THESIS

INTEGRATING BASIC REMOTE SENSING, TERRAIN ANALYSIS AND GEOSTATISTICAL METHODS TO GENERATE SPATIALLY EXPLICATE CONTINIOUS SOIL ATTRIBUTE MAPS FOR FRASER EXPERIMENTAL FOREST

Submitted by

John Barstow Norman III

Department of Soil and Crop Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2010

Master's Committee:

Department Head: Gary Peterson

Advisor: Eugene Kelly

Charles Rhoads
Robin Reich

ABSTRACT OF THESIS

INTEGRATING BASIC REMOTE SENSING, TERRAIN ANALYSIS AND GEOSTATISTICAL METHODS TO GENERATE SPATIALLY EXPLICATE CONTINUOUS SOIL ATTRIBUTE MAPS FOR FRASER EXPERIMENTAL FOREST

Hans Jenny's *Factors of Soil Formation*, a system of quantitative pedology (1941), concisely summarized and illustrated many of the basic principles of pedology utilized to date (Jenny, 1941). This state factor model became the backbone for soil survey research and production because it proposed that a limited number of environmental factors could largely explain the distribution of soils within and among ecosystems.

Advances in soil chemistry, soil physics, soil mineralogy, and soil biology, as well as in the basic sciences have helped increase our fundamental understanding of the spatial distribution of soil. In addition, new tools and new dimensions to the study of soil formation have evolved with the increasing power and utility of Geographical Information Systems (GIS) and geostatistical analysis to further quantify the complex spatial relationships of soils and landscapes. These advances have resulted in a new field of study termed pedometrics, which focuses on the application of mathematical and statistical methods for the study of the distribution and evolution of soils.

This study implements pedometric principles and methods to develop high resolution and spatially explicate soil attribute maps for Fraser Experimental Forest (FEF) based on simple terrain, remote sensing and geostatistical analyses. The soil attribute models developed for this study provided a continuous representation of soil properties (Total soil depth, A-horizon and O-horizon thickness) at a fine scale (0.001 ha). These spatial models can be used as inputs to hydrological and ecological models to further evaluate the soil's influence on water chemistry and vegetation distributions, and to provide an initial platform for future soil survey activities in FEF. In addition to developing soil attribute surfaces for FEF, I tested the statistical, spatial and cost efficiencies of the Spatially Balances Survey (SBS) design developed to sample soils and inform the geostatistical models for FEF.

AKNOWLEDGEMENTS

I would like to begin by extending an over arching thank you to friends and colleagues that I have had the privilege working and studying with at the Short Grass Step Long Term Ecological Research program (SGS-LTER), the Pedology and Soils Information Systems (PASIS) laboratory, Natural Resource Ecology Laboratory (NREL), and more recently Dr. Dave Theobald's lab. Through the course of working and studying at Colorado State University three individuals have played key roles that have "honed" my academic, scientific, technological and profession stills required to finish this thesis: my advisor, Dr. Eugene F. Kelly who gave me the opportunity to be involved with this project and sprinted with me at the end of thesis with critical edits that streamlined the thesis and helped me gather my thoughts; my boss, Dr. David M. Theobald whom I've had the extreme privilege to work with since 2004. This opportunity has provided me with an understanding of geospatial analysis that spans basic concepts of geography to graph theory, which will guide me throughout my career path. Finally, I would like to thank Robert O. Coleman, Colorado State University, who gave me a chance to succeed by mentoring me early on in my career, by hiring me as a teaching assistant and giving me office space during my undergraduate years. I would also like to thank my committee, Dr. Chuck C. Rhoades and Robin M. Reich who have for their advice in their respective area of specialty.

Last but not least, I would like to extent my deepest and sincere appreciation and gratitude to my parents (John B. Norman jr. and Roberta J. Norman), wife Georgian L.

Norman and my daughter Terra J. Norman, for they, more than anyone else, were my biggest supports and were to do whatever was necessary to help me achieve my goals. I dedicate all the efforts I made to finish this work to them.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

LIST OF FIGURES

Chapter 1

Introduction

1.1 – Background

The worldwide interest in collecting, categorizing and mapping soils is increasing with the recognition that soil properties and the processes governing their development are critical to exchanges between terrestrial, aquatic and atmospheric systems. Soil science and soil classification originated in the mid 1800s largely to address differences in land productivity. It is now clear that reliable assessments of current and future ecosystem status (e.g. health) requires detailed, spatially-explicit information about soil properties across natural and managed landscapes.

Our ability to integrate soils information across landscapes is hindered by the lack of theoretical models that incorporate the influence of topography on soil properties in complex terrain. For individual sites our knowledge allows us to quantify and describe soil properties and relationships. We currently have the ability extrapolate site-level information through space where terrain variables such as landform-type and topography remain relatively constant.

Jenny 1941, formalized the state factor model to better account for how environmental variables such as climate, organisms, relief (topography), parent material and time and topography conditioned soil properties through space and time. With the exception

of topography, the spatial scales at which the other four state factors (organisms, climate, time, and parent material) vary through space are easily quantified. Maps of potential natural vegetation, major land resource areas (USDA-NRCS, 1998), ecoregions of the United States (Bailey 1995) and geology illustrate our understanding of this variability and our ability to map it at broad geographic extents. Topography, however, may influence the development of soil properties at a much finer spatial scale (Weitz et al., 1993). Soil organic carbon and clay content variability within a soil profile are influenced by topographic relationships that vary within a hill slope (Aguilar et al., 1988; Kelly et al., 1988). Much research has been directed towards the quantification of soil properties as a function of topographic controls; attention must now be focused on how to reasonably and appropriately incorporate this variability at large spatial extents (e.g., National Forests, watersheds or counties). As such, the need to link soils to landscapes is of practical and scientific importance for the inventory and management of resources for federal, state, county and private lands.

Current data requirements for process-driven ecological or hydrologic models has sparked the need for high resolution spatially explicate soil attribute data that varies at a much finer spatial scale than current soil databases (e.g., STATSGO and SSURGO). To compile such spatial databases, soil scientists are beginning to adopt geostatistical methodologies and beginning to develop spatially explicit soil property models based on multiple sample locations and terrain attributes as predictor variables. This methodology also allows users of these data to quantitatively assess model performance as a global metric (e.g., R^2, MSE, Moran's I) or a spatial metrics (Confidence Intervals and MSE surfaces). In addition, this approach relies on a spatial sampling scheme with a large number of field based observations and in wilderness and remote settings of the western U.S. this can make this

methodology cost prohibitive. There is a critical need to collect these data necessary for forecasting global change in remote areas. Collecting these data can be costly and time consuming and new approaches and tools are needed. This, in essence, is the focus of the research presented in my Master's Thesis.

1.2 - Objectives

The main objective of this thesis is to develop high resolution (0.001 ha) spatially explicate soil depth, A-horizon and O-horizon thickness maps (surfaces) for Fraser Experimental Forest (FEF) utilizing geostatistical techniques. The development of the soil attribute surfaces required three components: 1.) geospatial data acquisition and analysis, 2.) development of an efficient sampling design to sample soils in remote inaccessible terrain, 3.) geostatistical techniques to develop soil attribute surfaces and quantify model performance. First, the geospatial acquisition and analysis entails downloading basic terrain and remote sensed data, as wells as performing simple spatial analysis to inform the geostatistical models. Second, the sampling component of this thesis consists of selecting intensive sampling units in FEF, testing the sampling design's efficiency performance, and field sampling techniques. Finally, the geostatistical model development incorporates simple statistical modeling techniques utilizing terrain and remote sensing analysis and field based observations to generate the soil attribute surfaces. The geostatistical model section also includes calculating soil attribute surface model performance summary statistics.

Chapter 2

Methods

2.1 - Study Site

The Fraser Experimental Forest (FEF) provides a suitable test bed to investigate advanced techniques in the inventory of soils. FEF was established in 1937 in the heart of the central Rocky Mountains located 50 miles (as the crow flies) from Denver, Colorado (Figure 2.1). This 9,300 ha (36 sq. mi) research facility (managed by the United States Forest Service Rocky Mountain Forest and Range Experiment Station) mission is to investigate how to better manage high elevation sub-alpine coniferous forest ecosystems to enhance timber, water and wildlife resources.

