MAPPING SOIL ORGANIC CARBON (SOC) IN A SEMI-ARID MOUNTAINOUS WATERSHED USING VARIABLES FROM HYPERSPECTRAL, LIDAR AND TRADITIONAL DATASETS

by

Ryan Matthew Will

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Hydrologic Sciences

Boise State University

December 2017

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Ryan Matthew Will

Thesis Title: Mapping Soil Organic Carbon (SOC) in a Semi-Arid Mountainous Watershed using Variables from Hyperspectral, Lidar and Traditional Datasets

Date of Final Oral Examination: 2 August 2017

The following individuals read and discussed the thesis submitted by student Ryan Matthew Will, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Shawn Benner, Ph.D.	Chair, Supervisory Committee
Jennifer Pierce, Ph.D.	Member, Supervisory Committee
Nancy Glenn, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Shawn Benner, Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Shawn Benner, as well as my committee members, Dr. Nancy Glenn and Dr. Jennifer Pierce for their support throughout this process. In addition, I would like to thank the Agricultural Research Service and Reynolds Creek Experimental Watershed staff for providing logistical support that was essential to this research. I also would like to thank, Jaron Adkins, Cody Black, Ben Bruck, Mayara Cizina, Jenna Duffin, Sam Evans, Alison Good, Mariah Holloway, Aihua Li, Peter Lindquist, Marion Lytle, Jason McKinnon, Mike Poulos, Zach Shephard, Amy Smith, Lucas Spaete, Dani Terhaar, Chris Thorton, Kerrie Weppner and Peter Youngblood for all of their help in the field and lab. The National Science Foundation Reynolds Creek Critical Zone Observatory [EAR 1331872] and Boise State University funded this project.

ABSTRACT

Quantifying soil organic carbon (SOC) in complex terrain is challenging due to its high spatial variability. Generally, limited discrete observations of SOC data are used to develop spatially distributed maps of SOC by developing quantitative relationships between SOC and available spatially distributed variables. In many ecosystems, remotely sensed information on aboveground vegetation can be used to predict belowground carbon stocks. In this research, we developed maps of SOC across a semi-arid watershed based on discrete field observations and modeling using a suite of variables inclusive of hyperspectral and lidar datasets; these observations provide insights into the controls on soil carbon in this environment. The Reynolds Creek Experimental Watershed (RCEW), in SW Idaho, has a strong elevation gradient that controls precipitation and vegetation. Soil samples were collected to 30 cm depth using a nested sampling approach, across the watershed (samples, 279 data points, in 28 plots, discretized with depth, total n=1344) and analyzed for SOC content. Point SOC data was combined with a suite of predictor variables from traditional, lidar and hyperspectral datasets to calibrate Random Forest and Stepwise Multiple Linear Regression models that predict SOC distribution across RCEW. In this study, SOC generally increased along the precipitation-elevation gradient corresponding with an increase in the diversity and abundance of vegetation. We found that variable soil bulk densities and areas of high rock content strongly influenced mass/unit area SOC values. Interestingly, rock content was also negatively correlated with percent SOC. Local variability of SOC in this study was high with the variability at

the plot scale about 1/3 of that observed at the watershed scale. Our research suggests that vegetation indices calculated from spectral data are the best predictors of SOC storage in this system. Roughly 60% of the variance in SOC data is explained using Normalized Difference Vegetation Index while two hyperspectral vegetation indices, Modified Red Edge Simple Ratio and Modified Red Edge Normalized Difference Vegetation Index explain over 70%. The addition of Lidar variables modestly improved SOC prediction, explaining 75% of variability in SOC.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SOC	Soil Organic Carbon (in units of kg*m-2)
%SOC	Soil Organic Carbon (in units of mass/mass)
CF/%CF	Coarse Fraction/%Coarse Fraction
BD	Soil bulk density
RCEW	Reynolds Creek Experimental Watershed
RF	Random Forest
IAA	Internal Accuracy Assessment
SMLR	Stepwise Multiple Linear Regression
DSM	Digital soil mapping
NDVI	Normalized Difference Vegetation Index
lidar	Light Detection and Ranging
CV	Cross-validation
SIC	Soil Inorganic Carbon
DEM	Digital Elevation Model
BCAL	Boise Center Aerospace Laboratory
WELD	Web-enabled LANDSAT data
PRISM	Parameter-elevation Regressions on Independent Slopes Model
SAGA	System for Automated Geospatial Analysis
MRESR	Modified Red-Edge Simple Ratio

R2	Coefficient of Determination
RMSE	Root-mean-square error
PPT	Precipitation
PET	Potential Evapotranspiration
WUE	Water-use efficiency
NPP	Net Primary Productivity
TPI	Topographic position index
PSRI	Plant Senescence Reflectance Index
MRENDVI	Modified Red Edge Normalized Difference Vegetation Index
NIR	Near-infrared
EM	Electro-magnetic
RSD	Relative standard deviation

INTRODUCTION

Soil organic carbon (SOC) is a large, dynamic carbon reservoir that stores more carbon than both the atmosphere and vegetation globally (Schlesinger, 1997; Lal, 2004). Constraining the distribution, and associated controls, on SOC storage is the first step to evaluating the potential positive and negative feedbacks with atmospheric composition and climate change (Raich, 1995; Trumbore, 1996; Woodwell, 1998; Jobbagy, 2000; Kulmatiski, 2004). The response of the SOC pool to current and projected changes in climate is one of the largest sources of uncertainty in climate and carbon cycling models (Minasny, 2013).

One of the most significant obstacles in constraining SOC distribution is the high spatial variability and associated uncertainty (Minasny, 2013). In recent years, modeling approaches have been developed that combine point SOC data observations with spatially distributed predictor variables to provide spatially distributed estimates of SOC distribution (Moore, 1993; McBratney, 2003; Simbahan, 2006; Gomez, 2008; Minasny, 2013). However, significant uncertainty remains regarding best the modeling methodology and environmental covariates, both of which may vary with scale and ecological system.

Estimating Soil Organic Carbon Storage in Semi-Arid Ecosystems

Interaction between climate, topography and vegetation in semi-arid mountainous regions produces high spatial heterogeneity in the processes influencing soil organic carbon accumulation, making estimating SOC storage especially difficult (Joggaby,

2000; Kulmatski, 2004; Kunkel, 2011; Hoffmann, 2014). These influences over water, energy and carbon flux create spatial heterogeneity across the landscape. This is especially evident at the hillslope scale as topographic alteration of the water and energy balance controls vegetation and SOC distribution (Kunkel, 2011; Smith, 2011; Hoffman, 2014; Patton, 2016). In addition, proportions of soil (< 2 mm) to coarse material (> 2 mm) and soil bulk density that are also highly variable in space (Hoffmann, 2014; Patton, 2016) influence the amount of SOC at a location. In semi-arid regions, vegetation is often a proxy for climate conditions and SOC contents tend to follow trends in aboveground vegetation. Our research focuses on mapping soil organic carbon distribution in a semi-arid, mountainous watershed by utilizing a suite of spatially distributed environmental covariates that represent vegetation, topography and climate. This approach is likely to provide more accurate estimates of the size and distribution of the SOC reservoir while revealing controlling mechanisms and constraining global and local biogeochemical models.

Statistical Approaches for Creating Distributed Maps of SOC Using Point Data

Field measurements of soil properties, including SOC, are typically collected at discrete locations. While the resulting data can provide useful information regarding how SOC is spatially distributed, field sample density is often not sufficient to create accurate maps of SOC across the landscape. For this reason, mapping of SOC often involves modeling its distribution using spatially distributed variables that are correlated to SOC.

Previous researchers have developed spatially distributed maps of SOC using a variety of statistical modeling approaches and environmental predictor variables. Some of the most commonly used modeling approaches include: linear regression/multiple linear

regression (LR/MLR) (Gessler, 2000; Mueller, 2003; Thompson, 2005; Selige, 2006; Kunkel, 2011), partial least squares regression (PLSR) (Gomez, 2008; Stevens, 2010; Bartholomeus, 2011), random forests (RF) (Grimm, 2008; Wiesmeier, 2011), artificial neural networks (ANN) (Minasny, 2006; Malone, 2009) and Kriging/regression kriging (Simbahan, 2006; Vasques, 2010; Kunkel, 2011).

Environmental Covariates

The most common environmental covariates for SOC mapping are terrain attributes (Mueller, 2003; Thompson, 2005; Minasny, 2006; Grimm, 2008; Malone, 2009) and variables derived from Landsat imagery, such as NDVI (Normalized Difference Vegetation Index) (Minasny, 2006; Malone, 2009; Vasques 2010; Kunkel, 2011). Past studies on SOC mapping have utilized both spectral information and terrain attributes at a variety of scales and ecosystems (Minasny, 2013). In addition, many studies have used geologic and soil maps (McKenzie, 1999; Simbahan, 2006; Grimm, 2008; Wiesmeier, 2011), land cover/use aerial imagery (Mora-Vallejo, 2008, Wiesmeier, 2011) and climate and meteorological data (McKenzie, 1999; Mishra, 2010; Martin, 2011).

More recently, researchers have utilized hyperspectral (Selige, 2006; Gomez, 2008; Stevens, 2010; Bartholomeus, 2011) and Lidar (or high-resolution field studies that mimic Lidar) (Gessler, 1995; Gessler, 2000; Mueller, 2003; Thompson, 2005; Grimm, 2008; Mulder, 2011; Lacoste, 2014) datasets for SOC prediction. A previous study in this region have observed strong relationships between SOC and NDVI (Kunkel, 2011). One of the goals of this research is to improve the accuracy of our SOC model by

incorporating variables from higher resolution, hyperspectral (AVIRIS-ng) and LIDAR datasets.

Hyperspectral data has been used extensively for mapping vegetation characteristics (Vogelmann, 1993; Gitelson, 1994; Curran, 1995; Datt, 1999; Daughtry, 2000; Broge, 2000; Sims, 2002; Haboudane, 2004). But, somewhat less research has focused on mapping SOC with hyperspectral data (Gomez, 2008; Stevens, 2010; Bartholomeus, 2011). Utilization of hyperspectral data for SOC mapping is relatively new and methodologies vary between studies. Many studies have focused on the spectral information of the soils to determine the SOC contents of soils (Bartholomeus, 2011; Peng, 2014). In several of these studies, the SOC contents of soil samples are determined using spectroscopy in field and laboratory environments (Bartholomeus, 2008; Gomez, 2008; Peng, 2014).

A common issue with airborne hyperspectral data is spectral mixing or mixed pixels, which occur when pixels include spectral information from soils and vegetation (Selige, 2006; Gomez, 2008). Many studies employ spectral un-mixing approaches to isolate soil signatures from vegetation (Stevens, 2010; Bartholomeus, 2011). Another approach for isolating vegetation signatures is to utilize narrowband indices, calculated over a five-nanometer window, allowing for detection of subtle differences in vegetation reflectance. While, other studies utilize vegetation changes to map SOC, by calculating vegetation indices similar to NDVI, or using spectra from airborne hyperspectral data to calibrate a PLSR model (Selige, 2006).

Much of the previous research involving SOC mapping with airborne hyperspectral data was conducted at smaller geographic scales and focused on precision

agriculture applications (Gomez, 2008; Hbirkou, 2012). In semi-arid regions, spectra from airborne hyperspectral data are heavily influenced by soil reflectance in low cover environments whereas spectra in higher biomass ecosystems are dominated by vegetation. For this reason, determination of soil spectral characteristics via airborne hyperspectral is difficult in high biomass ecosystems.

Mapping studies focused on hyperspectral data also vary with regard to how they use the large number of correlated bands as a tool for prediction. Some researchers use methods for reducing the dimensionality of hyperspectral data. The most common approaches are PCA (Principle Component Analysis) and MNF (Minimum Noise Fraction) transforms which are both used for a range of applications (Chang, 2001; Williams, 2002; Datt, 2003). Utilizing highly correlated and high dimension datasets such as hyperspectral often requires more advanced quantitative methods for model calibration and prediction. Previous studies using hyperspectral, often use Partial Least Squares Regression for variable selection and prediction but other approaches such as random forest, kriging, neural networks and SVMR.

Many SOC studies use Lidar or detailed field surveys to create high resolution DEMs that can be used to create topographic attributes of the land surface such as elevation, aspect, slope, curvature and topographic position (Gessler, 1995; Gessler, 2000; Mueller, 2003; Thompson, 2005; Grimm, 2008; Mulder, 2011; Lacoste, 2014). Lidar can also be used to model vegetation characteristics such as height, cover, LAI and biomass (Ni-Meister, 2010; Li et al, 2015) but these variables are rarely used for SOC prediction. Much of the research on SOC mapping with Lidar is conducted at smaller spatial scales in the absence of large climate gradients. In this research, we explore the

predictive capability of both Lidar vegetation and Lidar topographic variables at larger scales where strong climate gradients exert considerable influence on SOC.

Limitations of Previous Research

More recently, researchers have incorporated hyperspectral and Lidar data for SOC prediction. But, little work has been done with these data types at larger scales in semi-arid complex terrain in the presence of strong climate gradients. Many SOC mapping studies do not provide uncertainty estimates and some have no external validation of the model used (Minasny, 2013).We argue that validation methodology and uncertainty become increasingly important when mapping a variable like SOC in complex terrain where high spatial variability and low sample density increases the likelihood of a biased sample distribution. For this reason, it is important to utilize robust modeling approaches that employ cross-validation (CV) or bootstrap sampling, especially when using high-resolution datasets that provide an abundance of potential predictors.

The research described in this paper has 3 goals:

1. To identify spatially distributed variables (traditional, as well as Lidar and hyperspectral) that predict soil carbon distribution;
2. To develop spatially distributed maps of SOC while evaluating the influence of the modeling approach Random Forests (RF) v. Stepwise Multiple Linear Regression (SMLR) and attempting to provide a framework for evaluating multiple models in the context of variable selection, model calibration and validation approaches, and accuracy and uncertainty;
3. To provide insight into the controls on SOC distribution in semi-arid mountainous landscapes.

METHODS

Site Description

We conducted this study at the Reynolds Creek Experimental Watershed (RCEW), a semi-arid watershed (239 km^2), located in the Owyhee Mountains of southwest Idaho (Figure 1). There is a strong elevation gradient in the watershed (1099-2144m) and precipitation is highly correlated with more precipitation input at higher elevations (250-1100 mm. /yr. precipitation). Lower elevation ecosystems are hot, dry, and rain-dominated. In contrast, higher elevation ecosystems are cold and wet, and the majority of precipitation occurs during the winter months as snow. Vegetation type and abundance closely follows the elevation gradient with sagebrush-dominated ecosystems in the lowlands and a mix of sagebrush species and trees at higher elevations (Seyfried, 2001). RCEW contains a diverse range of soil parent materials including granite (Cretaceous), basalt (Miocene), rhyolite (Miocene), welded tuff, loess, and alluvial deposits. Soil orders found in the watershed include aridisols, inceptisols, andisols, vertisols and mollisols.

Field Methods

Site Selection

We collected soil samples along an elevation gradient, with site selection focused on representing all of the dominant vegetation functional groups present in the watershed. The primary vegetation functional groups are Wyoming Sage *(Artemisia tridentate subsp. Wyomingensis)*, Mountain Sage (*Artemisia Tridentate subsp. Vaseyana)*, Low Sage

(Artemisia arbuscula), Bitterbrush (*Purshia stansburyana*), Greasewood, Juniper (*Juniperus occidentalis*), Aspen (*Populus tremuloides)* and Conifer (Douglas fir (*Pseudolsuga menziesii glauca*) and alpine fir (*abies lasiocarpa*)). We also considered proximity to long-term climate stations and accessibility when selecting soil sampling sites. We collected field soil samples during the 2014 and 2015 field seasons.

Due to the high spatial variability in SOC, we utilized a stratified sampling approach (Figure A1). Our field sampling design involved sampling at 28 sites, with 10 sample locations at each of the sites. At each site, we dug one soil pit and extracted nine soil cores. Due to heterogeneous vegetation distribution, we used a paired sampling scheme. We paired the sample locations according to whether they were in the vegetation canopy or in the interspace. At each site, we sampled five canopy locations and five interspace locations with paired samples being within a meter of each other.

We randomly selected the sample locations relative to the pit, which was located in the center of the site or plot. Relative to the pit in the center, we determined core sampling locations by randomly generating distances (5-30 meters) and orientations (0-360°) relative to the pit. If a sample location is in the canopy, it was paired with the nearest interspace location and vice versa.

Field Sampling

We collected soil samples in the field using primarily a soil knife and sand auger. We collected soil samples according to the following depth increments, 0-5, 5-10, 10-20 and 20-30 cm. We collected soil cores to a depth of 30 cm, while pits range from 30 cm to 1 m depending on soil depth at the site. In addition, we sampled soil pits on three sides to evaluate uncertainty at a point.

We also collected samples for bulk density analysis from the pit at each site. We collected these samples from three walls of the pit and at 1-3 depths (depending on soil depth within the pit). We extracted the bulk density samples by inserting a hammer core, perpendicular to the pit wall.

We also analyzed soil samples from one wall of each pit for soil pH and texture (sand, silt and clay percentages) according to the depth increments above. At each site, we also took soil descriptions, and pictures of soil pits, site setting, and bird's eye view images of sample locations prior to sampling.

We also collected GPS coordinates at all sample location (pits and cores). However, only 200 of 279 samples have RTK (Real Time Kinematic, 0.5-meter horizontal accuracy) coordinates. If we were unable to obtain coordinates using RTK GPS, we collected coordinates with handheld GPS (5-meter horizontal accuracy).

Laboratory Analysis of Soils

Sample Preparation

After collection, we stored samples at low temperatures until we could begin laboratory processing. First, we dried soil samples in an oven for 24 hours at 50 degrees C. Then, we sieved soil samples using the #10 (2 mm) sieve and removed the coarse fraction (CF) (gravel, cobbles,> 2mm diameter) as well as roots and particulate organic material. We also weighed each soil sample and the removed coarse fraction and roots. We then split soil samples using a riffle splitter until we achieved a desired mass of 50-100 g. Then, we stored the remaining sample in our archive.

Methodology for Obtaining Soil Carbon and Nitrogen Data

Next, we thoroughly mixed and homogenized the split samples, and then collected a 5-gram subsample. We then completely removed any roots and particulate organic material (>2 mm length) in the subsample. Then, we powdered the sample on a ball mill for 48 hours. We then dried ground soil samples again to remove moisture that accumulated during processing. Then, we packed roughly 60 mg of powdered soil was into 5x9 mm aluminum tins. We analyzed these soil samples for percent carbon and percent nitrogen by mass using a Thermo Electron Flash EA 1112 CN analyzer (CE Elantech, Inc., Lakewood, NJ).

Quantification of Uncertainty in Soil C and N Data

We quantified instrument uncertainty by running triplicates every tenth sample as well as a soil standard and aspartic acid. We calculated instrument uncertainty to be 1.72% for Aspartic acid, 1.8% for soil standard and 2.17% for triplicates (equivalent to RSD of 2 for standards and 8 for triplicates)(Table A5a).

Removal of Soil Inorganic Carbon (Sic)

A portion of samples in our study contained SIC (Soil Inorganic Carbon), which we determined by reaction with Hydrochloric acid (HCl). Some of these samples went through a carbonate removal process where we mixed samples with 10% HCl, centrifuged, and rinsed for several iterations (McCorkle, 2015). We determined SIC contents of the remaining samples, using the methods described in Stanbery (2016).

Methodology for Other Datasets

Calculation of Soil Bulk Density in the Lab

Reasonably accurate soil bulk densities are critical when converting soil carbon concentrations (percent SOC by mass) to an estimate of carbon stored per unit area (SOC (kg/m^2)) on the land surface. In the following section, we describe calculations and laboratory processing for soil bulk density. However, we describe collection of bulk density samples in the field sampling section above.

While, bulk density samples collected in the field include both soil and coarse material, only fine fraction (< 2 mm) bulk density is required for SOC calculations. However, we need the density of the total sample ($\boldsymbol{BD_t}$) and/or the density of the coarse material ($\boldsymbol{BD_r}$) to calculate fine fraction soil bulk density($\boldsymbol{BD_s}$).

The first step for quantifying soil bulk density from field samples is to measure the mass of coarse material ($\boldsymbol{m_r}$) and soil ($\boldsymbol{m_s}$). Then, the volume of coarse material ($\boldsymbol{v_r}$) can be calculated using the mass of rock ($\boldsymbol{m_r}$) and the density of rock ($\boldsymbol{d_r}$) in ***Eq. 1***.

$$\boldsymbol{Eq.\ 1:\ v_r = \frac{m_r}{d_r}}$$

$\boldsymbol{v_r}$: $volume\ of\ rock$
$\boldsymbol{m_r}$: $mass\ of\ coarse\ material,\ measured\ in\ lab (> 2\ mm\ diameter)$
$\boldsymbol{d_r}$: $density\ of\ coarse\ material,\ assumed\ density\ of\ 2800\ kg * m - 3$

Once, a volume of rock ($\boldsymbol{v_r}$) is calculated using ***Eq.1***, the volume of soil can be calculated using the known volume of the hammer core ($\boldsymbol{v_t}$) and the estimated volume of the coarse material ($\boldsymbol{v_r}$) using ***Eq. 2***.

$$\boldsymbol{Eq.\ 2}: \quad \boldsymbol{v_s = v_t - v_r}$$

$\boldsymbol{v_s} = volume\ soil$
$\boldsymbol{v_r} = volume\ rock,\ calculated\ from\ Eq.\ 1$
$\boldsymbol{v_t} = total\ sample\ volume,\ equal\ to\ volume\ of\ hammer\ core\ (known)$

Once the volume of soil ($\boldsymbol{v_s}$) is calculated using ***Eq.2***, soil bulk density ($\boldsymbol{BD_s}$) can be calculated using ***Eq.3*** which also requires the mass of soil ($\boldsymbol{m_s}$), which we measured in the lab. In addition to soil bulk density, the density of the field sample ($\boldsymbol{BD_t}$) as well as just the coarse material ($\boldsymbol{BD_r}$) can be calculated using ***Eq. 3a*** and ***Eq. 3b***.

$$\boldsymbol{Eq.\ 3}: \quad \boldsymbol{BD_s = \frac{m_s}{v_s}}$$

$\boldsymbol{BD_s} = soil\ bulk\ density\ (< 2\ mm\ fraction)$
$\boldsymbol{m_s} = measured\ in\ lab,\ equal\ to\ m_t - m_r$
$\boldsymbol{v_s} = volume\ soil,\ calculated\ from\ Eq.\ 2$

$$\boldsymbol{Eq.\ 3a}: \boldsymbol{BD_t = \frac{m_t}{v_t}}$$

$\boldsymbol{BD_t} = bulk\ density\ of\ total\ sample, field\ bulk\ density\ (< and > 2\ mm\ fraction)$
$\boldsymbol{m_t} = mass\ of\ total\ sample,\ measured\ in\ lab,\ equal\ to$
$\boldsymbol{v_t} = volume\ of\ total\ sample,\ equal\ to\ volume\ of\ hammer\ core$

$$\boldsymbol{Eq.\ 3b}: \boldsymbol{BD_r = \frac{m_r}{v_r}}$$

$\boldsymbol{BD_r} = (bulk)\ density\ of\ rock\ (> 2\ mm\ fraction)$
$\boldsymbol{m_r} = mass\ of\ rock,\ measured\ in\ lab$
$\boldsymbol{v_r} = volume\ of\ rock\ from\ Eq.\ 1$

Soil Bulk Density Modeling

In this study, we observed considerable variability and uncertainty in our soil bulk density data. For this reason, we used our soil bulk density data, along with other samples

from RCEW, to calibrate a statistical model that predicts soil bulk density using SOC concentrations (percent SOC) and soil parent material. Other studies have observed strong relationships between soil organic material and soil bulk density (Perie, 2008; Patton, 2016). In addition to organic material, soil parent material exerts an important but secondary influence on soil bulk density, which Patton 2016 describes in detail.

The model, developed by Patton, 2016, allowed for distributed soil bulk density data instead of extrapolating limited soil bulk density data in pits to adjacent core samples. There was generally good agreement between bulk density estimates from the field and modeled values. However, there is a likely a considerable amount of error involved in estimating the soil bulk density of some soils, especially those with more rock.

Below, we show the statistical models developed by Patton in (***Eq.4, Eq. 5***). There is also a figure in the appendix that shows SOC plotted with BD for felsic and mafic parent materials (Figure A2). For a small number of samples, we used the mean of felsic and mafic bulk densities (***Eq. 6***) since parent material was neither felsic or mafic.

$$\boldsymbol{Eq.\ 4:}\ \boldsymbol{mBD_{felsic}} = 1.4178 \times SOC^{-0.148}$$

$$\boldsymbol{Eq.\ 5:}\ \boldsymbol{mBD_{mafic}} = 1.0322 \times SOC^{-0.324}$$

$$\boldsymbol{Eq.\ 6:}\ \boldsymbol{mBD_{other}} = \frac{mBD_{mafic} + mBD_{felsic}}{2}$$

$Mbd = modeled\ bulk\ density$
$mafic = model\ for\ soils\ of\ mafic\ parent\ material$
$felsic = model\ for\ soils\ of\ felsic\ parent\ material$
$other = model\ for\ soils\ of\ neither\ mafic\ or\ felsic\ parent\ material$
$SOC = soil\ organic\ carbon\ concentration,\ same\ as\ \%SOC$
***note:** soil bulk density data in Table A2a, **note2:** all soil parent materials indicated in Table A2b*

Once, bulk densities were calculated for every sample in this study, SOC (kg*m-2 was calculated using (***Eq.7)*** which also includes percent SOC, CF and depth.

$$\boldsymbol{Eq.\ 7}: \quad \boldsymbol{SOC_{kg*m-2}} = (\%\boldsymbol{SOC} * \boldsymbol{bulk\ density_{soil}} * \boldsymbol{depth_m}) * (\boldsymbol{1 - CF})$$

$\%\boldsymbol{SOC} = measured\ soil\ carbon\ concentrations\ (methods\ described\ in\ 2.3.2)$
$\boldsymbol{bulk\ density_{soil}} = from\ Eq.\ 3\ (field)\ or\ 4-6\ (model)$
$\boldsymbol{CF} = \frac{mass_{rock}}{mass_{total}},\ measured\ from\ field\ SOC\ samples$
$\boldsymbol{depth} = depth\ represented\ by\ sample\ (ie.\ 0-5\ cm = 0.05\ m)$

Soil Bulk Density and Percent Coarse Fraction (Cf)

We quantified percent coarse fraction (CF) as the ratio of coarse material relative to the total mass of sample collected. The coarse fraction includes both large rocks/cobbles as well as gravels. The resulting data is percent soil (equal to 1- percent CF) and percent CF.

pH

We determined the soil pH of our samples using a 1:1 soil to water mix with 25 g of each. Prior to analysis, we thoroughly mixed samples then left them to sit for an hour. Soil pH was then measured using a probe for three trials, where measurements were repeated if deviation was greater than 0.1 (units of pH) between the three trials. Then, we took the mean pH of three trials, which was utilized as the pH data for that each sample.

Soil Texture

We also measured soil texture, proportions of sand, silt and clay using a hydrometer method (McCorkle, 2015). We combined soil samples with Sodium hexametaphosphate (SHMP) and left the mixture to sit overnight. We then took triplicate readings of the hydrometer at zero and seven hours. Then, we wet sieved the mixture using a 0.05 mm sieve, which allows silt and clay to pass through while retaining the

sand-sized particles. We then took the sand-sized particles left from wet sieving and put them into muffle furnace for 8 hours at 450 degrees C. We took the mass of these samples before and after their time in the muffle furnace and calculated the difference.

The combustion step of texture analysis is for removing organic material that does not factor into soil texture analysis. Hydrometer readings of the sample allow for determination of relative proportions of silt and clay sized particles, which we calculated using Stoke's Law. While, the wet sieving and combustion allows us to calculate a mass of sand.

SOC Mapping and Modeling

We created distributed maps of SOC by generating statistical models based on our SOC dataset and a suite of predictor variables from hyperspectral, lidar and traditional datasets. In an attempt to quantify increases in prediction provided by LiDAR and hyperspectral, we added sets of variables sequentially (Figure A3, Table A1). The resulting predictor variables were split into three groups (traditional, traditional and lidar and traditional, lidar and hyperspectral variables) to evaluate the increase in predictive capability provided by the inclusion of lidar and hyperspectral variables.

We used each set of predictor variables to calibrate a model for two SOC datasets (percent SOC and SOC (kg/m2)) and two modeling approaches Random Forest (RF) (http://www.salford-systems.com) and Stepwise Multiple Linear Regression (SMLR) (MATLAB R2016a)). Upon providing calibration SOC dataset and suite of predictor variables, RF and SMLR perform a variable selection exercise that chooses the variables that predict SOC with the greatest accuracy. Although, the methodology for

determination of model accuracy and a variables predictive ability differ substantially, both were successful in selecting predictor variables with high predictive capacity.

Datasets

Hyperspectral and LiDAR Data Collection

The BCAL research group collected the airborne hyperspectral dataset utilized in this study in the spring of 2015 (AVIRIS-ng, spatial resolution: 2.5/1 meters, spectral range: 400-2500 nanometers, spectral sampling: five nanometers). The BCAL research group also collected the airborne lidar data used in this study in 2014 (Ilangakoon, 2016).

Other Datasets (base set of variables (non-LiDAR/non-hyperspectral)

In this section, we describes the sources of our other datasets, which we call the base set of variables, which we define as variables not derived from LiDAR or hyperspectral. We obtained estimates of annual and monthly precipitation for RCEW from the PRISM Precipitation model **(**PRISM Climate Group**)**. Development of the PRISM precipitation model involved some of the climate and meteorological data from RCEW. We used an NDVI (Web Enabled Landsat Data (WELD)) dataset as a comparison for the hyperspectral vegetation indices. The WELD NDVI dataset is the maximum NDVI at each 30 m Landsat pixel for the entire year (2012). Also in the base set of variables is a 30-meter DEM, provided by USGS, which we used as a comparison for the high-resolution lidar DEM.

Spatial Data Preparation

We used the .las files from the lidar data with the BCAL Lidar tools in ENVI (BCAL Lidar Tools) to calculate a suite of variables that represent vegetation structural parameters such as height and cover (Table A1c). We calculated lidar topographic

variables using the BCAL Lidar DEM (Illangakoon, 2016) and the SAGA GIS 2.2.5 Basic Terrain Analysis Module (Table A1b). We used the AVIRIS-ng hyperspectral data from 2015 to calculate a suite of vegetation indices, detailed on Harris Geospatial (Table A1d).

Once we created the distributed datasets described above, we extracted data values for each variable at each plot (60-meter diameter) using the ROI tool in ENVI Classic 5.3. Then, we calculated the mean of all pixels within the plot to obtain one value for each predictor variable at each site. We used a spatial resolution of 2.5-meters for the hyperspectral data and 3-meter for the lidar variables. This decision was based on the spatial resolution of the source datasets but variables were resampled to 3-meter if both lidar and hyperspectral were included in an SOC model.

Modeling Approach

Rf Variable Selection and Accuracy Assessment

We conducted RF model calibration using the program, Salford Systems Predictive Modeler (http://www.salford-systems.com). Using this program, we imported SOC data and predictor variables as a spreadsheet (.csv). Then, the GUI (Graphical User Interface) allows the user to set parameters for RF model simulations. During RF model calibration, SOC data and predictors variables are randomly selected over one thousand iterations in a process known as, bootstrap sampling. After 1000 simulations, Salford Systems calculates an out-of-bag error (OOB) based on left out samples through the simulations. These error estimates provide the metrics, R2 and RMSE along with a metric that measures a variables predictive ability, known as variable importance.

Rf Imputation

Once, we determined the most predictive variables, we aggregated them using the layer stack function in ENVI. This stack of .tiff files from ENVI was provided to the R wrapper, yaImpute (Crookston, 2007), which uses the random forest machine learning algorithm to produce spatially distributed maps of SOC using the point SOC data which is in a spreadsheet with data values for selected predictors, along with maps of the selected predictors in the layer stack.

