDVI.tif"))
u<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2011.11.17.mosaic.250m_16_days_NDVI.tif"))
v<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2011.12.03.mosaic.250m_16_days_NDVI.tif"))
w<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2011.12.19.mosaic.250m_16_days_NDVI.tif"))
x<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2012.01.01.mosaic.250m_16_days_NDVI.tif"))
y<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2012.01.17.mosaic.250m_16_days_NDVI.tif"))
z<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2012.02.02.mosaic.250m_16_days_NDVI.tif"))
aa<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2012.02.18.mosaic.250m_16_days_NDVI.tif"))
ab<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2012.03.05.mosaic.250m_16_days_NDVI.tif"))
ac<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-2014/2012.03.21.mosaic.250m_16_days_NDVI.tif"))
```

```
ad<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.04.06.mosaic.250m_16_days_NDVI.tif"))
ae<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.04.22.mosaic.250m_16_days_NDVI.tif"))
af<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.05.08.mosaic.250m_16_days_NDVI.tif"))
ag<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.05.24.mosaic.250m_16_days_NDVI.tif"))
ah<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.06.09.mosaic.250m_16_days_NDVI.tif"))
ai<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.06.25.mosaic.250m_16_days_NDVI.tif"))
aj<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.07.11.mosaic.250m_16_days_NDVI.tif"))
ak<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.07.27.mosaic.250m_16_days_NDVI.tif"))
al<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.08.12.mosaic.250m_16_days_NDVI.tif"))
am<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.08.28.mosaic.250m_16_days_NDVI.tif"))
an<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.09.13.mosaic.250m_16_days_NDVI.tif"))
ao<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.09.29.mosaic.250m_16_days_NDVI.tif"))
ap<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.10.15.mosaic.250m_16_days_NDVI.tif"))
aq<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.10.31.mosaic.250m_16_days_NDVI.tif"))
ar<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.11.16.mosaic.250m_16_days_NDVI.tif"))
as<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.12.02.mosaic.250m_16_days_NDVI.tif"))
at<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2012.12.18.mosaic.250m_16_days_NDVI.tif"))
au<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.01.01.mosaic.250m_16_days_NDVI.tif"))
av<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.01.17.mosaic.250m_16_days_NDVI.tif"))
aw<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.02.02.mosaic.250m_16_days_NDVI.tif"))
ax<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.02.18.mosaic.250m_16_days_NDVI.tif"))
ay<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.03.06.mosaic.250m_16_days_NDVI.tif"))
az<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.03.22.mosaic.250m_16_days_NDVI.tif"))
```

```
ba<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.04.07.mosaic.250m_16_days_NDVI.tif"))
bb<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.04.23.mosaic.250m_16_days_NDVI.tif"))
bc<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.05.09.mosaic.250m_16_days_NDVI.tif"))
bd<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.05.25.mosaic.250m_16_days_NDVI.tif"))
be<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.06.10.mosaic.250m_16_days_NDVI.tif"))
bf<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.06.26.mosaic.250m_16_days_NDVI.tif"))
bg<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.07.12.mosaic.250m_16_days_NDVI.tif"))
bh<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.07.28.mosaic.250m_16_days_NDVI.tif"))
bi<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.08.13.mosaic.250m_16_days_NDVI.tif"))
bj<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.08.29.mosaic.250m_16_days_NDVI.tif"))
bk<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.09.14.mosaic.250m_16_days_NDVI.tif"))
bl<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.09.30.mosaic.250m_16_days_NDVI.tif"))
bm<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.10.16.mosaic.250m_16_days_NDVI.tif"))
bn<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.11.01.mosaic.250m_16_days_NDVI.tif"))
bo<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.11.17.mosaic.250m_16_days_NDVI.tif"))
bp<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.12.03.mosaic.250m_16_days_NDVI.tif"))
bq<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2013.12.19.mosaic.250m_16_days_NDVI.tif"))
br<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.01.01.mosaic.250m_16_days_NDVI.tif"))
bs<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.01.17.mosaic.250m_16_days_NDVI.tif"))
bt<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.02.02.mosaic.250m_16_days_NDVI.tif"))
bu<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.02.18.mosaic.250m_16_days_NDVI.tif"))
bv<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.03.06.mosaic.250m_16_days_NDVI.tif"))
bw<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.03.22.mosaic.250m_16_days_NDVI.tif"))
```

```
bx<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.04.07.mosaic.250m_16_days_NDVI.tif"))
by<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.04.23.mosaic.250m_16_days_NDVI.tif"))
bz<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.05.09.mosaic.250m_16_days_NDVI.tif"))
ca<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.05.25.mosaic.250m_16_days_NDVI.tif"))
cb<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.06.10.mosaic.250m_16_days_NDVI.tif"))
cc<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.06.26.mosaic.250m_16_days_NDVI.tif"))
cd<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.07.12.mosaic.250m_16_days_NDVI.tif"))
ce<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.07.28.mosaic.250m_16_days_NDVI.tif"))
cf<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.08.13.mosaic.250m_16_days_NDVI.tif"))
cg<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.08.29.mosaic.250m_16_days_NDVI.tif"))
ch<- values(raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2014.09.14.mosaic.250m_16_days_NDVI.tif"))

# add raster values to dataframe and transpose df so each row represents a different
time period
t.df <- t(data.frame(r,     s,    t,    u,    v,    w,    x,    y,    z,    aa,
    ab,   ac,   ad,   ae,   af,   ag,   ah,   ai,   aj,   ak,   al,
    am,   an,   ao,   ap,   aq,   ar,   as,   at,   au,   av,   aw,   ax,
    ay,   az,   ba,   bb,   bc,   bd,   be,   bf,   bg,   bh,   bi,   bj,
    bk,   bl,   bm,   bn,   bo,   bp,   bq,   br,   bs,   bt,   bu,   bv,
    bw,   bx,   by,   bz,   ca,   cb,   cc,   cd,   ce,   cf,   cg,
    ch))

##############################################################

#  smooth data across the 2011-2014 time span at each pixel
#first, transpose data in data frame so each day is a column (aka the y data) ; then add
the julian days as the first column
#use this FILE to get julian dates for each pixel and create an 'x' column in the data
frame
FILE = 'F:/MScThesis/IndStd/Rproj/NDVItable.txt'
FILE
mydata <- read.table(FILE, header=TRUE, as.is=TRUE)
mydata
x <- mydata$x1[18:86] #julian dates
y <- as.matrix(t.df) #NDVI values
```

```
a = FALSE
for (i in 1:ncol(y)){
   ndvi.loess <- loess(formula(y[,i]~x), type='smooth',span = 0.1)
          ndvi.predict <- predict (ndvi.loess, data.frame(x=x))
          plot(x, ndvi.predict)
          write(ndvi.predict, file="F:/MScThesis/IndStd/Rproj/ndvi.loess.bc.dat", sep=",",
append=a, ncol=length(ndvi.predict))
          a=TRUE
}

####notes: I had to use write() instead of print() so that each iteration of the loop
would write one new row of data into the file. Print() creates a variable that grows and
must be rewritten each iteration. Since there are 28mil pixels in the raster files and
therefore 28mil iterations of the loop, the print() function eventually crashed the
computer.

##############################################################
#now read the table written in the loop, back in
# use colClasses and nrows so table reads in faster
tab5rows <- read.table("F:\\MScThesis\\IndStd\\Rproj\\ndvi.loess.bc.dat", header =
FALSE, sep=",", nrows = 5)
tab5rows
classes <- sapply(tab5rows, class)
result <- read.table("F:\\MScThesis\\IndStd\\Rproj\\ndvi.loess.bc.dat", sep=",",
header=FALSE, colClasses=classes, nrows = 27757120)

#apply .0001 conversion factor and transform log(n+1)
#get growing season averages for each year
#to create rasters with the appropriate projection, dimensions, and reference, import a
filler raster
#then populate it with the NDVI GS averages and organize it in the correct order
r1 <- log((result[,14:23]*.0001)+1)
NDVI_GS_2012 <- rowMeans(r1,na.rm=TRUE)
a<- raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2011.01.01.mosaic.250m_16_days_NDVI.tif")
values(a) = NDVI_GS_2012
writeRaster(a, "F:/MScThesis/IndStd/Rproj/NDVI_GS_2012_1", format = "GTiff")
r2 <- log((result[,37:46]*.0001)+1)
NDVI_GS_2013 <- rowMeans(r2,na.rm=TRUE)
b<- raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2011.01.17.mosaic.250m_16_days_NDVI.tif")
values(b) = NDVI_GS_2013
writeRaster(b, "F:/MScThesis/IndStd/Rproj/NDVI_GS_2013_2", format = "GTiff")
r3 <- log((result[,60:69]*.0001)+1)
NDVI_GS_2014 <- rowMeans(r3,na.rm=TRUE)
```

```
c<- raster("F:/MScThesis/IndStd/Rproj/mosaicNDVI2011-
2014/2011.02.02.mosaic.250m_16_days_NDVI.tif")
values(c) = NDVI_GS_2014
writeRaster(c, "F:/MScThesis/IndStd/Rproj/NDVI_GS_2014_2", format = "GTiff")

###############################################################

##Final notes: It would be better to first run 'Set Null' tool in ArcGIS to remove error
codes before running loess function. Also, batch clip all mosaics in ArcGIS to smallest
extent possible before running this function. In my case, it would have saved days.