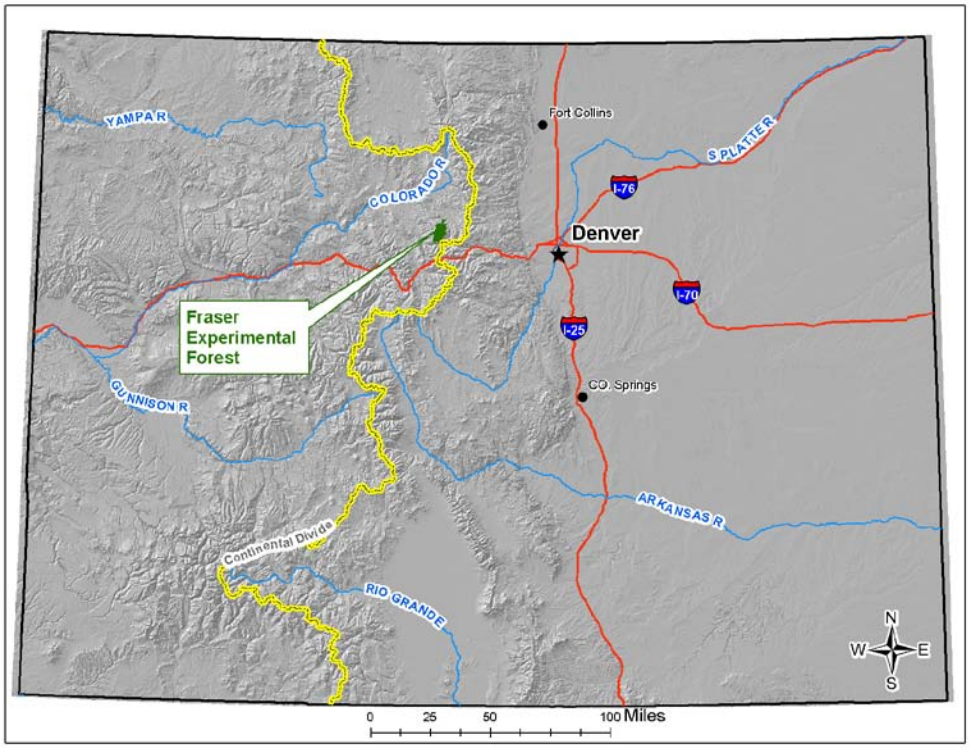


Figure 2.1 - Location of Fraser Experimental Forest in North Central, Colorado

The FEF is bounded by the St. Louis Creek watershed, a tributary of the Fraser River and part of the Upper Colorado River system. The St. Louis watershed is representative of Colorado and Wyoming high elevation headwaters which provide 85% of annual stream flow in Colorado and accounts for 20 million acre-feet of stream discharge annually. The climate is cool and humid with long winters and short, cool summers (Popovich et al. 1993). Precipitation averages 74 cm annually; about 75% falls as snow. Average annual temperature is 0.5° C (33° F) with an average annual range of 72° C (min. -40°, max. 32°) at FEF headquarters (9,000 feet). The vegetation is typical of the sub-alpine forest zones of the central Rocky Mountains with elevation separating vegetation associations. Engelmann spruce (Picea englmannii) and sub-alpine fir (Abies lasiocarpa) dominate at higher elevations that are north facing and along streams in lower elevations; lodgepole pine (Pinus contorta)

dominates lower elevations and drier upper slopes; herbaceous vegetation is sparse except along streams and disturbed sites; and barren rocks intermix with alpine tundra, meadows, and wetlands above timberline. The major geologic features of FEF are Proterozic metamorphic, calcium-rich gneisses, granodiorite and schist (Retzer 1962; Eppinger et al. 1985). There are a few isolated inclusions of Cretaceous Dakota sandstone and Jurassic Morrison shale and limestone. Topography of FEF is composed of steep, high mountain slopes with narrow, small flood plains. Elevation ranges from 2,682 to 3,902 meters with three-fourths of the total area above 3,048 meters (10,000 feet) and one-third above timberline (3,505 meters or 11,500 feet). Due to the irregular "boot" shape at the base of FEF, the watershed boundary of St. Louis Creek (APPENDIX 1) was chosen as the spatial extent of the analysis. This was done to increase the spatial leverage of the intensive sampling units within FEF and to maintain hydrologic connectivity of the spatial models.

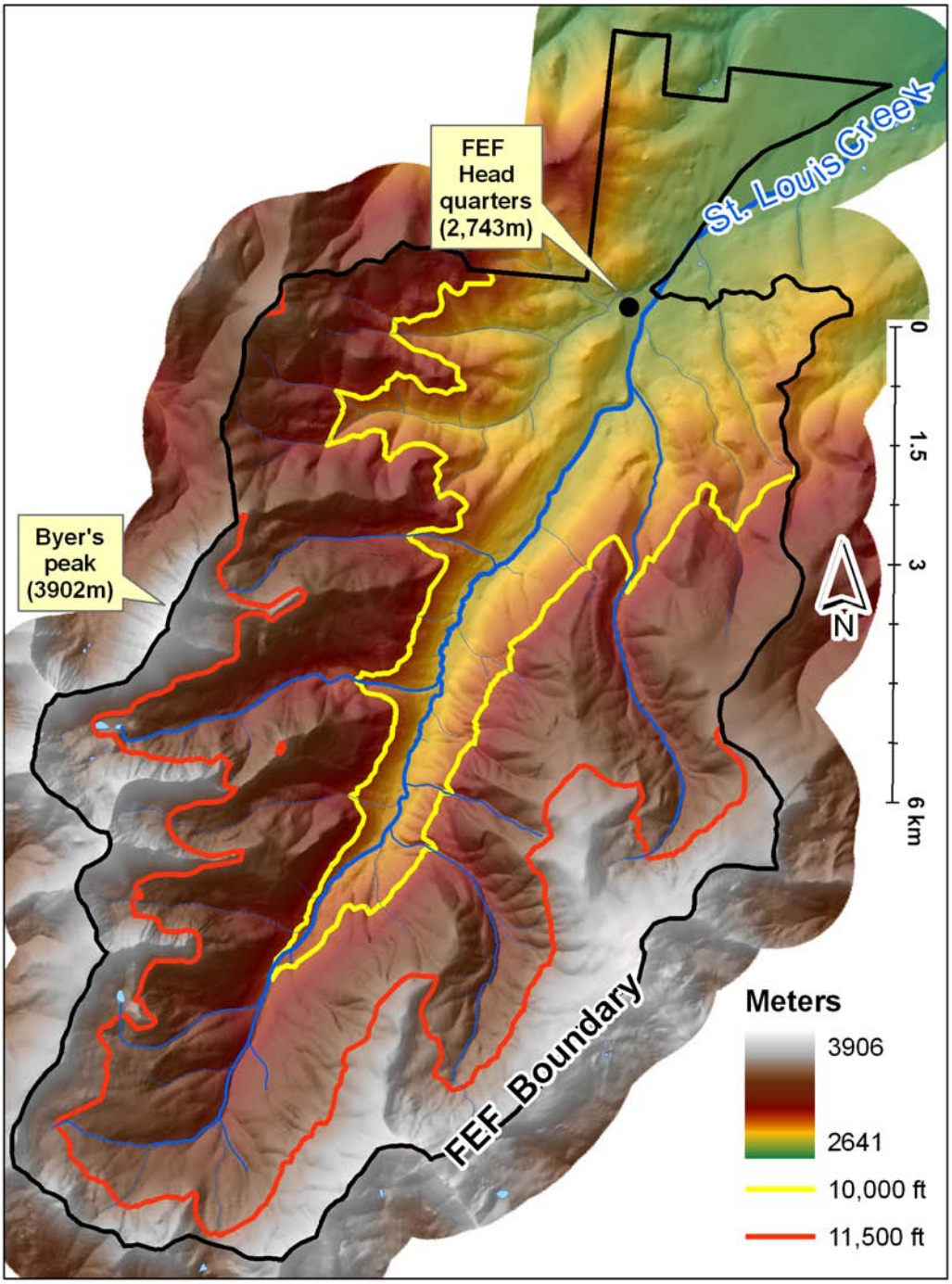


Figure 2.2 - Spatial extent of Fraser Experimental Forest

The only inventory of the soils at FEF was conducted during the mid-1950s (Retzer 1962). That survey provided basic soils information that assisted with development of research and natural resource management activities in the Fraser valley. The current soil

survey is available in two forms—the traditional soil survey document and a spatially explicit digital format. The survey document provides spatial (paper maps), taxonomic, chemical and physical data for 20 map units (APPENDIX 2) within FEF. In 2005, the Retzer (1962) paper maps were georectified and digitized into a GIS, with basic soil properties (soil depth, percent sand silt and clay, O- horizon thickness, and A-horizon thickness) attributed for each map unit (APPENDIX 3). The digital information was used to gain an understanding of the soil landscape within FEF for field sampling purposes and to provide a comparison against the soil attribute models.

2.2 - Geospatial data acquisition and analysis

Data collection entailed acquiring topographic, vegetation, spectral, and vector (streams and roads) spatial datasets from federal, state, and academic web-based data gateways (Table 2.1). These datasets were chosen because they have national coverage and can be post-processed into datasets that capture biotic, abiotic, local climate, and hydrologic variability. Landsat 7 ETM+ was chosen as the remote sensing platform for this study due to its vast spatial coverage (global), high temporal resolution (14 day repeat) and its multi-sensor platform, which has been used to develop a wide array custom multi-band vegetation and soil indices. The June 23, 2002 imagery was selected for this study because its cloud and snow cover free providing maximum contrast between alpine rock and alpine vegetation.