Stepwise Multiple Linear Regression (SMLR) Variable Selection and Accuracy Assessment

We also developed maps of SOC using a Stepwise Multiple Linear Regression (SMLR) approach in MATLAB. We developing these maps by providing SOC data and our suite of predictor variables to the MATLAB function, 'stepwisefit' (Draper, 1998), which selects the most predictive variables and calculates regression coefficients for selected variables.

For variable selection, SMLR uses a methodology where inclusion of a variable in a model is dependent on the result of an f-test, which calculates a p-value for each predictor provided. The results of the f-test are based on a variables fit with the variable of interest, in this case, SOC. If the p-value for a variable is above/below a certain threshold, the variable will be included in the model. In this study, we calibrated SMLR models using the entire calibration dataset. For this reason, the error metrics we calculated for SMLR models are from the models fit with the entire field dataset. This is in contrast to the error metrics calculated by RF that are calculated from the left out

samples over one thousand simulations. For this reason, we cannot use the error metrics R2 and RMSE to compare between RF and SMLR models.

SMLR Imputation

To Impute the SMLR models, we arranged selected variables in a layer stack in ENVI. Then, we used the band math function to calculate SOC distribution using data values for predictor variables and calculated regression coefficients from MATLAB.

Additional Considerations Regarding SMLR Models

In this study, SMLR models appear to over fit field data due to inclusion of entire calibration SOC dataset and lack of bootstrap sampling. However, SMLR models could be modified to include bootstrap sampling. This would offset the influence of overfitting field data and it is likely the preferred approach if SMLR models are used for mapping SOC.

Overfitting was also evident in the selection of variables in SMLR models. SMLR models will continue to select variables as long as the fit between modeled SOC and field data improves. In this study, we limited SMLR models to five variables since increases in predictive ability were negligible and adding additional variables appeared to increase likelihood of overfitting. To limit overfitting, we attempted to select variables using the AIC (Akaike Information Criterion), which penalizes models for adding additional variables. However, this approach was ineffective as the selected variables were the same as the SMLR models where variable selection is determined by the p-values calculated during the f-test.

RESULTS

Introduction to Results

Review of Methods

This project developed maps of shallow soil carbon (30 cm depth) across the Reynolds Creek Experimental Watershed (RCEW), in SW Idaho, where a gradient in elevation and precipitation produces dramatic differences in vegetation across the watershed. We collected soil samples to 30 cm depth along this gradient using a nested sampling approach, (1344 samples, 279 data points, in 28 plots, discretized with depth) and analyzed the samples for SOC content. Using this large soil carbon dataset, we first evaluated a suite of spatially distributed variables (including traditional, as well as lidar and hyperspectral) for their capacity to predict soil carbon distribution. Then, we used those identified predictor variables to produce spatially distributed maps (using both RF and SMLR) of SOC across the watershed. Finally, based on field data and mapping products, we made a number of observations of key variables and trends that influence the soil carbon mapping effort.

Note: The datasets and products described in this document are available for download; for more information, find the citation, Will, 2017 in the references section.

Summary of Results

In this study, SOC generally increased along the precipitation-elevation gradient where there is an increase in the diversity and abundance of vegetation. Our research suggests that vegetation indices calculated from spectral data are the best predictors of SOC storage in semi-arid mountainous regions. Landsat NDVI explained roughly 60% of

the variance in SOC data and two hyperspectral vegetation indices, MRESR and MRENDVI, individually explain over 70%. SOC models that include both lidar and hyperspectral can explain up to 75% of variability in SOC but lidar alone is less effective without spectral vegetation information.

Dataset Summary

SOC data in this study ranged from 1.4-7.8 kg*m-2 in the top 30 cm with a mean of 3.35 kg*m-2 and a standard deviation of 1.94 kg*m-2. SOC concentrations ranged from 0.4%-4.9% with a mean of 1.79% and a standard deviation of 1.46%. Coarse fraction had a range of 7.1-61 with a mean of 27.5 and a standard deviation of 14.4. Soil bulk density ranged from 471-1641 kg*m-3 with a mean of 997 kg*m-3 and a standard deviation of 322 kg*m-3 (Figure 2).

Soil bulk density as well as coarse fraction (CF) vary dramatically across the watershed, strongly influencing mass/unit area SOC values. We also observed a negative correlation between coarse fraction and percent SOC at higher elevations.

The scale of observation illustrates how soil carbon varies at multiple scales. Plot scale variability (60 meter diameter, n=10) of SOC in this study was equal to roughly 1/3 of the variability in SOC observed at the watershed scale. In many cases, highly predictive variables of SOC at the watershed scale were not predictive at the plot scale.

Random Forest and Multiple Linear Regression Variable Identification

We analyzed a suite of traditionally used predictor variables (i.e. slope, aspect, precipitation, NDVI) as well parameters derived from airborne LIDAR and Hyperspectral analysis (Table A1). We conducted this analysis using both Random Forest (RF) and

Simple Multiple Linear Regression (SMLR) to predict SOC as mass per unit area (kg/m2) as well as concentration (mass %).

While a range of variables in all categories were predictive, hyperspectral vegetation indices were most predictive, significantly improving models of both percent SOC and SOC (kg/m2) (Table 1). RF R2 increased from 0.69 to 0.75 for percent SOC and from 0.65 to 0.75 for predicting SOC when the hyperspectral indices were included. Similar increases were observed in SMLR models when hyperspectral variables were introduced (R2=0.61 to 0.89 for SOC and from R2=0.8-0.94 for percent SOC). In the absence of hyperspectral data, NDVI was universally selected as the best predictor of SOC. Only four different highly correlated vegetation indices (Table A6) were selected as the best predictor of SOC through the 12 model calibration efforts. While, lidar variables were never selected as best predictors of SOC, several of the best models included lidar topographic and/or vegetation predictor variables.

While multiple variables commonly identified as good predictors, there were differences between both selection methods (RF vs SMLR) as well as target (SOC vs percent SOC). These differences in variable selection between RF and SMLR arise from the methodology of the statistical approaches, while differences between SOC and percent SOC are a result of differences in the distribution of the two SOC datasets. The best predictor of SOC and percent SOC in SMLR models was MRESR but RF selected the two variables, MRENDVI and PSRI. Topographic position index was also only selected by SMLR models while aspect was only selected in RF model calibration.

Linear Regression Analysis

Simple linear regression analysis for SOC found vegetation indices most predictive, but also identified elevation and precipitation as useful predictors (Figure 3, Table A1). NDVI had an R2 of 0.56 and 0.71 for %SOC and SOC while MRESR (hyperspectral) explained 88% of the variability in percent SOC and 71% in SOC. Relationships are not as strong for elevation (R2= 0.30 for SOC, 0.59 for %SOC) and precipitation (R2= 0.31 for SOC and 0.60 for %SOC). Topographic variables are even less effective at the watershed scale with R2 of 0.22, 0.16, 0.06 and 0.02 for aspect (northness), aspect (eastness), slope and insolation respectively.

Resulting Maps

The SOC maps generated with Random forest and stepwise multiple linear regression approaches produced maps that have many similarities (Figure 4). All maps exhibit an increase in SOC with increasing elevation and vegetation abundance. Similarly, all maps also show highest soil carbon in areas with heavy vegetation cover (conifer, aspen, juniper, high biomass sage).

The most significant difference between maps at the watershed scale is the rates of SOC increase and contrast between high and low SOC environments (gradual v. sharp). Some models produced a consistent, gradual change in SOC moving up the elevation gradient while others produced a sharp change in SOC when moving from sagebrush dominated to forested ecosystems. The degree of homogenization or heterogeneity in high and low SOC environments also varies between maps. Some models produced an abundance of high SOC values at high elevations, producing a relatively homogenous

SOC distribution while variability in high SOC environments can be observed in other maps.

Maps do, however, exhibit different patterns in soil carbon that relate to how different predictor variables represent SOC variability at smaller scales (Figure 4, Figure 5). One major difference is how SOC maps represent topographic differences in SOC. Strong aspect related differences in SOC are observed in maps where aspect was selected as a predictor while maps where topographic position index was selected show more influence from the proximity to channels and valleys versus ridges. Although, datasets with high spatial resolution (lidar and hyperspectral), provide increases in model accuracy, maps created using these datasets are more susceptible to extreme values, speckled characteristics or artifacts.

Composite and Standard deviation maps

Composite and difference maps were created to evaluate similarities and differences between models (Figure 6). SOC data values from all of the SOC and %SOC models were averaged to create composite maps. These composite maps including many of the same trends as the maps in Figure 4 while smoothing out differences between SOC and percent SOC maps.

The composite maps for SOC and percent SOC differ slightly, mostly at higher elevations where topographic influences on SOC vary depending on which variable was included in the model. Differences between models of both SOC and percent SOC are most significant at intermediate to high elevations while lower elevation and lower SOC portions of the watershed are relatively similar between models. The difference maps for SOC have more consistent variation at high elevations. However, the most significant

differences in percent SOC models appear to be driven by proximity to channels and valley bottoms. These trends are related to the selection of aspect and TPI (Topographic Position Index) as predictors as the SOC difference map is responding to aspect related differences in SOC while the percent SOC difference map is responding to differences in SOC concentrations caused by topographic position.

Spatial distribution of error in SOC model outputs

Generally, the spatial distribution of error in SOC models is dominated by a few outliers (Figure 7). Although, relatively high error was observed at a few low elevation/low SOC sites, error is generally greater at high elevations. While, error in %SOC models is relatively consistent at high elevations, many of the highest errors in SOC were at high elevation sites with high CF and often shallow soils. In several SOC model residual maps, alternating under and over estimates are evident, suggesting topographic controls are especially important at intermediate to high elevations. The largest differences between modeled and measured for both SOC and percent SOC occur at high elevations where SOC storage is also the greatest. In addition to high error, SOC models are producing a larger range of values in these high SOC environments.

Distribution of SOC data values from model outputs

The relative distribution of values (histograms) as well as the minimum, maximum, mean and median values of SOC and percent SOC varies between models (Figure 8, Table 2). While, all of the distributions for percent SOC and SOC models display a decreased proportion of pixels at high SOC values. The relative proportions of low and high SOC values varies as does the rate of decrease in proportion moving from low SOC values to higher SOC values. However, these differences in the distribution of SOC and

percent SOC model outputs generally lead to relatively small differences in total SOC at the watershed scale.

The distribution of output SOC values may be effective for highlighting differences between models but are unlikely to lead to the preference of a particular model. Since, the distribution of calibration data are influenced by the distribution of sample locations, which has its own biases.

Bulk density and percent coarse fragment are important variables

Soil bulk density and the fraction coarse fragment fraction are highly variable across the watershed; this high variability strongly influences calculated mass/unit area SOC values (Figure 9). We calculated SOC (kg/m2) using percent SOC, percent soil (1-CF) and soil bulk density (kg/m3), which is described in the laboratory processing portion of the methods section. Soil bulk density has a strong inverse relationship with percent SOC and elevation; where soils with high SOC content have much lower BD due to the increase in SOM relative to mineral soil that has much greater density. The spatial variability of CF is complex but proportions of coarse fragment tend to increase with elevation where SOC concentrations are also the highest. Interestingly, there is also a negative correlation between CF and percent SOC at high elevations.

Scales of SOC spatial variability

SOC has a high degree of spatial variability at multiple scales (Figure 9, Table 3). When we evaluate point data at the watershed scale, SOC concentrations (percent SOC) range from 0.2-8.2% while values for SOC range from 0.45-11.4 kg/m^2 SOC. When we average point data by the plot, the range of percent SOC by site is 0.4-4.9% while the range of plot averaged SOC is 1.4-7.8 kg/m2. Standard deviations at the watershed scale

are 1.46 percent SOC and 1.94 kg/m2 for SOC. When the point SOC data is averaged by site, standard deviations decrease to 0.5 for percent SOC and 0.82 for SOC. At the watershed scale, relative standard deviation (RSD, equal to standard deviation/mean) in percent SOC is greater than SOC with RSDs of 82 and 58 for percent SOC and SOC respectively.

The RSD in percent SOC at the plot scale is roughly one third of the variability observed in the entire watershed, with RSD of 28 and 24 for percent SOC and SOC respectively. However, both absolute (standard deviation) and relative variability (relative standard deviation) at the plot scale varies considerably. Plot scale standard deviations range from 0.13-2.23 for percent SOC and 0.18-2.34 for SOC in kg/m2, while relative variability at the plot scale ranges from 6-93 for percent SOC and 11-59 for SOC.

Variability at the pit scale is slightly less than plot scale variability with RSDs of 19 for the pit scale and 28 for plot scale. The mean uncertainty in SOC at a point, calculated from pits, was 0.22 in units of percent SOC and 0.37 kg/m2, which is equal to RSD of 19 for percent SOC and 15 for SOC. However, pit scale variability varies considerably across the watershed with SDs ranging from 0.02-0.97 kg/m2 SOC or 0.02-0.98 in units of percent SOC.

Laboratory uncertainty is less than uncertainty in SOC concentrations at a point, with RSD of 19 for uncertainty at a point and 9 for laboratory uncertainty. Propagation of laboratory uncertainty in SOC concentrations leads to a laboratory uncertainty equal to 0.05 kg/m2 for SOC. Lab uncertainty in SOC concentrations is greater than instrument uncertainty, with RSD of 9 for laboratory uncertainty and 2(aspartic acid)/3(m soil and triplicates) for instrument uncertainty.

Mapping SOC at the plot scale

SOC variability at the plot scale appears to be controlled by a variety of factors. Although, none of the predictors consistently explain plot scale SOC variability. In some cases, fine scale SOC variability can be modeled with remote sensing datasets (Figure 11, Table 5). RF models predicted SOC at the plot scale with a mean R2 of 0.53 while SMLR models predicted SOC at the plot scale with a mean R2 of 0.70/0.75. Individually, predictors evaluated in this research have a mean R2 at the plot scale near 0.2 when all sites are considered. However, a variety of predictors have R2 ranging from 0.6-0.9 for individual sites at the plot scale. Also, there appears to be a slight increase in the selection of topographic variables at this scale relative to watershed scale.

SOC depth relationships

SOC concentrations are generally highest at the surface but SOC depth relationships are highly variable throughout the watershed (Figure 12, Figure A1). Generally, variability in SOC is greatest near the surface where differences in SOC between canopy and interspace are also the greatest (Table A4D-F, Table 8).

Correlation between SOC and predictors also varies a bit with depth as higher R2 for SOC and vegetation indices are observed near the surface (Figure 13). The hyperspectral vegetation indice, MRESR explains 91% of the variability in SOC concentrations in the top 10 cm. However, MRESR is also very effective when predicting SOC to 50 cm depth with an R2 of 0.71. Elevation and precipitation display a slight increase in predictive capability with depth, shown by higher R2 for SOC below 10 cm.

DISCUSSION

Which Variables Were Most Predictive and Why

Spectral vegetation indices, such as NDVI or MRESR, have a stronger relationship with SOC than the other traditional (non-lidar and non-hyperspectral, detailed in Table A1a) and lidar derived predictors evaluated in this research. Hyperspectral vegetation indices were also modestly better as predictors than LANDSAT NDVI. We attribute the increase in predictive capability of hyperspectral vegetation indices to a combination of factors including but not limited to continuous coverage of EM spectrum and finer spectral sampling, timing of data collection and finer spatial resolution.

Wavelength selection and spectral sampling

Hyperspectral data provides a suite of different vegetation indices, many of which are highly predictive of soil carbon (Table A1d). The predictive differences are small and these indices are often highly correlated. Of the 52 hyperspectral vegetation indices evaluated, 26 had higher R2 for both SOC and %SOC compared with LANDSAT NDVI. Similarly, 18 indices had an R2 greater than 0.8 for percent SOC and seven indices had an R2 greater than 0.65 for SOC.

In this study, the best predictors of SOC used reflectance in the visible and NIR portion of the EM spectrum. The correlation between SOC and reflectance peaks near the red edge (700-750 nm) and indices calculated using reflectance data near these wavelengths, like MRESR and MRENDVI (calculated using reflectance at 445, 705 and

750 nm) had the highest correlation with SOC (Figure 14, Figure 15). However, there was an abundance of good predictors calculated using a variety of wavelengths and spectral sampling densities. This suggests that these factors can modestly improve models but are not solely responsible for the observed increases in SOC prediction accuracy provided by hyperspectral datasets relative to LANDSAT.

What are we measuring with hyperspectral vegetation indices?

In this study, we utilized relationships between SOC and spectral vegetation data to produce maps of SOC. However, the relationship between SOC and aboveground vegetation is complex and determining the vegetation characteristics responsible for these correlations is difficult. There are several important processes and physical characteristics of vegetation that can be measured using reflectance data in the visible and NIR portion of the EM spectrum. Many of the indices that were most predictive of SOC utilize reflectance data from this portion of the EM spectrum.

One important characteristic of vegetation reflectance data is the red edge (700-750 nm) which separates the EM spectrum into wavelengths that are dominated by absorption (< 700 nm) and wavelengths that are dominated by reflectance (>750 nm) in vegetation (Figure 16). The visible portion of the EM spectrum is heavily influenced by absorption features, which are present when photosynthesis is actively occurring, as shown in Figure 16. Vegetation indices in this portion of the spectrum may be used to estimate concentrations of photosynthetically active leaf pigments like chlorophyll a and b and carentoids (VREI2, PSRI, MCARI) (Figure 15). These wavelengths may also be indicative of water stress and the timing and duration of photosynthesis (MRENDVI, PSRI). The near-IR portion of the EM spectrum is dominated by reflectance

characteristics that are most closely related to metrics like LAI and leaf water content (MRESR, MRENDVI, MCARI2, MSI).

In this study, SOC was more highly correlated with absorption features (R2=0.53) relative to reflectance features (R2 = 0.3) (Figure 14). However, the most predictive vegetation indices generally use both absorption and reflectance features. The higher correlation for SOC in absorption-dominated wavelengths suggests that metrics related to the absorption of incoming solar radiation such as NPP are closely related to SOC. However, there is likely a strong correlation between the strength of absorption features in the visible and reflectance in the near-infrared. For example, greater rates of absorption or photosynthesis are observed in trees relative to sagebrush and trees also have greater leaf area and stronger reflectance features (Figure 16). Although absorption and reflectance is closely related to vegetation species, it likely varies within species as well as throughout the growing season.

The results of this analysis suggest that SOC is an ecosystem property, closely linked to water and energy fluxes and the associated vegetation response (carbon flux). The strength of these relationships suggest that other important metrics related to ecosystem fluxes could be estimated or spatially distributed using these vegetation indices or other variables from hyperspectral data.

Timing of Spectral Data Collection

In semi-arid regions, vegetation absorption and reflectance vary seasonally in response to growing season and loss of biomass as water limitation begins in summer. When the growing season ends, leaves begin to change color and reflectance characteristics may become less representative of the peak aboveground vegetation.

However, the timing of peak greenness and growing season length vary in both space and time. The peak in the growing season occurs in spring at low elevations and migrates up elevation, occurring in mid-summer at high elevations (Figure 17). This suggests the potential value of using different timed data for different elevations

These seasonal fluctuations in vegetation reflectance have a significant impact on the quality of vegetation indices and their ability to predict SOC. Vegetation indices collected near the peak in the growing season have a strong relationship with SOC. In contrast, spectral datasets collected outside of the growing season are generally ineffective. Figure 18 shows a comparison of three NDVI datasets. Both NDVI WELD (max NDVI of year) and hyperspectral NDVI collected in May were effective predictors with R2 of 0.71 and 0.75 respectively, but NDVI collected in September was much less effective with R2 of 0.34.

Resolution of remote sensing datasets and field observations

The scales of both the predictor variables and the field observations matter. Higher resolution spatial (spectral) datasets tend to have a stronger relationship with SOC. The increased predictive capability provided by high resolution datasets likely relates to improved characterization of sample plots by limiting influence from portions of pixels located outside of the plot.

Interestingly, the predictive capacity of remote sensing datasets and SOC declines when applied to individual points (cores and pit data) compared to plot-averaged values for SOC and predictors (Table 5). This observation suggests that the benefit of having more confidence in data points by taking the mean of multiple samples outweighs the potential benefit of representing SOC variability at a higher resolution.

Limitations of spectral data for predicting SOC at the watershed scale

Vegetation indices were very effective for predicting SOC at the watershed scale in this study (Table 1). However, there are limitations in their ability to predict the distribution of SOC (Figure 19, Figure 20). SOC varies in response to many interrelated factors that in some cases are not represented by trends in aboveground vegetation (Doetterl, 2015). Vegetation indices had a stronger relationship with percent SOC than SOC. Interestingly, Lidar topographic variables were only selected by models predicting SOC (percent SOC) concentrations. This is surprising considering that SOC (in kg/m2), appears to be more influenced by topography, which appears to be closely related to coarse fraction (percent CF).

Models calibrated with spectral vegetation data may have the tendency to over/under estimate soils in certain types of ecosystems (Figure 20). In forested areas, we observe a range of SOC and %SOC contents, yet vegetation indices consistently predict high SOC values. Similarly, species differences appear to produce SOC differences not captured by vegetation indices. Specifically, Aspen and conifer ecosystems produce similar values for vegetation indices but SOC storage tends to be higher in aspen ecosystems. Sagebrush ecosystems produce a range of values for vegetation indices but they are consistently lower than forested areas. In some cases, sagebrush sites stored more carbon than forested areas and were underestimated by models. In addition, Bitterbrush-dominated sites store relatively little SOC relative to sage brush, but have similar vegetation indices, causing overestimation of SOC. These vegetation trends appear to explain many of the mapped areas with elevated error.

Comparison of SOC Maps

All of the maps display a general increase in SOC moving up the elevation gradient, consistent with increases in aboveground biomass at high elevations. This trend is driven within the map modelling by spectral vegetation indices being selected as the best predictor. However, there are noticeable differences between the maps.

Some of these most visible differences between maps can be attributed to variables were selected as 2nd and 3rd best predictor. Perhaps the most striking difference between maps is driven by the selection of topographic related variables like aspect, elevation and/or topographic position (Figure 4). Selection of aspect as a predictor created blocky slopes with sharp aspect related differences in SOC while models that selected topographic position tend to be veiny with proximity to valleys and ridges controlling SOC at the hillslope scale.

Topographic variables are likely good predictors because the terrain modulates the energy and water balance and, by extension the generation and consumption of SOC. For example, in water limited environments, north-facing slopes typically have more SOC because these slopes often store more water (Kunkel, 2011; Patton, 2016). Similarly, slope position dictates insolation intensity and often is a good indicator of soil thickness, both important variables governing water availability. While, variables such as slope, aspect and topographic position, influence these processes, they do not uniquely capture them. This is further complicated by the fact that the influence of topographic alteration of water energy balance varies across the watershed in response to the gradients in the energy and moisture flux. For example, at higher elevations water limitation may be less influential than at lower elevations. Similarly, specific vegetation types are likely

to have distinct responses to the water-energy balance. For this reason, topographical variables are likely to be better predictors when energy and water flux is held constant. For example, at a fixed elevation in the watershed where climatic influences are more constant, these variables are likely to be effective predictors of differences in SOC.

Comparison of Random Forest and Stepwise Multiple Linear Regression Ii Approaches

In recent years, SOC and soil mapping research has focused increasingly on uncertainty and validation approaches, which is one of the limitations of early SOC mapping studies (Minasny; 2013). While, both random forest (RF) and stepwise multiple linear regression (SMLR) utilize environmental covariates to distribute point data and provide a mechanism for evaluating large quantities of highly correlated environmental covariates. The methodology for selecting variables and developing a model varies between the two approaches. Due to the robust modeling approach provided by bootstrap sampling, we would argue the RF is likely the preferred approach, especially when using data-rich datasets like hyperspectral imagery. This is in agreement with many other recent studies that have focused on soil mapping with hyperspectral and lidar datasets, which have used more robust modeling approaches (Minasny, 2006; Gomez, 2008; Grimm, 2008; Wiesmeier, 2011; LaCoste 2014).

During RF model calibration, field SOC data is randomly split into calibration and validation datasets while random combinations of predictor variables are also selected over thousands of iterations (Breiman, 2001). This approach is applied during model calibration (known as bootstrap sampling), allowing all of the field SOC data to be used for both calibration and validation. This approach improves confidence in model accuracy metrics and variable selection while limiting overfitting. In the more traditional

SMLR approach, the f-test (p-value) determines whether the addition of a variable improves an SMLR model. In this study, model accuracy and variable selection were based solely on the fit of regression equation with field data. The SMLR approach could be modified to include bootstrap sampling, which would decrease the likelihood of overfitting.

Controls on SOC in Semi-Arid Complex Terrain

The SOC content, both mass per unit area (SOC) and mass per soil mass (SOC%) varies dramatically over the Reynolds Creek Watershed (Table 6). Similarly, the controls on the generation and consumption (photosynthesis and respiration) of SOC vary, often with complex relationships with climate, topography and vegetation. Our mapping exercise was not explicitly designed to identify controls of SOC distribution. However, the richness of the observational data does provide some insight into the processes that control SOC in this semi-arid, mountainous watershed.

Scale and Ecosystem Controls on SOC

In this study, SOC generally increased moving up the elevation gradient from hot/dry low elevation ecosystems to higher elevation ecosystems with lower temperatures and more precipitation. The complexity of SOC variability can be observed in varying predictive capability (R2) for predictors depending on both scale and ecosystem of interest (Table 6, Table A4).

Importance of Water Availability

Water availability is a primary control on ecosystem processes in semi-arid regions (Noy-Meir, 1973; McFadden, 2013) and, by extension, exerts a strong influence on SOC distribution. High rates of insolation often lead to water stress (Kunkel, 2011) as

water availability limits rates of photosynthesis for the many intermediate and low elevation sagebrush ecosystems (PET>PPT, WUE=90-5%) (Fellows, 2016). In semi-arid regions, moisture availability sets a limit on the potential for C flux and storage on an annual basis and increases in moisture generally lead to increased rates of photosynthesis and an associated increase in SOC. This observation suggests that photosynthesis is more responsive to moisture availability than respiration in this ecosystem; a trend hinted at by previous researchers (Schlesinger, 1977; Schlesinger, 1997; Joggaby, 2000). Carbon input to soil via photosynthesis is also more likely to exhibit response to not just amount, but also timing of moisture availability (Smith, 2011).

SOC follows similar trends to modeled NPP data (Figure 21) but correlations between SOC and NPP are not particularly strong (Figure 22). There is likely a stronger correlation between these two variables than what is observed in Figure 21. But, there is significant hillslope scale variability in NPP (Adams, 2014; Swetnam, 2017), which is not captured by coarse resolution data (250 meter pixel size).

Topography Exerts More Control on SOC at High Elevations

While SOC has a strong relationship with precipitation at low elevations, the interplay between climate, topography and vegetation is more complex at high elevations. At higher elevations, SOC moisture relationships are not captured by measured precipitation data and SOC storage appears to be more heavily influenced by topographic and geomorphic controls.

This trend is illustrated in Table 6, where precipitation data explains 2% of the variability in SOC above 1800 m, while CF explains 90% of variability in SOC (kg/m2)

and 41% of SOC concentrations. In addition, topographical position index (TPI) can predict 60% of the variability in SOC and 55% of variability in CF.

The strong relationships between SOC and topography at high elevations likely relate to increased importance of snow as a source of moisture. Snow (re)distribution and residence time into the summer season is strongly influenced by topography (Stephenson, 1965; Winstral, 2002; Williams, 2009). Field observations from this study are consistent with other studies that have observed shallower soils (In this study, all pits shallower than 50 cm depth were at high elevations or local topographic high points) and more coarse material at higher elevations and/or higher topographic positions (Schimel, 1985; Hoffmann, 2014) (Figure 9).

These topographic and geomorphic controls on moisture availability create significant differences in SOC at the hillslope scale (Kunkel, 2011; Patton, 2016) (Figure 23 and Figure 24). Soils on ridges and some slopes tend to be erosive environments with less SOC, more coarse material and shallower soils, while soils on lower topographic positions are more likely to be depositional environments characterized by deeper, finer grained soils with moisture and SOC (Stephenson, 1965). Where soils are thinner, limited soil moisture storage, likely results in soil moisture deficits earlier in the season (Smith, 2011). It is likely that photosynthesis is more strongly influenced by moisture limitation in these environments, depressing SOC production.

The controls on SOC in valley environments are quite different. These thicker soils can store more moisture and also experience increased moisture flux since saturated soils upslope are likely sources of moisture for soils at lower topographic positions (Stephenson, 1965). In addition, topographic controls on the redistribution of snow may

also lead to increased moisture storage in valleys relative to ridges. This is probably most dramatically evident by the strong tendency for aspen to grow on the leeward, and more northern facing sides of ridges where large snow drifts generally accumulate in the winter and soils are often disproportionately thick and fine-grained (Stephenson, 1965).

The above observations illustrate the positive feedback between the soil hydrologic characteristics and aboveground vegetation (Table 7, Figure 25 and Figure 26) The high predictive capability of hyperspectral data, which is measuring vegetation differences, also appears to capture the influence of soil properties (Ballabio, 2012), in particular soil moisture availability. Rates of photosynthesis over longer time scales are heavily influenced by the composition of vegetation communities which vary in terms of their ability to generate fluxes of C (Schlesinger, 1997). Areas with more moisture are more likely to support vegetation communities that increase potential for greater net gains in C over time and for this reason SOC storage is greater in these environments.

Importance of Soil Texture and Parent Material Controls on SOC

Soil texture and amount of coarse material strongly influence SOC in the RCEW. This influence is illustrated in Figure 27, where soil texture appears to be regulated by parent material. Relative to felsic soils, soils with mafic parent material are generally have more coarse material (CF for mafic sites= 33%, felsic=21%.), lower soil bulk densities (BD for mafic sites= 0.94 g/cm3, felsic=1.12 g/cm3) (Patton, 2016) and finer soil textures. These soil parent material characteristics are also related to differences in weathering rates as well as the particle size and mineralogy of weathering products (Aguilar 1988). These soil characteristics in turn affect soil hydrologic characteristics (like water holding capacity), which alter the composition of associated vegetation

communities (Low sage is found mostly on mafic soils while bitterbrush and juniper appear to have an affinity for coarser grained felsic soils) as well as SOC contents (Aguilar 1988).

It is worth noting that this study was not designed to evaluate differences in SOC caused by parent material. For this reason, the distribution of felsic and mafic sites may be biased. Areas of potential bias in this analysis are the distribution of felsic and mafic sites with respect to elevation (Mean elevation=1660 meters for mafic sites and 1509 meters for felsic sites), precipitation (584 mm/yr. for mafic sites and 501 mm/yr. for felsic sites) and NDVI (3975 for mafic sites and 3780 for felsic sites).

Importance of Soil Conditions and Covariation with SOC

Several other key soil parameters have strong relationships with SOC (Table A2). Soil nitrogen was correlated with SOC to an R2 of 0.97. While, C: N ratios were also highly correlated with SOC as higher SOC soils had a higher ratio of carbon to nitrogen, with an R2 of 0.71. In addition, strong relationships were observed between SOC and pH as more acidic soils were found in high SOC soils while lower SOC soils, often containing SIC (soil inorganic carbon) have basic or neutral pH. SOC concentrations explained roughly 51% of the variability in soil pH.

Trends in Plot Scale SOC Data

Biota is responsible for much of the nutrient cycling in arid and semi-arid regions. In these ecosystems, widespread roots systems accumulate nutrients, which are concentrated near vegetation (Schlesinger, 1996; Titus 2002). The results of this study appear to be consistent with this observation. In this study, proximity to vegetation, as measured by paired sampling in canopy and interspace, showed the greatest differences

in SOC in the hottest/driest ecosystems, primarily composed of Wyoming sagebrush (Table 8). However, juniper ecosystems also displayed significant differences in SOC between interspace and canopy. These types of ecosystems with less continuous vegetation cover and more bare ground have the greatest plot scale SOC variability and localized C input heavily influences fine scale SOC variability.