################################################################
################################################################

##5) Soil Carbon Predictions with Random Forest
##Purpose: load all raster layers, create data frame for Random Forest to call 'predict'
function on, and generate predictive maps for soil carbon and soil organic carbon
##Created by Heather Richardson

################################################################
#

#set preferences and install packages
install.packages('rgdal')
library(rgdal)
install.packages('raster')
library(raster)
install.packages('randomForest')
library(randomForest)
set.seed(415) ## Because the process bags and bootstaps data, it is a good idea to set
the random seed in R before you begin. This makes your results reproducible next time
you load the code up, otherwise you can get different classifications for each run.
memory.limit() #increase memory or data frame cant be created

################################################################

#load all raster layers
#NDVI has been pre-processed in previous code and all other layers have been pre-
processed in ArcGIS
#all layers have been transformed, clipped to the same extent, and projected into BC
Albers projection with ModelBuilder
NDVI2013GS_AV = values(raster("F:\\MScThesis\\MapData\\tc_NDVI20131.tif"))
NDVI2014GS_AV =values(raster("F:\\MScThesis\\MapData\\tc_NDVI20141.tif"))
MAP = values(raster("F:\\MScThesis\\MapData\\tc_map1.tif"))
MAT = values(raster("F:\\MScThesis\\MapData\\tc_mat1.tif"))
SLOPE = values(raster("F:\\MScThesis\\MapData\\tc_slope1.tif"))
```

```
ELEVATION = values(raster("F:\\MScThesis\\MapData\\tc_elevation1.tif"))
ASPECT = values(raster("F:\\MScThesis\\MapData\\tc_aspect1.tif"))
GrassComm = values(raster("F:\\MScThesis\\MapData\\t_vegcomm1.tif"))
SoilDrain = values(raster("F:\\MScThesis\\MapData\\tc_soildrain1.tif"))
SoilType = values(raster("F:\\MScThesis\\MapData\\tc_soiltype1.tif"))

##############################################################

df <- data.frame(NDVI2013GS_AV,NDVI2014GS_AV,MAP, MAT, SLOPE, ELEVATION,
ASPECT, GrassComm, SoilDrain, SoilType)
write.table(df, file="F:/MScThesis/IndStd/Rproj/RFdf.txt", sep=",")

##############################################################

#load training data for RF
FILE = paste (WD,'2013master_transform_training.txt', sep = '')
FILE
mydata <- read.table(FILE, header=TRUE, sep="\t")
mydata

##############################################################

#run RF and get predicted values
rffit <- randomForest(SOC_0.10_LW_2013 ~ ASPECT + SLOPE + ELEVATION+ MAT+
MAP + SoilType + SoilDrain + GrassComm +  NDVI2013GS_AV, data=mydata,
importance=TRUE, ntree=501, na.action=na.roughfix)
#varImpPlot(rffit)
print(rffit)
importance(rffit)
Prediction <- predict(rffit, df)# prediction based on dataframe of all raster layers