Table 2.1 - Spatial datasets downloaded from state and federal websites

Spatial dataset	Source link
Digital Elevation Model (DEM)	http://seamless.usgs.gov/website/seamless/viewer.htm
Hydrology (1:24,000)	http://nhdgeo.usgs.gov/viewer.htm
Colorado Vegetation Model (CVM)	http://www.nrel.colostate.edu/~davet/cvm.html
Landsat 7 ETM+	http://edc.usgs.gov/products/satellite/landsat7.php
Roads and Trails	http://svinetfc4.fs.fed.us/vectorgateway/index.html

The geospatial analysis consisted of processing the base layers (Table 2.1) utilizing terrain, spectral, and cost-analysis techniques to develop the sampling design inclusion probability surface and as covariates for the soil property models. The terrain analysis consisted of simple metrics, such as slope and curvature (Table 2.2) to complex hydrologically-weighted metrics that account for movement of soil as a function on gravity and water.

Table 2.2 – Primary terrain analysis methods ArcGIS 9.2 (ESRI 2007) tools

Terrain attribute	Tool
Profile curvature	CURVATURE
Slope	SLOPE
Solar Radiation	AREA SOLAR RADIATION

The spectral analyses are based on three widely used indices and transformations that spectrally separate vegetation from exposed soil and rock:

1.) Normalized Difference Vegetation Index

The Normalized Difference Vegetation Index (NDVI) is the mostly widely used vegetation index developed for Landsat 7 ETM+ because it's simple to calculate, robust across biomes and is good proxy of above ground bio mass. NDVI is calculated as the ratio between the red and near infrared (NIR) portions of the

spectrum (Equation 2.1). These two spectral bands are most affected by the absorption of chlorophyll in leafy green vegetation, and by the density of green vegetation (Lillesand and Kiefer, 2000).

$$NDVI = \frac{NIR - \mathrm{Re}\,d}{NIR + \mathrm{Re}\,d} \quad \text{(Equation 2.1)}$$

2.) Tassel Cap Transformation

The Tassel Cap Transformation (TCT) (Crist and Cicone, 1984; Kauth and Thomas, 1976) for Landsat 7 ETM+ takes advantage of the high degree of correlation that exists between visible, near infrared, and mid-infrared spectrums. This transformation reduces the seven Landsat bands into three orthogonal indices called brightness, greenness, and wetness, which account for 97% of spectral variability of the original seven bands. The three (Brightness, Greenness, and Wetness) TCT indices were calculated by summing the seven Landsat 7ETM+ bands weighted by the coefficients listed in Table 2.3.

Table 2.3 - Tasseled Cap Transformation coefficients for Landsat 7 ETM+ bands

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Brightness	0.33183	0.33121	0.55177	0.42514	0.48087	0.25252
Greenness	-0.24717	-0.16263	-0.40639	0.85468	0.05493	-0.11749
Wetness	0.13929	0.2249	0.40359	0.25178	-0.70133	-0.45732

a) Brightness (TCT1), is a weighted sum of all bands defined in the direction of the principle variation in soil reflectance (Lillesand and Kiefer, 2000)

b) Greenness (TCT2), is orthogonal to brightness and is a contrast between near-infrared and reflectance (Lillesand and Kiefer, 2000). It measures the presence and density of green vegetation.

c) Wetness (TCT3), is a contrast between shortwave-infrared (SWIR) and visible/near-infrared reflectance, providing a measure of soil and canopy moisture (Crist and Kauth, 1986)

3.) Modified Soil-Adjusted Vegetation Index

The Modified Soil-Adjusted Vegetation Index (MSAVI) adjusts for soil background reflectance allowing for a better representation of green healthy vegetation (Rondeaux et al. 1996). This index is an important consideration in this study because alpine vegetation is intermixed with brightly reflecting parent material. Qi et al. (1994) developed several MSAVI indices ranging from simple straight forward equations to very difficult abstract equations. The one used in this study is the simplest form of MSAVI and doesn't require pre-calculation of other vegetation indices or the slope of the soil reflectance line (Equation 2.2) (Qi et al., 1994).

$$MSAVI = \frac{2(NIR+1)}{2} - \sqrt{(2NIR+1)^2 - 8(NIR - \mathrm{Re}\,d)} \quad \text{(Equation 2.2)}$$

The terrain analysis focused on primary and secondary terrain attributes derived from a 10 meter Digital Elevation Model (DEM). Primary terrain attributes are derived directly from the DEM (Table 2.2) which capture local topographic phenomena that influences the soil landscape. Secondary terrain attributes are not derived directly from the DEM, but rather

from combinations of primary attributes to account for movement of hill slope alluvium from source to sink.

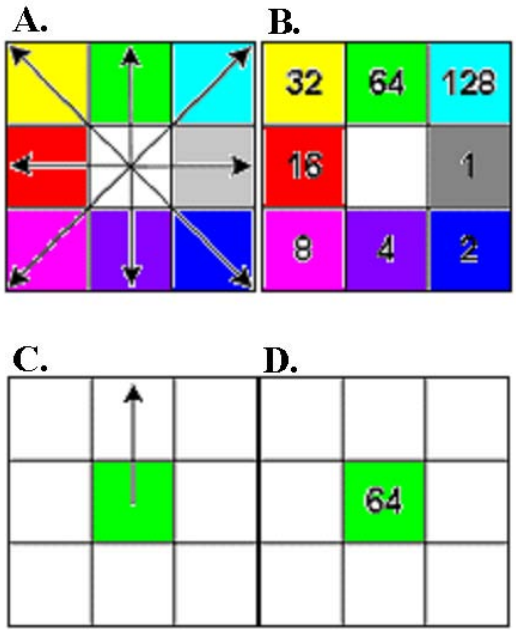


Figure 2.3 - Flow direction example A) Demonstrates the eight directions of hydrologic flow from center cell; B) contains the direction coding system for the eight directions; C) Demonstrates the direction of flow from the center cell if flow direction is north; D) The coding scheme if the flow direction is north.

The secondary terrain analyses used for this study is the Topographic Wetness Index (TWI) (Sorensen et al. 2006) which correlates with areas of greater soil moisture and shallow groundwater levels (Equation 2.3) where α is upslope area and β is slope (degrees). Upslope area (α) was calculated by generating a flow direction raster (Figure 2.3) form the DEM and accumulating the number of hydrologicaly connected cells (pixels) down slope. The TWI equation produces small values for topographic positions that are steep with small upslope areas (sources) and large values for topographic positions that have large upslope areas with shallow slopes (sinks).

$$TWI = \ln\left(\frac{\alpha}{\tan\beta}\right) \quad \text{(Equation 2.3)}$$

2.3 - Selecting Intensive Sampling Units

To capture soil variability at a scale less than 10 hectares for the entire FEF intensive sample units were confined to sub-watersheds that make up the upper St. Louis Creek watershed. Selection of intensive sampling units began by identifying 11 sub-watersheds that capture substantial terrain, vegetation, and physiographic variability in FEF. The 11 sub-watersheds indentified (Figure 2.4) are on average 436 hectares account for 56% of FEF area and are made up mostly of Spruce-fir with, Lodgepole pine, and Alpine meadow cover types. Lexen Creek watershed was selected to provide detailed soil information for current and future research. Iron Creek was selected as an intensive sampling due to its size, accessibility, composition of existing soil map units (Retzer 1962) and land cover types, as well as its fit with Lexen Creek in representing FEF. Iron Creek was selected over Range and East St. Louis Creeks because it is smaller than East St. Louis Creek and its land cover is proportionally more representative of FEF (Table 2.4). This allowed for the land cover types to be sampled more intensively and equally based on their respective proportions. This is especially true for alpine meadow, lodgepole pine and spruce-fir land cover types, which are the top thee land cover types in FEF.