Higher moisture environments, usually at higher elevations, tend to have less variability and canopy/interspace differences in SOC are not as pronounced. This likely occurs because ecosystems with more moisture often support more interspace vegetation which appears to reduce SOC plot-scale variability. In some cases, large vegetation canopies spread detritus over larger areas or across an entire site and proximity to individual trees is less significant. Although, the relative difference in plot scale SOC is lower at higher elevations, higher SOC contents in these ecosystems may lead to greater absolute differences in SOC contents at the plot scale.

Discussion of Plot Scale SOC Mapping Results

Trends in aboveground biomass have strong relationships with SOC at the watershed scale but these relationships often breakdown at the plot scale. In some cases, SOC variability at the plot scale can be modeled using high-resolution lidar and hyperspectral datasets (Figure 28). However, at many sites, SOC distribution at the plot scale is not well predicted (Table 4, Figure 11).

In general, applying the approach we used at the watershed to map SOC at the plot scale was not particularly effective; we attribute this failure to three reasons. First, the controls at the plot scale are often observed at the scale of individual plants (often sub-meter). The resolution of remote sensing data is not always able to capture vegetation

with sufficient accuracy to capture SOC variability. Second, plant distribution at the plot scale is not driven by stationary climatic variables and likely varies in time. Soil carbon residence times are often longer than most plant lifetimes, so SOC contents are likely the influenced by prior vegetation distributions. This signal is not measured by current vegetation distribution. Third, plot scale prediction is likely hampered by data limitations. The influence of decreased sample size relative to predictor variables is evident in several places. The difference in R2 and RMSE between RF and SMLR models is greater at the plot scale than the watershed scale (Table 4). In addition, a greater range of R2 and RMSE was observed for RF models. This variability in SOC models shows the sensitivity of model accuracy metrics to sample size as individual samples have more influence on selection of predictors and model accuracy.

Heterogeneous ecosystems with larger vegetation such as conifers or juniper, are the most likely to be mapped with accuracy at fine scales considering the size of vegetation relative to the spatial resolution of remote sensing datasets. In low elevation sagebrush ecosystems, SOC generally increases near sagebrush, but distributing SOC data is complicated by the influence of soil reflectance and scale/resolution issues. At low elevations, bare soil has considerable influence on reflectance spectra and differences in vegetation reflectance are less pronounced. In addition, sagebrush are smaller than the resolution of remote sensing datasets, making it difficult to distribute SOC data based on proximity to vegetation. In contrast, at many high elevation sites, vegetation cover is relatively continuous and differences in vegetation reflectance are less significant, yet SOC varies below the canopy, illustrating the limitations of hyperspectral data as a SOC predictor in these settings.

Implications for SOC Modeling

While this work confirms the value of hyperspectral remote sensing data for predicting SOC in semi-arid environments, we have also identified influences on SOC distribution that are unlikely to be captured by hyperspectral data alone. Of particular importance are soil bulk density and rock content, which strongly influence SOC in the Reynolds Creek watershed. These variables, along with soil thickness, are more likely to be captured by topographical variables, highlighting the value of a hybrid approach. Identification of spatially distributed datasets that capture rock content could also improve predictive capacity. The strong influence of water availability on vegetation distribution in semi-arid systems and the high correlations between vegetation and SOC in this work, suggest that it could also be productive to explore use of soil moisture, or evapotranspiration, parameters to directly predict SOC distribution.

Sensitivity of SOC and Ecosystem C Flux to Projected Changes in Climate

Some of the observations above may assist in understanding the sensitivity of SOC storage in projected climate scenarios. Since the beginning of the 20th century, the Earth's mean surface temperature has increased by about 1°C. (IPCC, 2007). This is consistent with research from RCEW which states that the mean surface temperature in western North America has increased by 1–3°C over the last 50 years (Nayak, 2010), with rates of temperature increase being even greater in mountainous regions (Lemke, 2007). The influence of increasing temperatures can also be observed in research from RCEW which shows movement of the rain snow transition over the last 50 years (Klos, 2014).

The carbon balance and SOC storage are in dynamic equilibrium with ecosystem energy and water balance and changes in SOC will likely be apparent in shifting vegetation patterns (Ballabio, 2012). Examples of vegetation trends in the Great Basin that are potentially affecting ecosystem carbon dynamics are the expansion of juniper (Rau, 2011a; Kormos, 2016), and the decrease in cold temperature species like Douglas fir and Alpine fir (Beedlow, 2013; Restaino, 2016). In addition, more frequent fires and larger fires (Westerling, 2006), increase the probability of net C loss through ignition and natives being replaced by an invasive such as cheatgrass which generally store less SOC (Rau, 2011b; Austreng 2012).

Importance of high elevation ecosystems

In this research, error in SOC models and differences between SOC models were the greatest at high elevations where SOC contents are also the highest. Much of this is due to the complexity of these ecosystems, difficulty with scale and the abundance of related topographic variables, such as aspect, curvature and topographic position, which all relate to energy and water related controls on SOC as well as geomorphic controls, while not explicitly representing them.

Although, much research has went into understanding SOC variability at the hillslope scale (Schimel, 1985; Kunkel, 2011; Bameri, 2015; Patton, 2016), there is still considerable uncertainty regarding this issue. Topographically-induced hillslope scale variability in SOC is likely one of the largest source of uncertainty in SOC estimates currently. It may also be one of the most important to improve since ecosystems in arid/semi-arid regions and/or complex terrain are the most sensitive to alterations to water and energy balance (Hoffmann, 2014).

These ecosystems also appear to have greater potential to be sources of C to atmosphere given the high SOC contents and hydro climatic setting of these environments relative to lower-precipitation sagebrush ecosystems that have lower SOC contents and soil nutrient and climatic conditions that do not favor high rates of respiration. The changes in NPP in lower elevation ecosystems are also likely to be less sensitive to warming since rates of SOC accumulation and NPP are already severely water-limited in these ecosystems.

SOC uncertainty in complex terrain

The spatial variability of SOC, operating at multiple scales is often the largest source of uncertainty in SOC estimates (Minasny, 2013). This is especially true in complex terrain where significant changes in soil forming factors over short distances leads to uncertainty in SOC (Hoffmann, 2014). Efforts to understand SOC variability are often complicated by issues of scale (Wang, 2009; Miller, 2015) and feedbacks (Ballabio, 2012) which make it difficult to establish relationships between SOC and environmental covariates.

These complex influences on SOC necessitate evaluation of bias. Robust modeling approaches that employ bootstrap sampling allow for detailed uncertainty analysis and may be the best approach for evaluating bias. This is an important consideration in terms of model quality since bias in SOC and predictor variable datasets may lead to flawed interpretation of SOC controls.

Considering the issues of scale and uncertainty described in this research, it may be useful to employ non-quantitative analysis of model options. We attempt to provide a simple framework for evaluating tradeoffs between multiple models in the context of

these issues with Figures 5-7. Figure 5 shows where models are similar as well as where they are different in space, highlighting areas of high sensitivity to model choice and potential uncertainty. Figure 6 shows the spatial distribution of error based on field data fit, allowing for highlighting areas where models are less effective in SOC prediction. These figures could be used for selection of a particular model if it is more important to be accurate in certain areas. In addition, a study by (Weismeier, 2011), suggested using the spatial distribution of error to improve models with regression kriging based approaches. Figure 7 shows the distributions of model outputs and how much total SOC varies between models, which could be useful for determining how sensitive watershed or regional estimates are to selection of a particular model or to evaluate the distribution to see if it matches field data or expected results of the modeling effort. Although, this framework may not lead to the selection of a model, it may be useful for evaluating the pros and cons of different models, understanding the error and bias and determining how to potentially improve to the models.

CONCLUSIONS

Our research suggests that, the best predictors of SOC storage are vegetation indices calculated from spectral data. NDVI explains roughly 60% of the variance in SOC data while two hyperspectral vegetation indices, MRESR and MRENDVI explain over 70%. SOC models that include both lidar and hyperspectral can explain up to 75% of variability in SOC but the Lidar-derived variables evaluated in this study are less effective without spectral vegetation information.

In semi-arid mountainous Reynolds Creek Watershed, SOC has high degree of spatial variability at multiple scales. Local variability of SOC in this study was quite high with roughly 1/3 the variability in the watershed observed at the plot scale. At the plot scale, the large-scale climatic drivers that control SOC at the watershed scale are absent and the degree of variability and its controls are more complex. The amount of SOC at a location is heavily influenced by BD (soil bulk density) and CF (coarse fraction) that are also highly variable in space in complex terrain. The interactions of these variables make SOC (kg/m2) harder to predict than percent SOC. Coarse fraction exerts even more influence on SOC distribution at high elevations where SOC contents tend to be the highest

SOC has a strong, positive, relationship with the watershed-scale variables of elevation and precipitation. However, trends in vegetation abundance at the hillslope scale, as estimated from remote sensing datasets, is a better predictor of SOC distribution. This observed relationship between SOC and aboveground biomass likely reflects the

fact that vegetation is the primary source of C input to soil, suggesting that inputs (photosynthesis), rather than outputs (respiration) may more strongly influence soil carbon accumulation in this environment. Although, precipitation was less effective as a predictor of SOC, there is evidence that moisture is a primary influence on SOC distribution. In this study, locations with higher soil moisture (north-facing and thicker soils) have higher vegetation densities and generally store the most SOC. In the rain-snow transition, the relationship between SOC and precipitation is complex and measured precipitation is not always representative of available moisture.

APPENDIX (FIGURES AND TABLES)

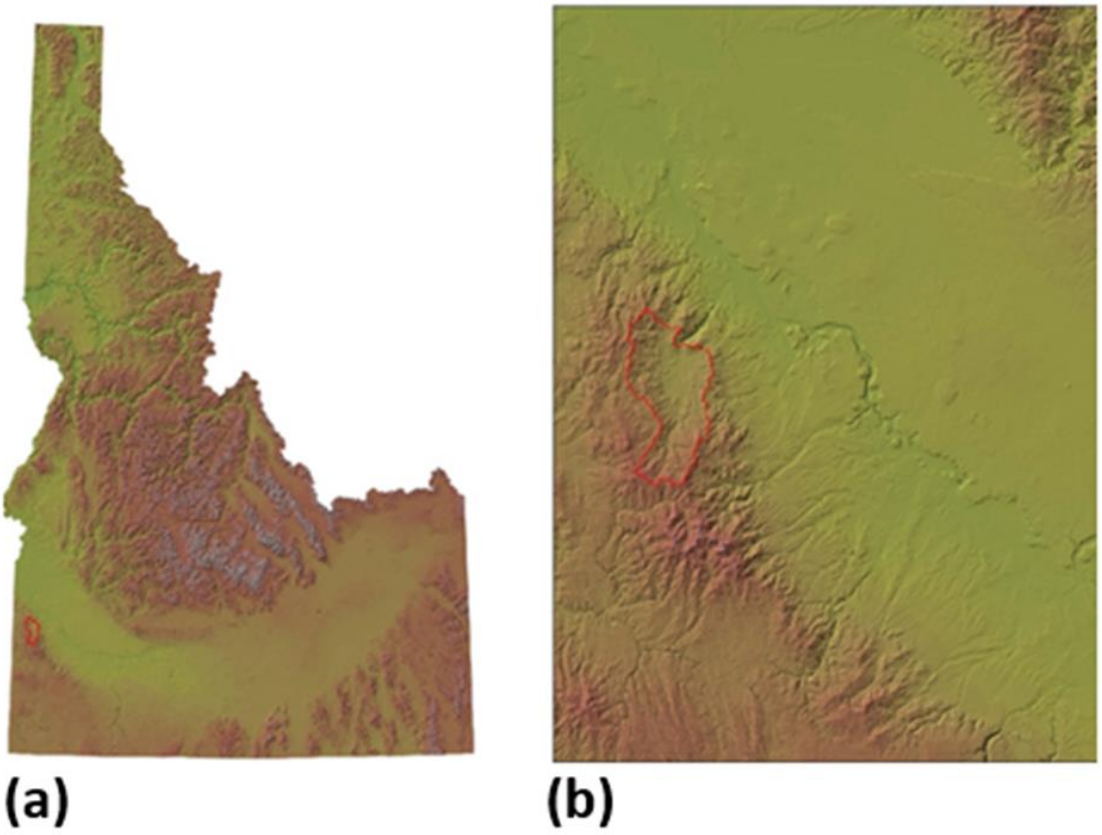

Figure 1a. Location of RCEW within Idaho. b. Close-up of RCEW location in Idaho.

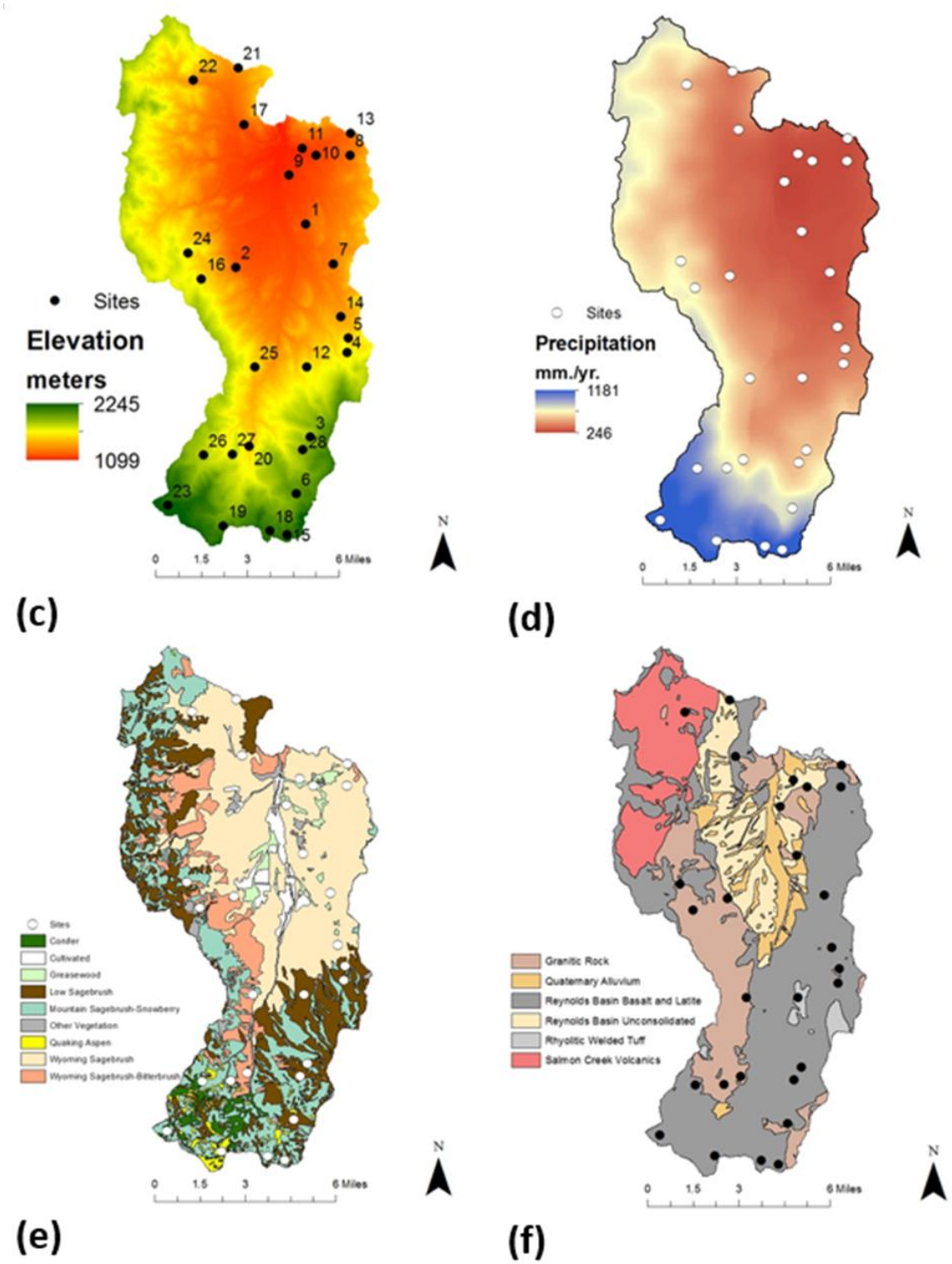

Figure 1c. Map of elevation (meters) for RCEW with sites numbered. d. Map of precipitation for RCEW with sites shown. e. Map of vegetation species for RCEW with sites shown. f. Map of surface geology for RCEW with sites shown.

Field Observations

	SOC kg*m-2	%SOC	CF	Bulk density Kg*m-3
Range	1.4-7.8	0.4-4.9	7.1-61	472- 1641
Mean	3.35	1.79	29	1010
Median	2.81	1.24	27.5	997
Std. Dev	1.94	1.46	14.4	322

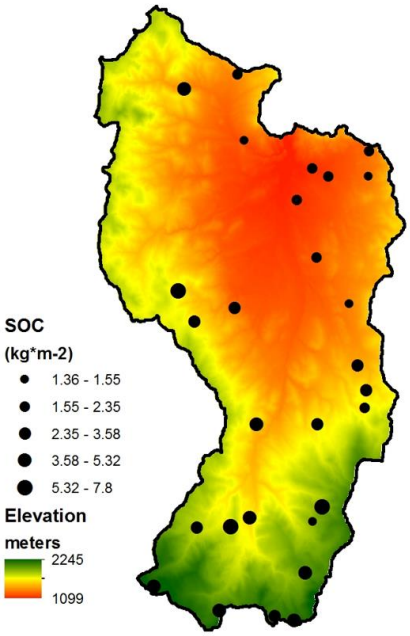

Figure 2. Summary of SOC and percent SOC data, soil bulk density and coarse fraction

Table 1. Error metrics and selected predictors for every model and their associated map. For Random Forest (RF) models, both field data fit and internal accuracy assessment (bootstrap sampling) metrics were provided

Map#	Data	Approach	Datasets	R2	RMSE	Variables
1	SOC	RF	T	0.65	1.01	NDVI, aspect (by class)
2	SOC	RF	TL	0.57	1.12	NDVI, *curvature minimum*
3	SOC	RF	ALL	0.75	0.8	***MRENDVI, PRI,*** *%V1_2.5m*
4	SOC	SMLR	T/TL	0.58	1.15	NDVI
5	SOC	SMLR	ALL	0.79	0.81	***MRESR, NLI,*** *elevation*, *V%2.5_10*, *vegetation density*
6	%SOC	RF	T	0.68	0.77	NDVI, aspect (by class)
7	%SOC	RF	TL	0.69	0.77	NDVI, *elevation*, *vegetation height kurtosis*
8	%SOC	RF	ALL	0.75	0.63	***PSRI***
9	%SOC	SMLR	T	0.69	0.79	NDVI
10	%SOC	SMLR	TL	0.8	0.63	NDVI, *TPI*
11	%SOC	SMLR	ALL	0.88	0.44	***MRESR***, *TPI*, ***PSRI,*** *V%1_2.5*

Note 1: ***Bold and italic:*** hyperspectral, **Bold:** LiDAR topographic, *Italic:* LiDAR vegetation. If not specified, the variable is considered traditional.
Note 2: Spatial resolution of predictor variables: Traditional datasets: 30-100 m, LiDAR and hyperspectral: 3 m. **Note 3:** Map 4 represents both the traditional and traditional + LiDAR approaches since LiDAR variables were not selected when provided.

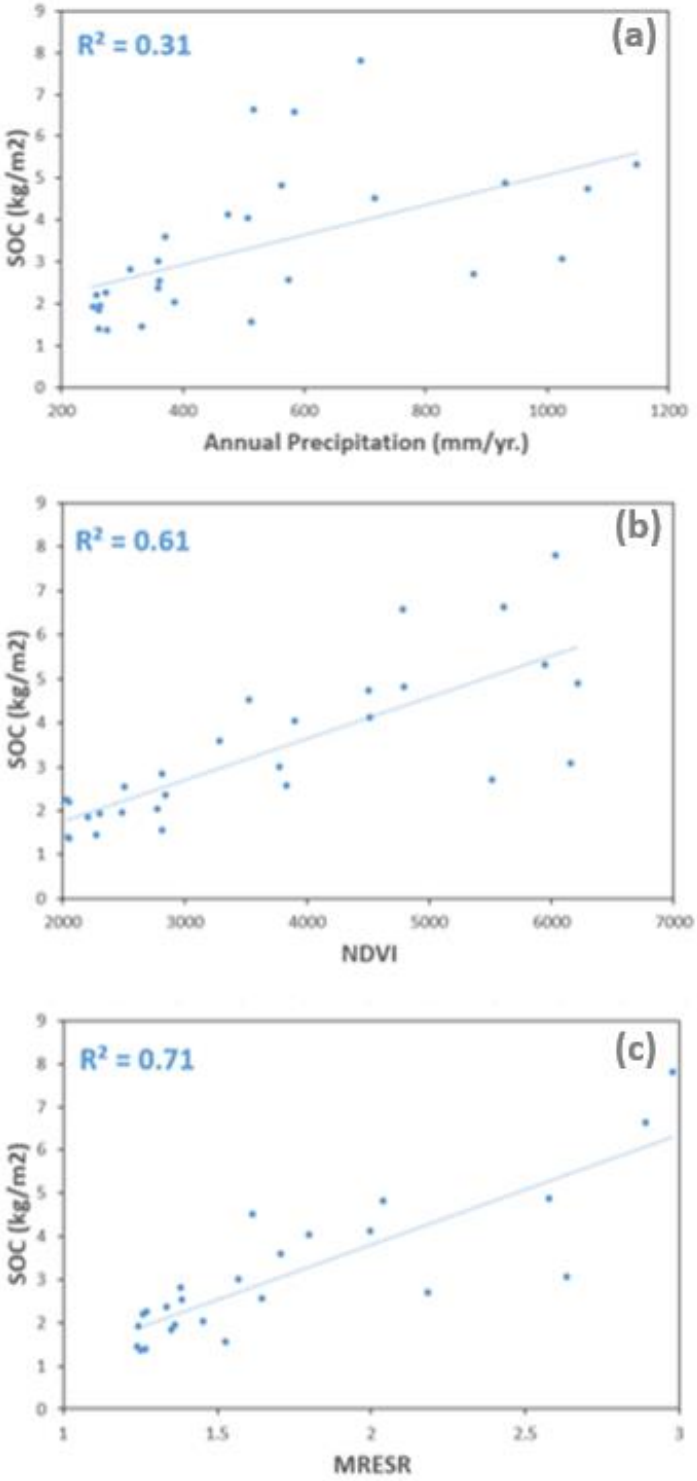

Figure 3. 1:1 Linear Regression plots of SOC and the predictors (precipitation (a), NDVI (b) and MRESR (c)). Data for this figure was generated using site averaged values for SOC and predictors at each site.

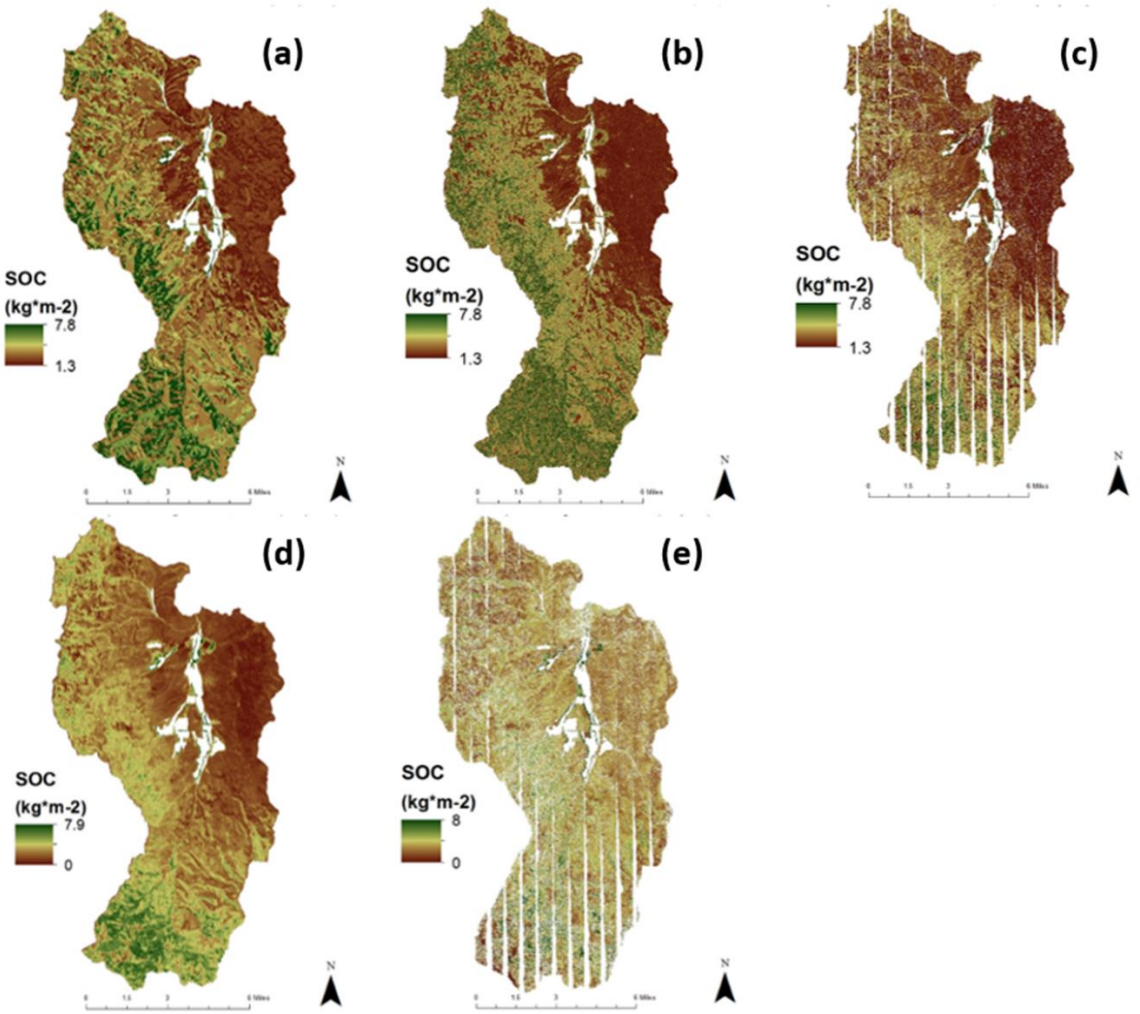

Figure 4a. Map 1 (SOC) was created using only traditional predictor variables in a Random Forest model. b. Map 2 (SOC) was created using traditional and LiDAR-derived predictor variables in a Random Forest model. c. Map 3 (SOC) was created using traditional, LiDAR-derived and hyperspectral predictor variables in a Random Forest model. d. Map 4 (SOC) was created using traditional variables in a Stepwise Multiple Linear Regression model (LiDAR-derived variables provided but not selected). e. Map 5 (SOC) was created using traditional, LiDAR-derived and hyperspectral predictor variables in a Stepwise Multiple Linear Regression model.

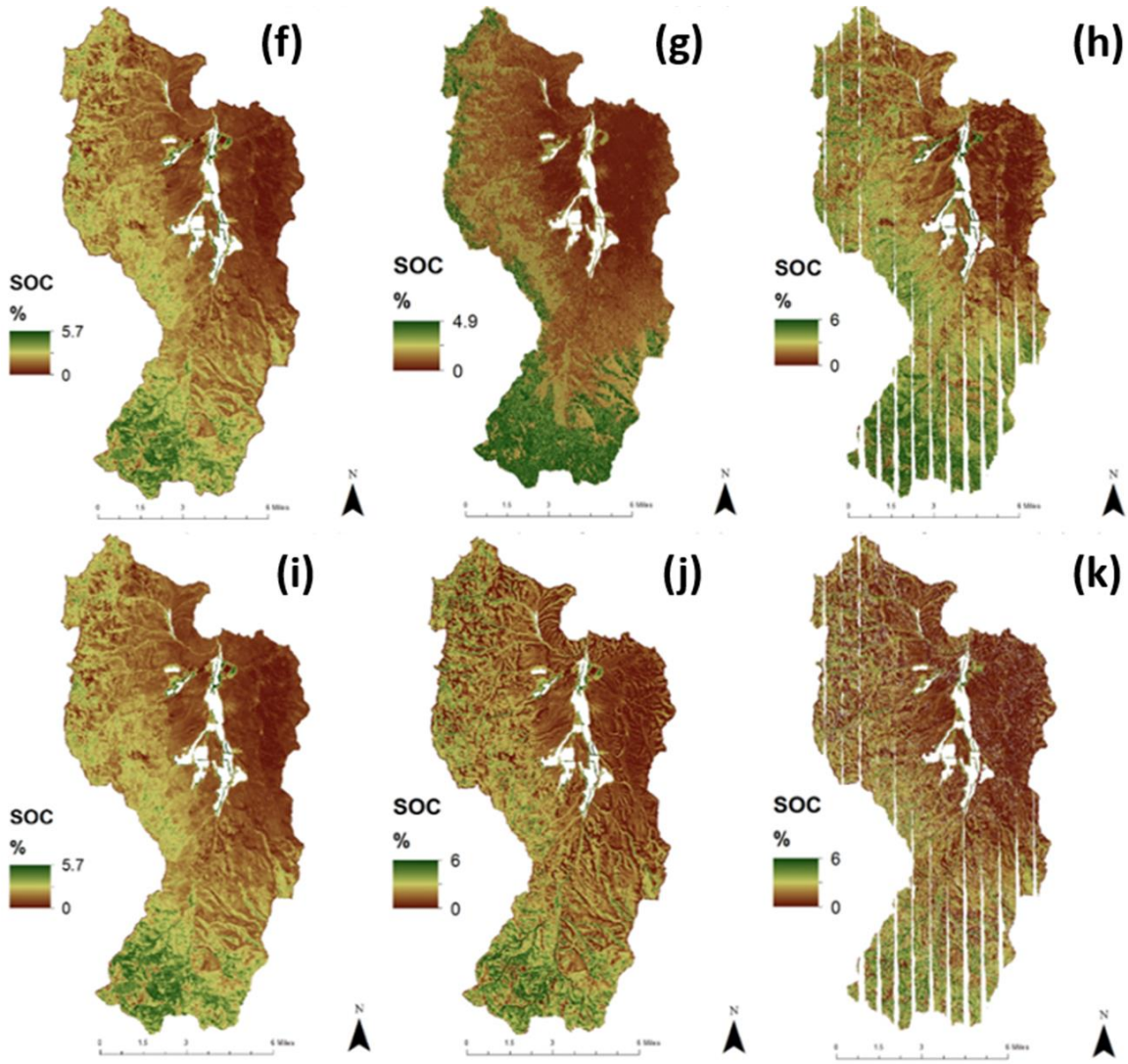

Figure 4f. Map 6 (%SOC) was created using only traditional predictor variables in a Random Forest model. g. Map 7 (%SOC) was created using traditional and LiDAR-derived predictor variables in a Random Forest model. h. Map 8 (%SOC) was created using traditional, LiDAR-derived and hyperspectral predictor variables in a Random Forest model. i. Map 9 (%SOC) was created using traditional variables in a Stepwise Multiple Linear Regression model. j. Map 10 (%SOC) was created using traditional and LiDAR-derived predictor variables in a Stepwise Multiple Linear Regression model. k. Map 11 (%SOC) was created using traditional and LiDAR-derived predictor variables in a Stepwise Multiple Linear Regression model.