##############################################################

#write new raster layers based on Predictions from RF
#use a random raster layer to set the appropriate extent and project
#population raster with values from prediction
a<- raster("F:\\MScThesis\\MapData\\tc_soiltype1.tif")
values(a) = Prediction
writeRaster(a, "F:/MScThesis/IndStd/Rproj/RF_socf13_3", format = "GTiff")
```

APPENDIX C: PREDICTIVE MAPS

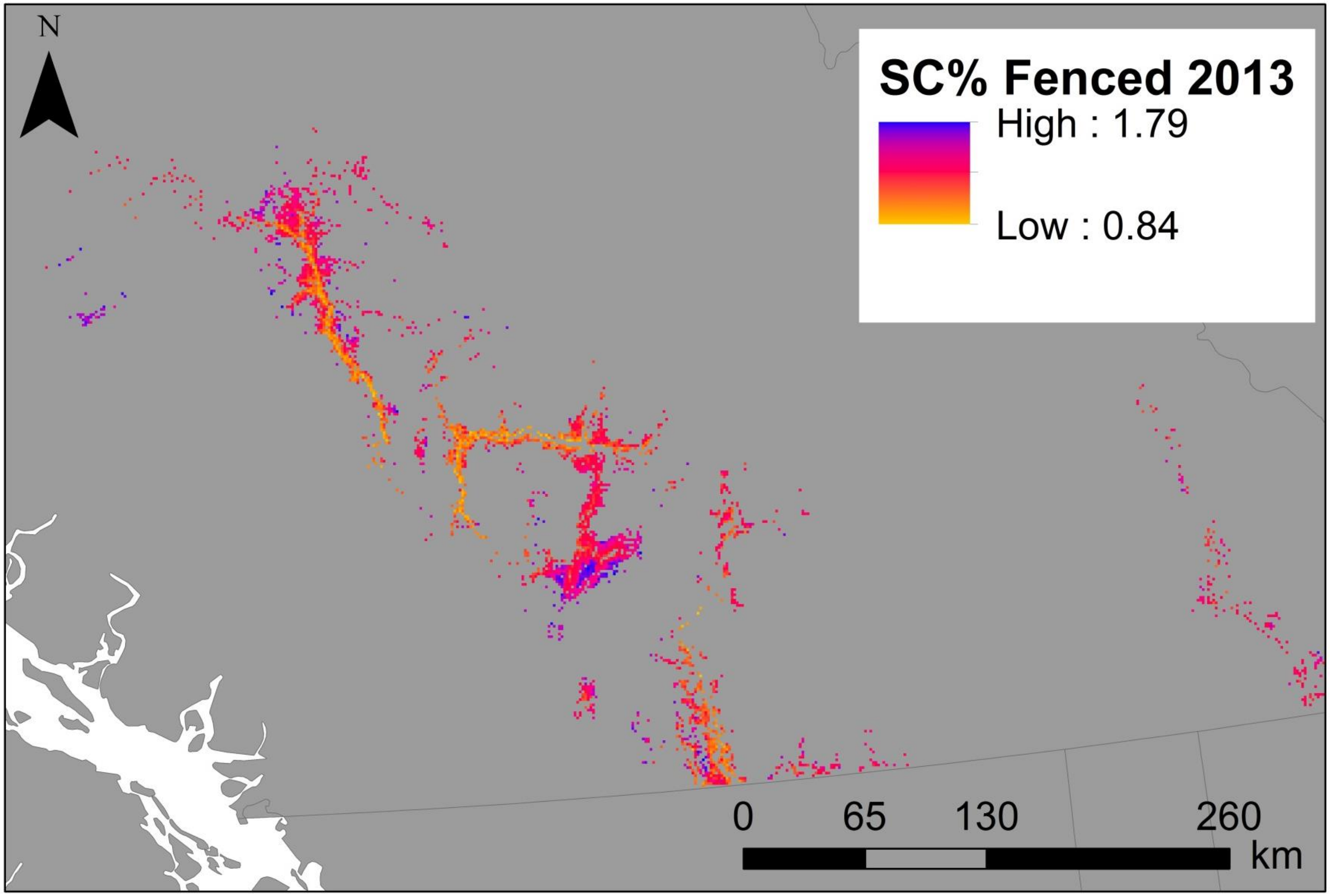

Figure C.1: Predicted Soil Carbon (SC(%)) values, based on the Random Forest results for 2013 fenced systems.

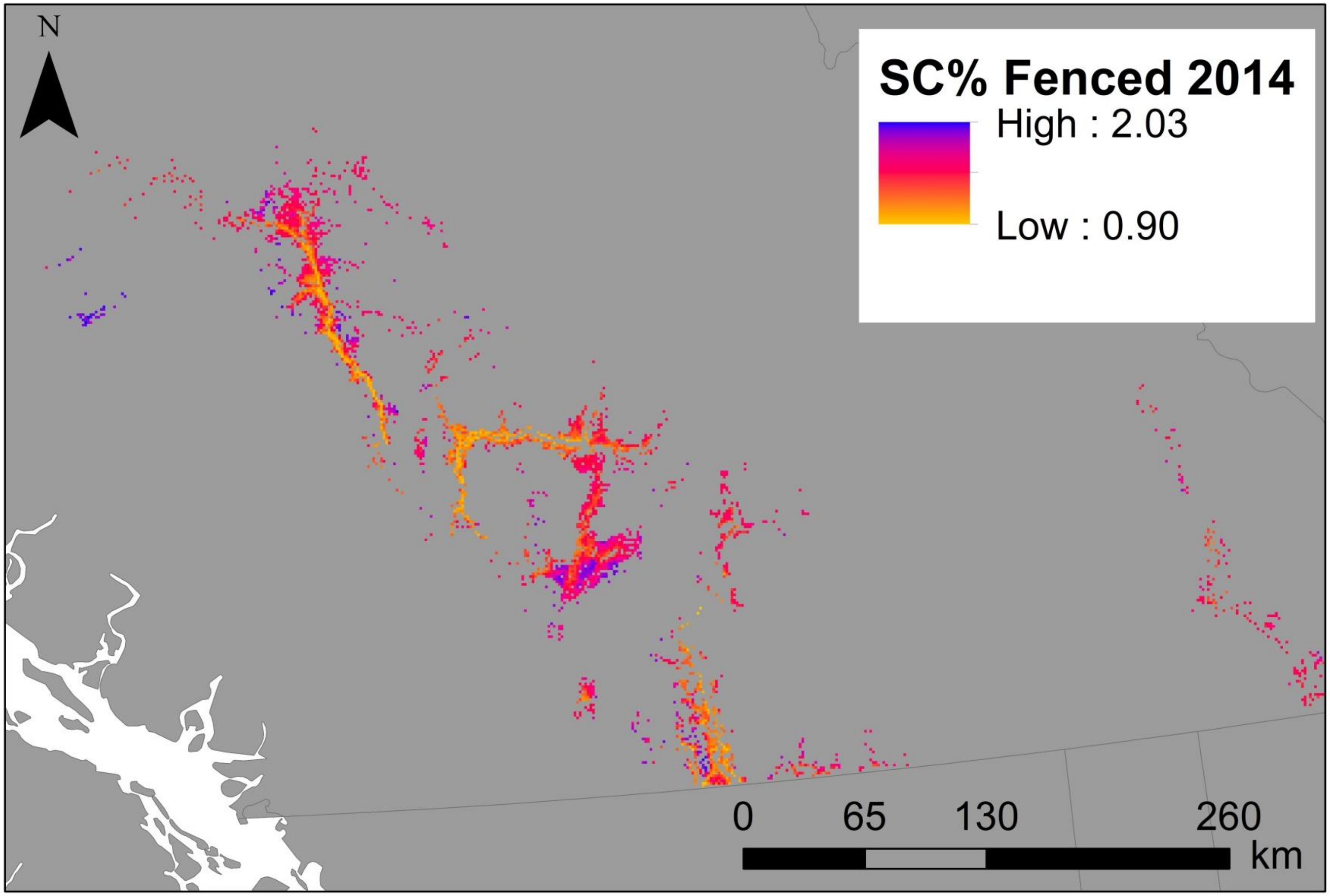

Figure C.2: Predicted Soil Carbon (SC(%)) values, based on the Random Forest results for 2014 fenced systems.