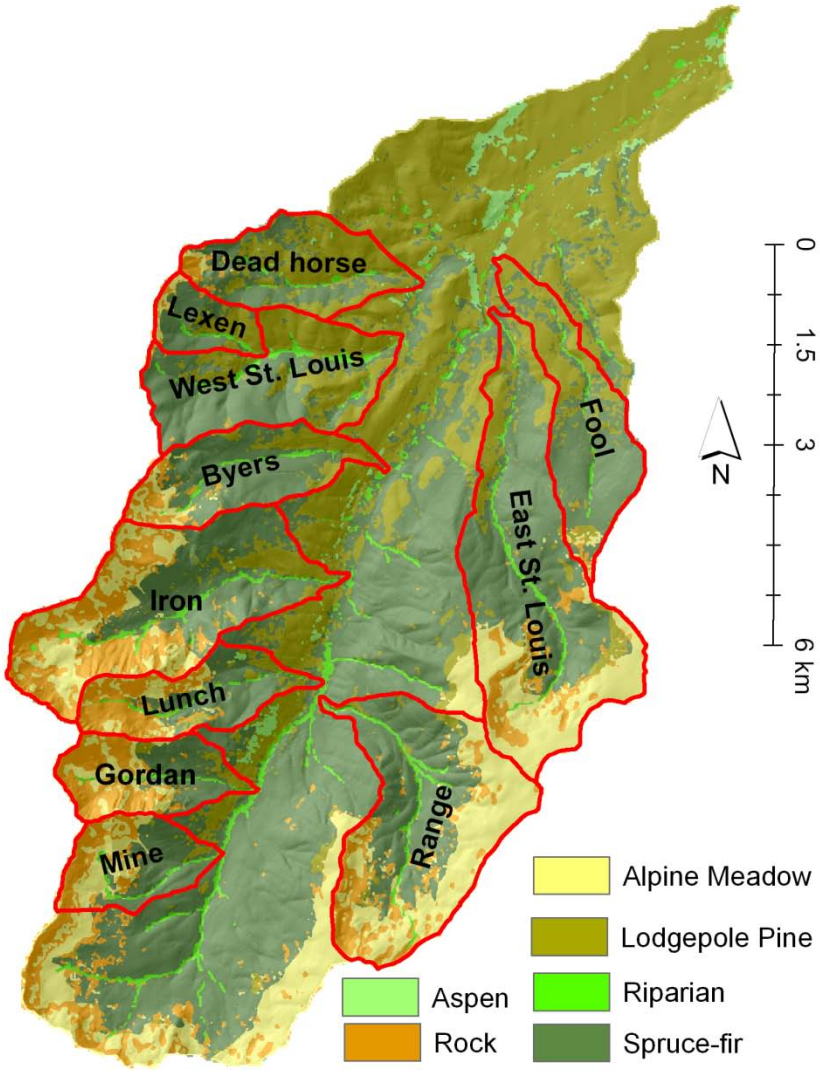


Figure 2.4 – Sub-watersheds involved in intensive sample unit selection with land cover from the Colorado Vegetation Model (CVM) database (Theobald et. al 2004)

Iron Creek contains 70% of the map units delineated in FEF which is between East St. Louis and Range Creeks with East St. Louis having the highest proportion (Table 2.4). Iron Creek's topographic characteristics capture 82 percent of elevation and 97 percent of slope variability of FEF by having an elevation gain of 1028 meters and a slope range of 58.3 degrees (Table 2.5). Iron Creek is more accessible than East St. Louis and Range by having

shorter distances between sample points than East St. Louis Creek, and is closer to FEF headquarters and passable roads than Range Creek.

Table 2.4 - Vegetation and Soil survey map unit summary statistics for 11 sub-watersheds involved in intensive sampling unit selection.

Watershed	Alpine Meadow (%)	Aspen (%)	Lodgepole Pine (%)	Riparian (%)	Rock (%)	Spruce-fir (%)	Soil units (%)
Byers	7.7	0.2	18.3	3.3	10.5	60.0	61.0
Dead horse	1.6	0.7	53.3	2.9	4.2	37.9	43.5
East St. Louis	20.0	0.1	13.9	2.9	5.4	57.7	91.3
Fool	2.0	0.3	32.8	2.7	2.9	59.3	39.1
Gordan	18.4	0.0	6.8	3.1	38.7	33.0	52.2
Iron	**21.6**	**0.1**	**19.2**	**3.2**	**17.8**	**38.1**	**69.6**
Lexen	**6.5**	**0.4**	**31.3**	**4.2**	**6.6**	**51.0**	**17.4**
Lunch	13.0	0.6	9.8	3.7	31.6	41.3	52.2
Mine	25.1	0.0	3.0	4.3	23.9	43.7	65.2
Range	34.4	0.0	2.7	3.9	16.5	42.9	55.6
West St. Louis	0.6	0.0	35.6	4.0	0.4	59.4	25.2
Iron and Lexen	**16.7**	**0.2**	**24.7**	**3.5**	**13.7**	**41.2**	**75.2**
FEF	**14.2**	**1.4**	**28.2**	**0.1**	**11.7**	**44.4**	**100.0**

The number of sample sites to sample in Lexen and Iron Creek was determined to be 137, which provides an overall sampling intensity of 7.4 hectares per sample. Lexen Creek was allocated 37 sample points to provide detailed soil characteristics (6 hectares) for research purposes. Iron Creek was allocated 100 sample points to meet sampling time constraints, the sampling density threshold, and statistical model requirements.

Table 2.5 - Topographic and physical characteristics of the 11 sub-watersheds involved in the intensive sampling unit selection.

		Elevation (meters)			Slope (degrees)		
Watershed	**Area(ha)**	**Min.**	**Mean**	**Max.**	**Min.**	**Mean**	**Max.**
Byers	356	2851	3289	3781	0.6	20.3	47.1
Dead horse	356	2800	3095	3526	0.1	19.5	49.8
East St. Louis	975	2787	3343	3847	0.1	17.7	48.9
Fool	402	2763	3117	3503	0.4	12.9	37.0
Gordan	275	3017	3457	3856	1.8	26.1	54.5
Iron	**783**	**2876**	**3414**	**3904**	**0.0**	**25.2**	**58.3**
Lexen	**136**	**2961**	**3256**	**3527**	**5.0**	**21.5**	**42.6**
Lunch	297	2913	3400	3861	0.2	23.9	55.4
Mine	250	3065	3436	3784	0.0	22.9	57.3
Range	774	2922	3502	3889	0.1	21.1	50.2
West St. Louis	498	2835	3234	3524	0.0	18.3	49.7
FEF	**9078**	**2646**	**3312**	**3904**	**0.0**	**18.2**	**60.0**

2.4 - Development of the sampling design

The sampling design is based on a relatively new technique called Spatially-Balanced Sampling (SBS) (Stevens and Olsen (2004), which is a probabilistic-based approach that generates points that are spatially regular and can be parameterized to optimize sampling efforts. The SBS design for Lexen and Iron Creek watersheds was developed using the Reversed Randomized Quadrant-Recursive Raster (RRQRR) algorithm (Theobald et al. (2007); ArcGIS 9.2 Toolbox, ESRI 2007). To optimize sampling efforts an inclusion probability surface (raster map) (Theobald et al. 2007) was developed to capture soil development variability and reduce sampling travel time. This was accomplished using terrain, remote sensing, and travel time analysis that captures *hydrologic, biotic*, and *human movement* processes. The terrain analysis consisted of implementing TWI (Equation 2.3),

which takes into account the hill slope area above a given location and the local slope. The remote sensing analysis relied on NDVI (Equation 2.1) which estimates plant vigor and density. The combination of TWI and NDVI is a proxy for soil development. This assumes that areas with high above ground biomass that are lower in the drainage are more productive thus having a higher degree of soil development. The human movement analysis entailed generating an accessibility (average one way travel time (hr)) surface which encompasses roads, trails, slope, and vegetation to from the headquarters of the FEF. This was calculated using a land cover enhanced Tobler's travel time function (Equation 2.4; Tobler 1961):

$$\omega = 6^{\exp(-3.5 \bullet |\tan(\beta)+0.05|)} + \nu \qquad \text{(Equation 2.4)}$$

where ω is travel velocity (km/hr), β is slope in degrees and ν is a land cover residence factor. These three surfaces (NDVI, TWI, travel time) were normalized between 0 (least desirable) and 1 (most desirable) and then averaged. The averaged inclusion probability surface is based on a weighting scheme that weights NDVI and TWI with 0.45 each and travel time 0.1. Giving NDVI and TWI priority over travel time in the placement of sample locations, but travel time does have an underlying effect (Figure 2.5).