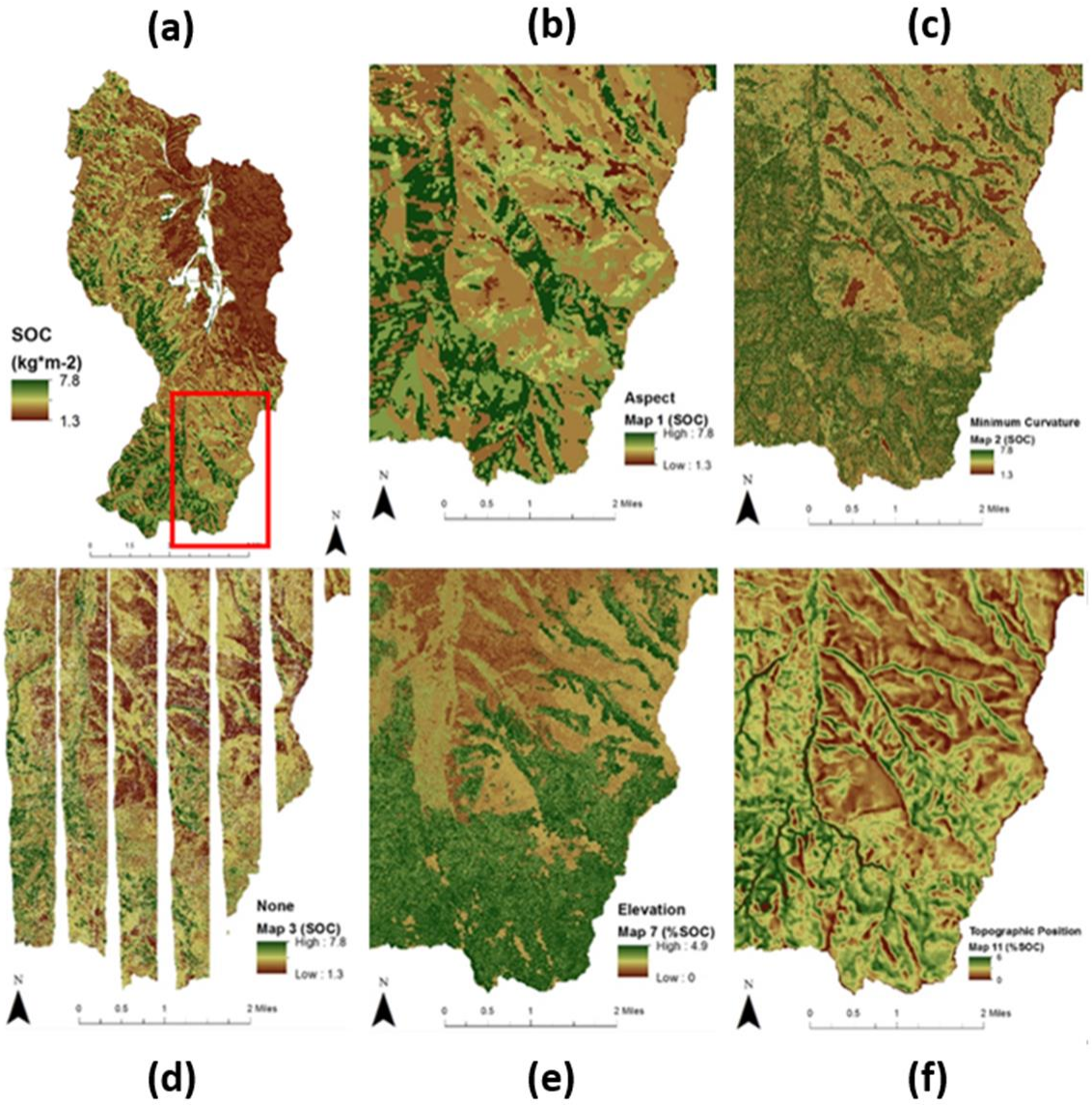

Figure 5a. Extent of close-ups shown in Figure 4b-f. b. Close up of Map 1 (SOC) where aspect was selected as a predictor. c. Close up of Map 2 (SOC) where minimum curvature was selected as a predictor. d. Close up of Map 3 (SOC) where no topographic variables were selected. e. Close up of Map 7 (%SOC) where elevation was selected as a predictor. f. Close up of Map 11 (%SOC) where topographic position index was selected as a predictor.

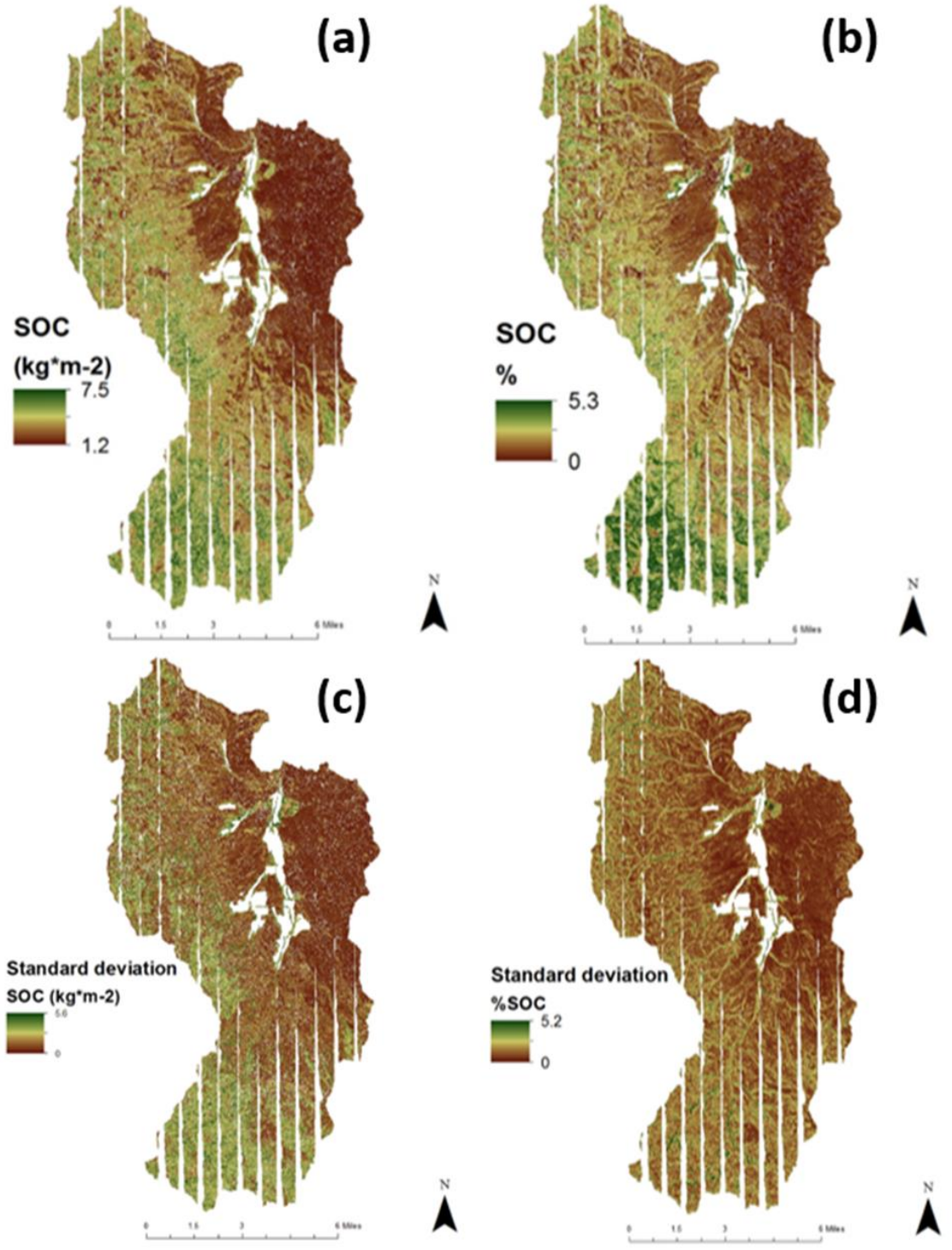

Figure 6a. Composite map for SOC, created by averaging the model outputs of all the SOC models. b. Composite map for %SOC, created by averaging the model outputs of all the %SOC models. c. Standard deviation map for SOC, created by calculating the standard deviation of all of the SOC models. d. Standard deviation map for %SOC, created by calculating the standard deviation of all of the %SOC models.

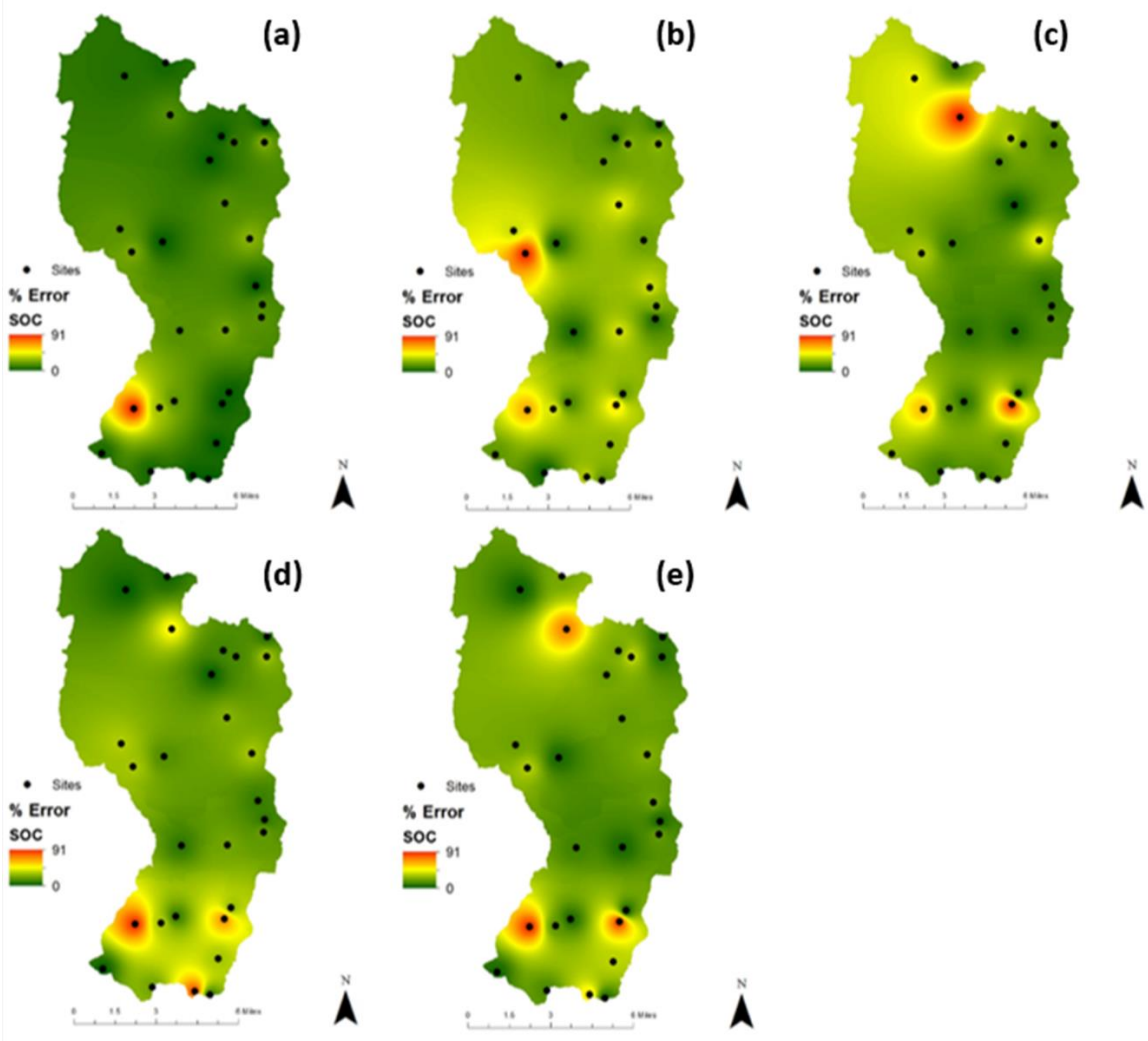

Figure 7a. %Error (absolute value of (Modeled SOC-Measured SOC/Measured SOC)*100) map for Map 1. b. %Error map for Map 2. C. %Error map for Map 3. d. %Error map for Map 4. e. %Error map for Map 5

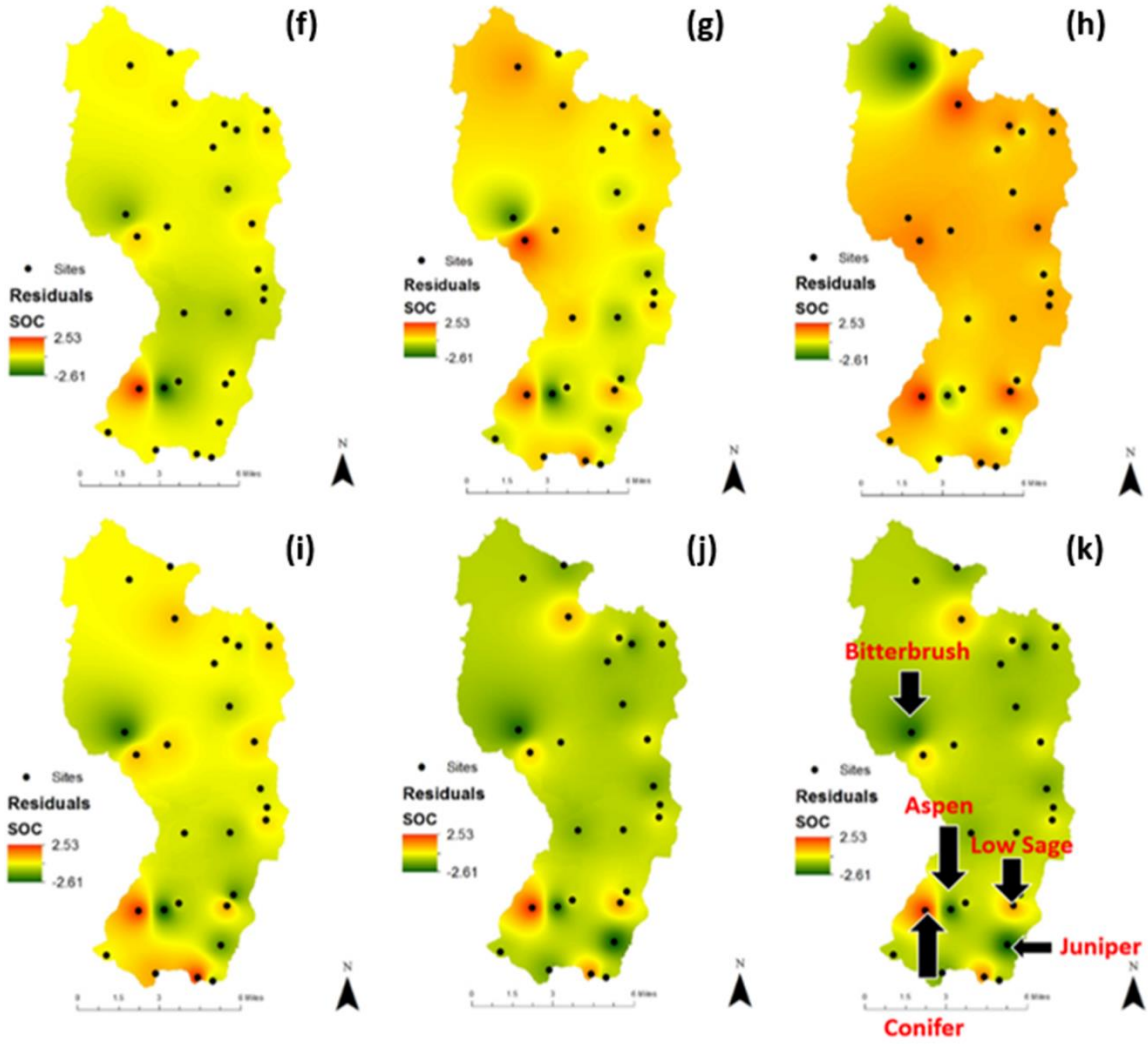

Figure 7f. Residual map (observed SOC-predicted SOC) for Map 1 (positive (red) = underestimation, negative (green) = overestimation). g. Residual map for Map 2. h. Residual map for Map 3. i. Residual map for Map 4. j. Residual map for Map 5. k. Residual map for Map 5 with vegetation species labeled at sites with considerable over and underestimation

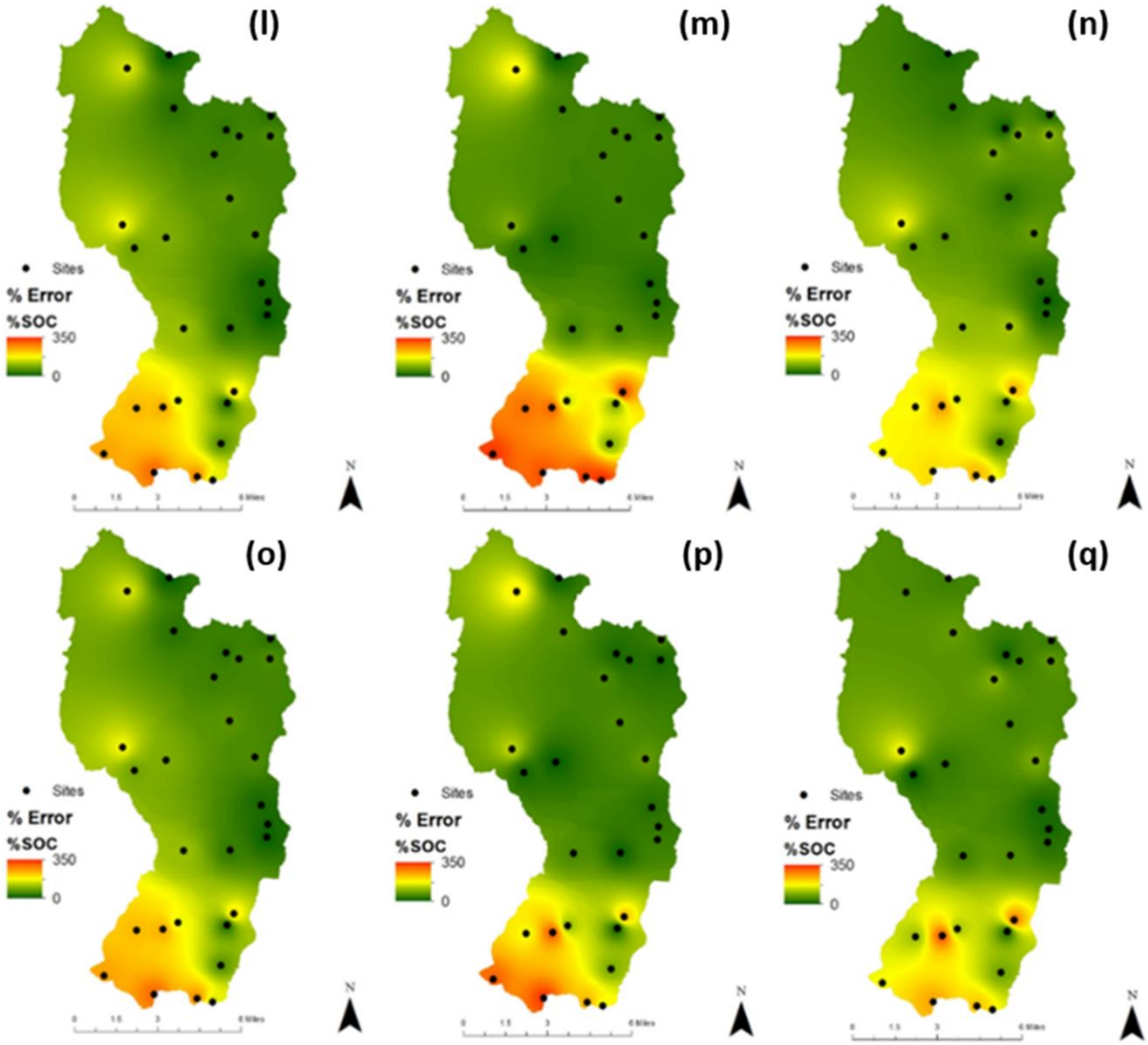

Figure 7l. %Error (absolute value of (Modeled %SOC-Measured %SOC/Measured %SOC)*100) map for Map 6. m. %Error map for Map 7. n. %Error map for Map 8. o. %Error map for Map 9. p. %Error map for Map 10. q. %Error map for Map 11

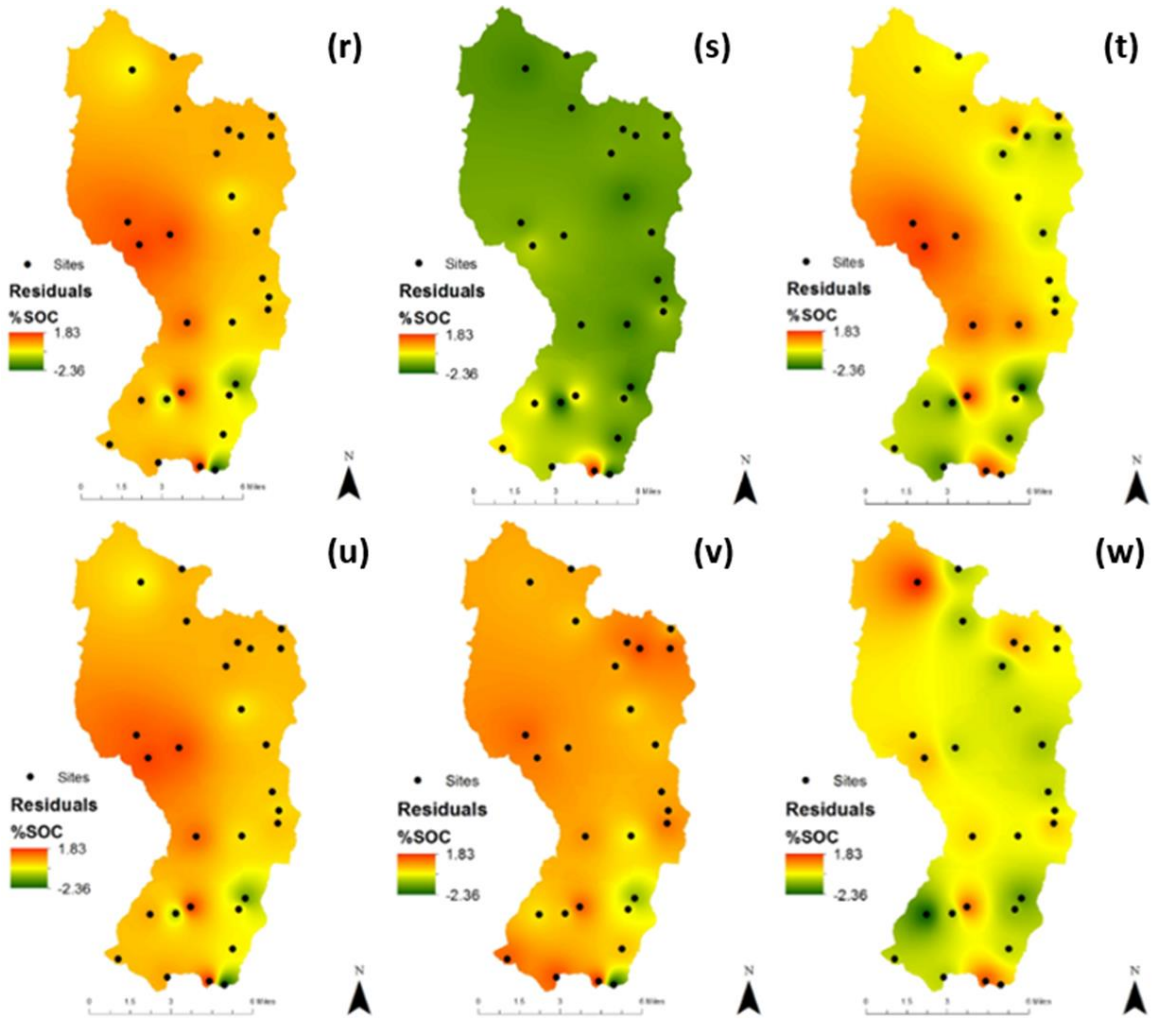

Figure 7r. Residual map (observed %SOC-predicted %SOC) for Map 6 (positive (red) =underestimation, negative (green) =overestimation). s. Residual map for Map 7. t. Residual map for Map 8. u. Residual map for Map 9. v. Residual map for Map 10. w. Residual map for Map 11.

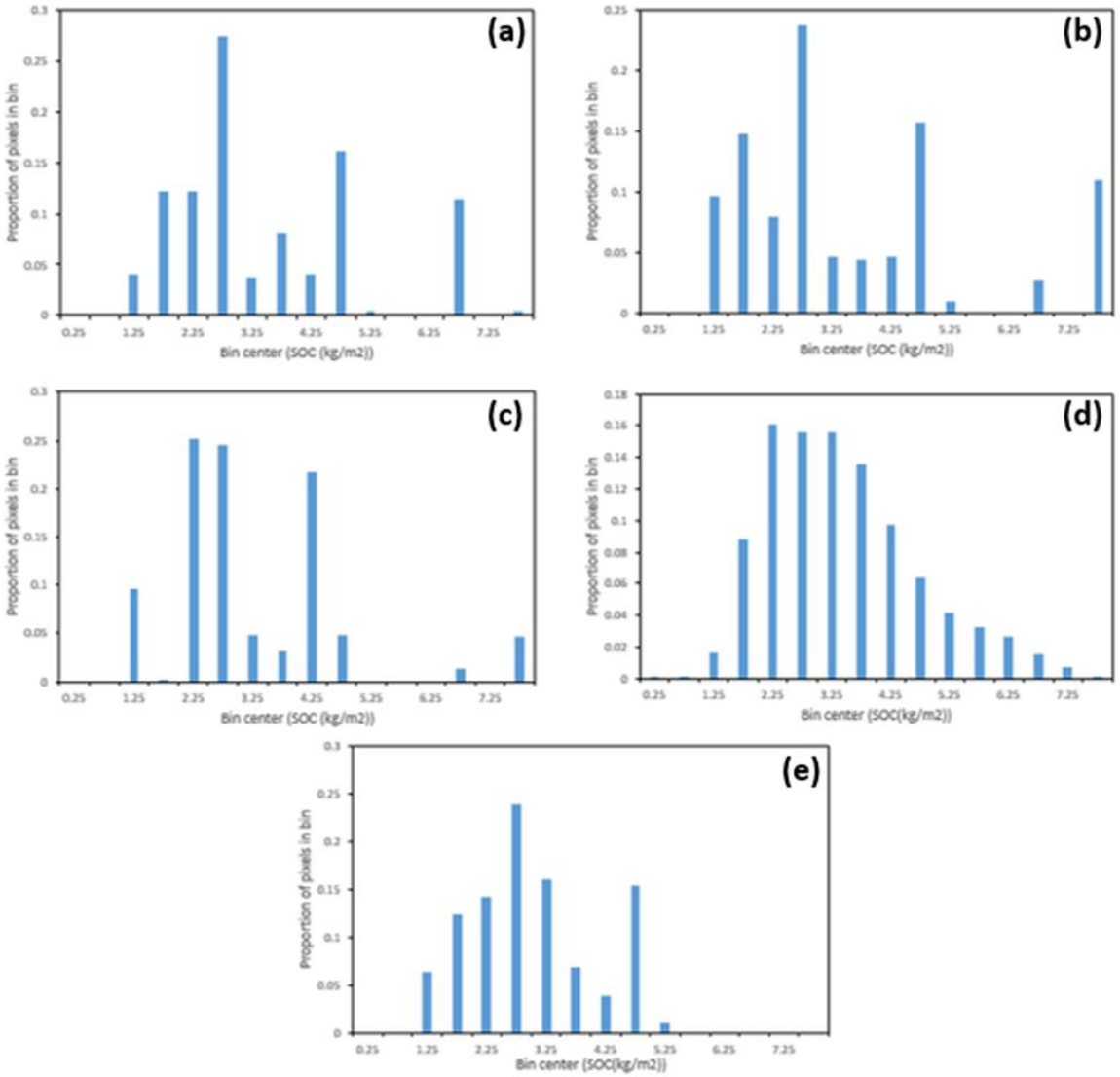

Figure 8a. Histogram of modeled SOC data from Map 1. b. Histogram of modeled SOC data from Map 2. c. Histogram of modeled SOC data from Map 3. d. Histogram of modeled SOC data from Map 4. e. Histogram of modeled SOC data from Map 5.

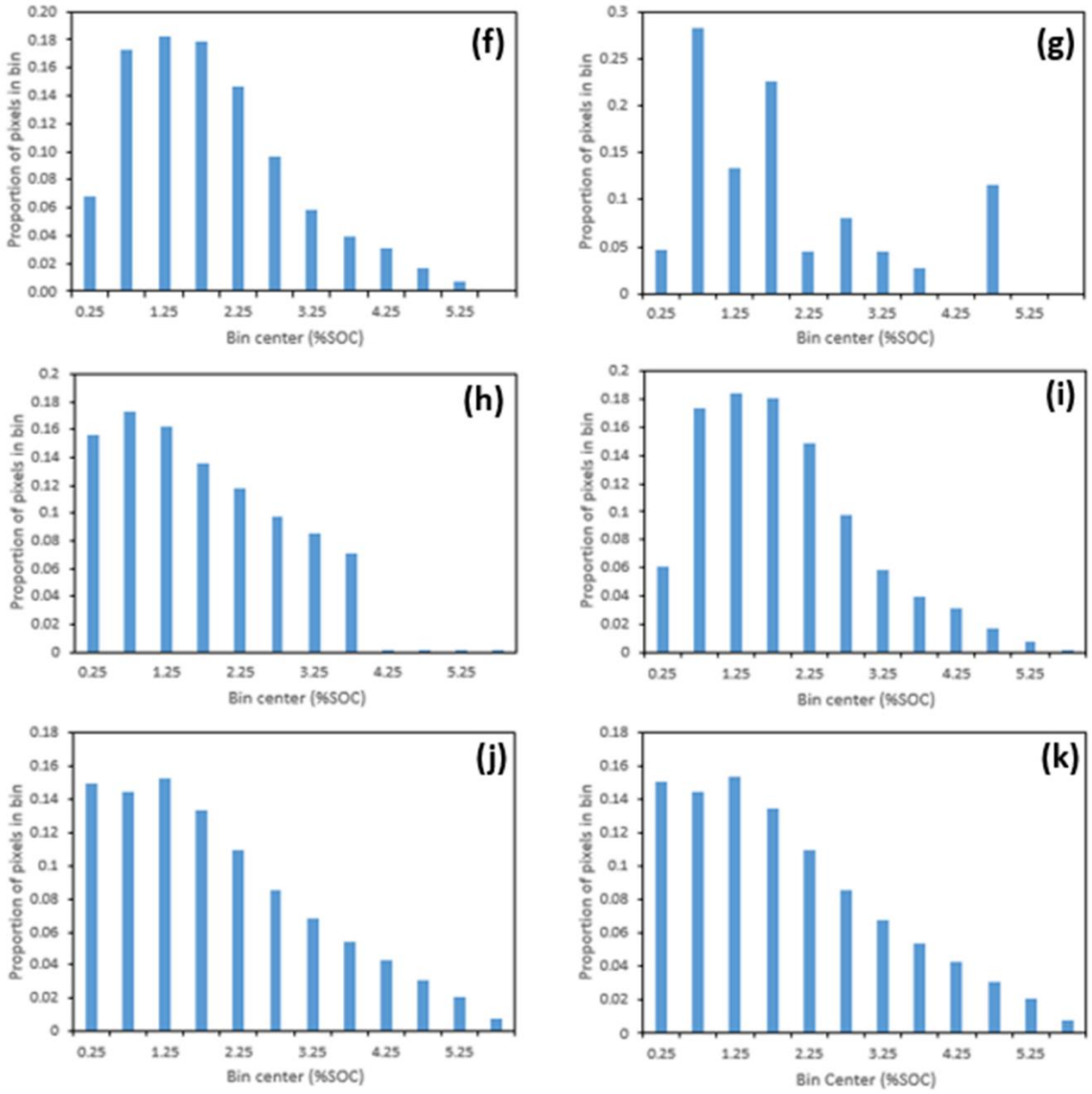

Figure 8f. Histogram of modeled %SOC data from Map 6. Figure 7g. Histogram of modeled %SOC data from Map 7. h. Histogram of modeled %SOC data from Map 8. i. Histogram of modeled %SOC data from Map 9. j. Histogram of modeled %SOC data from Map 10. k. Histogram of modeled %SOC data from Map 11.