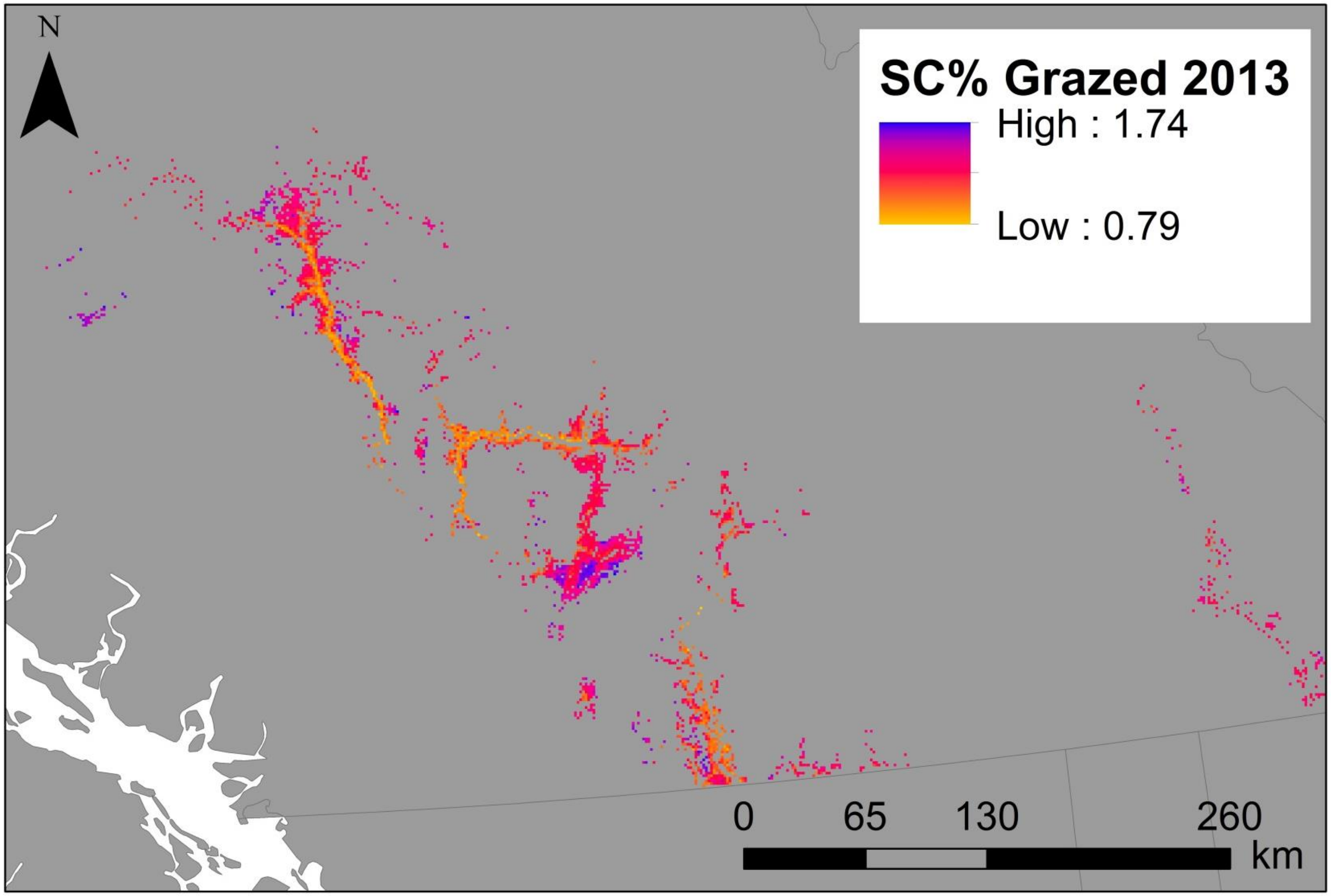

Figure C.3: Predicted Soil Carbon (SC(%)) values, based on the Random Forest results for 2013 grazed systems.

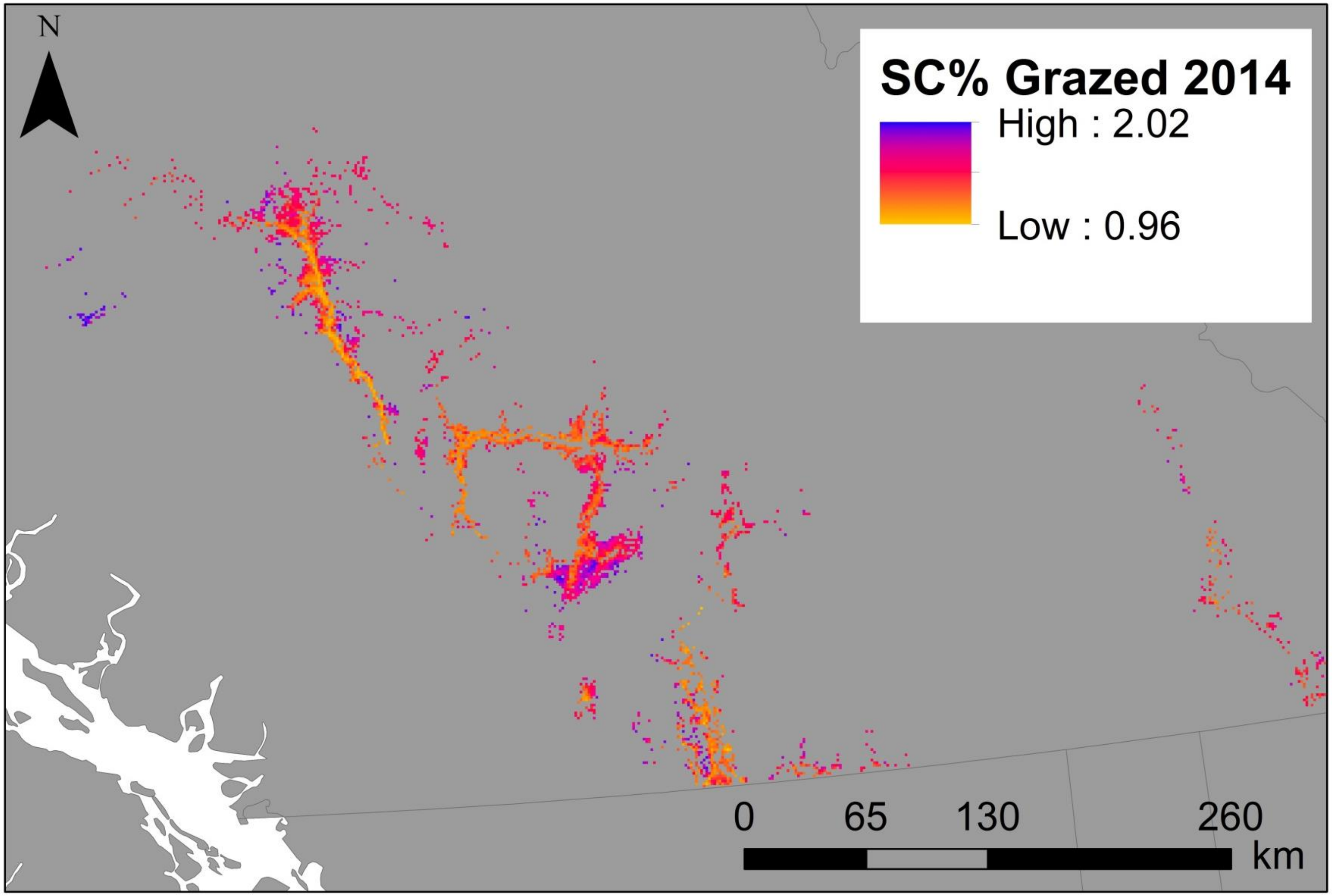

Figure C.4: Predicted Soil Carbon (SC(%)) values, based on the Random Forest results for 2014 grazed systems.

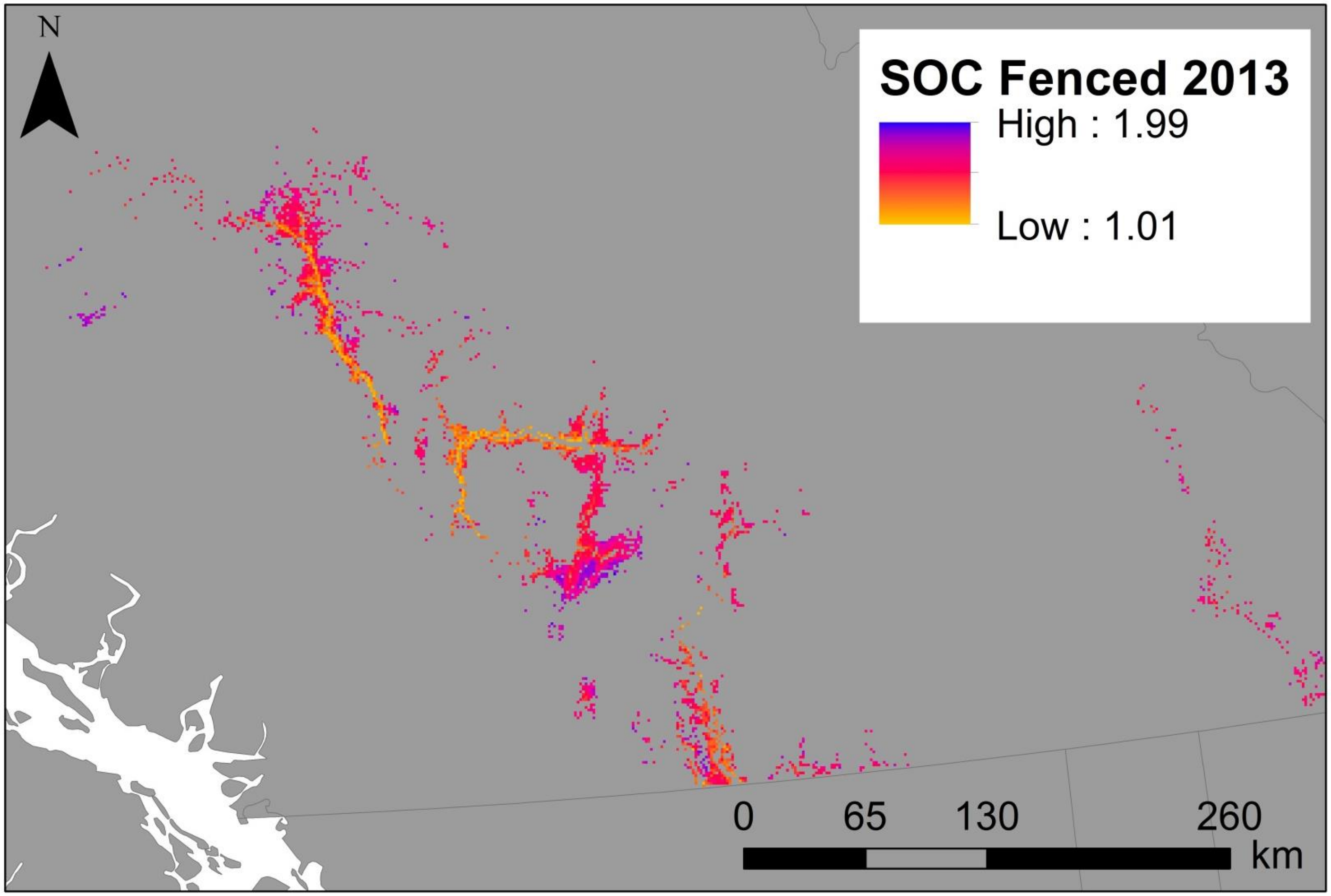

Figure C.5: Predicted Soil Organic Carbon (SOC(g/kg)) values, based on the Random Forest results for 2013 fenced systems.

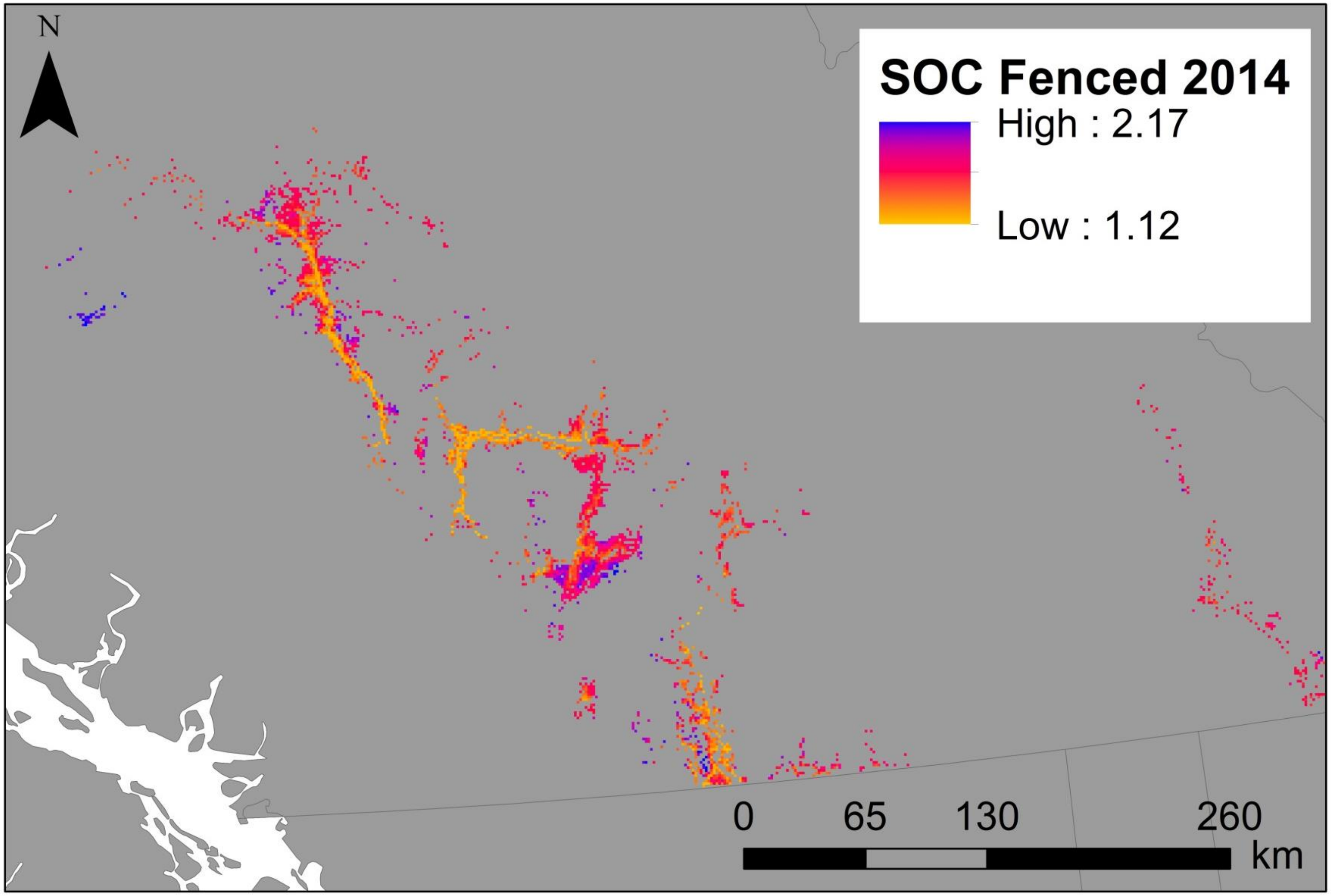

Figure C.6: Predicted Soil Organic Carbon (SOC(g/kg)) values, based on the Random Forest results for 2014 fenced systems.

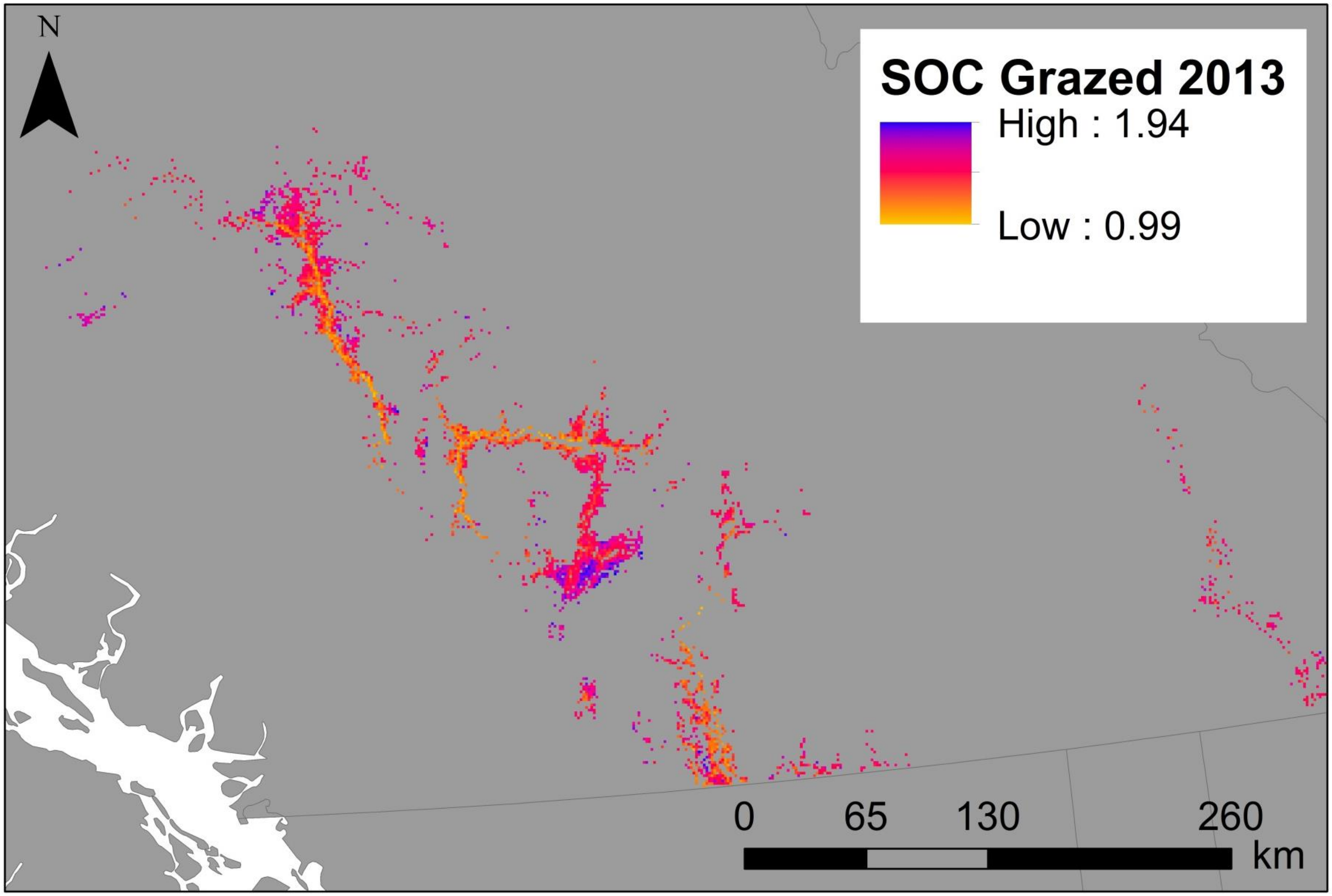