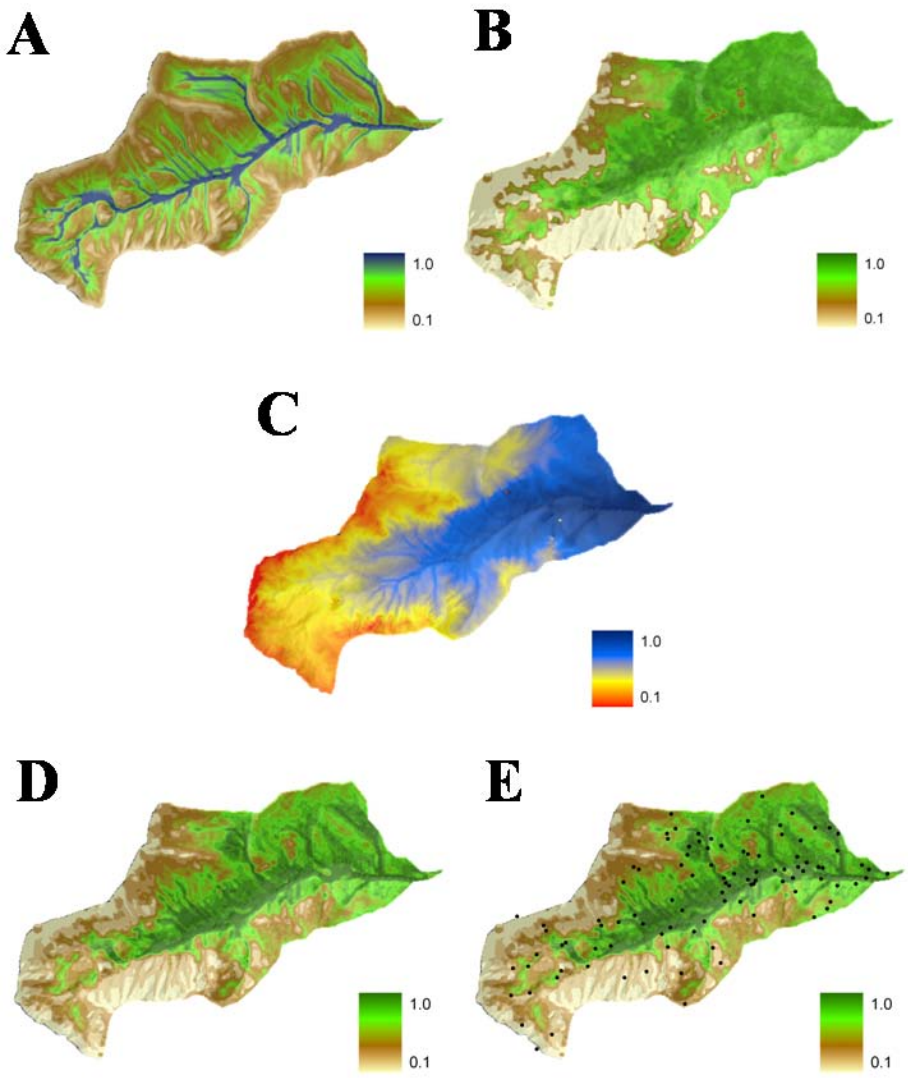


Figure 2.5 - Surfaces used to develop the inclusion probability surface for Iron creek: A) Normalized TWI surface; B) Normalized NDVI surface; C) Normalized travel time surface; D) Final inclusion probability surface; E) Inclusion probability surface with 100 RRQRR sample points

Calculating the 137 sample site x,y coordinates for Lexen and Iron Creek watersheds with the RRQRR algorithm required three basic steps. The first step was to generate a "Sequence Raster" for Lexen and Iron Creeks using the RRQRR tool Generate Sequence Raster, which gives every raster cell a unique spatially balanced address (Theobald et. al 2007). The second step is to filter the "Sequence Raster" against the inclusion probability surface to alter sequenced values based on the inclusion probability using the RRQRR tool

Filter Sequence Raster. The third step is to extract the 137 sample points from the filtered sequence raster using the RRQRR tool Extract Sample Points.

2.5 - Spatially Balanced Sampling Efficiency Simulation (SBSES)

Theobald (2007) tested the RRQRR algorithm for spatial efficiency (configuration) using the Efficiency Ratio (ER) metric (Stevens and Olsen 2004) and found that it produces point patterns that are spatially balanced (ER < 0.8) and that the proportion of samples that fall within the different inclusion probability zones "mimics" the inclusion probability surface. The objective of the SBSES is to evaluate the spatial, statistical and economic efficiencies of sample points generated by the RRQRR algorithm in conjunction with inclusion probability surface developed to sample Lexen and Iron Creek watersheds. The SBSES efficiency categories (Table 2.7) focus on metrics commonly used to quantify statistical model performance (Statistical efficiency), test the spatial configuration of a sample design (Spatial efficiency), and placement of points on the landscape (Economic efficiency). The statistical efficiency metrics test how well a sampling design captures variability by producing models that are accurate (R^2), precise (MSE), and are not influenced by spatial autocorrelation (Moran's I). The Spatial efficiency metric (ER) evaluates the spatial dispersion of sample points within FEF. A sample design that positions sample points that are evenly dispersed is considered to be more spatially efficient (Stevens and Olsen 2004). Evaluating the economic efficiency focuses on the placement of the sample points in relation to topography and travel time from FEF headquarters.

At the heart of the SBSES analysis is a known soil property that is continuous and complete for FEF to sample, model and evaluate performance. Since such a spatial dataset doesn't exist, NDVI was used because it is continuous (30 x 30 meter resolution) and complete for FEF and it is highly correlated with soil depth (Figure 2.6).

The SBSES involves producing 137 SBS and Simple Random Sampling (SRS) points, extracting dependant (NDVI), independent, and economic spatial information for the points, developing a multiple regression linear model (LM), producing a NDVI prediction surface from the LM, and populating a database with the model efficiency metrics (Figure 2.7 and APPENDIX 3). These five basic steps where repeated 1000 times for SBS and SRS similar to Theobald (2007). This was accomplished with a Python script implemented within ArcGIS (ESRI 2007) that utilizes GIS and R statistical language operations (R Development Core Team 2006).

The SBSES inputs are based on spatial information used for this project to sample soils and build geostatistical soil attribute models. The spatial information parameters for the simulation are broken into four types: 1.) Sampling frame 2.) Dependant surface 3.) Independent surfaces 4.) Economic cost surface. The sampling frame inputs are the RRQRR sequence raster (Theobald et al., 2007) and the inclusion probability raster, developed for FEF. These two surfaces also provide the sampling frame to calculate random x,y locations for the SRS iterations. The dependant surface is NDVI to mimic soil depth and is used to calculate Mean Squared Error (MSE). The independent surface inputs consist of many surfaces that are used to predict NDVI (Table 2.5). The economic cost input is the travel time (hr) from FEF headquarters surface, which is used to evaluate economic efficiency of sample point placement.

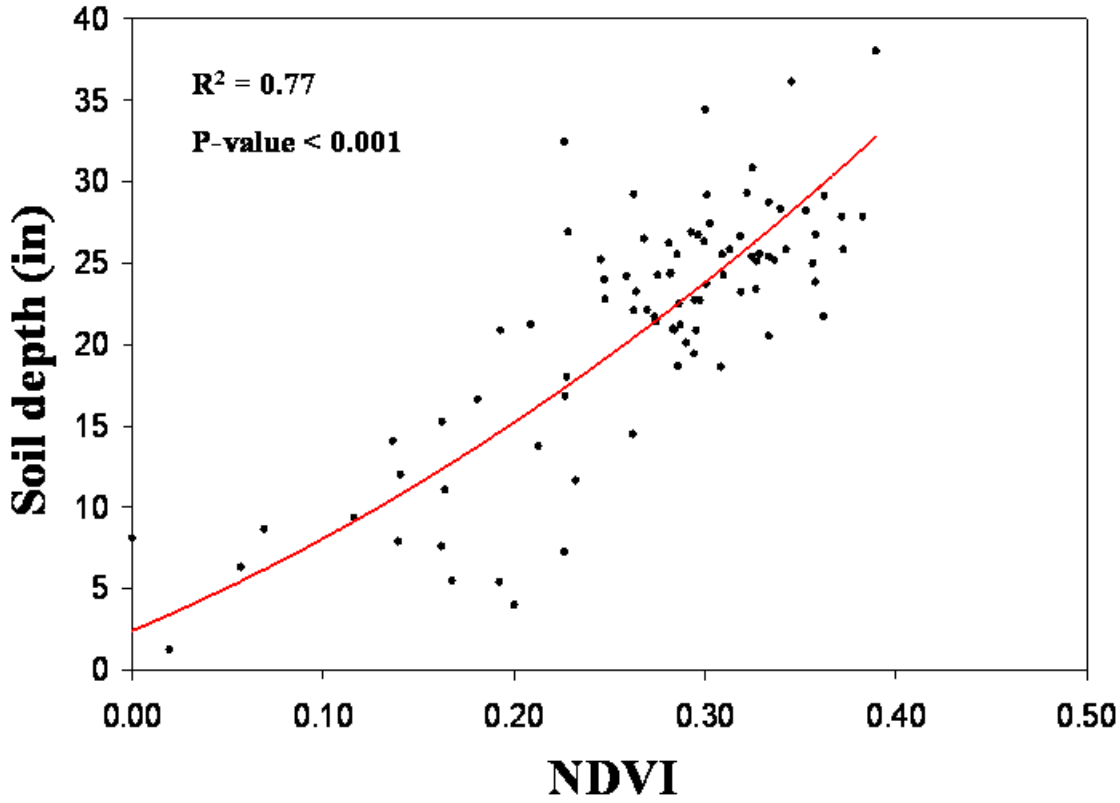


Figure 2.6 - Correlation between soil depth and NDVI based on 137 soil samples collected in Iron and Lexen Creek watersheds.