Table 2. The minimum, maximum, median, median and total C of each model

Map#	Min	Max	Median	Mean	Sum
1	1.36	7.8	2.99	3.4	8.99E+05
2	1.36	7.8	2.99	3.51	9.29E+05
3	1.36	7.8	2.81	3.16	7.26E+05
4	0.0027	7.9	3.25	3.44	9.10E+05
5	1.36	7.8	2.99	3.57	9.44E+05
6	1.60E-04	5.71	1.72	1.89	N/A
7	0.41	4.88	1.58	1.91	N/A
8	4.70E-07	6	1.74	1.92	N/A
9	2.00E-04	5.71	1.72	1.89	N/A
10	2.40E-07	6	1.82	2.08	N/A
11	2.40E-07	6	1.82	2.08	N/A

Note: Data for maps 1-5 is SOC (in kg*m-2). Data for maps 6-11 is percent (%SOC).

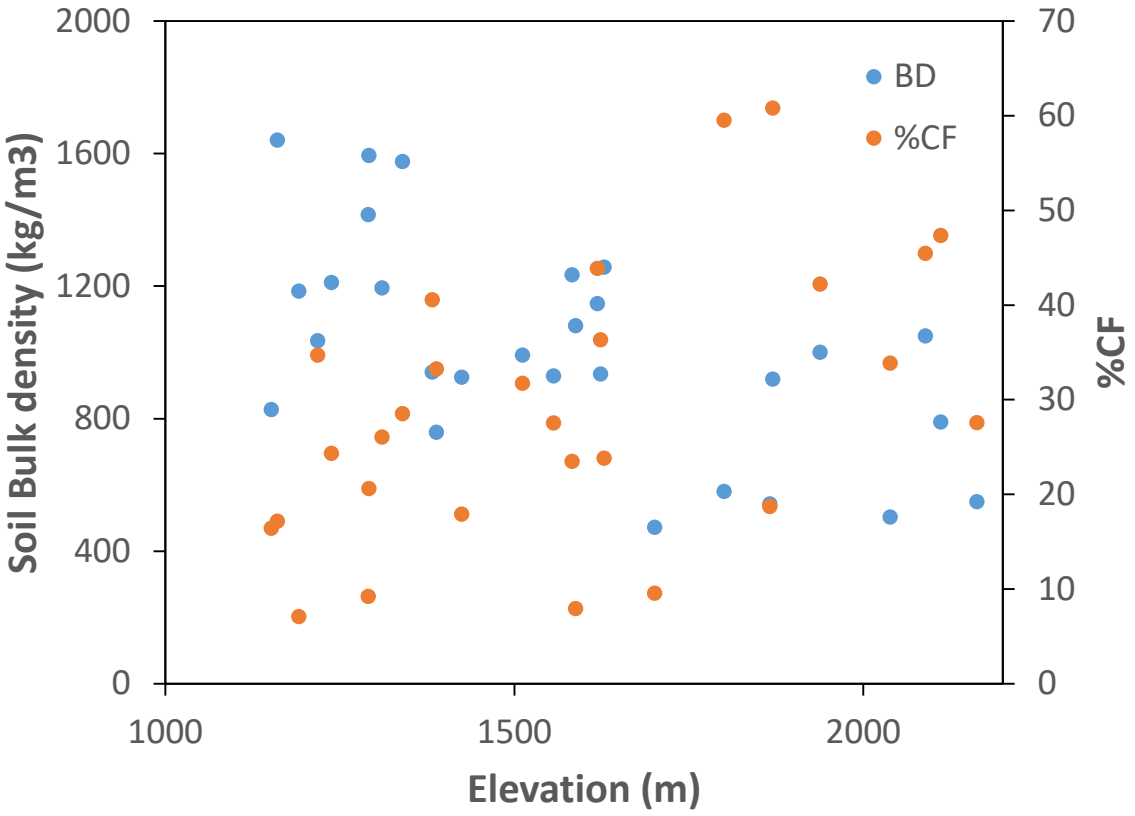

Figure 9. Soil bulk density and %CF measured at each of the sites in this study on an elevation gradient.

Table 3. Several metrics that illustrate the scales of SOC variability.

Scale		SOC (kg/m2)	%SOC
Watershed	Range	0.45-11.4	0.2-8.2
	Mean	3.35	1.79
	Median	2.81	1.24
	SD	1.94	1.46
	RSD	58	82
Plot	Range of Means	1.4-7.8	0.4-4.9
	Range of SD	0.18-2.34	0.13-2.23
	Range of RSD	11-59	6-93
	Mean SD	0.82	0.50
	Mean RSD	24	28
Point	SD	0.37	0.22
	RSD	15	19
	Mean	2.88	1.44

Note: At a point is defined as all of the cores and pits (mean SOC of 3 pit walls) (*n=279*) in this study. Uncertainty at a point is equal to the standard deviation in SOC calculated using samples from three adjacent pit walls less than 1 m apart. Plot averaged is equal to the mean of all 10 points at each site (n=28). The representative scales for each group are: point=1 m2, plot=3600 m2, watershed= 2.4e8 m2.

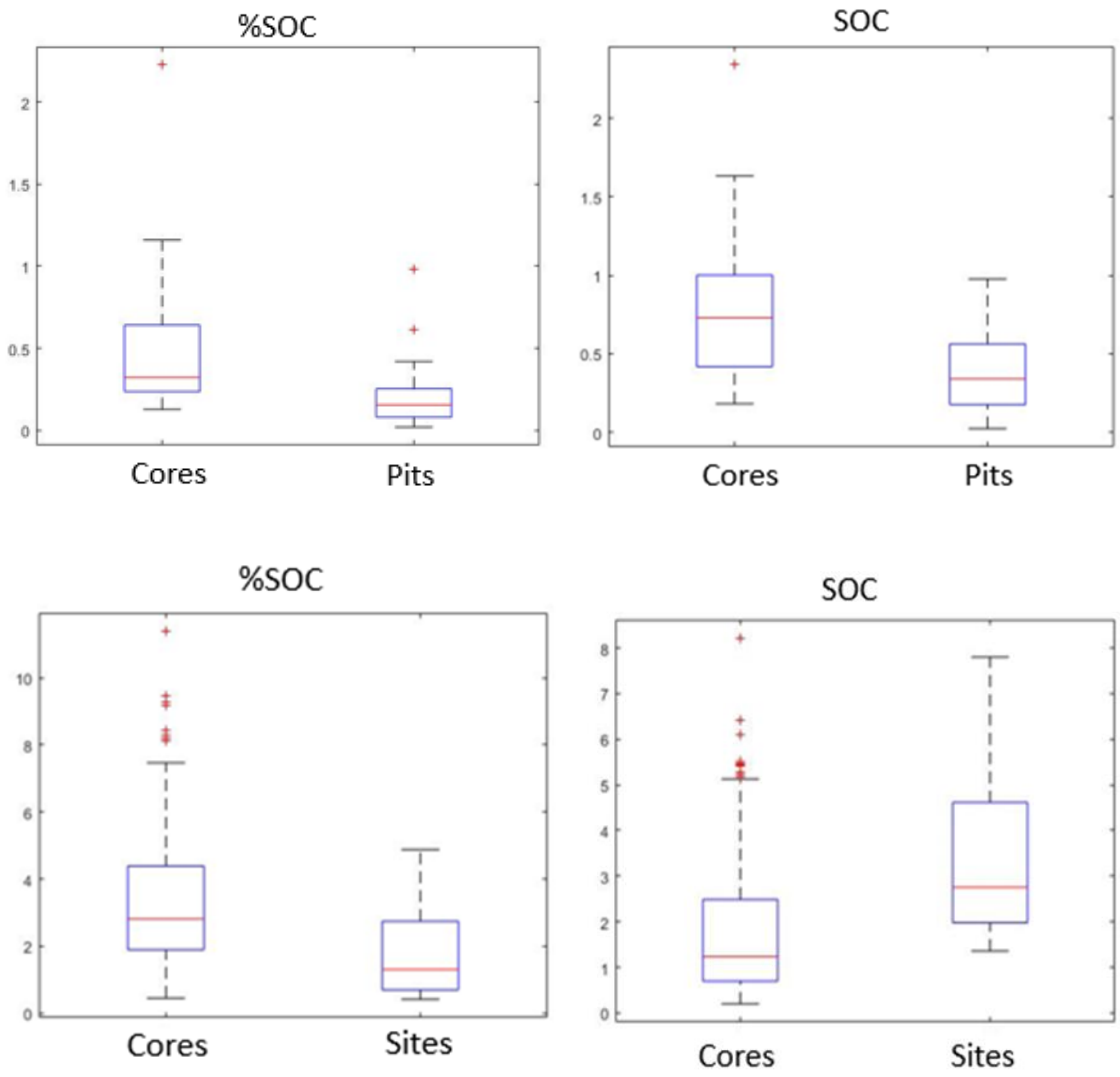

Figure 10. Box plots comparing variability (standard deviations) in SOC and %SOC between cores and pit walls (10 cores vs. 3 sides of pit) (above) with data from all 28 sites. The boxplot below show how SOC data (average of 10 data points vs. 279 individual data points)

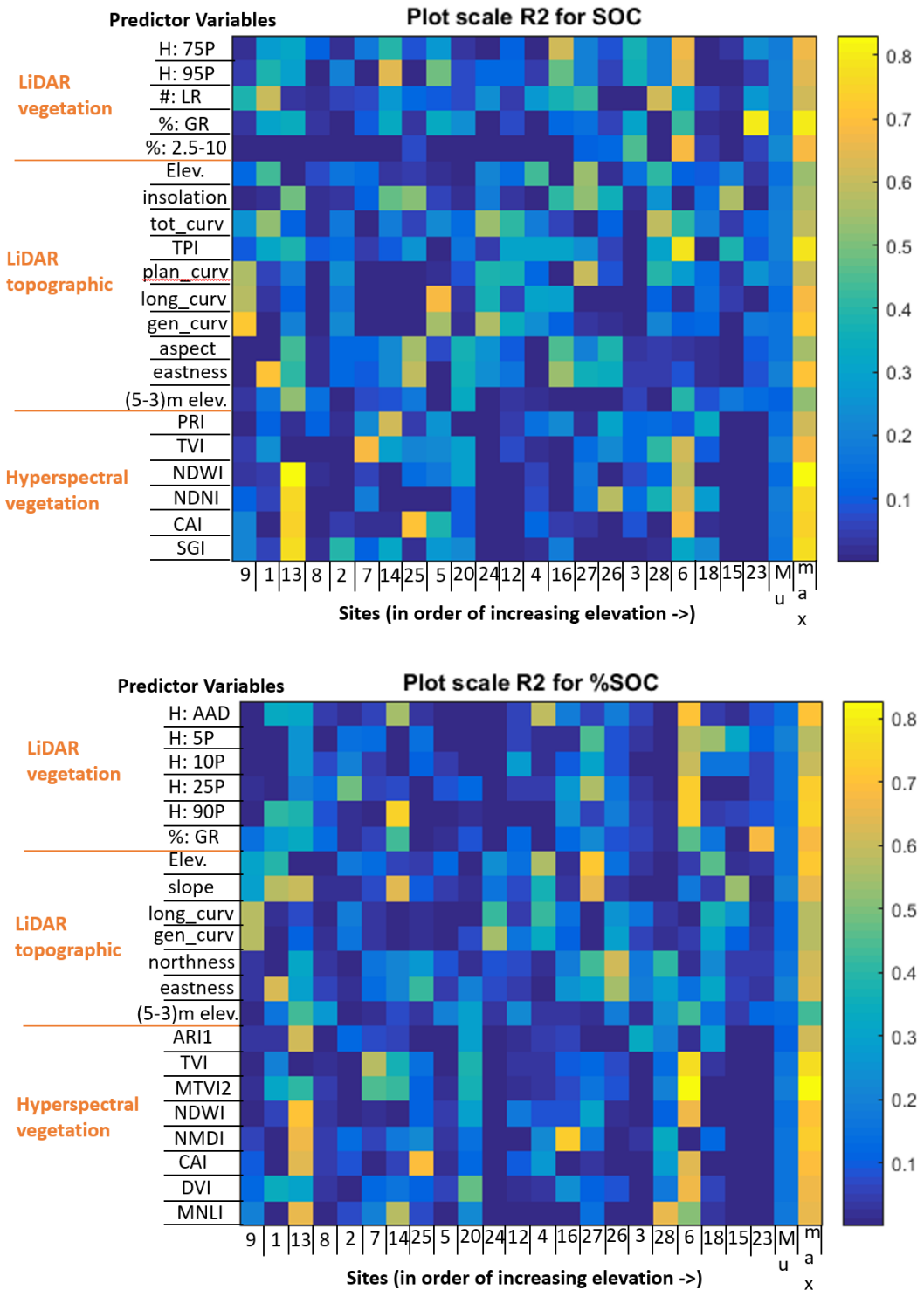

Figure 11. 1:1 relationships for SOC/%SOC and predictor variables at the plot scale and the mean and maximum R2 for each predictor at the plot scale.

Table 4. Error metrics and selected variables for plot scale RF and SMLR models, where 10 SOC data points were used to calibrate model using LiDAR and hyperspectral predictor variables.

	SOC			%SOC		
site	R2	RMSE	Variables	R2	RMSE	Variables
1	0.70	0.51	eastness	0.63	0.20	eastness
2	N/A		N/A	0.49	0.08	H: 25P
3	0.45	1.84	%: 2.5-10	N/A		N/A
4	0.43	0.26	elev	0.59	0.15	H: AAD
5	0.68	0.21	long_curv	N/A		N/A
6	0.96	0.38	tpi, wvii	1.00	0.03	H: MAD, TPI, MTVI2, CR
7	0.69	0.24	TVI	0.57	0.14	TVI
8	N/A		N/A	N/A		N/A
9	0.73	0.47	gen_curv	0.58	0.15	long_curv
12	0.41	0.48	tot_curv	N/A		N/A
13	0.83	0.19	NDWI	0.70	0.09	NDWI
14	0.97	0.17	H: 95P, TPI, DVI	0.90	0.15	H: 90P, #: LGR
15	0.56	0.52	insolation	0.54	0.27	slope
16	0.90	0.22	H: 75P, NMDI, WVSI	0.85	0.07	gen_curv, NMDI
18	N/A		N/A	0.81	0.28	H: 5P, min_curv
20	N/A		N/A	0.48	0.21	DVI
23	0.89	0.43	H: 10P, %: GR	0.68	0.31	%: GR
24	0.56	0.50	gen_curv	0.55	0.14	gen_curv
25	0.97	0.29	H: 75P, CAI, GARI	0.92	0.10	min_curv, CAI
26	0.58	1.10	NDNI	0.61	0.46	northness
27	0.98	0.30	tan_curv, slope, plan_curv	0.97	0.23	H: range, H: 10P, elev
28	0.61	0.24	#: LR	0.65	0.39	MNLI

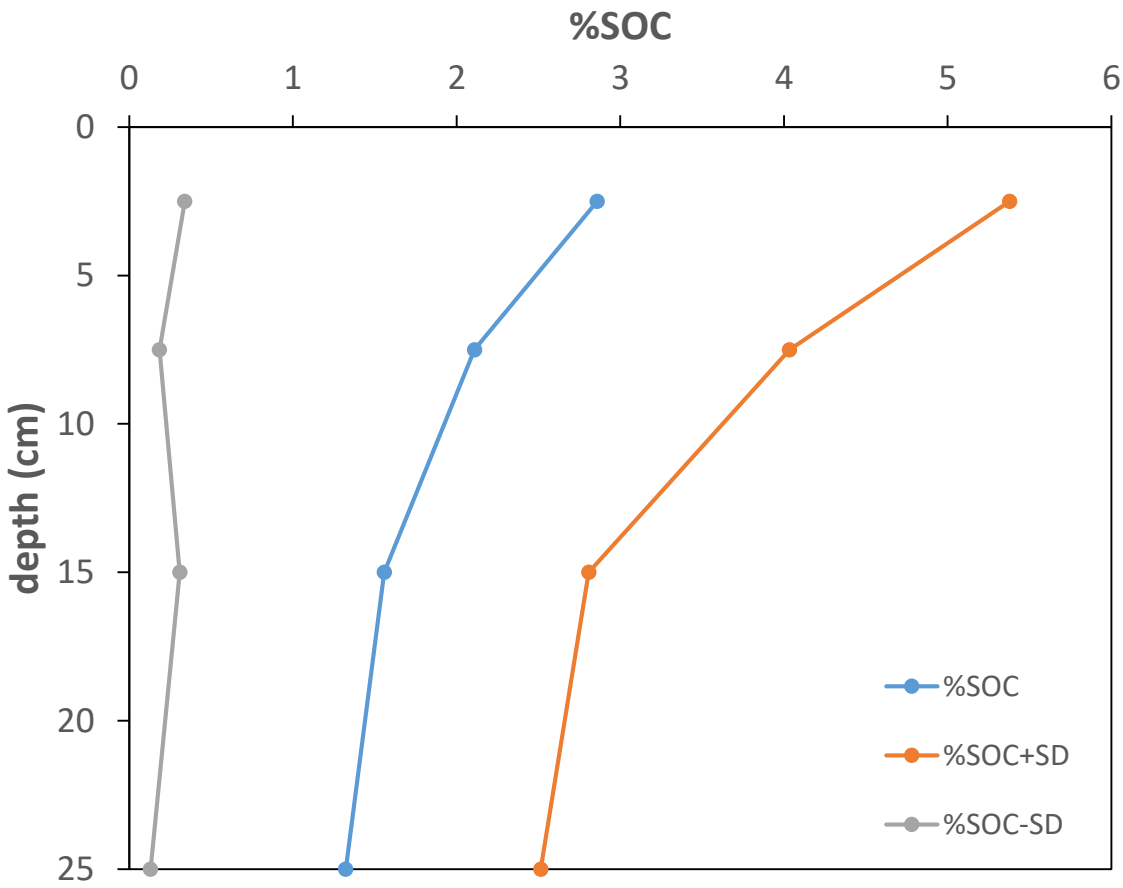

Figure 12. Mean SOC plus or minus one standard deviation as a function of depth, calculated using SOC data at all 279 cores.

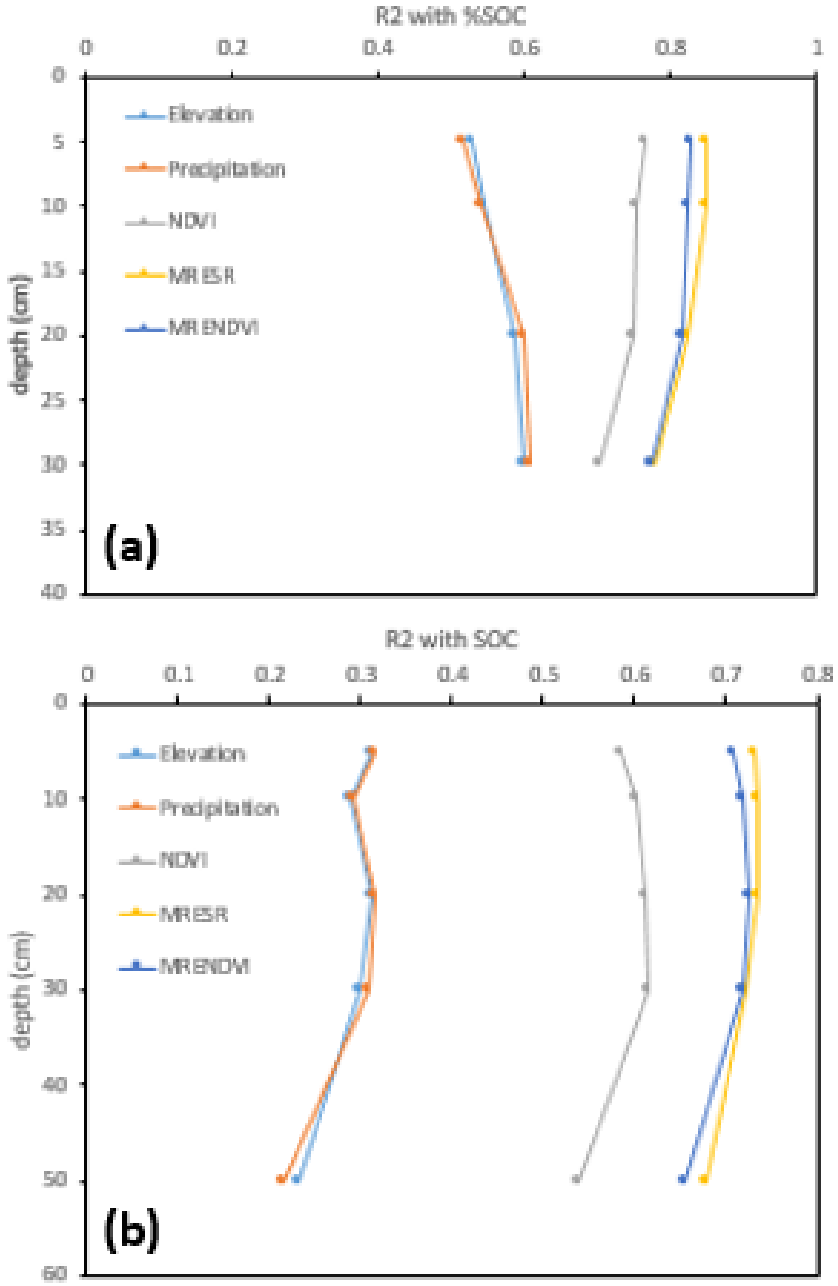

Figure 13. The predictive capability (R2) of selected predictor variables with %SOC (a) and SOC (b) varies as a function of depth.

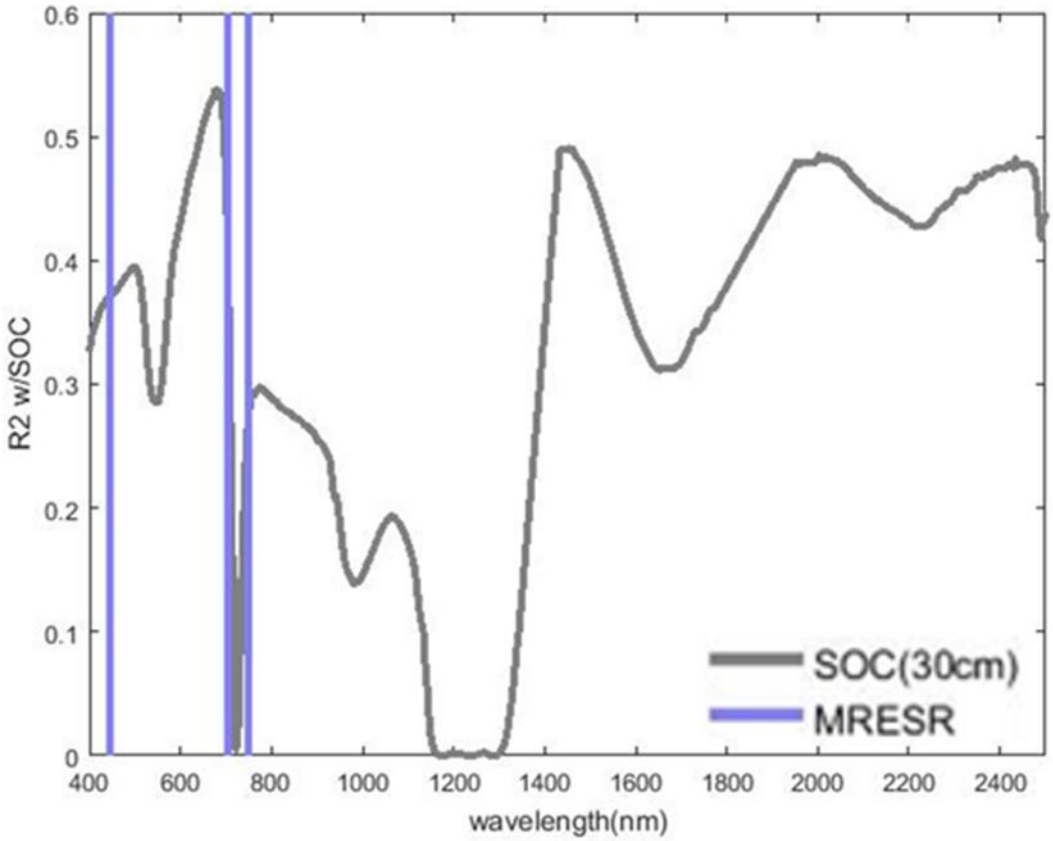

Figure 14. Relationship (R2) between SOC and reflectance at every wavelength in the spectrum, plotted with the best predictor (MRESR) in this study.

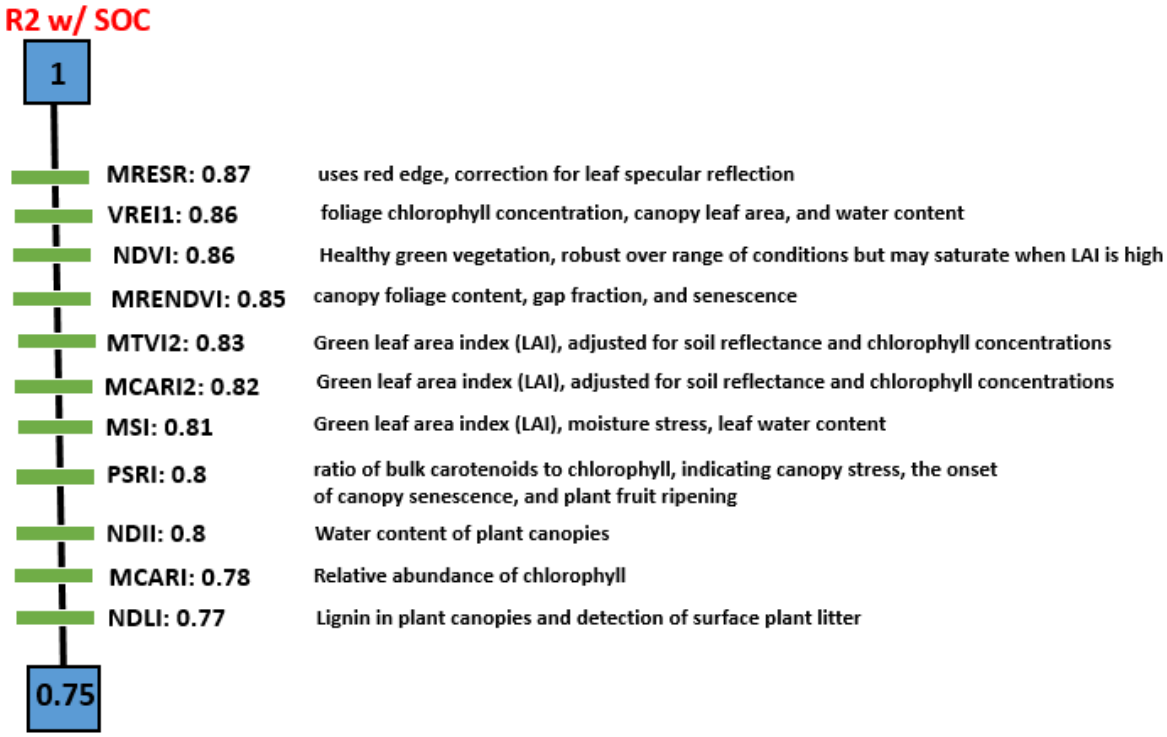

Figure 15. Some of the most predictive hyperspectral vegetation indices and what physical characteristics of vegetation they were designed to measure

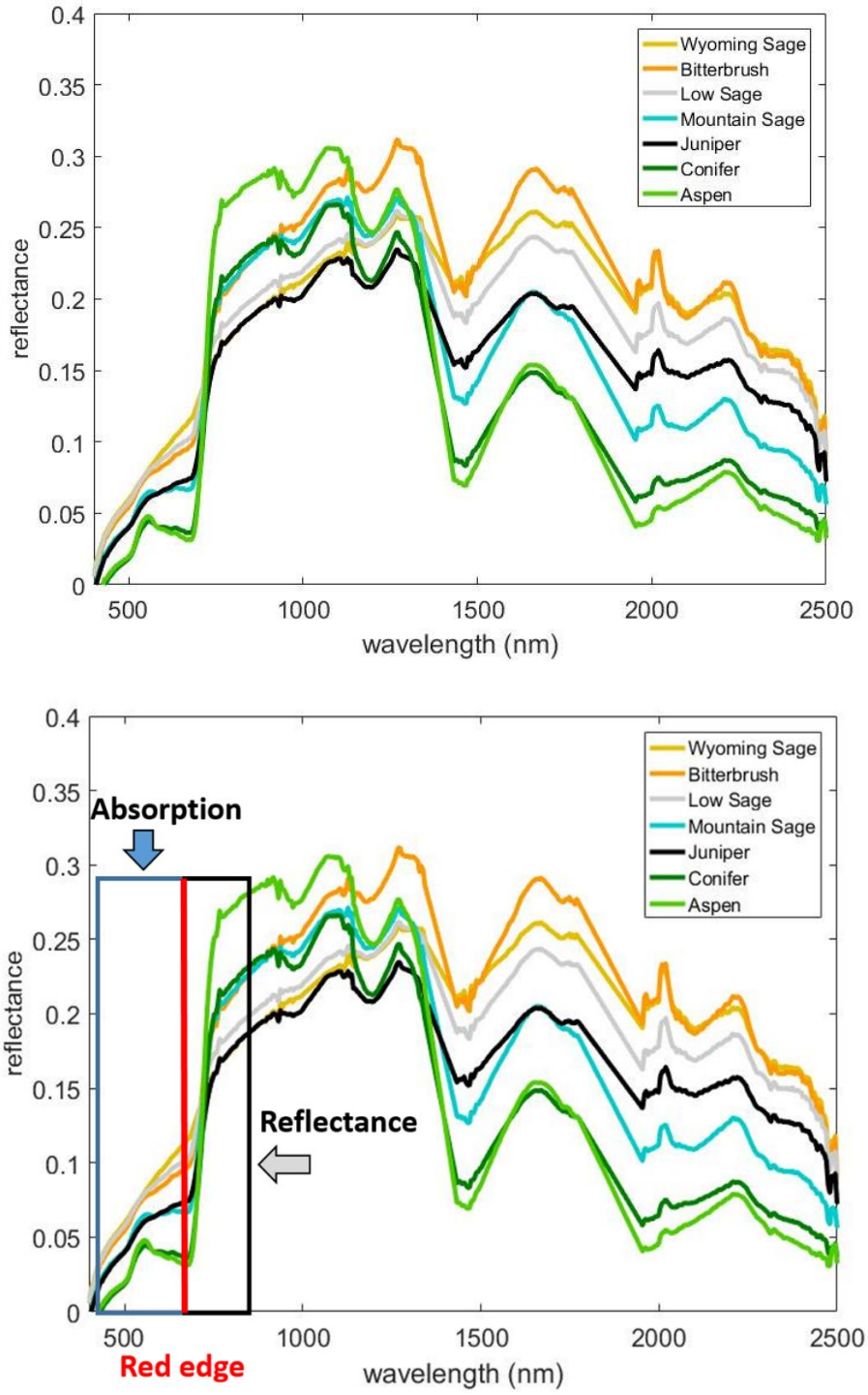

Figure 16. Spectral information for each site in this study grouped by dominant vegetation type. The plot below shows the location of the red edge on the electromagnetic spectrum, as well as the areas where the EM spectrum is heavily influenced by vegetation absorption and reflectance.