Figure C.7: Predicted Soil Organic Carbon (SOC(g/kg)) values, based on the Random Forest results for 2013 grazed systems.

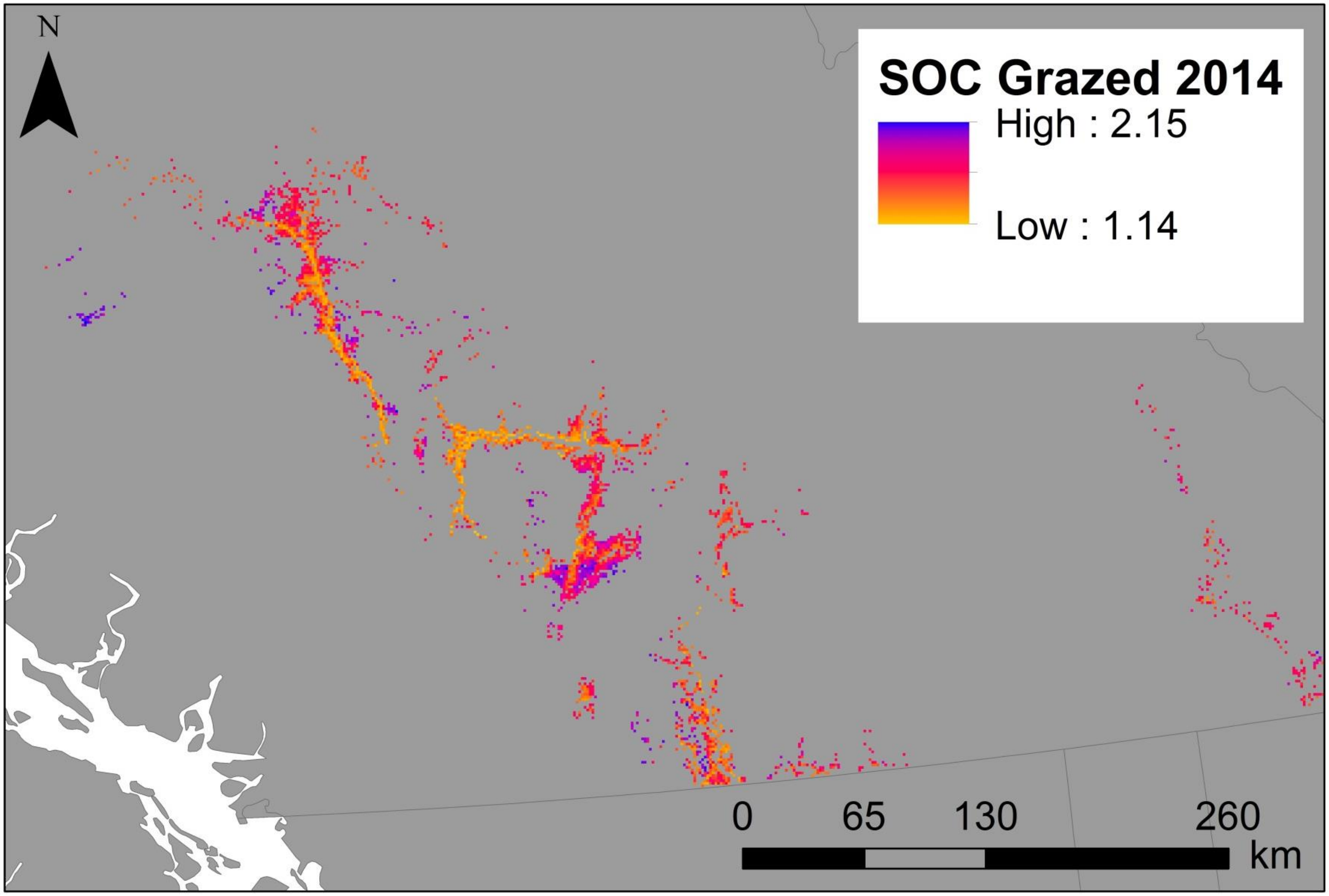

Figure C.8: Predicted Soil Organic Carbon (SOC(g/kg)) values, based on the Random Forest results for 2014 grazed systems.

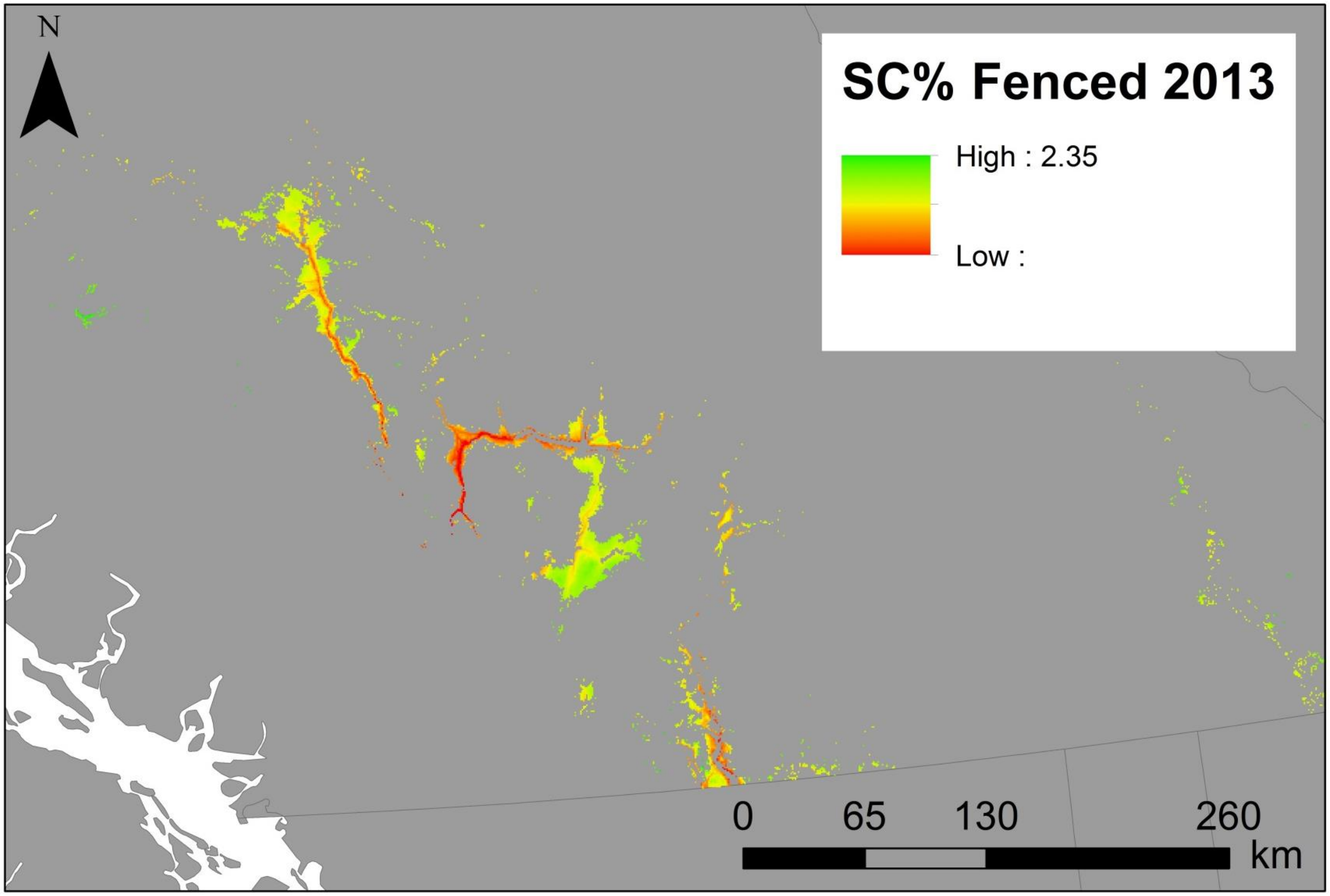

Figure C.9: Predicted Soil Carbon (SC(%)) values, based on the Linear Stepwise Regression results for 2013 fenced systems. SC= -8.22 + 0.78(Elevation) +0.71(MAP)

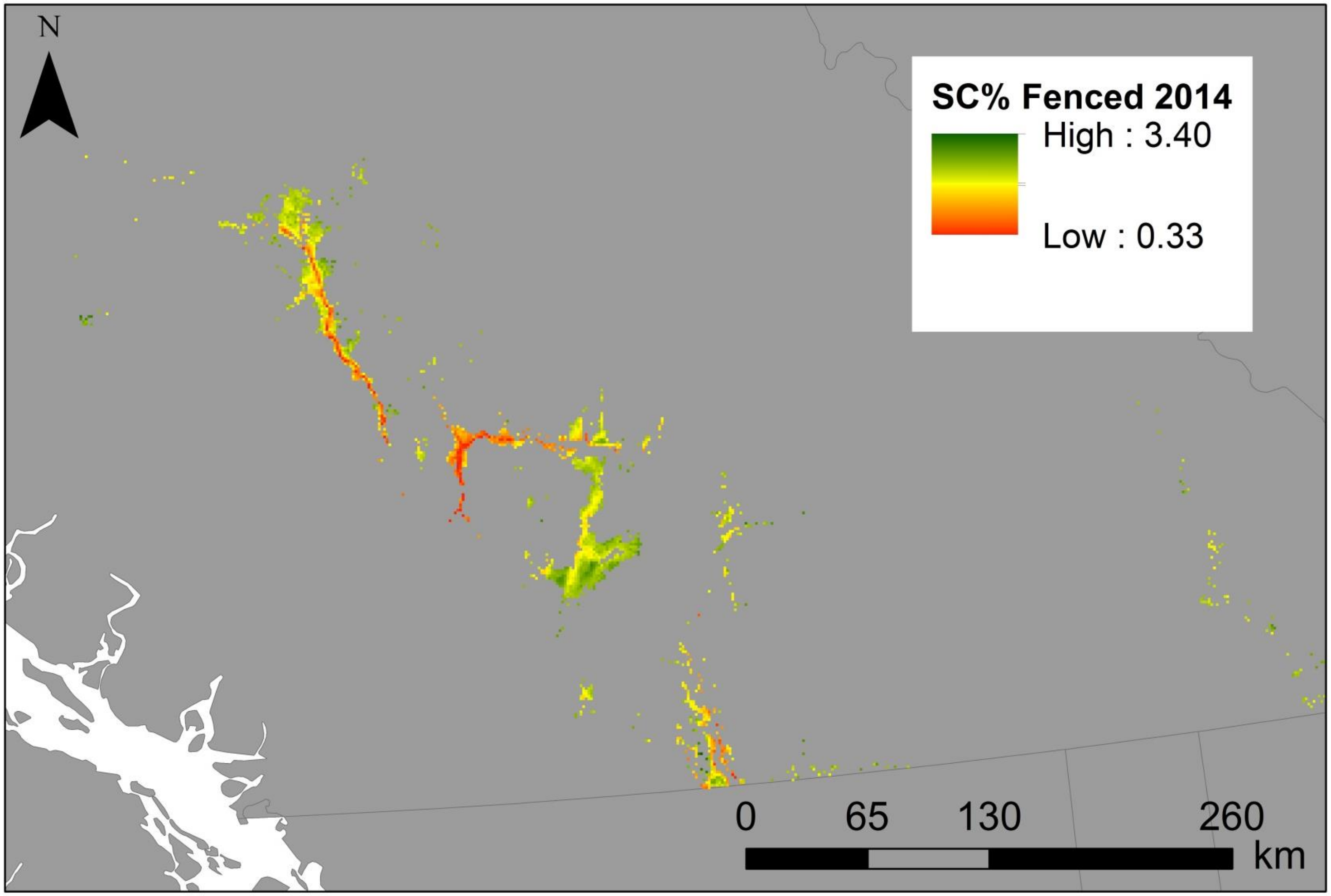

Figure C.10: Predicted Soil Carbon (SC(%)) values, based on the Linear Stepwise Regression results for 2014fenced systems. SC= -13.50 + 1.11(Elevation) +0.50(MAT) + 0.99(MAP) + 1.74($NDVI_{MODIS}$)