Table 2.6 - Spatial variables used in the simulation

Variable	Units
Elevation	Meters
Slope	Degrees
Solar insulation	Index value (0-1)
TWI	Index value (0 -18)
LS 7 ETM+ Band 1	Brightness (0-255)
LS 7 ETM+ Band 2	Brightness (0-255)
LS 7 ETM+ Band 5	Brightness (0-255)
LS 7 ETM+ Band 7	Brightness (0-255)

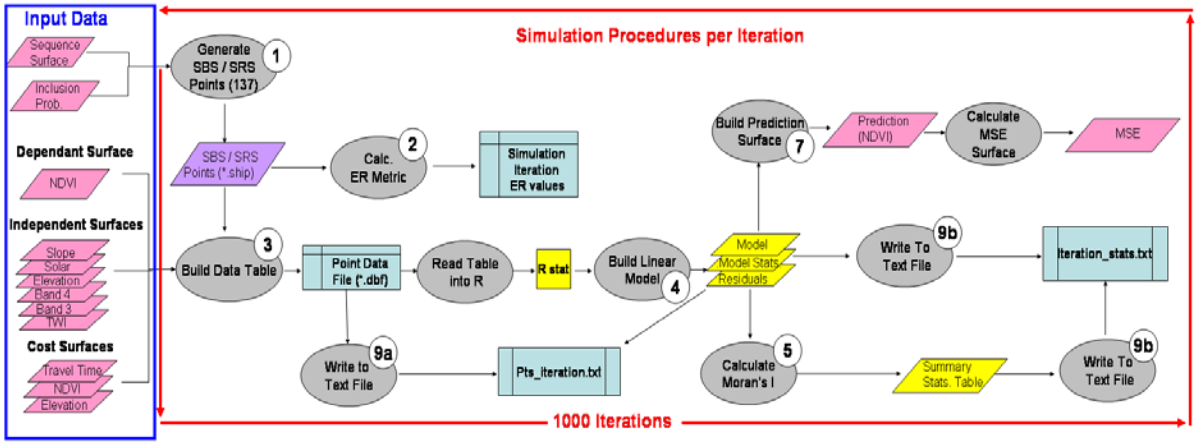


Figure 2.7 - Simulation inputs, procedures and outputs

The simulation generated two ASCII files (pts_iteration.txt and iteration_stats.txt) that capture the spatial configuration and attributes of the points, as well as iteration summary statistics for SBS and SRS. These two files were used to generate the efficiency metrics (Table 2.7) to evaluate SBS performance. To evaluate SBS performance a single ASCII file was compiled that contains the efficiency metrics (Table 2.7) for SBS and SRS, which was necessary for the post simulation analysis.

Table 2.7 - Simulation efficiency type and metric involved in the post simulation analysis

Efficiency category	Metric
Statistical Efficiency	Adjusted R^2
Statistical Efficiency	Mean Squared Error
Statistical Efficiency	Moran's I (p-value)
Statistical Efficiency	Modified Moran's I
Economic Efficiency	Average travel time (hrs)
Economic Efficiency	Average elevation (m)
Spatial Efficiency	Efficiency Ratio (ER)

The post simulation analysis utilized the data file compiled from the SBS and SRS simulations to evaluate efficiency differences and test if the differences are significant. Excluding the ER and Moran's I p-value metrics, efficiency differences were evaluated based on average, median and standard deviation statistics, percent difference between SBS and

SRS averages, and testing if the SBS and SRS efficiency metric distributions are significantly different. To test significance, the nonparametric Kolmogtov-Smironov test (Davis 1986) was used because of the large population size (1000) and it focuses on the distribution of the two populations instead of their means. The ER metric is a ratio based on Voronoi polygon area variances formed by the SBS and SRS designs across all simulations (Stevens and Olsen 2004). If ER <1.0, then the SBS design is more spatially efficient than SRS. The Moran's I p-value analysis compares the proportion of SBS and SRS LM models that had a p-value less than or equal to 0.05. The sampling design cost efficiency analysis also involves evaluating spatial variance captured to travel time. This analysis is a ratio between SRS and SBS with the numerator and denominator being the product of average travel time to sample points and MSE. If the ratio is >1.0, then the SBS is more cost efficient, because it is capturing more spatial variability in a more cost effective manner.

2.6 - Sampling soils in the field

The ability to accurately navigate to predetermined sample sites in the field is essential to tie geographical information with field-based observations. Sample sites were located to within three meters using a Geographical Positioning System unit (GPS) (Garmin GPSMAP 60CSx) that was Wide Area Satellite System (WASS) enabled using a Universal Transverse Mercator WGS84, Zone 13 coordinate system. Navigation to the sample sites was executed on foot with some sites taking up to 5 hours to reach, which required developing a sampling order strategy before each sampling trip to optimize sample collection and minimize hiking time. This was done using a 1:24,000 Digital Raster Graphic (DRG) in

conjunction with the travel time surface to manually define sampling path to optimize the number of samples collected each outing. Figure 2.8 is an example of how the DRG and the travel time raster were used to define sampling paths.

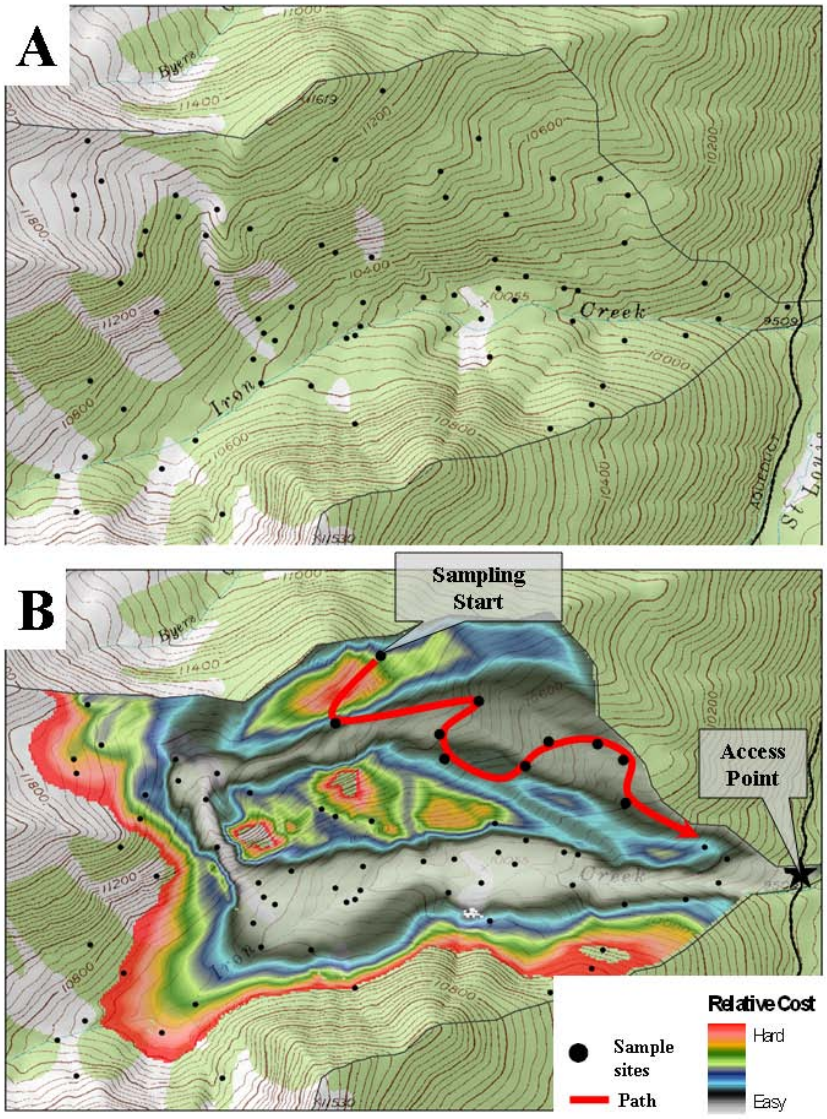


Figure 2.8 - Example of sample site path selection in lower Iron creek: A) Demonstrates sample points over laid on a DRG; B) Shows how the use of the accessibility surface in conjunction with the sample points and DRG to enhance sampling loop selection.

To reduce sampling time, field methods were made up of seven simple produces: 1.) Locate sample point using a (GPS), 2.) Dig a soil pit to the C or R horizon, 3.) Define genetic

horizons , 4.) Make field based observations, 5.) Complete field based sampling sheet (see Appendix B), 6.) Collect soil samples for every horizon, 7.) Take photographs of the soil pit and surrounding terrain and vegetation. Post sampling procedures involved updating the GIS database for the sampled points with field based observations, air drying collected horizon samples and storing for future use.

2.7 - Developing Soil Attribute Spatial Models

Environmental correlation (geostatistical methods) have a form similar to Jenny's (1941) functional factorial model (Equation 2.5) with factor interactions approximated using terrain and remote sensing techniques (Equation 2.6):

$$S = f(cl, o, r, p, t, \ldots) \quad \text{(Equation 2.5)}$$

$$S_i = f(cl_i, o_i, r_i, hy_i, k_i \ldots k_j) \quad \text{(Equation 2.6)}$$

where for each site, S_i is the soil property observed in the field. The explanatory factors cl_i, o_i, r_i and hy_i in (Equation 2.5) are geospatial representations (Table 2.8) that approximate climate, organisms, topographic dynamics and hydrologic processes that influence a soil property at observation S_i, with ($k_i \ldots k_j$) representing other miscellaneous site specific environmental predictors that may be available for a survey area (e.g., Lidar, geology, high resolution multi-spectral imagery, etc.).