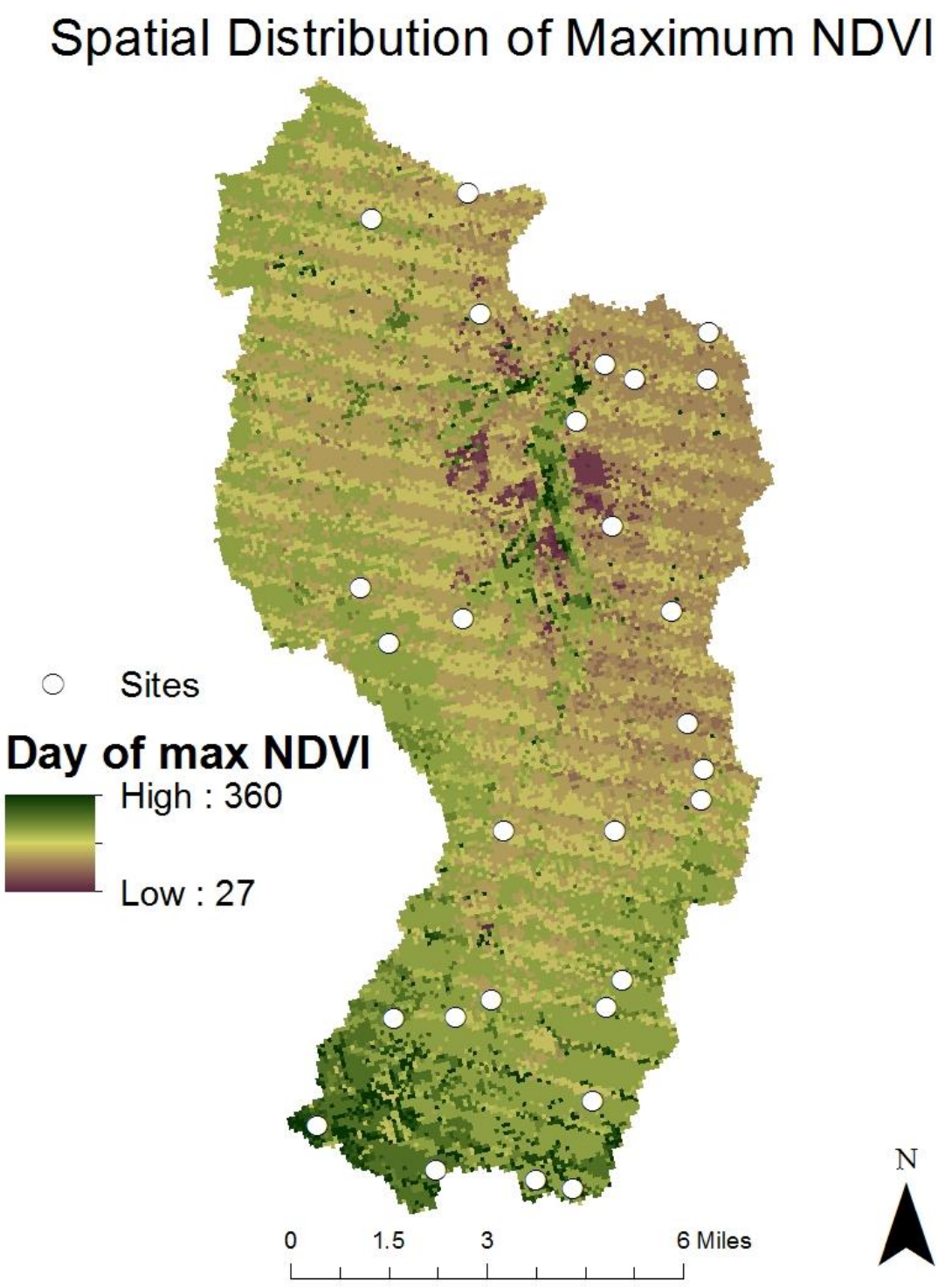

Figure 17. The timing of maximum NDVI varies in space.

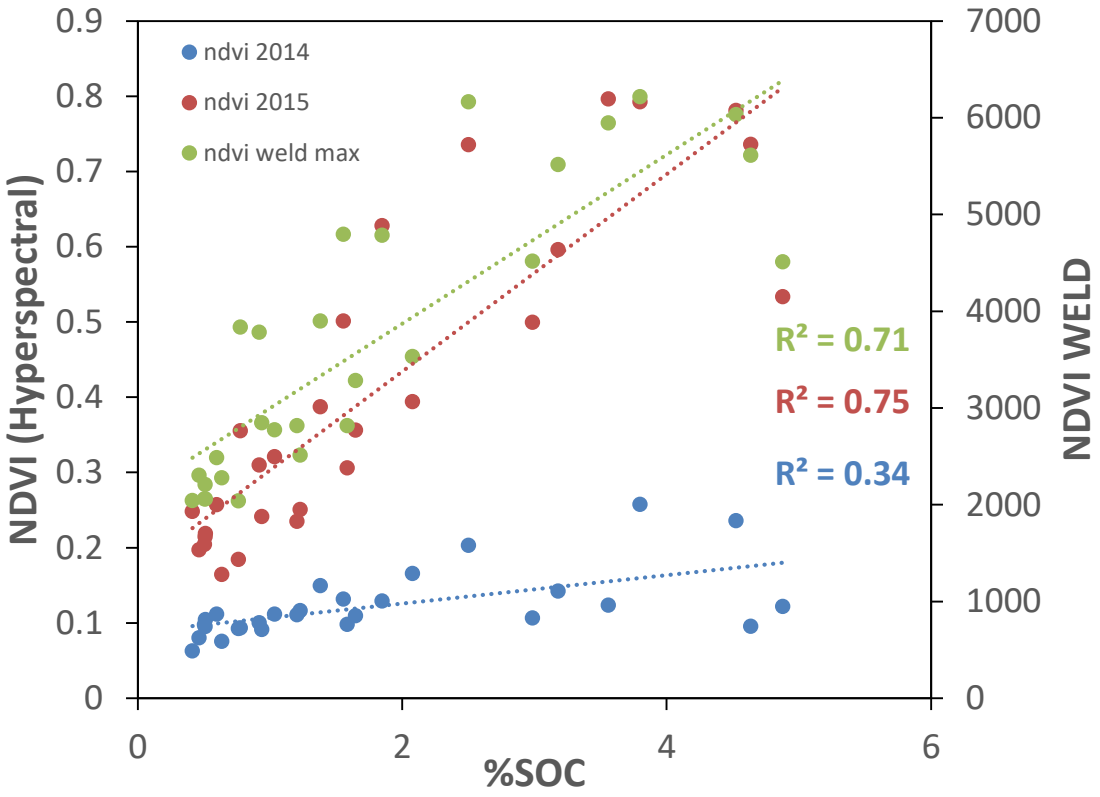

Figure 18. The predictive capability of NDVI datasets for predicting SOC varies depending on timing of data collection. As shown by 3 different NDVI datasets, Hyperspectral NDVI 2014 (collected in September), Hyperspectral NDVI 2015 (collected in June) and NDVI WELD (Web-enabled LANDSAT) (max NDVI of entire year).

Table 5. R2 varies for SOC and predictors, depending on whether data is averaged over the plot or as individual points

	SOC		**%SOC**	
	plot	cores	plot	cores
Elev.	0.3	0.23	0.6	0.5
PPT	0.31	0.24	0.61	0.52
NDVI	0.61	0.44	0.71	0.59
MRESR	0.71	0.42	0.87	0.64
MRENDVI	0.7	0.45	0.85	0.65

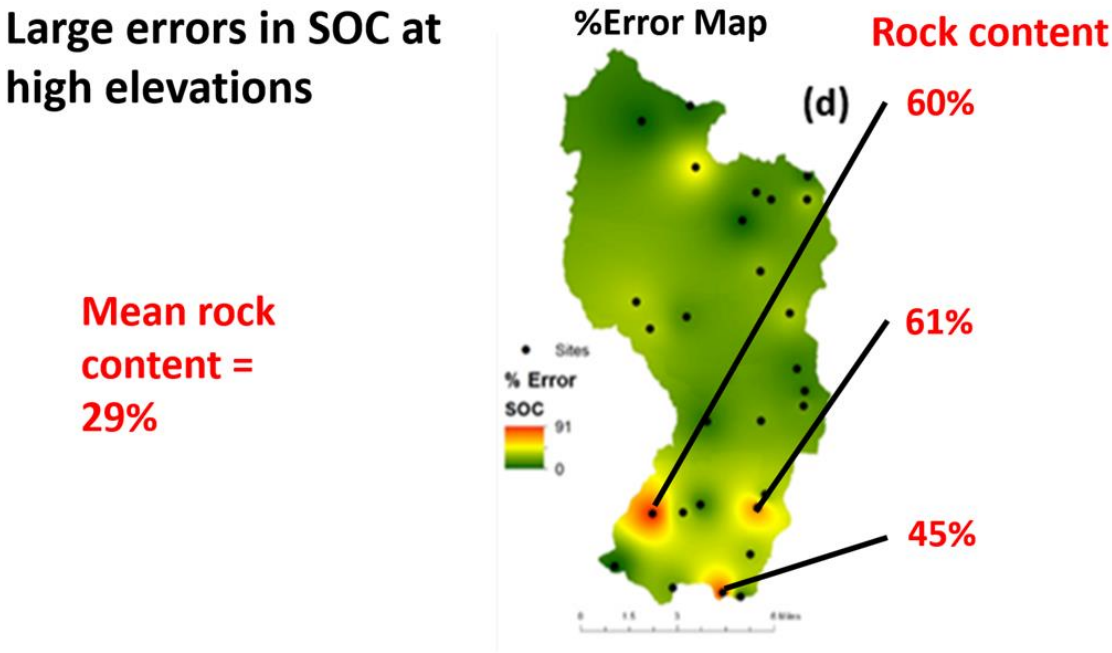

Figure 19. Rock content, also known as CF, has considerable influence on SOC storage and is not well represented by hyperspectral vegetation indices, leading to error in even the most accurate SOC models from this study

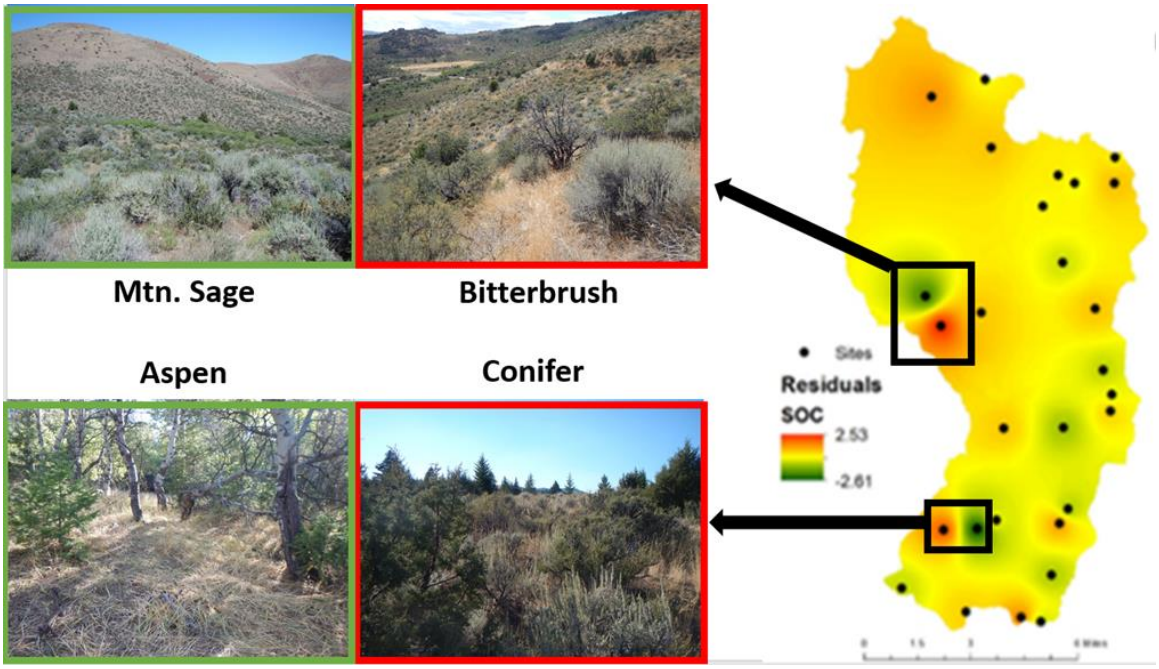

Figure 20. SOC models based on hyperspectral vegetation indices have the tendency to over/under estimate SOC in certain types of ecosystems

Table 6. SOC contents and SOC-predictor variable relationships vary depending on the scale and ecosystem of interest. Table shows R2 values for 1:1 fit between percent SOC and the predictor variables, precipitation (PPT), NDVI, MRESR, Insolation and coarse fraction (CF).

		PPT	NDVI	MRESR	Insolation	TPI	CF
Low	SOC	0.66	0.83	0.84	0.59	0	0.06
Low	%SOC	0.82	0.75	0.89	0.73	0.02	0.19
Mid	SOC	0.39	0.72	0.89	0.17	0.25	0.3
Mid	%SOC	0.03	0.27	0.51	0.17	0.03	0.03
High	SOC	0	0.24	0.5	0.55	0.61	0.91
High	%SOC	0.04	0.31	0.47	0.59	0.29	0.4
ALL	SOC	0.31	0.61	0.71	0.06	0.02	0.04
ALL	%SOC	0.61	0.71	0.87	0.02	0.06	0.06

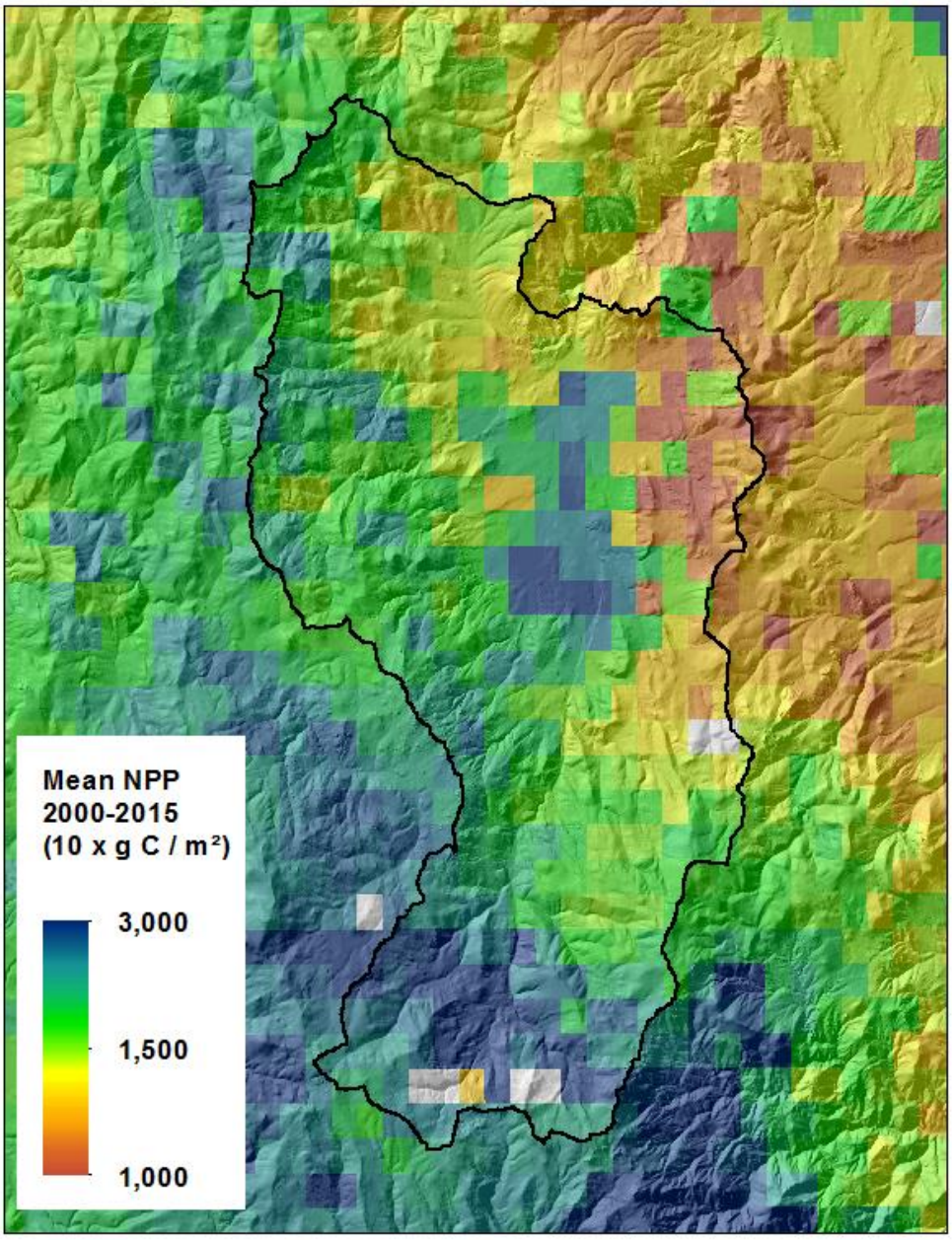

Figure 21. Spatial distribution of Net Primary Productivity (NPP) in RCEW (Credit: Poulos)

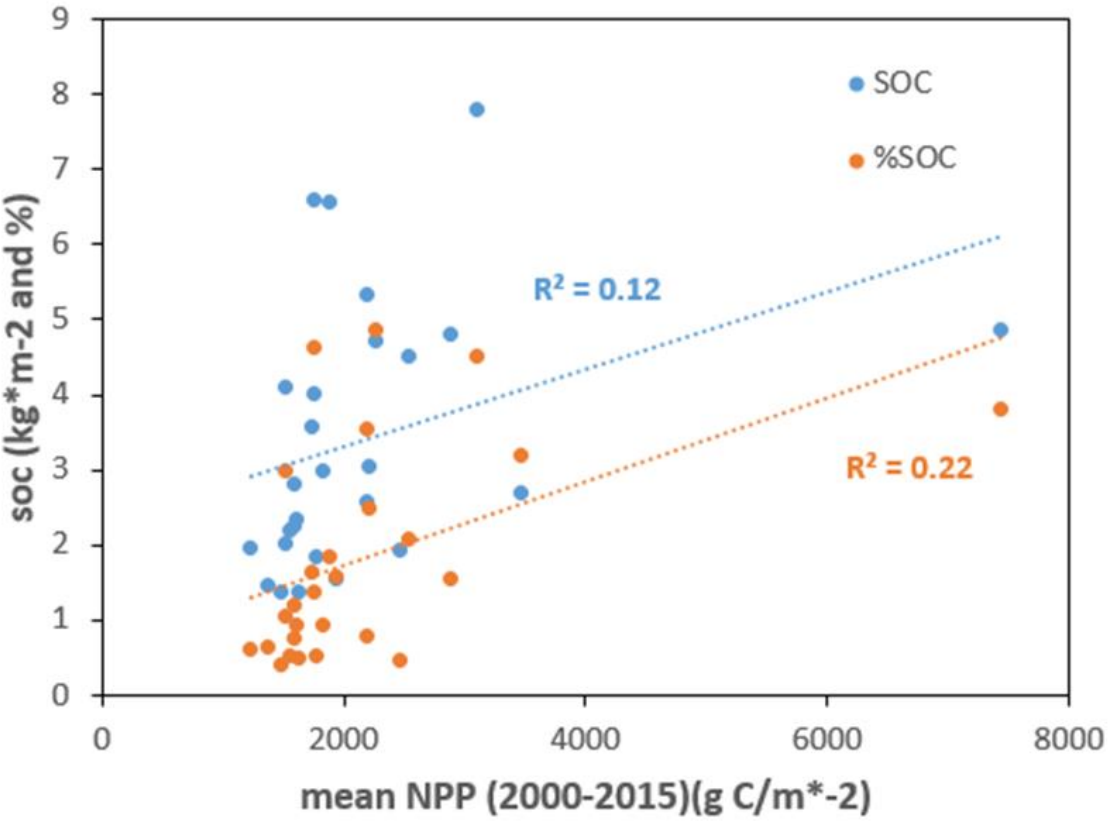

Figure 22. Relationship between SOC and NPP (data from Figure 20)

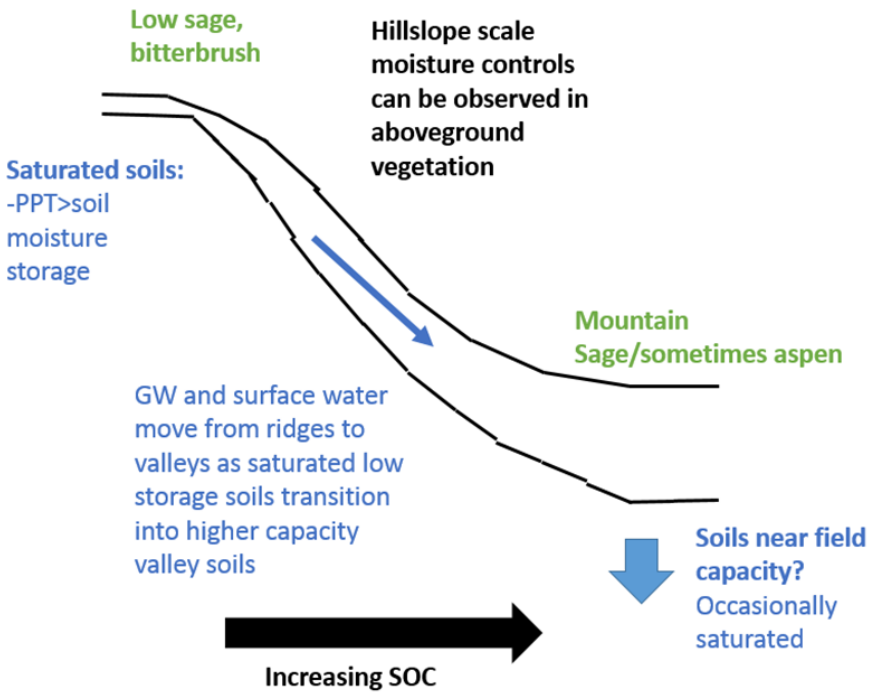

Figure 23. Hillslope hydrology influences on vegetation and SOC (conceptual diagram)

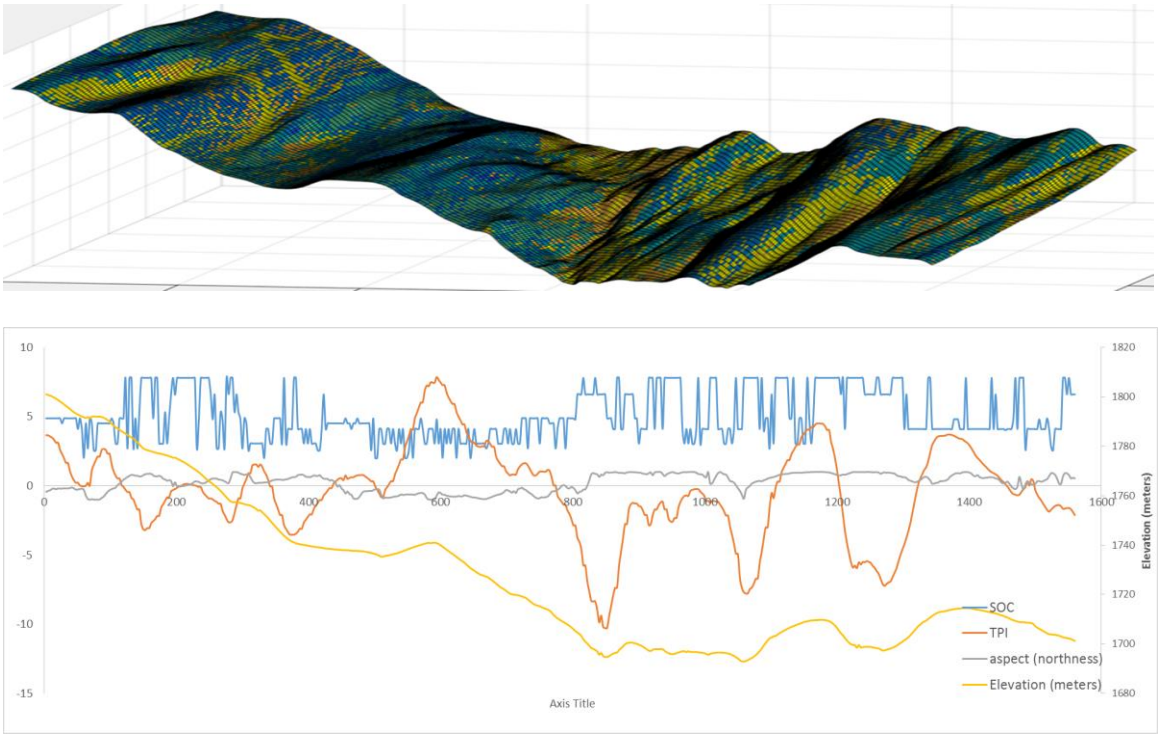

Figure 24. Transect of modeled SOC concentrations plotted with aspect, elevation and topographic position (TPI) (left side of x-axis= site 28, right side of x-axis= site 3)

Table 7. SOC, spectral vegetation and ecosystem characteristics

	SOC	%SOC	NDVI	MRESR	Vegetation/ ecosystem characteristics
Wyoming Sage	1.96	0.67	2326	1.29	Dominant below 4,500 ft.
Bitterbrush	3.45	1.08	4138	1.75	Found at all elevations, often with other species, most common on coarse-grained grantic soils
Low Sage	2.42	1.37	2846	1.52	Dominant above 4,500 ft., S/W facing slopes, poorly drained soils, high wind limits snow acccumulation (15 in), super saturated by mid winter
Mountain Sage	5.13	3.24	4604	2.12	Above 6,000 ft., N/E facing slopes, high snow accumulation (>5ft) which persists into late spring. Soils are well drained and commonly deep
Juniper	4.27	1.73	3716	1.71	Actively expanding range, 1400-2000 m elev., primarily granitic soils that are too dry for mtn. sage or other tree species (field ob.)
Conifer	3.99	3.26	5961	2.52	Found mostly on N/E facing slopes, 6,000 ft.: closed canopy, douglas fir above aspen. 6,500 ft: alpine fir above aspen, 7,000 ft.: dense stands of alpine fir exclude all other major species since only they can withstand cold soil temp. and short growing season
Aspen	7.21	4.58	5823	2.93	N/E facing slopes, convex valleys often below large snow drifts, very deep fine grained soils

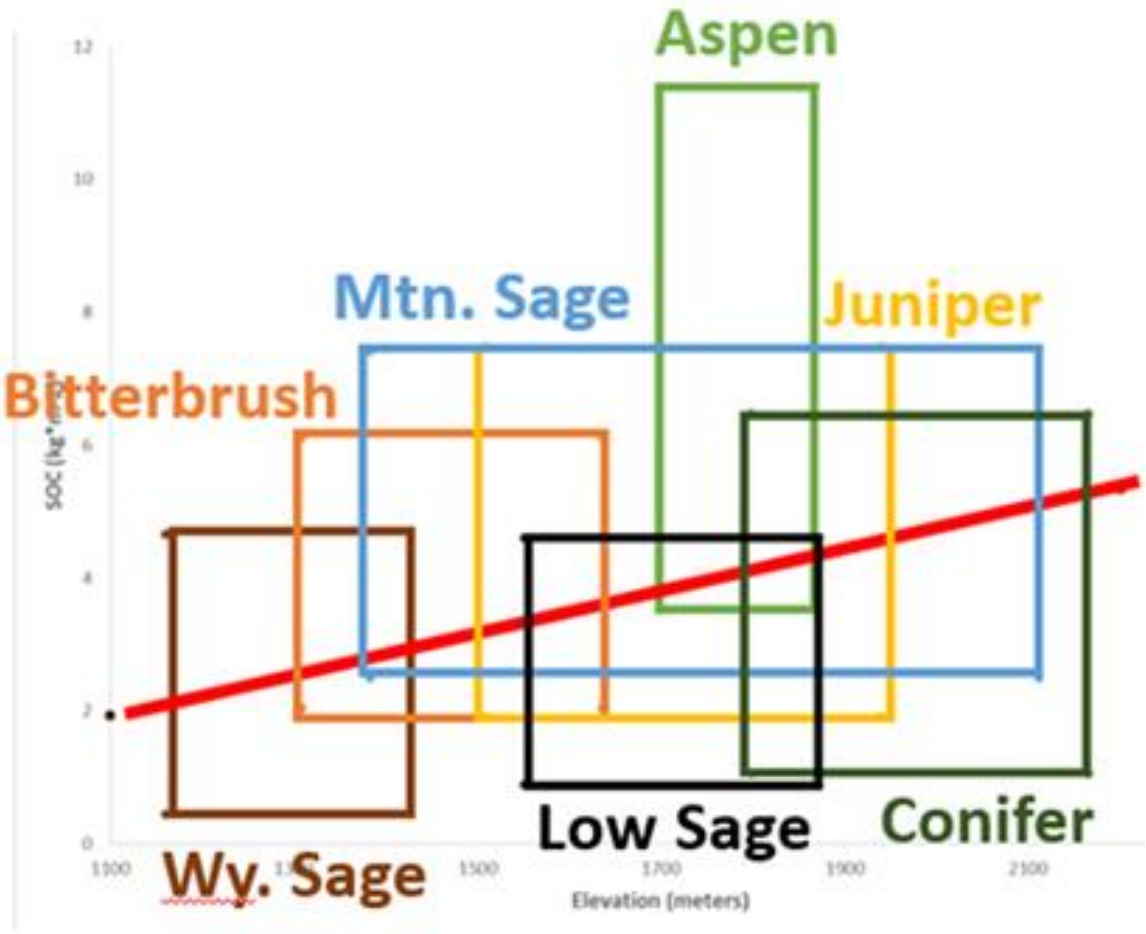

Figure 25. Distribution of vegetation species and their relation to elevation and SOC.