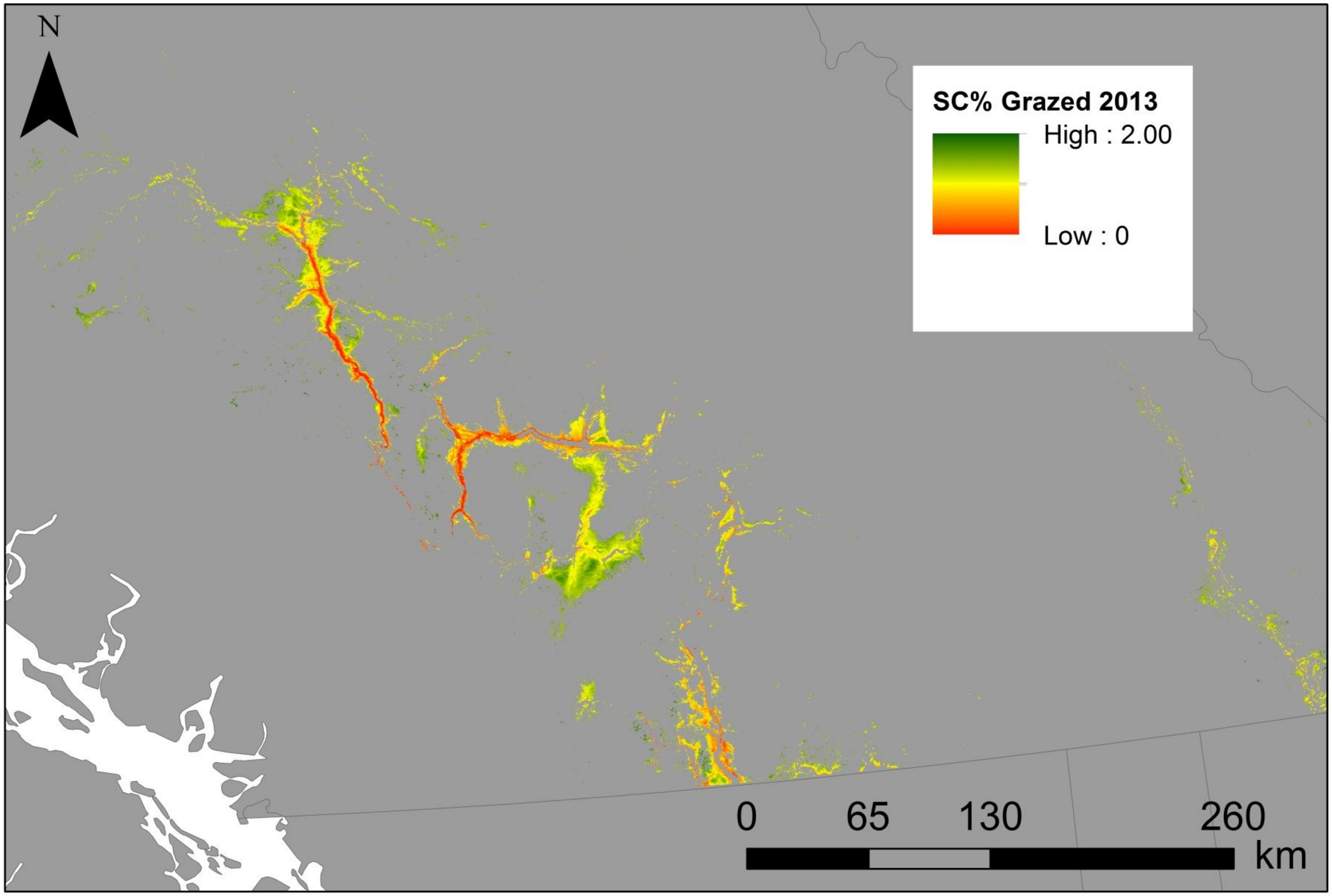

Figure C.11: Predicted Soil Carbon (SC(%)) values, based on the Linear Stepwise Regression results for 2013 grazed systems. SC= -4.06 + 0.52(Elevation) + 1.89($NDVI_{MODIS}$) +0.52(Soil Drainage)

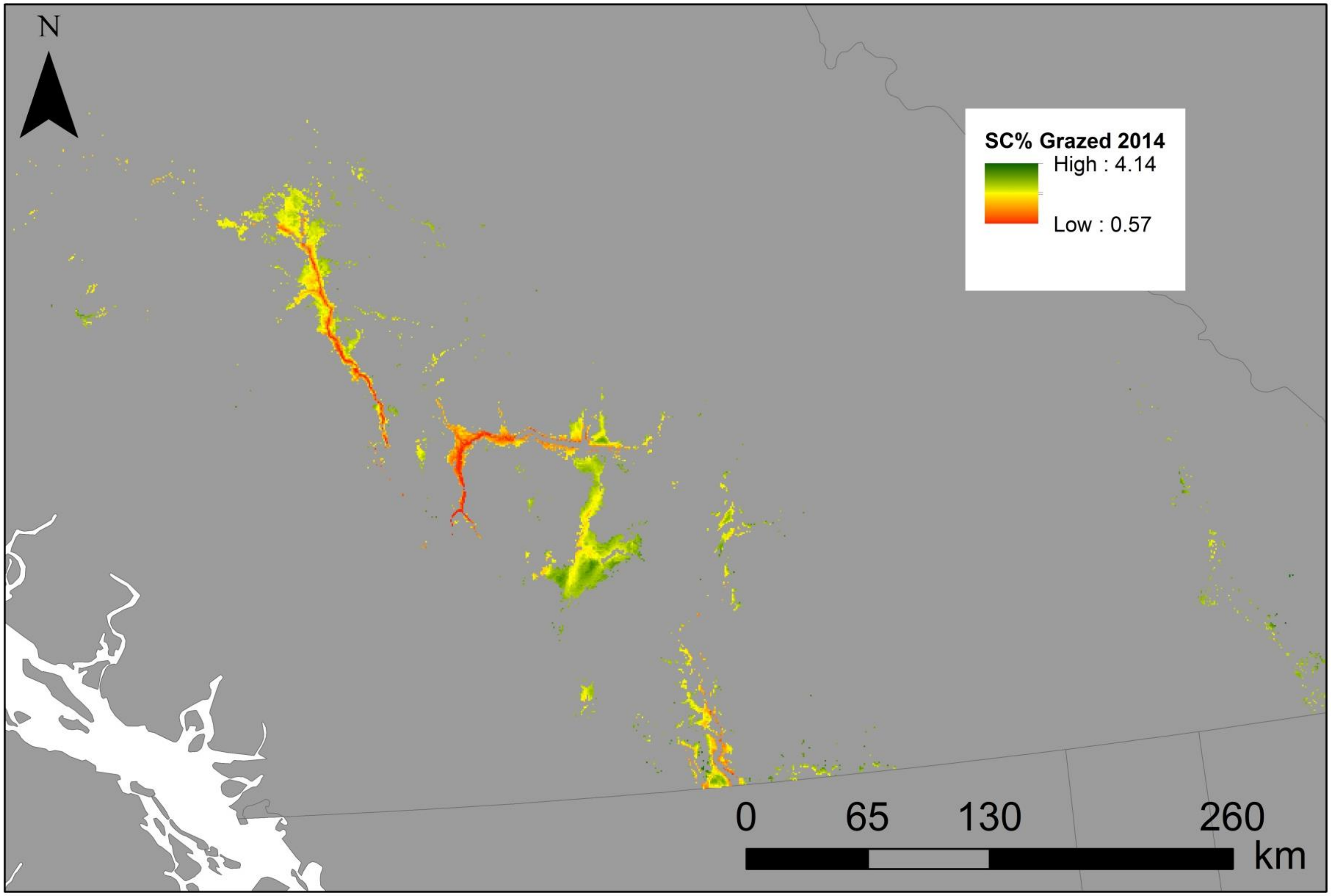

Figure C.12: Predicted Soil Carbon (SC(%)) values, based on the Linear Stepwise Regression results for 2014 grazed systems. SC= -15.86 + 1.31(Elevation) +0.75(MAT) + 1.02(MAP) +2.05($NDVI_{MODIS}$)

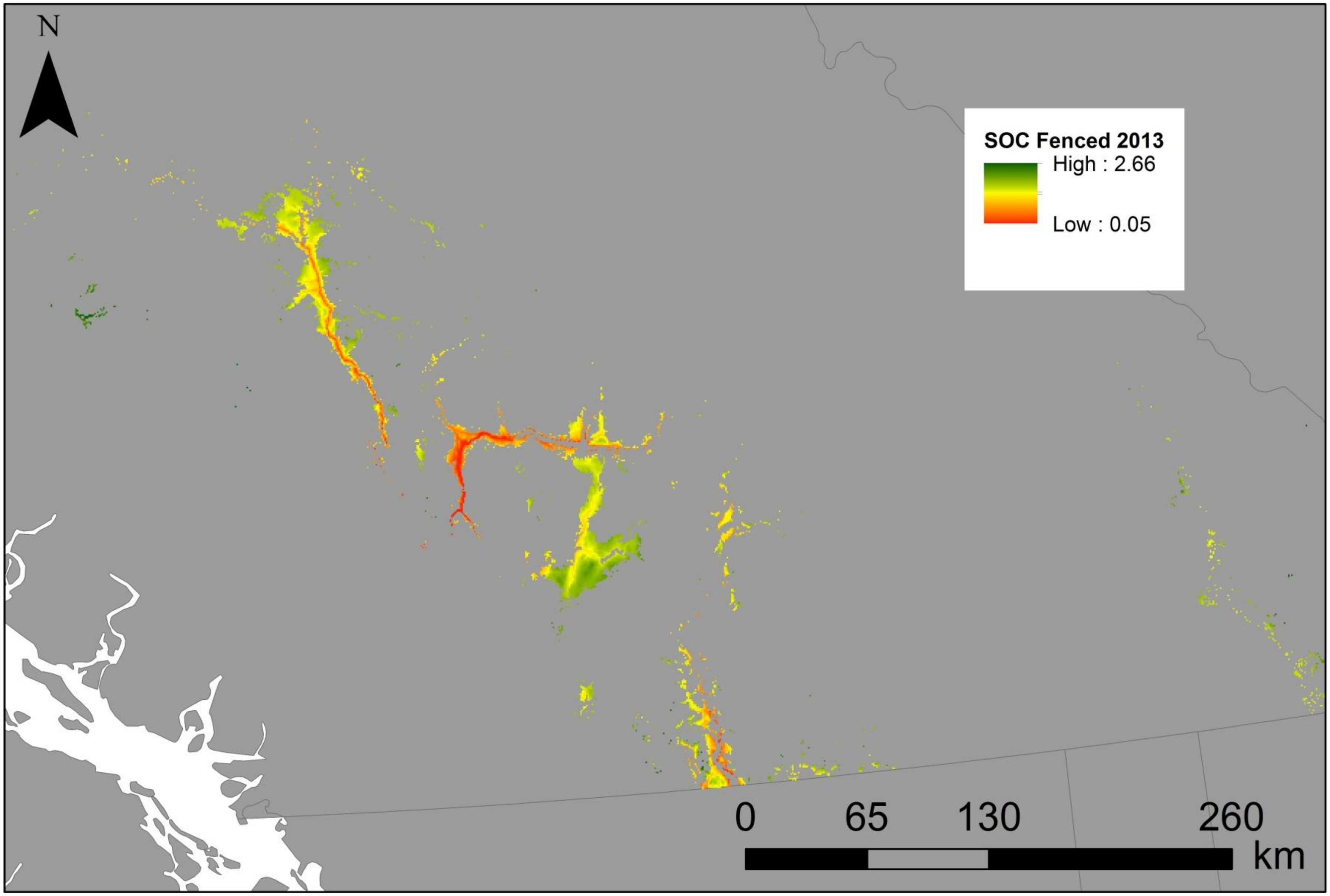

Figure C.13: Predicted Soil Organic Carbon (SOC(g/kg)) values, based on the Linear Stepwise Regression results for 2013 fenced systems. SC= -4.06 + 0.52(Elevation) + 1.89($NDVI_{MODIS}$) + 0.52(Soil Drainage)