Table 2.8 - Soil attributes modeled for FEF

Soil Property (S_i)	Units
Total soil depth	Inches
A-horizon Thickness	Inches
O-horizon Thickness	Inches

A range of statistical analyses can be applied to develop models for spatial prediction using environmental correlation. These include Bayesian rule-based systems (Cook et al., 1996; Skidmorre et al., 1996), neural nets, fuzzy logic (Xhu et al., 1997), generalized linear models (McKenzie and Austin, 1993; Gessler et al., 1995), tree-based methods and co-kriging (Odeh et al., 1994). A thorough analysis of the advantages of different strategies for environmental correlation has yet to be done in soil survey although Austin et al. (1995) have undertaken such a study for vegetation prediction. They concluded that a combination of generalized linear models and generalized additive models was superior to tree-based procedures but all were acceptable for practical applications.

In an attempt to streamline the soil property modeling process, a generalized linear model (GLM) was chosen to model large scale variability and a tree-based method to model small scale variability. These methods were chosen because they are spatially robust to noisy conditional relationships that are common is soil sample data, they do not require specialized skills to parameterize, and they produce results that are easy to interpret and implement using spatial data.

Table 2.9 - Independent spatial covariates used for soil property spatial models

Independent variable	Units
Local Slope	Degrees (0 - 90)
TWI	Index (-10 –10)
Curvature	Index (-13 – 12)
Solar Insulation	Index (0 – 1)
Wetness	Index (-5 – 167)
Brightness	Index (130 – 373)
Greeness	Index (-72 – 20)
NDVI	Index (-10 – 10)
MSVAI	Index (-236 – 16)
TM Band 7	Index (0 - 255)

The GLM / tree-based modeling procedures (Figure 2.9) entail GLM prediction model to capture large scale soil variability. The GLM residuals are modeled using a tree-based method to capture small scale error not captured by the GLM. The two surfaces GLM and tree-based are added together to get the final model.

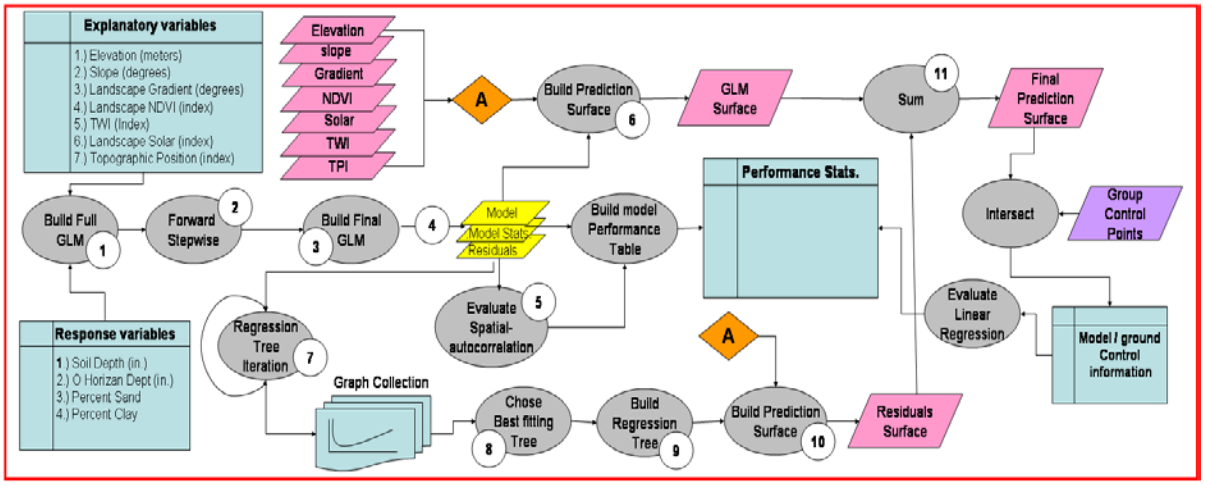


Figure 2.9 - Soil property modeling procedures

The GLM is used to model large scale variability because it is effective at handing variables that are continuous (e.g., elevation) or discrete (e.g., land cover) and that are not normally distributed. As a consequence, predictions are more realistic because they portray soil variation as being either gradual or discontinuous. The residuals (model errors) from the GLM prediction surface analysis are modeled using a regression tree method, which accounts for non-parametric relationships by successively splitting the data into increasingly homogenous groups.

The GLM soil property modeling process was executed in R (R Development Core Team 2006) requiring six basic steps:

1) Generate a full GLM soil property (Table 2.8) regression model involves all independent variables (Table 2.6).

2) Perform a backward stepwise selection procedure, *stepAIC MASS* Package (Ripely 2008) to identify significant predictor covatiates optimizing on Akaike Information Criteria (AIC) (Akaike 1969).

3) Execute GLM model based on covariates selected by *stepAIC*.

4) Calculate summary statistics, Mean Squared Error (MSE), and extract model residuals for the final GLM.

5) Calculate partial and global Moran's I on residuals to determine and record spatial bias of model.

6) Build the GLM soil property prediction raster in ArcGIS 9.2 using spatial analysis tools *SingleoutputMapAlgebra* (ESRI 2007) equation string.

The resulting soil property GLM accounts for parametric relationships between the dependant variables (Table 2.8) and covariates (Table 2.9). Since the objective of this modeling project is to develop soil property surfaces for the entire FEF, performing spatial interpolation analysis on the residuals to account for spatial autocorrelation is not useful due to the limited spatial extent of the samples. A tree-based modeling technique was used to capture small scale variability of non-parametric relationships contained in the residuals of the GLM. The tree-based modeling of the residuals where executed in R (R Development Core Team 2006) in five steps (Figure 2.9; numbers 7 – 11):

7) Model GLM residuals using regression tree analysis.

8) Generate a 10-fold cross validation analysis on the full regression tree and select the best fitting regression tree based on the 10-fold cross validation and regression statistic.

9) Prune the best fitting tree based on 10-fold cross validation statistics.

10) Calculate model performance stats for the regression tree

11) The final step is to sum the GLM and tree-based prediction surfaces together resulting in the final soil property surface.

2.8 - Soil property model validation

Validation is one of the crucial parts of the modeling process. It can be achieved by comparing observed and predicted responses using an independent data set (Guisan et al., 1998) or by cross-validation when a dataset can be split in several subsets (Lehmann et al. 2002). It is not clear whether independent datasets are really preferable to cross-validation. Additionally, test of spatial autocorrelation of residuals also provides insight of the model's ability to capture spatial patterns independent of environmental predictors. The soil attribute models were evaluated by splitting the original data set (n= 137) into a model training dataset (80%, n=110) and a model validation dataset (20%, n=27) (split-sample approach; see Guisan and Zimmermann, 2000) Guisan and Zimmermann (2000) suggest a 70% training and 30% evaluation split. This degree of splitting would have resulted in 96 training points and 41 validation points, but due to the limited sample size, an 80/20 split was chosen. The model performance was evaluated by calculating the correlation between the training and validation points using a linear regression, as well as quantifying the spatial autocorrelation of the residuals based on both a global Moran's I p-value and Modified Moran's I.

Chapter 3

Results and Discussion

3.1 - Watershed characterization and sampling efficiency

A total of 137 soil pits were described and sampled in the Iron and Lexen Creek watersheds (Table 3.1). The Iron Creek sampling sites had an average elevation of 3133 meters with an average slope of 20.7 degrees and were on average 2.4 hours from the road. Lexen Creek's sampling sites were at a lower elevation, on steeper slopes but generally more accessible (Table 3.1). Spatially, the distribution of Lexen Creek's sample sites is 25 percent more dense with each site representing an average area of 6.1 hectares (Table 3.1). The average accessibility to sample locations from the nearest access point was 2.1 hours making it 85 percent more accessible than Iron Creek (Table 3.1).

Table 3.1 - Sample plot description summary statistics

		Total Area	**Average Area (ha)**		**Elevation (m)**		**Slope (deg.)**		**Accessibility (hr)**	
	n	**(ha)**	Mean	STD	Mean	STD	Mean	STD	Mean	STD
Iron	100	790	7.8	5.0	3313	200.6	20.7	10.7	2.4	1.5
Lexen	37	190	6.1	4.0	3229	139	21	5.4	1.3	0.86
Both	137	9.8	7.4	5.0	3293	191.3	21	9.7	2.1	0.94

The results of the field sampling are split into two components that describe soil property and sampling time variability. Soil property data include soil morphological properties. Sampling time data explores the time required to navigate, dig and collect soil

properties at each sample site providing further insight to the relative accessibility of the Lexen and Iron Creek watersheds for soil survey activities.

Table 3.2 - Soil sampling summary statistics

	n	# trips (day)	Avg. Samples / Trip	Avg. Sampling Time (min.)	Avg. Travel Time (min.)
Iron	100	11	8	20.5	73.8
Lexen	37	4	5	18.6	27.0
Both	137	15	6	19.3	57.0

Of the 137 sample sites, 118 were collected in 142 hours over three months during 2005 and 2006. The large size and low accessibility of the Iron Creek watershed resulted in a 20% longer travel time between sites and a slightly longer sampling time than Lexen Creek (Table 3.2). At a sample site, the time spent digging and describing a soil pit was 19.3 minutes on average (Table 3.2). Sampling time in Iron Creek was slightly higher than Lexen Creek due to GPS problems and also because the field sampling protocols became more time efficient as the study progressed. The post processing of field samples required 30 hours to air dry and sort soil samples, as well as data entry and data management.