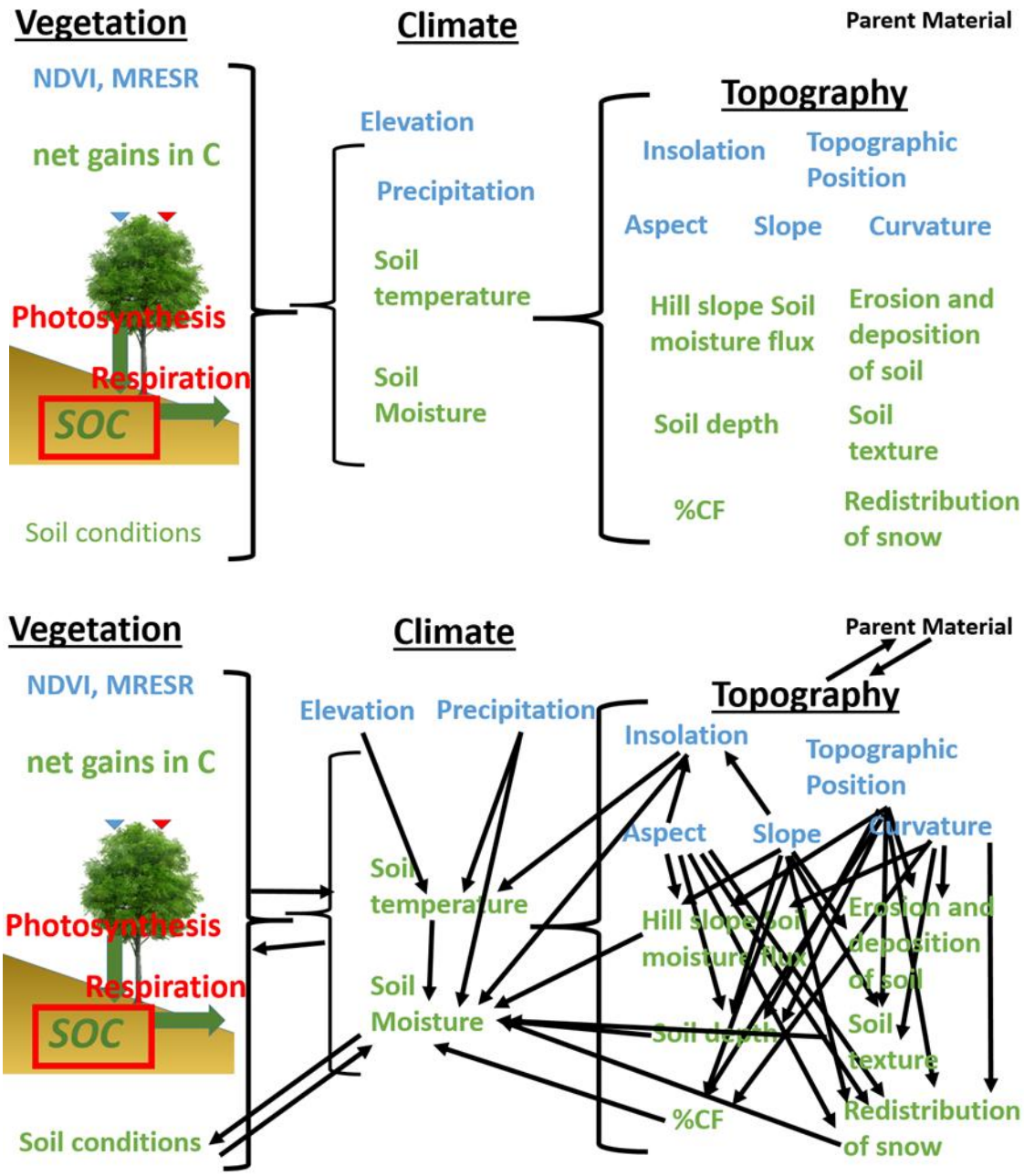

Figure 26. SOC and ecosystem carbon storage is influenced by a myriad of interrelated climatic and topographic controls

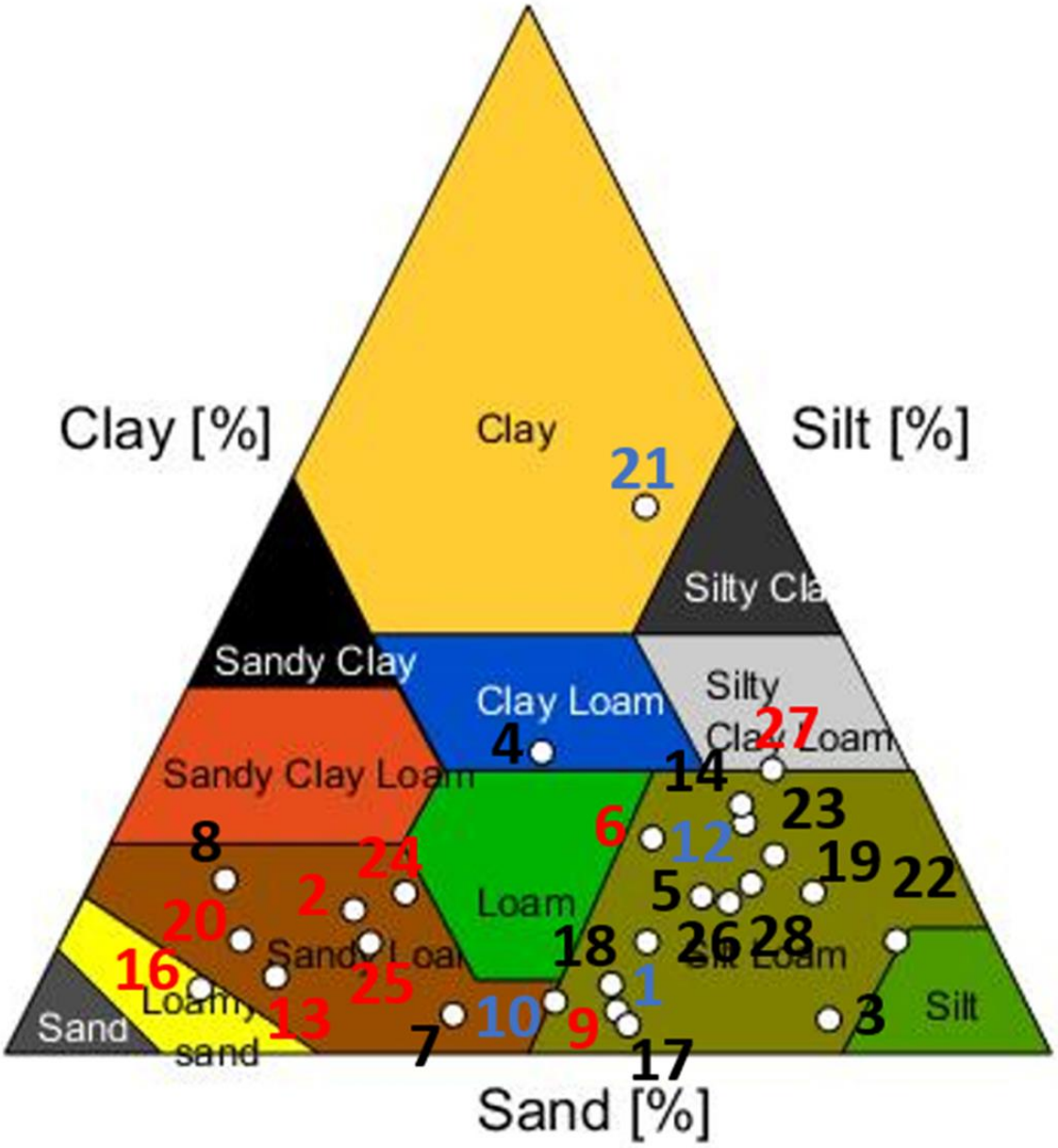

Figure 27. Soil texture data, grouped by soil parent material. Color of site number indicates soil parent material (red = felsic, black = mafic, blue = other)

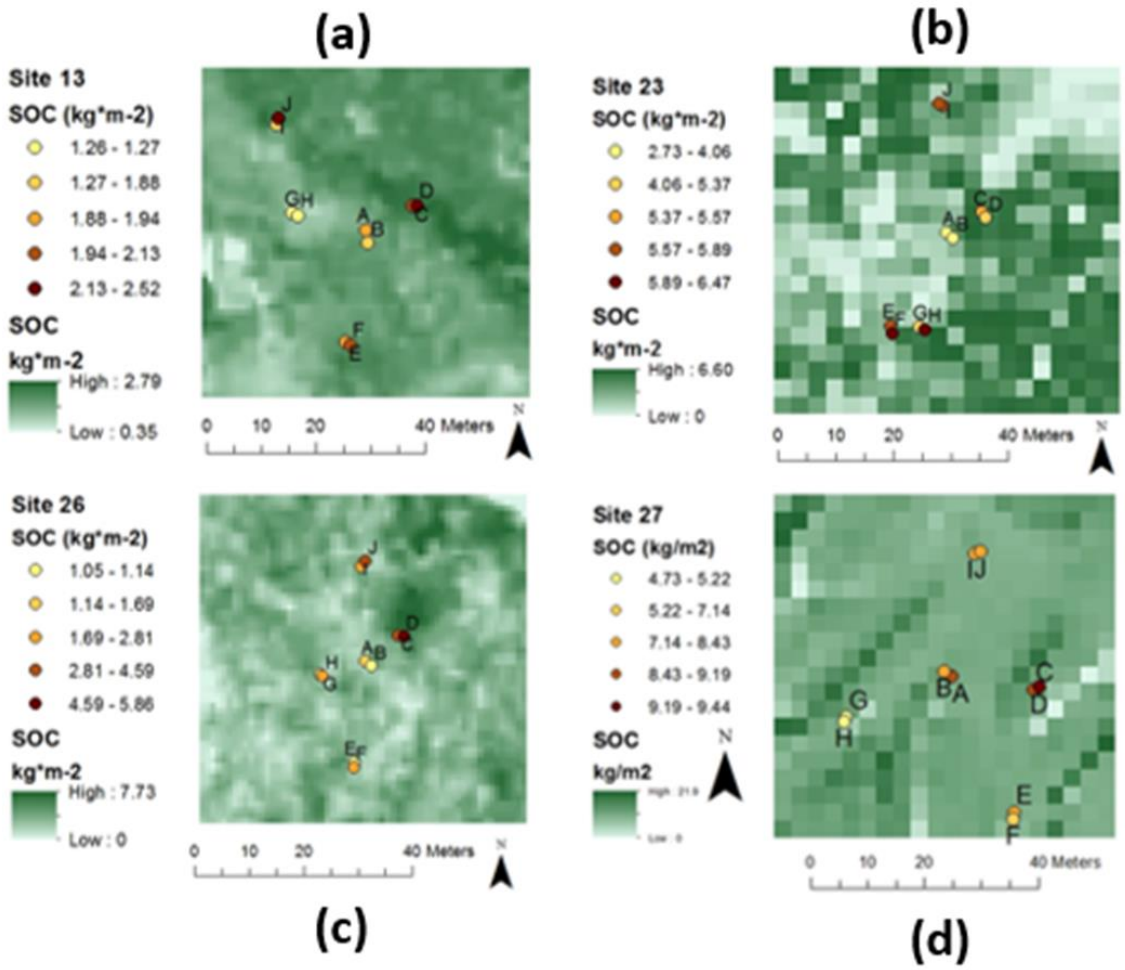

Figure 28. Plot scale SOC maps for selected sites a. Site 13. b. Site 23. c. Site 26. d. Site 27

Table 8. SOC variability at the plot scale is influenced by proximity to vegetation and the influence of proximity to vegetation varies depending on the dominant species at a site.

	30cm	**30cm**	**10cm**	**10cm**
	%increase	%SOC	%increase	%SOC
Wyoming Sage	43.47	0.18	71.21	0.38
Bitterbrush	12.12	0.11	22.32	0.23
Low Sage	11.40	0.15	21.65	0.34
Mountain Sage	5.36	0.44	5.04	0.52
Juniper	29.17	0.44	55.74	1.02
Conifer	14.34	0.25	19.23	0.32
Aspen	4.04	0.12	9.88	0.58

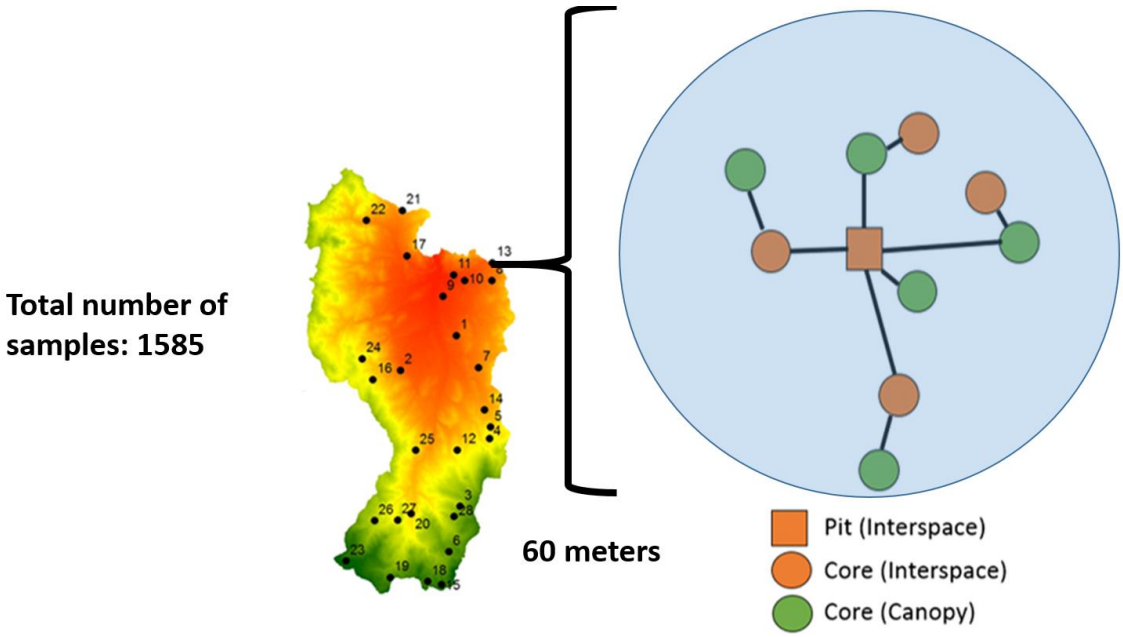

Figure A1. Field sampling design schematic

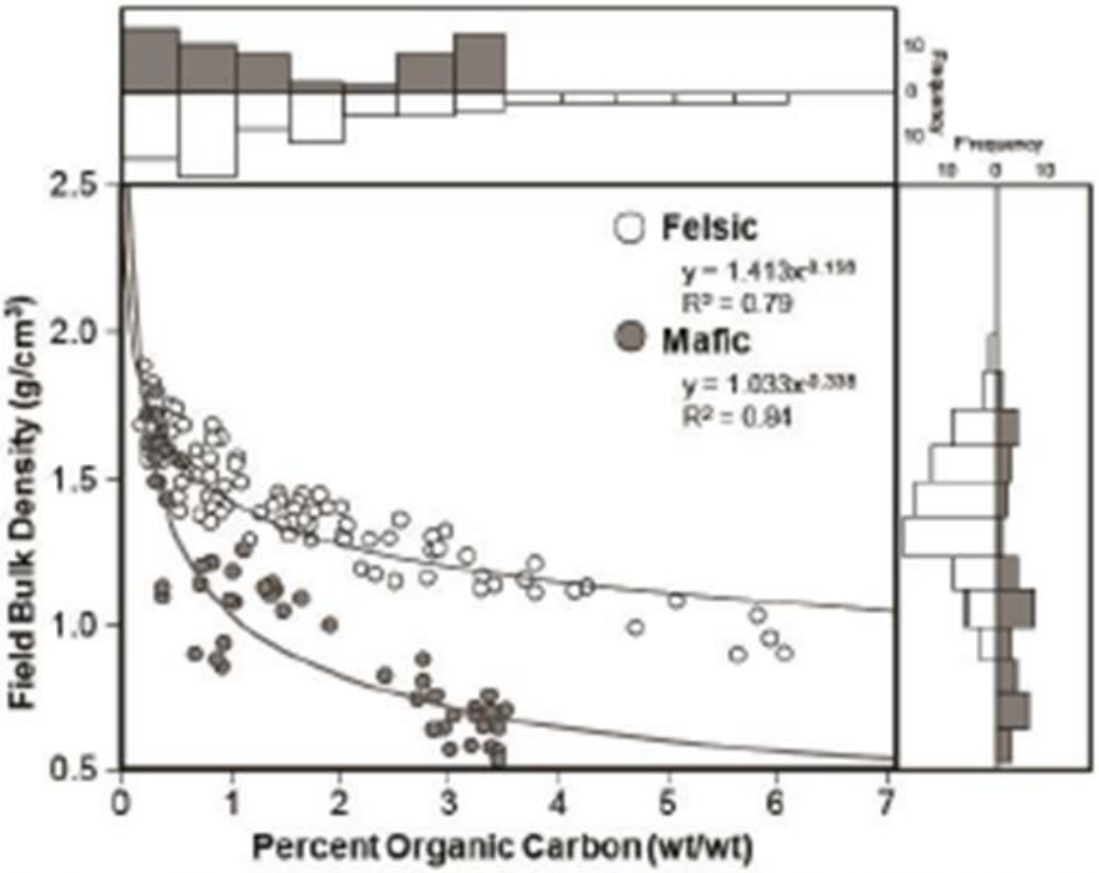

Figure 2: Soil organic carbon (SOC) to field bulk density function derived from 90 felsic samples (white dots) and 52 mafic samples (gray dots).

Figure A2. Patton 2016 soil bulk density model

Table A1a. Suite of predictor variables: Traditional

Predictor Variable	R2 SOC	R2 %SOC	Description	Ref.
NDVI	0.61	0.71	Normalized Difference Vegetation Index. 640-670nm (red), 850-880nm (NIR)	1
Annual Precip.	0.31	0.61	Annual Precipitation (mm/yr.)	2
Elevation	0.3	0.59	Land surface elevation (m)	3
Aspect (northness)	0.01	0.06	SIN(aspect) (in radians)	4
Aspect (eastness)	0.40	0.22	COS(aspect) (in radians)	4
Slope	0.02	0.07	degrees	4
Insolation	0.05	0.02	Potential insolation (yearly (by month?)	4
ARS veg	N/A	N/A	Vegetation species	5
ARS soils	N/A	N/A	Non-NRCS soil classification	5
ARS geology	N/A	N/A	Parent material	5

Note 1: source: WELD NDVI
Note 2: source: PRISM precipitation model
Note 3: USGS DEM
Note 4: calculated from USGS DEM using spatial analyst tools in ArcMap
Note 5: source: ARS soils, geology, veg

Table A1b. Suite of predictor variables: LiDAR topographic

Predictor	R2 SOC	R2 %SOC	Description	Ref.
Elevation	0.30	0.60		6
C_total	0.02	0.00	C= curvature	7
C_tangential	0.01	0.05		7
C_profile	0.00	0.00		7
C_plan	0.00	0.01		7
C_minimum	0.07	0.01		7
C_maximal	0.03	0.00		7
C_longitudinal	0.00	0.01		7
C_general	0.00	0.03		7
C_flowline	0.00	0.02		7
C_cross-sectional	0.01	0.05		7
Topographic Position index	0.02	0.06		7
slope	0.02	0.00		7
Aspect (northness)	0.22	0.25		7
Aspect (eastness)	0.16	0.05		7
Insolation	0.06	0.02		8

Note 6: source= BCAL LiDAR DEM
Note 7: calculated using BCAL LiDAR DEM in SAGA GIS Terrain Analysis Module
Note 8: calculated using BCAL LiDAR DEM in ArcMap using solar radiation tool

Table A1c. Suite of predictor variables: LiDAR vegetation

Predictor	R2 SOC	R2 %SOC	Description
minimum height	0.02	0.08	H_min
maximum height	0.20	0.24	H_max
height range	0.21	0.24	H_range
mean height	0.17	0.23	H_mean
mad	0.24	0.29	
aad	0.25	0.26	
variance	0.15	0.18	
std dev	0.24	0.25	
skewness	0.08	0.30	
kurtosis	0.16	0.28	
interquartile range	0.26	0.28	
coefficient of variation	0.08	0.00	
0.05	0.03	0.09	H_5th percentile
0.1	0.04	0.11	
0.25	0.06	0.13	
0.5	0.15	0.22	
0.75	0.21	0.25	
0.9	0.21	0.24	
0.95	0.21	0.24	
lidar returns	0.05	0.00	
lidar vegetation returns	0.01	0.08	
lidar ground returns	0.10	0.10	
total vegetation density	0.03	0.15	
vegetation cover	0.15	0.19	
% of ground returns	0.04	0.15	
% 0-1m	0.32	0.32	% of veg returns between 0 and 1 m
%1-2.5m	0.30	0.25	
%2.5-10m	0.30	0.30	
%10-20m	0.06	0.10	
%20-30m	0.03	0.08	
%>30m	N/A	N/A	
crr	0.05	0.26	
texture of heights	0.04	0.02	

Note: calculated using BCAL LiDAR tools and .las files from BCAL

Table A1d. Suite of predictor variables: Hyperspectral vegetation

Type	Predictor	R2 SOC	R2 %SOC	Description	Wavelengths used
BBG	ARVI	0.62	0.84	Atmospherically Resistant Vegetation Index	NIR, Red, Blue
BBG	DVI	0.59	0.70	Difference Vegetation Index	NIR, Red
BBG	EVI	0.64	0.78	Enhanced Vegetation Index	NIR, Red, Blue
BBG	GARI	0.63	0.78	Green Atmospherically Resistant Index	NIR, red, blue, green
BBG	GDVI	0.51	0.59	Green Difference Vegetation Index	NIR, green
BBG	GNDVI	0.55	0.75	Green Normalized Difference Vegetation Index	NIR, green
BBG	GRVI	0.45	0.69	Green Ratio Vegetation Index	NIR, green
BBG	IPVI	0.63	0.84	Infrared Percentage Vegetation Index	NIR, Red
BBG	LAI	0.64	0.78	Leaf Area Index	NIR, Red, Blue
BBG	MNLI	0.59	0.73	Modified Non-Linear Index	NIR, Red
BBG	MSR	0.61	0.82	Modified Simple Ratio	NIR, Red
BBG	NLI	0.62	0.81	Non-Linear Index	NIR, Red
BBG	NDVI	0.64	0.86	Normalized Difference Vegetation Index	NIR, Red
BBG	OSAVI	0.64	0.82	Optimized Soil Adjusted Vegetation Index	NIR, Red
BBG	RDVI	0.64	0.80	Renormalized Difference Vegetation Index	NIR, Red
BBG	SAVI	0.63	0.77	Soil Adjusted Vegetation Index	NIR, Red
BBG	SR	0.58	0.79	Simple Ratio	NIR, Red
BBG	SGI	0.35	0.51	Sum Green Index	green
BBG	TDVI	0.63	0.84	Transformed Difference Vegetation Index	NIR, Red
BBG	VARI	0.68	0.79	Visible Atmospherically Resistant Index	Green, red, blue
NB	MCARI	0.59	0.78	Modified Chlorophyll Absorption Ratio Index	550, 670, 700 nm
NB	MCARI2	0.65	0.82	Modified Chlorophyll Absorption Ratio Index Improved	550, 670, 800 nm
NB	MRENDVI	0.70	0.85	Modified Red Edge Normalized Difference Vegetation Index	445, 705, 750 nm
NB	MRESR	0.71	0.87	Modified Red Edge Simple Ratio	445, 705, 750 nm
NB	MTVI	0.64	0.78	Modified Triangular Vegetation Index	550, 670, 800 nm
NB	MTVI2	0.65	0.83	Modified Triangular Vegetation Index - Improved	550, 670, 800 nm
NB	RENDVI	0.67	0.86	Red Edge Normalized Difference Vegetation Index	705, 750 nm
NB	TCARI	0.59	0.78	Transformed Chlorophyll Absorption Reflectance Index	550, 670, 700 nm
NB	TVI	0.64	0.79	Triangular Vegetation Index	550, 670, 750 nm
NB	VREI1	0.66	0.86	Vogelmann Red Edge Index 1	720, 740 nm

Type	Predictor	R2 SOC	R2 %SOC	Description	Wavelengths used
NB	VREI2	0.63	0.83	Vogelmann Red Edge Index 2	715, 726, 734, 747 nm
LUE	PRI	0.30	0.23	Photochemical Reflectance Index	531, 570 nm
LUE	SIPI	0.57	0.68	Structure Insensitive Pigment Index	445, 680, 800 nm
LUE	RGR	0.62	0.73	Red Green Ratio Index	600-699 (red), 500-599(green) nm
CN	NDNI	0.59	0.61	Normalized Difference Nitrogen Index	1510, 1680 nm
DSC	NDLI	0.61	0.77	Normalized Difference Lignin Index	1680, 1754 nm
DSC	CAI	0.30	0.49	Cellulose Absorption Index	2000, 2100, 2200 nm
DSC	PSRI	0.64	0.80	Plant Senescence Reflectance Index	500, 680, 750 nm
LP	ARI1	0.09	0.21	Anthocyanin Reflectance Index 1	550, 700 nm
LP	ARI2	0.31	0.54	Anthocyanin Reflectance Index 2	550, 700, 800 nm
LP	CRI1	0.18	0.33	Carotenoid Reflectance Index 1	510, 550 nm
LP	CRI2	0.17	0.32	Carotenoid Reflectance Index 2	510, 700 nm
CWC	MSI	0.55	0.81	Moisture Stress Index	819, 1599 nm
CWC	NDII	0.54	0.80	Normalized Difference Infrared Index	819, 1649 nm
CWC	NDWI	0.51	0.74	Normalized Difference Water Index	857, 1241 nm
CWC	NMDI	0.02	0.06	Normalized Multi-band Drought Index	860, 1640, 2130 nm
CWC	WBI	0.51	0.74	Water Band Index	900, 970 nm

Note: source data is AVIRIS hyperspectral 2015 data from BCAL
Note 2: Vegetation indices from Harris Geospatial were calculated using this data in MATLAB and ENVI Band Math
BBG: Broadband Greenness
NB: Narrowband Greenness
LUE: Light Use Efficiency
CN: Canopy Nitrogen
DSC: Dry or Senescent Carbon
LP: Leaf Pigments
CWC: Canopy Water Content

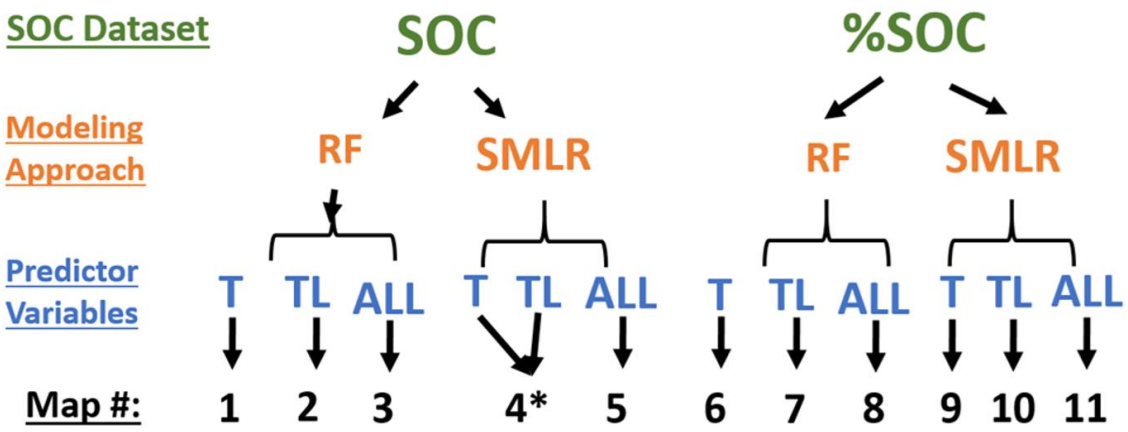

Figure A3. Schematic of SOC modeling iterations

Table A2a. Watershed SOC Summary

	SOC (kg/m2)			%SOC		%CF	BD (kg/m3)
Site	***10 cm***	***30 cm***	***50 cm***	***10 cm***	***30 cm***	***30 cm***	***30 cm***
1	1.04	2.25	2.62	1.08	0.76	24.3	1211
2	1.29	3.00	3.74	1.19	0.92	26.1	1194
3	3.00	6.61	10.45	6.45	4.63	18.7	543
4	0.94	2.02	2.78	1.27	1.03	36.3	935
5	1.10	2.53	2.53	1.62	1.23	27.5	929
6	1.82	4.51	5.87	3.01	2.08	42.2	1001
7	0.62	1.36	1.86	0.67	0.50	28.5	1576
8	0.61	1.37	1.92	0.59	0.41	20.6	1594
9	0.80	1.92	2.37	0.63	0.46	17.2	1641
10	0.96	2.20	3.14	0.71	0.51	7.1	1185
11	0.86	1.83	2.10	0.69	0.51	16.4	828
12	1.69	3.58	4.77	2.29	1.65	43.9	1148
13	0.82	1.94	2.60	0.94	0.60	9.2	1416
14	1.33	2.81	3.97	1.45	1.20	17.9	925
15	2.01	4.72	6.30	5.44	4.88	47.4	790
16	1.05	2.56	2.56	1.10	0.77	23.8	1258
17	0.66	1.45	2.08	0.81	0.63	34.7	1035
18	1.37	3.06	4.02	3.91	2.50	45.5	1049
19	2.03	4.87	7.51	5.76	3.80	33.9	503
20	1.81	4.80	5.36	2.52	1.56	23.5	1235
21	1.04	2.35	2.35	1.18	0.94	40.6	941
22	1.95	4.11	5.72	4.12	2.98	33.2	759
23	2.51	5.32	5.32	4.17	3.56	27.6	550
24	2.82	6.58	8.77	2.65	1.85	7.9	1081
25	1.64	4.02	5.53	1.89	1.38	31.7	992
26	1.15	2.69	2.86	3.88	3.18	59.5	580
27	3.55	7.80	11.41	7.25	4.53	9.6	472
28	0.71	1.55	1.62	2.02	1.58	60.8	919

Table A2b. Watershed SOC site info

Site	*Vegetation Species*	*Parent Material*	*C:N*	*%N*	*Ph*	*Soil depth*	*Texture*
1	Wyoming sage	Alluvium	12.5	0.08	7.4	90 cm	Silt loam
2	Bitterbrush	Felsic	10.7	0.09	6.6	95 cm	Sandy loam
3	Aspen	Mafic	12.2	0.36	6.0	> 50 cm	Silt loam
4	Low Sage	Mafic	9.2	0.11	7.5	90 cm	Clay loam
5	Low Sage	Mafic	9.1	0.13	6.9	30 cm	Silt loam
6	Juniper	Felsic	11.8	0.17	6.1	75 cm	Silt loam
7	Wyoming sage	Mafic	7.5	0.07	7.8	100 cm	Sandy Loam
8	Wyoming sage	Mafic	6	0.07	7.9	60 cm	Sandy Loam
9	Wyoming sage	Felsic	7.5	0.06	7.7	80 cm	Silt loam
10	Wyoming sage	Felsic	7.6	0.07	8.4	> 1m	Silt loam
11	Greasewood	Alluvium	8.1	0.07	9.7	> 1m	N/A
12	Low Sage	Welded tuff	9	0.18	6.6	1 m	Silt loam
13	Wyoming sage	Mafic	6.8	0.09	8.1	1 m	Sandy loam
14	Wyoming sage	Mafic	8.5	0.14	7.5	1 m	Silt loam
15	Mountain Sage	Mafic	11.6	0.42	n/a	>50 cm	N/A
16	Bitterbrush	Felsic	7.6	0.1	7.0	30 cm	Loamy sand
17	Wyoming sage	Mafic	7.7	0.1	7.2	>50 cm	Silt loam
18	Conifer	Mafic	10.7	0.23	5.8	>50 cm	Silt loam
19	Conifer	Mafic	13.4	0.27	5.7	>50 cm	Silt loam
20	Bitterbrush	Felsic	9.6	0.15	6.7	40 cm	Sandy loam
21	Wyoming sage	Felsic	8.5	0.12	7.0	30 cm	Clay
22	Mountain Sage	Mafic	10.3	0.29	6.5	> 50 cm	Silt loam
23	Conifer	Mafic	10.5	0.34	n/a	40 cm	Silt loam
24	Mountain Sage	Felsic	10.2	0.18	6.5	>50 cm	Sandy loam
25	Juniper	Felsic	8.4	0.16	6.7	>50 cm	Sandy loam
26	Conifer	Mafic	10.3	0.31	6.1	40 cm	Silt loam
27	Aspen	Felsic	12.1	0.37	6.5	>50 cm	Silty clay loam
28	Low Sage	Mafic	9.7	0.16	6.9	40 cm	Silt loam