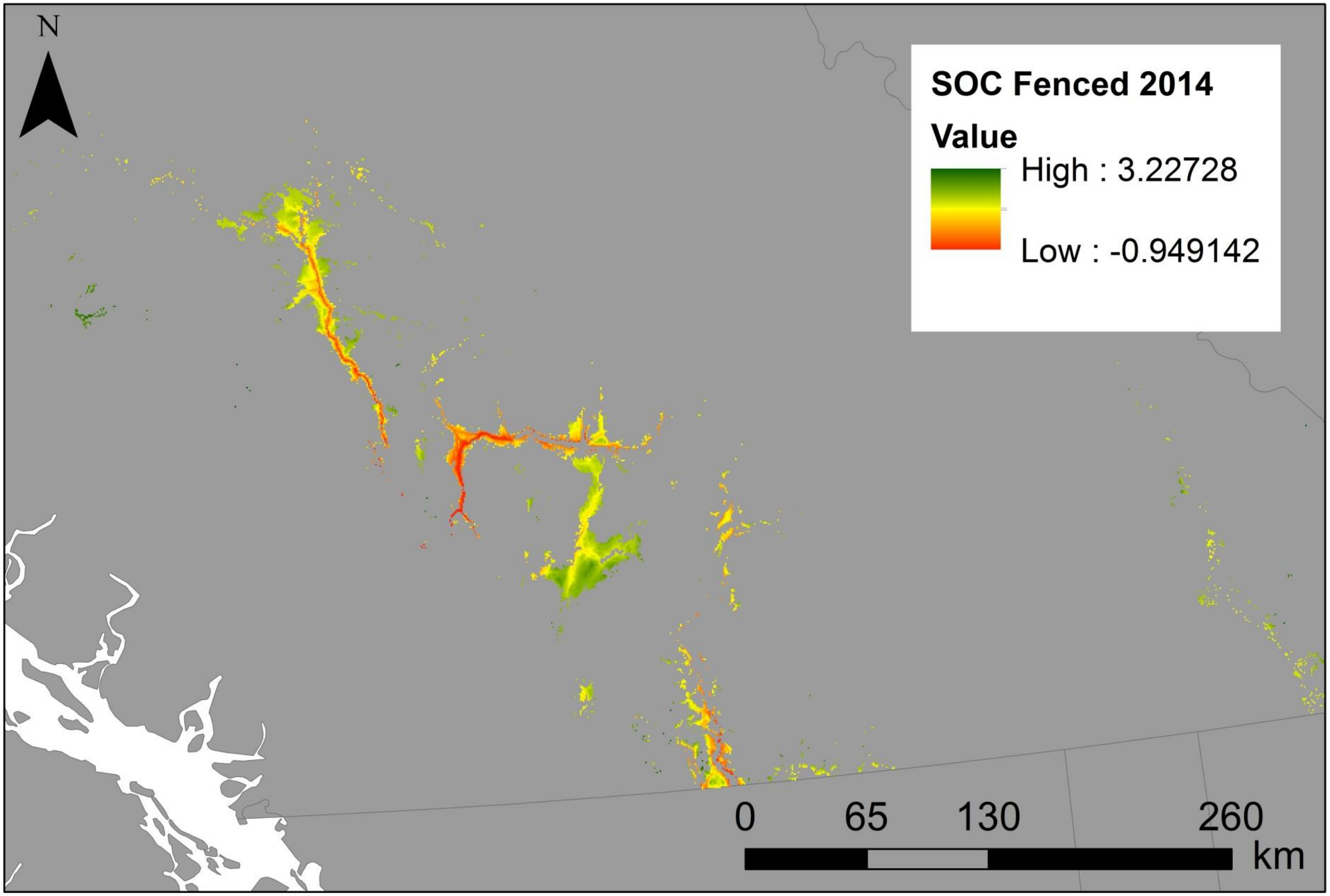

Figure C.14: Predicted Soil Organic Carbon (SOC(g/kg)) values, based on the Linear Stepwise Regression results for 2014 fenced systems. SC= -13.96 + 1.50(Elevation) + 0.89 (MAP)

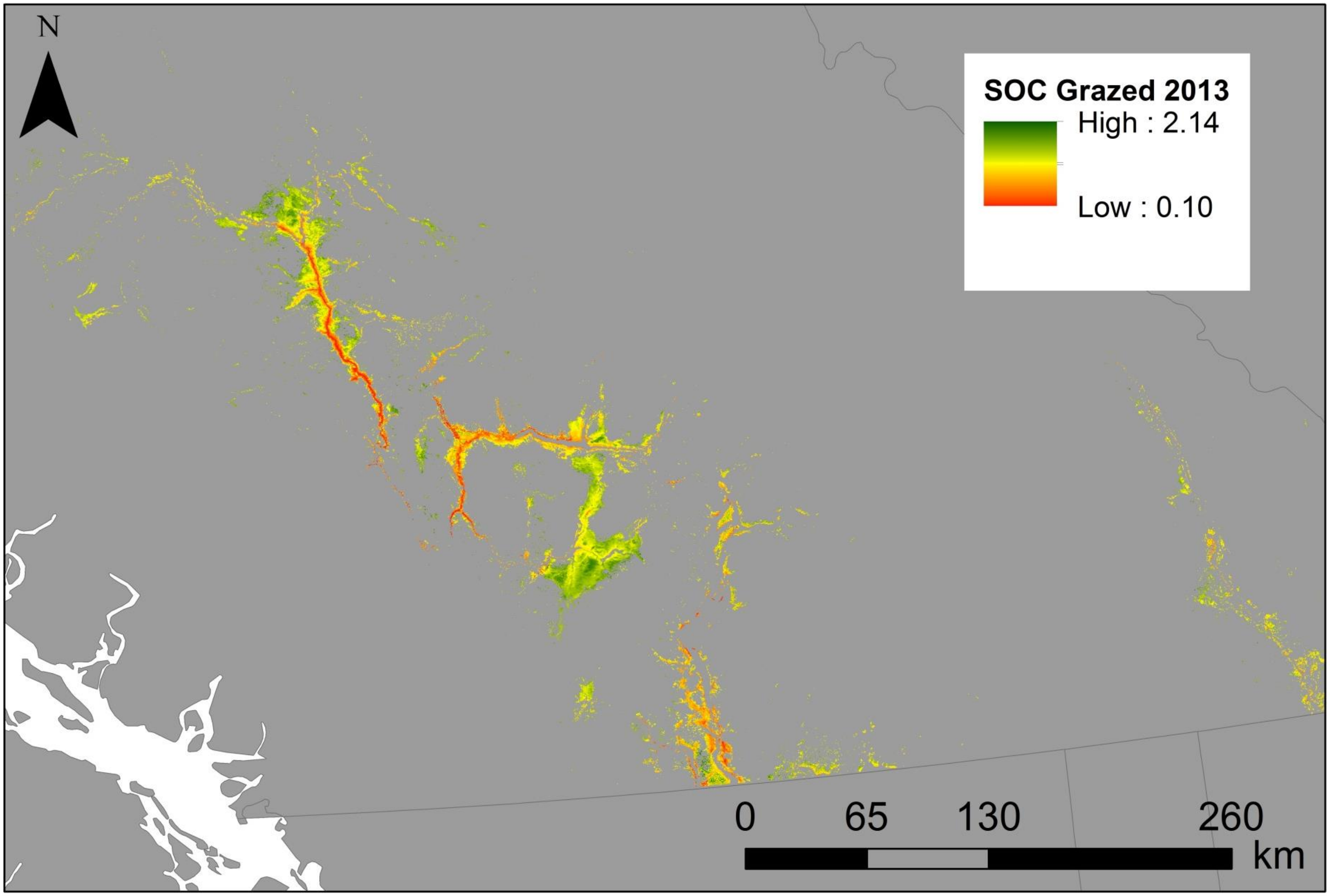

Figure C.15: Predicted Soil Organic Carbon (SOC(g/kg)) values, based on the Linear Stepwise Regression results for 2013grazed systems. SC= -2.23 + 0.38(Elevation) + 0.82($NDVI_{MODIS}$)

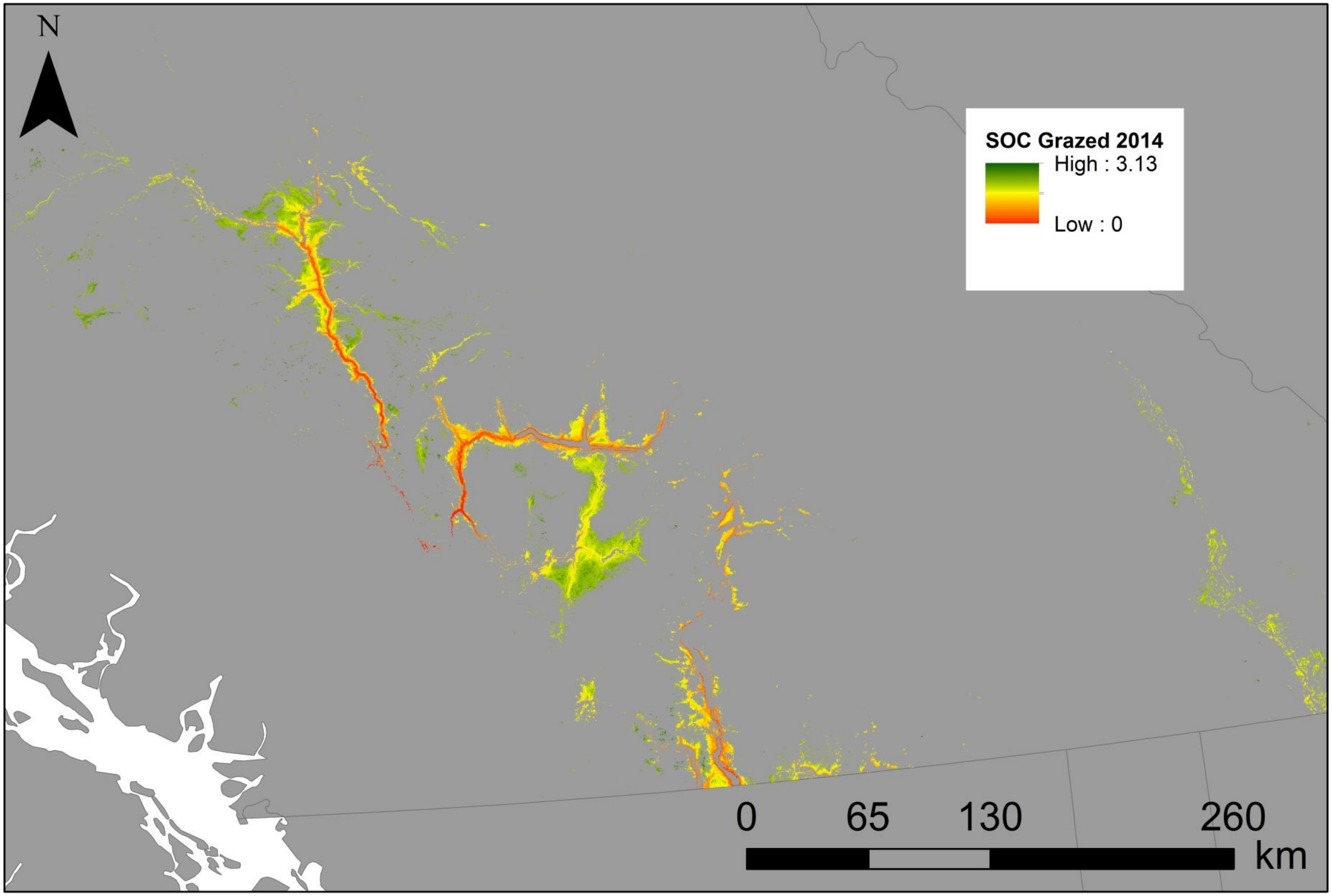

Figure C.16: Predicted Soil Organic Carbon (SOC(g/kg)) values, based on the Linear Stepwise Regression results for 2014grazed systems. SC= -8.60 + 1.55(Elevation) -0.28(Slope)

APPENDIX D: GRASS SPECTRAL SIGNATURES

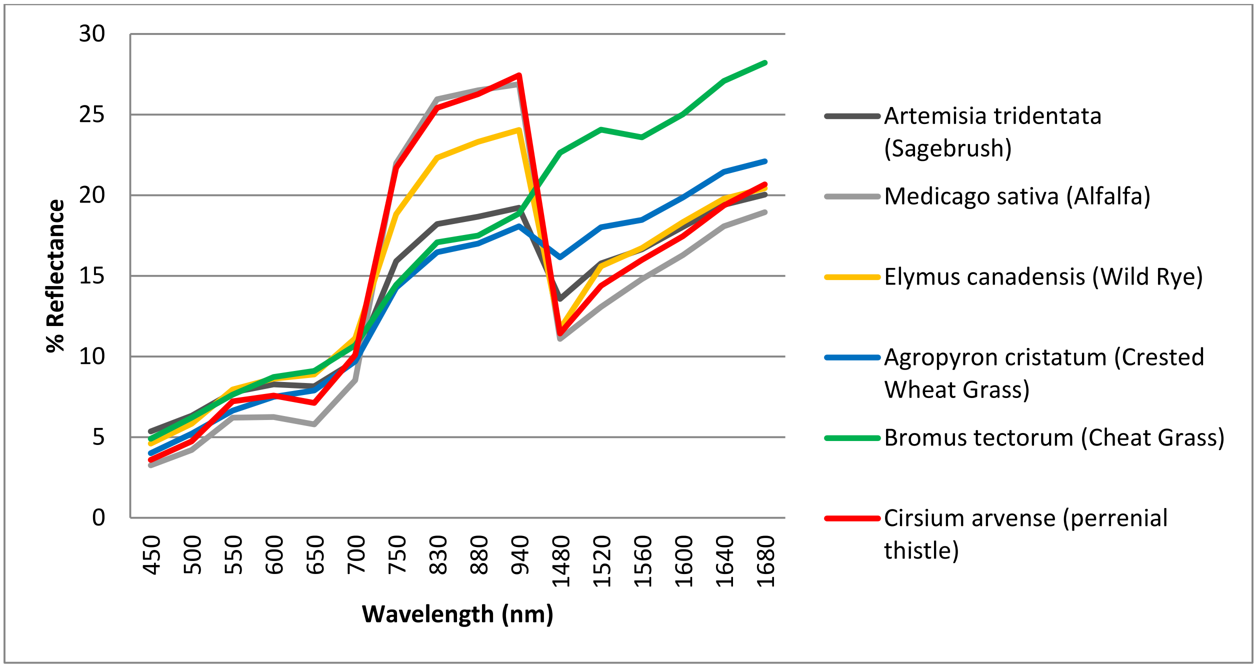

Figure D: Spectral Signatures of Grassland Monocultures in Lac du Bois, Kamloops, BC, Canada. Signatures derived from MSR (MSR16R, Cropscan Inc). Data collected August 2014 with Simon Oliver.

Appendix E: Supplemental Resources

Project website:

https://grazingmgtandclimatechange.wordpress.com/

Interactive map of site locations online:

http://soilcarbonsequestrationproject.websitesofcanada.com/

Interactive map of site locations kml file:

https://sites.google.com/site/googsitemap/home/google-earth-maps/heathsitemap2.kml?attredirects=0&d=1

High resolution maps of predicted soil carbon and soil organic carbon:

https://sites.google.com/site/predictionmaps/maps