Table A2c. Values for relevant predictors

	Elev. (m)	PPT	NDVI	MRESR	Eastness	Northness	aspect	slope
1	1238	274	2041	1.27	-0.84	-0.18	W/SW	4.8
2	1310	360	3782	1.57	0.65	-0.16	E/SE	7.3
3	1866	518	5614	2.89	-0.27	0.92	N	15.5
4	1623	388	2774	1.46	-0.90	-0.43	W/SW	10.8
5	1556	363	2513	1.39	1.00	0.00	E	15.7
6	1938	716	3532	1.62	-0.60	-0.64	S/SW	8.1
7	1340	276	2060	1.25	-0.96	-0.19	W/SW	5.7
8	1292	263	2043	1.27	-0.94	-0.31	W/SW	5.8
9	1160	252	2304	1.24	-0.72	0.64	NW/N	4.6
10	1191	259	2061	1.26	-0.78	0.14	E/NE	9.6
11	1152	262	2209	1.35	-0.04	-0.96	S	8.0
12	1619	372	3283	1.71	-0.11	0.99	N	4.6
13	1291	263	2488	1.36	-0.73	0.56	W/NW	10.8
14	1424	314	2818	1.38	0.29	0.95	N	6.3
15	2111	1067	4511	2.11	-0.48	0.87	N	7.2
16	1629	575	3837	1.65	0.93	-0.34	E/SE	22.6
17	1218	333	2278	1.24	-0.54	-0.84	S	12.2
18	2089	1026	6165	2.64	-0.44	0.89	N	6.2
19	2038	932	6218	2.58	-0.34	0.36	NW	7.1
20	1583	564	4796	2.04	-0.05	-0.94	S	8.5
21	1370	360	2849	1.34	-0.09	-0.97	S	14.5
22	1389	476	4519	2.00	-0.05	1.00	N	22.3
23	2163	1149	5947	2.67	0.46	0.87	N	7.6
24	1588	584	4784	2.27	0.71	0.66	N	14.7
25	1512	507	3900	1.80	0.70	-0.50	SE	14.1
26	1800	879	5515	2.18	0.76	-0.49	SE	8.9
27	1701	694	6033	2.98	0.62	0.65	N	7.9
28	1870	513	2816	1.53	-0.20	-0.98	S	7.0

Table A3a. Plot scale SOC

Metric	**Standard deviations**					**Relative standard deviation**				
Variables	SOC		%SOC		%CF	SOC		%SOC		%CF
Depth	10	30	10	30	30	10	30	10	30	30
1	0.53	0.89	0.64	0.34	7.94	51	40	58	45	33
2	0.17	0.39	0.19	0.14	8.14	13	13	16	15	30
3	1.24	2.34	3.47	2.23	13.24	41	35	55	48	67
4	0.26	0.33	0.50	0.29	9.60	27	16	38	28	26
5	0.18	0.35	0.32	0.18	10.25	16	14	19	15	35
6	1.21	1.59	1.76	1.03	12.43	66	35	57	50	29
7	0.26	0.41	0.44	0.26	8.82	41	30	66	52	31
8	0.16	0.18	0.23	0.13	10.51	26	13	38	31	55
9	0.47	0.85	0.29	0.23	17.65	58	44	46	50	96
10	0.30	0.70	0.34	0.48	5.00	31	32	34	93	68
11	0.30	0.66	0.31	0.24	11.02	35	36	45	48	73
12	0.30	0.59	0.45	0.26	13.81	18	17	18	16	32
13	0.26	0.43	0.38	0.20	7.01	32	22	36	33	78
14	0.42	0.75	0.79	0.47	9.70	32	27	54	39	51
15	0.37	0.74	0.91	0.73	7.34	18	16	15	15	15
16	0.38	0.56	0.41	0.23	12.01	36	22	35	30	47
17	0.21	0.29	0.26	0.18	13.72	32	20	32	29	40
18	0.35	0.43	1.24	0.64	5.70	25	14	32	25	13
19	0.60	0.98	2.45	1.16	5.41	29	20	43	31	16
20	0.68	0.74	0.77	0.38	6.13	38	15	29	25	25
21	0.42	1.01	0.51	0.30	27.33	40	43	41	32	67
22	0.56	0.99	0.95	0.95	21.79	29	24	23	32	66
23	0.46	1.13	0.88	0.60	12.93	18	21	21	17	46
24	0.33	0.71	0.37	0.26	4.63	12	11	16	14	57
25	0.61	1.28	0.58	0.39	14.86	37	32	36	28	51
26	0.53	1.59	1.04	0.70	24.63	46	59	26	22	41
27	0.91	1.63	0.30	0.29	8.28	26	21	30	06	126
28	0.11	0.36	1.20	0.65	15.94	16	23	60	41	27

Table A3b. Canopy/Interspace controls on SOC

	C SOC	I SOC	C-I	C SOC	I SOC	C-I
depth	30 cm	30 cm	30 cm	10 cm	10 cm	10 cm
1	0.8	0.7	0.2	1.3	0.9	0.4
2	1	0.9	0.1	1.3	1.1	0.2
3	4.8	4.5	0.3	7.4	5.5	1.9
4	1	1.1	-0.1	1.1	1.4	-0.3
5	1.2	1.2	0	1.7	1.5	0.2
6	2.4	1.8	0.6	3.8	2.3	1.5
7	0.5	0.5	0.1	0.8	0.6	0.2
8	0.5	0.3	0.1	0.8	0.4	0.3
9	0.6	0.4	0.2	0.8	0.4	0.4
10	0.6	0.4	0.1	0.8	0.7	0.1
11	0.6	0.4	0.2	0.9	0.5	0.4
12	1.7	1.6	0.1	2.5	2.1	0.4
13	0.7	0.5	0.1	1.1	0.8	0.4
14	1.4	1.1	0.3	1.8	1.1	0.7
15	5	4.7	0.3	6.2	5.9	0.3
16	0.8	0.8	0	1.2	1	0.2
17	0.7	0.6	0.1	1	0.6	0.4
18	2.5	2.5	0.1	4	3.8	0.2
19	3.7	3.7	0	5.6	5.9	-0.4
20	1.7	1.5	0.2	2.7	2.3	0.4
21	1.1	0.4	0.6	1.5	0.6	0.9
22	2.9	3.1	-0.2	4.1	4.2	-0.1
23	3.8	3.3	0.4	4.5	3.9	0.6
24	2	1.7	0.2	2.7	2.6	0.2
25	1.5	1.3	0.3	2.2	1.6	0.6
26	3.3	3	0.3	4.3	3.4	0.9
27	4.5	4.6	-0.1	6.9	7.6	-0.7
28	1.9	1.3	0.6	2.5	1.5	1
mu	1.9	1.7	0.2	2.7	2.3	0.4
std dev	1.4	1.4	0.2	2	2	0.5
RSD	70	80	110	70	90	130

C SOC: SOC for canopy samples
I SOC: SOC for interspace samples
C-I: SOC (canopy samples) – SOC (interspace samples)

Table A3c. Canopy/interspace controls II

	(C-I)/SOC	(C-I)/SOC	(C-I)/STD. DEV	(C-I)/STD. DEV	C std. dev	I std. dev	C std. dev	I std. dev
	30 cm	10 cm	30 cm	10 cm	30 cm	30 cm	10 cm	10 cm
1	0.2	0.4	0.5	0.6	0.4	0.3	0.7	0.6
2	0.1	0.1	0.6	0.8	0.2	0.1	0.2	0.1
3	0.1	0.3	0.1	0.5	2	0.9	4.7	1.6
4	-0.1	-0.2	-0.3	-0.5	0.3	0.3	0.4	0.6
5	0	0.1	-0.1	0.6	0.2	0.1	0.4	0.2
6	0.3	0.5	0.6	0.8	1.3	0.6	2.2	0.9
7	0.1	0.3	0.2	0.5	0.3	0.3	0.6	0.3
8	0.3	0.6	1.1	1.4	0.1	0.1	0.2	0.1
9	0.5	0.6	0.9	1.3	0.2	0.2	0.3	0.2
10	0.3	0.1	0.3	0.2	0.3	0.1	0.3	0.2
11	0.3	0.5	0.7	1.1	0.2	0.2	0.3	0.3
12	0.1	0.2	0.5	0.9	0.3	0.2	0.5	0.3
13	0.2	0.4	0.6	1	0.2	0.1	0.4	0.2
14	0.2	0.5	0.6	0.9	0.6	0.3	1	0.2
15	0.1	0.1	0.4	0.4	0.7	0.8	0.8	1.1
16	0.1	0.2	0.2	0.4	0.2	0.3	0.3	0.5
17	0.2	0.5	0.7	1.4	0.2	0.2	0.2	0.2
18	0	0	0.1	0.1	0.6	0.8	0.9	1.6
19	0	-0.1	0	-0.1	0.8	1.5	1.3	3.4
20	0.1	0.2	0.5	0.5	0.4	0.4	0.7	0.9
21	0.7	0.7	2	1.7	0.3	0.2	0.5	0.3
22	-0.1	0	-0.2	-0.1	0.9	1.1	0.9	1.1
23	0.1	0.1	0.7	0.6	0.6	0.5	1	0.8
24	0.1	0.1	0.9	0.4	0.2	0.3	0.3	0.5
25	0.2	0.3	0.7	1	0.4	0.3	0.6	0.4
26	0.1	0.2	0.4	0.9	0.6	0.8	1	0.9
27	0	-0.1	-0.3	-2.3	1.4	1.3	2.3	2.2
28	0.4	0.5	0.9	0.8	0.8	0.2	1.5	0.5
mu	0.2	0.3	0.5	0.6	0.5	0.5	0.9	0.7
std dev	0.2	0.2	0.5	0.8	0.4	0.4	0.9	0.8
RSD	100	90	100	130	90	90	110	100

C-I/SOC: (SOC (canopy samples) – SOC (interspace samples))/ site SOC
C-I/Std. dev.: (SOC (canopy samples) – SOC (interspace samples))/ site SOC std. dev.
C std. dev: SOC standard deviations for canopy samples
I std. dev.: SOC standard deviations for interspace samples

Table A3d. Canopy/Interspace by pair

	%increase						units of SOC					
	10cm			30cm			10cm			30cm		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
1	-15	256	73	-10	154	39	-0.3	1.5	0.4	-0.1	0.7	0.2
2	-10	36	15	-13	35	10	-0.1	0.4	0.2	-0.1	0.3	0.1
3	-16	91	28	-34	66	9	-0.9	7.3	1.9	-2.2	3.3	0.3
4	-54	69	-9	-29	19	-5	-1.1	0.7	-0.3	-0.4	0.2	-0.1
5	-25	36	14	-22	15	0	-0.5	0.6	0.2	-0.3	0.2	0.0
6	4	185	75	-14	107	36	0.1	4.8	1.5	-0.3	2.1	0.6
7	-55	309	94	-46	159	43	-0.6	1.0	0.2	-0.4	0.4	0.1
8	2	170	86	19	65	42	0.0	0.5	0.3	0.1	0.2	0.1
9	42	151	94	35	105	77	0.2	0.5	0.4	0.1	0.4	0.2
10	-32	63	13	-26	88	31	-0.2	0.4	0.1	-0.1	0.4	0.1
11	23	207	85	-6	128	55	0.2	0.6	0.4	0.0	0.5	0.2
12	-17	53	21	-6	33	10	-0.4	1.1	0.4	-0.1	0.5	0.1
13	13	115	48	-7	57	22	0.1	0.8	0.4	0.0	0.3	0.1
14	13	182	63	-1	69	25	0.1	2.2	0.7	0.0	0.9	0.3
15	-24	21	5	-6	15	4	-1.7	6.9	1.5	-0.3	5.4	1.3
16	-33	78	30	-8	31	12	-0.6	0.5	0.2	-0.1	0.2	0.1
17	20	172	71	-4	67	30	0.1	0.8	0.4	0.0	0.3	0.1
18	-26	43	12	-12	32	6	-1.7	1.2	0.2	-0.5	0.6	0.1
19	-55	127	18	-48	65	16	-6.5	4.2	-0.4	-2.9	1.7	0.0
20	-11	63	22	-5	51	14	-0.4	1.4	0.4	-0.1	0.7	0.2
21	12	365	97	2	310	82	0.1	1.6	0.6	0.0	0.9	0.3
22	-26	37	2	-21	23	-3	-1.4	1.2	-0.1	-1.0	0.6	-0.2
23	4	23	14	6	18	13	0.2	1.0	0.6	0.2	0.7	0.4
24	-12	49	8	1	37	14	-0.4	1.0	0.2	0.0	0.6	0.2
25	6	78	36	-9	57	23	0.1	1.2	0.6	-0.1	0.6	0.3
26	-16	81	34	-16	48	23	-0.8	2.6	0.9	-0.7	1.2	0.5
27	-30	16	-9	-14	5	-1	-2.3	1.4	-0.7	-0.7	0.3	-0.1
28	8	168	60	-4	110	41	0.1	3.0	1.0	-0.1	1.6	0.6

Note: All metrics relate to increase in SOC for canopy samples relative to interspace.

Table A4. SOC Data and variability by elevation

Elevation class	SOC Data	Mean	SD	RSD
Low	SOC	2.16	0.81	37
Low	%SOC	0.84	0.73	87
Mid	SOC	3.16	1.5	42
Mid	%SOC	1.33	0.35	26
High	SOC	4.57	1.94	43
High	%SOC	3.42	1.18	34
ALL	SOC	3.35	1.94	58
ALL	%SOC	1.79	1.46	82

Table A5a. Instrument uncertainty was quantified using the standards, aspartic acid and M soil, as well as triplicates on every tenth sample.

	Aspartic Acid	M Soil	Triplicates
True value	36.09	1.55	N/A
SD	0.86	0.03	0.03
RSD	2	3	3
SE	0.19	0.01	0.01
mean %error	1.72	1.80	2.17

Table A5b. Laboratory uncertainty was calculated by running 3 subsamples of the exact same sample through the entire laboratory process and this was repeated for 3 trials.

Metric	SOC (kg/m2)	%SOC
SD	0.05	0.11
RSD	5	9
Mean	0.98	2.84

Table A6. Correlation (R2) between predictors (vegetation indices)

	NDVI	**%1-2.5m**	**%2.5-10m**	**NLI**	**MRENDVI**	**MRESR**	**PSRI**
NDVI(max)	1.00	0.23	0.41	0.81	**0.95**	**0.92**	**0.90**
%1-2.5m	0.23	1.00	0.58	0.27	0.27	0.30	0.23
%2.5-10m	0.41	0.58	1.00	0.40	0.39	0.40	0.33
NLI	0.81	0.27	0.40	1.00	0.80	0.84	0.70
MRENDVI	**0.95**	0.27	0.39	0.80	1.00	0.98	0.95
MRESR	**0.92**	0.30	0.40	0.84	0.98	1.00	0.90
PSRI	**0.90**	0.23	0.33	0.70	0.95	0.90	1.00

.

REFERENCES

Adams, H. R., H. R. Barnard, and A. K. Loomis. 2014. Topography alters tree growth-climate relationships in a semi-arid forested catchment. Ecosphere 5 (11): 148. http://dx.doi.org/10.1890/ES14-00296.1

Aguilar, R., and R. D. Heil. 1988. Soil organic carbon, nitrogen and phosphorus quantities in Northern Great Plains rangeland. Soil Science Society of America Journal 52: 1076-1081.

Austreng, Andrew Connor, "The Carbon Budget Impact of Sagebrush Degradation" (2012). *Boise State University Theses and Dissertations*. 268. http://scholarworks.boisestate.edu/td/268

Ballabio, C., Fava, F., and Rosenmund, A.: A plant ecology approach to digital soil mapping, improving the prediction of soil organic carbon content in alpine grasslands, Geoderma, 187, 102–116, doi:10.1016/j.geoderma.2012.04.002, 2012.

Bameri, A., Khormali, F., Kiani, F., Dehghani, A.A., 2015. Spatial variability of soil organic carbon in different hillslope positions in Toshan area, Golestan province, Iran: geostatistical approaches. J. Mt. Sci. 12 (6), 1422–1433

Bartholomeus, H.M., et al., 2008. Spectral reflectance based indices for soil organic carbon quanitification. Geoderma 145, 9.

Bartholomeus, H., Kooistra, L., Stevens, A., van Leeuwen, M., van Wesemael, B., Ben-Dor, E., Tychon, B., 2011. Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. Int. J. Appl. Earth Obs. Geoinf. 13, 81–88.

Beedlow PA, Lee EH, Tingey DT, Waschmann RS, Burdick CA (2013) The importance of seasonal temperature and moisture patterns on growth of Douglas-fir in

western Oregon, USA. Agr Forest Meteorol 169: 174-185. doi:10.1016/j.agrformet.2012.10.010.

Breiman L (2001) Random forests. Mach Learn 45:5–32

Broge, N., and E. Leblanc. "Comparing Prediction Power and Stability of Broadband and Hyperspectral Vegetation Indices for Estimation of Green Leaf Area and Canopy Chlorophyll Density." *Remote Sensing of Environment* 76 (2000): 156-172.

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 Sci. Soc. Am. J. 65, 480– 490.

Crookston, Nicholas L.; Finley, Andrew O. 2007. yaImpute: An R Package for k-NN Imputation. Journal of Statistical Software. 23(10):1-16.

Curran, P., W. Windham, and H. Gholz. "Exploring the Relationship Between Reflectance Red Edge and Chlorophyll Concentration in Slash Pine Leaves." *Tree Physiology* 15 (1995): 203-206.

Datt, B. "A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests Using Eucalyptus Leaves." *Journal of Plant Physiology* 154 (1999):30-36.

Datt, B., T. R. McVicar, T. G. Van Niel, D. L. B. Jupp, and J. S. Pearlman, "Preprocessing EO-1 hyperion hyperspectral data to support the application of agricultural indexes," IEEE Trans. Geosci. Remote Sens., vol. 41, no. 6, pp. 1246–1259, Jun. 2003.

Daughtry, C., et al. "Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance." *Remote Sensing Environment* 74 (2000): 229–239.

Doetterl, Sebastian, Antoine Stevens, Johan Six, Roel Merckx, Kristof Van Oost, Manuel Casanova Pinto, Angélica Casanova-Katny, Cristina Muñoz, Mathieu Boudin, Erick Zagal Venegas, and Pascal Boeckx. "Soil carbon storage controlled by interactions between geochemistry and climate." *Nature Geoscience* 8.10 (2015): 780-83. Web.

Draper, N. R., and H. Smith. *Applied Regression Analysis*. Hoboken, NJ: Wiley-Interscience, 1998. pp. 307–312.

Fellows, Aaron, and Gerald Flerchinger. “CO2 Flux Data from RCEW CZO Core Sites.” Reynolds Creek Experimental Watershed, 2016.

Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil-landscape modelling and spatial prediction of soil attributes. Int. J. Geogr. Inf. Sci. 9 (4), 421–432.

Gessler, P.E., Chadwick, O.A., Chamran, F., Althouse, L., Holmes, K., 2000. Modeling soil–landscape and ecosystem properties using terrain attributes. Soil. Sci. Soc. Am. J. 64, 2046–2056.

Gitelson, A., and M. Merzlyak. "Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus Hippocastanum L.* and *Acer Platanoides L.* Leaves." *Journal of Plant Physiology* 143 (1994): 286-292.

Gomez, Cécile, Raphael A. Viscarra Rossel, and Alex B. Mcbratney. "Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study." *Geoderma* 146.3-4 (2008): 403-11. Web.

Grimm, R., Thorsten Behrens, M. Maerker, and H. Elsenbeer. "Soil Organic Carbon Concentrations and Stocks on Barro Colorado Island; Digital Soil Mapping Using Random Forests Analysis." *Geoderma* 2nd ser. 146.1 (2008): 102-13. *GeoRef*. Web. 7 May 2014.

Haboudane, D., et al. "Hyperspectral Vegetation Indices and Novel Algorithms for Predicting Green LAI of Crop Canopies: Modeling and Validation in the Context of Precision Agriculture." *Remote Sensing of Environment* 90 (2004): 337-352.

Hbirkou C, Pätzold S, Mahlein A-K, Welp G (2012) Airborne hyperspectral imaging of spatial soil organic carbon heterogeneity at the field-scale. Geoderma 175–176:21–28

Hoffmann, Ulrike, Thomas Hoffmann, E.a. Johnson, and Nikolaus J. Kuhn. "Assessment of variability and uncertainty of soil organic carbon in a mountainous boreal forest (Canadian Rocky Mountains, Alberta)." *Catena* 113 (2014): 107-21. Web.

Ilangakoon, Nayani; Glenn, Nancy F.; Spaete, Lucas P.; Dashti, Hamid; and Li, Aihua. (2016). 2014 Lidar-Derived 1m Digital Elevation Model Data Set for Reynolds Creek Experimental Watershed, Southwestern Idaho [Data set]. Boise, ID: http://doi.org/10.18122/B26C7X

Intergovernmental Panel on Climate Change (IPCC) (2007), Climate Change 2007: The Physical Science Basis. Contributions of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by S. Solomon et al., Cambridge Univ. Press, Cambridge, U. K.

Jobbagy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol. Appl. 10, 423–-436 (2000).

Klos, P.Z.; Link, T.E.; Abatzoglou, J.T. Extent of the rain-snow transition zone in the western U.S. under historic and projected climate. *Geophys. Res. Lett.* 2014, *41*, 4560–4568.

Kormos, P.R.; Marks, D.; Pierson, F.B.; Williams, C.J.; Hardegree, S.P.; Havens, S.; Hedrick, A.; Bates, J.D.; Svejcar, T.J. Ecosystem Water availability in juniper versus sagebrush snow-dominated rangelands. *Rangel. Ecol. Manag.* 2016, in press.

Kulmatiski, Andrew, Daniel J. Vogt, Thomas G. Saccama, and Joel P. Tilley. "Landscape Determinants of Soil Carbon and Nitrogen Storage in Southern New England." *Soil Science Society of America Journal* 68.6 (2004): 2014-022. *Science Direct*. Web. 27 Sept. 2013.

Kunkel, Melvin L., Alejandro N. Flores, Toni J. Smith, James P. McNamara, and Shawn G. Benner. "A Simplified Approach for Estimating Soil Carbon and Nitrogen Stocks in Semi-arid Complex Terrain." *Geoderma* (2011): 1-11. *Science Direct*. Web. 6 Oct. 2013.

Lacoste, Marine, Budiman Minasny, Alex Mcbratney, Didier Michot, Valérie Viaud, and Christian Walter. "High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape." *Geoderma* 213 (2014): 296-311. Web.

Lal, R. "Soil Carbon Sequestration Impacts on Global Climate Change and Food Security." *Science* 304.5677 (2004): 1623-627. Web.

Lemke, P., et al. (2007), Observations: Changes in snow and frozen ground, in Climate Change 2007: The Physical Science Basis. Contributions of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by S. Solomon et al., pp. 337–383, Cambridge Univ. Press, Cambridge, U. K

Li, A., Glenn, N. F., Olsoy, P. J., Mitchell, J. J., & Shrestha, R. (2015). Aboveground biomass estimates of sagebrush using terrestrial and airborne Lidar data in a dryland ecosystem. Agricultural and Forest Meteorology, 213, 138–147. http://dx.doi.org/10.1016/j. agrformet.2015.06.005.

Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152.

Martin, M.P., Wattenbach, M., Smith, P., Meersmans, J., Jolivet, C., Boulonne, L., Arrouays, D., 2011. Spatial distribution of soil organic carbon stocks in France. Biogeosciences 8, 1053–1065.

McBratney, A. B., M. L. Santos, and B. Minasny. "On Digital Soil Mapping." *Geoderma* 117 (2003): 3-52. *Science Direct*. Web. 2 May 2014.

McCorkle, E., 2015. Protocols, Reynolds Creek Critical Zone Observatory. http://info.reynoldscreekczo.org/dokuwiki/doku.php?id=protocols, October, 2016.

McFadden, L.D., 2013. Strongly dust-influenced soils and what they tell us about landscape dynamics in vegetated arid lands of the southwestern United States. In: Bickford, M.E. (Ed.), The Web of Geological Sciences: Advances, Impacts, and Interactions. Geological Society of America Special Paper, 500. http://dx.doi.org/10.1130/ 2013.2500(15) (chapter 15).

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

Miller, B.A., Koszinski, S., Hierold, W., Schröder, B., Wehrhan, M., Sommer, M., 2015. Mapping soil carbon landscapes: issues of sampling scale and transferability. Soil Till. Res.

Minasny, B., McBratney, A.B., Mendonça-Santos, M.L., Odeh, I.O.A., Guyon, B., 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. Aust. J. Soil Res. 44, 233–244.

Minasny, Budiman, Alex B. McBratney, Brendan P. Malone, and Ichsani Wheeler. "Digital Mapping of Soil Carbon." *Advances in Agronomy* 118 (2013): 1-47. *Science Direct*. Web. 8 Oct. 2013.

Mishra, U., Lal, R., Liu, D.S., Van Meirvenne, M., 2010. Predicting the spatial variation of the soil organic carbon pool at a regional scale. Soil. Sci. Soc. Am. J. 74, 906–914.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal 57, 443 – 452.

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

Mueller, T.G., Pierce, F.J., 2003. Soil carbon maps: enhancing spatial estimates with simple terrain attributes at multiple scales. Soil. Sci. Soc. Am. J. 67, 258–267.

Mulder, V.l., S. De Bruin, M.e. Schaepman, and T.r. Mayr. "The use of remote sensing in soil and terrain mapping — A review." *Geoderma* 162.1-2 (2011): 1-19. Web.

Nayak, A., D. Marks, D. G. Chandler, and M. Seyfried (2010), Long-term snow, climate, and streamflow trends at the Reynolds Creek Experimental Watershed, Owyhee Mountains, Idaho, United States, Water Resour. Res., 46, W06519, doi:10.1029/2008WR007525.

Ni-Meister, W., Lee, S., Strahler, A. H., Woodcock, C. E., Schaaf, C., Yao, T., et al. (2010). Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. Journal of Geophysical Research, 115, 1–12.

Noy-Meir, I., 1973, Desert ecosystems: Environment and producers: Annual Review of Ecology and Systematics, v. 4, p. 25-51.

Patton, Nicholas R., 2016, Topographic Controls on Total Mobile Regolith and Total Soil Organic Carbon in Complex Terrain, [M.S. Thesis] Pocatello, Idaho State University, 101 p.

Peng, Xiaoting, Tiezhu Shi, Aihong Song, Yiyun Chen, and Wenxiu Gao. "Estimating Soil Organic Carbon Using Vis/NIR Spectroscopy with SVMR and SPA Methods." *Remote Sensing* 6.4 (2014): 2699-717. Web.

Périé, C., Ouimet, R., 2008. Organic carbon, organic matter and bulk density relationships in boreal forest soils. Canadian Journal of Soil Science 88, 315–325

Raich, JW & Potter CS (1995) Global patterns of carbon dioxide emissions from soils. Global Biogeochem. Cycles 9: 23–36

Rau, B. M., D. W. Johnson, R. R. Blank, R. J. Tausch, B. A. Roundy, R. F. Miller, T. G. Caldwell, and A. Lucchesi. 2011a. Woodland expansion's influence on belowground carbon and nitrogen in the Great Basin U.S. Journal of Arid Environments 75:827-835.

Rau B. M., Johnson DW, Blank RR, Lucchesi A, Caldwell TG, Schupp EW. 2011b. Transition from sagebrush steppe to annual grass (Bromus tectorum): influence on belowground carbon and nitrogen. Rangel Ecol Manag 64:139–47.

Restaino, C.M.; Peterson, D.L.; Littell, J. Increased water deficit decreases Douglas fir growth throughout western US forests. *Proc. Natl. Acad. Sci. USA* 2016, *113*, 9557–9562.

Schimel, D., M. A. Stillwell, and R. G. Woodsmansee, Biogeochemistry of C, N, and P in a soil catena of the shortgrass steppe, *Ecology*, 66, 276–282, 1985.

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

Schlesinger WH, Raikes JA, Hartley AE, Cross AF. 1996. On the spatial pattern of soil nutrients in a desert ecosystem. Ecology 72:364–74

Schlesinger, W. H. 1997. Biogeochemistry, an analysis of global change. Academic Press, San Diego, California, USA.

Selige, T., Böhner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma 136, 235–244.

Seyfried, M., R. Harris, D. Marks, and B. Jacob (2001), Geographic Database, Reynolds Creek Experimental Watershed, Idaho, United States, Water Resour. Res., 37(11), 2825–2829,

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

Sims, D., and J. Gamon. "Relationships Between Leaf Pigment Content and Spectral Reflectance Across a Wide Range of Species, Leaf Structures and Developmental Stages." *Remote Sensing of Environment* 81 (2002):337-354.

Smith, T. J., J. P. McNamara, A. N. Flores, M. M. Gribb, P. S. Aishlin, and S. G. Benner. "Small Soil Storage Capacity Limits Benefit of Winter Snowpack to Upland Vegetation." Hydrological Processes 25.25 (2011): 3858-865. Print.

Stanbery, Christopher Allen, "Controls on the Presence, Concentration, Storage, and Variability of Soil Inorganic Carbon in a Semi-Arid Watershed" (2016). *Boise State University Theses and Dissertations*. 1233. http://scholarworks.boisestate.edu/td/1233

Stephenson, Gordon R. "Soil-Geology-Vegetation Inventories for Reynolds Creek Watershed." Agricultural Experiment Station, University of Idaho College of Agriculture. in cooperation with Agricultural Research Service and Soil Conservation Service (1965): n. pag. U.S. Department of Agriculture. Web. 19 June 2014.

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., van Wesemael, B.,2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. Geoderma 158, 32–45.

Swetnam, T. L., P. D. Brooks, H. R. Barnard, A. A. Harpold, and E. L. Gallo. 2017. Topographically driven differences in energy and water constrain climatic control on forest carbon sequestration. Ecosphere 8(4):e01797. 10.1002/ecs2.1797

Thompson, James A., and Randall K. Kolka. "Soil Carbon Storage Estimation in a Forested Watershed Using Quantitative Soil-Landscape Modeling." *Soil Science Society of America Journal* 69.4 (2005): 1086-093. *GeoRef*. Web. 1 Oct. 2013.

Titus, J.H., Nowak, R.S., Smith, S.D., 2002. Soil resource heterogeneity in the Mojave desert. J. Arid Environ. 52, 269–292.

Trumbore SE, Chadwick OA & Amundson R (1996) Rapid exchange between soil carbon and atmospheric carbon dioxide driven by temperature change. Science 272: 393–396

Vasques, G.M., Grunwald, S., Comerford, N.B., Sickman, J.O., 2010a. Regional modelling of soil carbon at multiple depths within a subtropical watershed. Geoderma 156, 326–336.

Vogelmann, J., B. Rock, and D. Moss. "Red Edge Spectral Measurements from Sugar Maple Leaves." *International Journal of Remote Sensing* 14 (1993): 1563-1575.

Wang, Y.Q., Zhang, X.C., Zhang, J.L., Li, S.J., 2009. Spatial variability of soil organic carbon in a watershed on the Loess Plateau. Pedosphere 19, 486– 495

Westerling A.L., H.G. Hidalgo, D.R. Cayan, and T.W. Swetnam, 2006: Warming and earlier spring increase western U.S. forest wildfire activity. Science, 313(5789), 940-943.

Wiesmeier, M., Barthold, F., Blank, B., Kögel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using random forest modeling in a semi-arid steppe ecosystem. Plant Soil 340, 7–24.

Will, Ryan M.; Benner, Shawn; Glenn, Nancy F.; Pierce, Jennifer; Lohse, Kathleen A.; Patton, Nicholas; Spaete, Lucas P.; and Stanbery, Christopher. (2017). *Mapping Soc Distribution in Semi-arid Mountainous Regions Using Variables From Hyperspectral, Lidar and Traditional Datasets* [Data set]. Retrieved from https://doi.org/10.18122/B2Q598

Williams AM, Hunt Jr ER (2002) Estimation of leafy spurge from hyperspectral imagery using mixture tuned matched filtering. Remote Sens Environ 82:446–456.

Williams, CJ, McNamara, JP, Chandler, DG, 2009. Controls on the temporal and spatial variability of soil moisture in a mountainous landscape: the signature of snow and complex terrain. Hydrology and Earth System Sciences 13 (7), 1325–1336.

Winstral, A, Marks, D, 2002. Simulating wind fields and snow redistribution using terrainbased parameters to model snow accumulation and melt over a semi-arid mountain catchment. Hydrological Processes 16 (18), 3585–3603.

Woodwell, G. M., F. T. Mackenzie, R. A. Houghton, M. Apps, E. Gorham, and E. Davidson. "Biotic Feedbacks in the Warming of the Earth." *Climatic Change* 40.3-4 (1998): 495-518. *Biotic Feedbacks in the Warming of the Earth*. Springer, 1999. Web. 13 Apr. 2015.