Table 3.3 - Horizon thickness and total soil depth thickness summary statistics

	O-horizon		A-horizon (in)		B-horizon (in)		E-horizon (in)		Total Soil depth (in)	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD
Iron	4.3	2.2	3.3	2.1	11.2	6.5	5.3	3.3	20.8	11.8
Lexen	2.2	1.5	4.7	2.3	12.3	3.1	4.17	1.7	28.2	9.8
Both	4.1	1.9	4.3	2.1	10.7	5.1	4.6	2.7	22.6	10.4

Lexen Creek soils have thicker A and B horizons than Iron Creek soils (Table 3.3). The O-horizon in Iron Creek is almost twice as deep on average than Lexen, due to large frequent wet lands, riparian zones and spruce-fir which have the deepest O-horizons (Table

3.10). The E- horizon in Iron creek was thicker due to cooler and wetter climatic conditions based on topographic position, aspect and elevation than Lexen Creek. The total soil depth (including the C-horizon) of Lexen Creek is 28 percent greater than Iron Creek (Table 3.3) by having no exposed rock sampled (Table 3.5) and largely composed (82%) of lodgepole pine and spruce-fir stands (Table 2.4) which have the deepest soils (Table 3.10).

Table 3.4 - Horizon texture summary statistics

	Sand (%)		Silt (%)		Clay (%)		Texture Class
	Mean	STD	Mean	STD	Mean	STD	
A-horizon	53	11.7	34	9.76	12	7.8	Loam
B-horizon	46	7.3	39	6.2	15	5.3	Loam
E-horizon	70	8.1	27	5.7	2.0	2.6	Loamy Sand

The soils in Lexen and Iron Creeks are coarse textured with the majority of the mineral matter consisting of sand size particles (Table 3.4). The B-horizon is the finest textured of the all horizons sampled (Table 3.4) and these horizons were identified most frequently (Table 3.5). The O-horizon is found more frequently than the E-horizon due to the higher number of organic soils found in Iron Creek. Sample sites classified as rock outcrop occurred only in Iron Creek making up 18 percent of the samples.

Table 3.5 - Horizon frequency

	A	B	E	O	Surface Rock
Iron	67%	77%	21%	65%	18%
Lexen	67%	93%	35%	48%	0%
Both	67%	77%	21%	65%	14%

3.2 - Spatially Balanced Survey Efficiency Simulation (SBSES) results

The main aim of the SBSES is to illustrate the effectiveness of the SBS in conjunction with the inclusion probability surface developed for sampling and spatially modeling soil properties in FEF.

Table 3.6 – SBSES summary statistics

Metric	Median		Mean		STD		Percent Difference
	SBS	SRS	SBS	SRS	SBS	SRS	
Adjusted R^2	0.92	0.91	0.92	0.9	0.015	0.02	2.0
Modified Moran's I	0.21	0.25	0.21	0.26	0.025	0.12	34.0
MSE	0.004	0.07	0.004	0.072	0.023	0.032	93.0
Elevation (m)	3198	3221	3198	3221	20	25	<1.0
Travel Time (hr)	3.0	3.2	3.0	3.2	0.15	0.20	6.0

The SBSES statistical efficiency results provide evidence that SBS produces statistical models that are better fitting (Adjusted R^2), 93 percent more spatially precise (MSE) (Table 3.6) with model residuals that are less influenced by spatial autocorrelation than SRS. The average Adjusted R^2 difference between SBS and SRS is slight, but the cumulative distributions (Figure 3.1) differed significantly (two-sample Kolmogorov-Smirnov $P< 0.001$) with a maximum percent difference of 33 percent (D = 0.33) (Table 3.7). The MSE cumulative distributions for SBS and SRS (Figure 3.1) along with the Kolmogorov-Smirnov test (Table 3.7) demonstrates that SBS generates more spatially precise models. The majority of SBS and SRS models satisfy the assumption that the residuals are spatially independent with 92 percent of SBS models and 60 percent of SRS models have residuals that are not spatially correlated (Moran's I $P>0.05$, Figure 3.2). The Modified Moran's I values show that the SBS model residuals are 34 percent less spatially correlated than SRS (Table 3.6), with a cumulative distribution (Figure 3.2) that is 43 percent

different (D=0.43) from SRS and are significantly different (two-sample Kologorov-Smiron test P<0.001, Table 3.7).

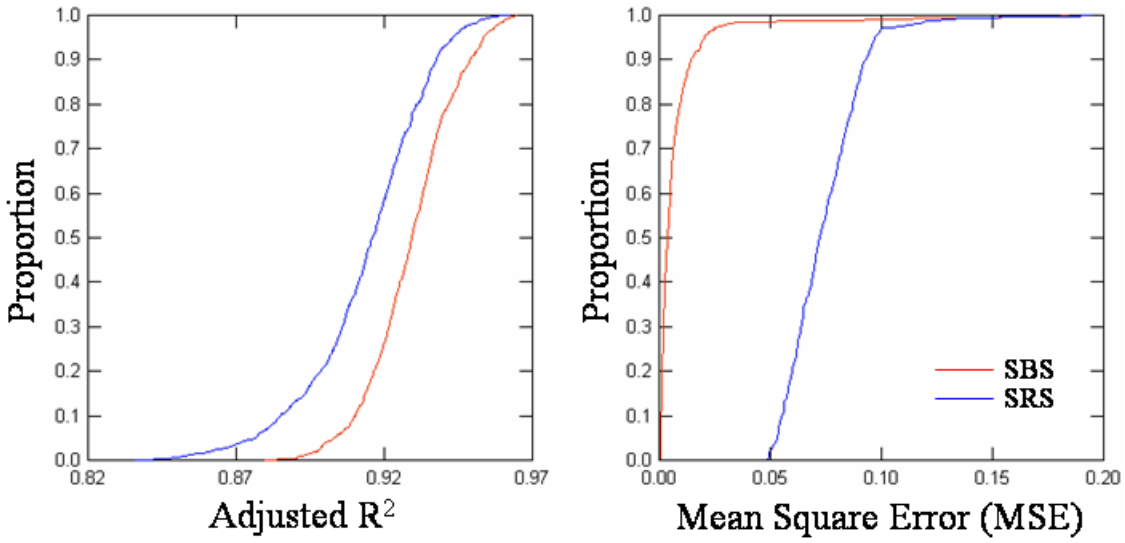


Figure 3.1 SBSES Statistical efficiency metrics Cumulative Distribution Frequency analysis graphs for Adjusted R^2 and MSE

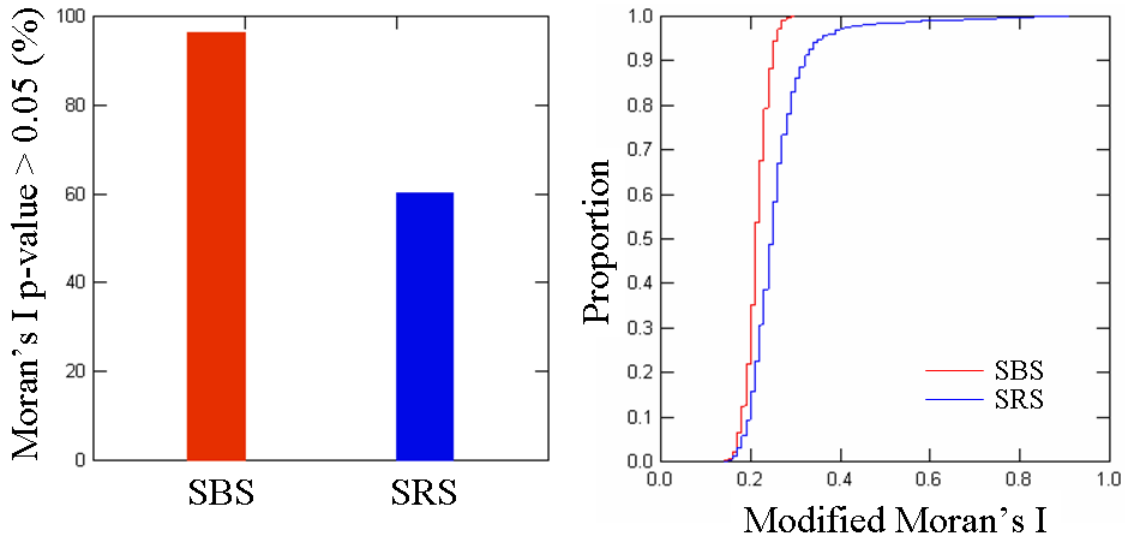


Figure 3.2 SBSES Statistical efficiency metrics Cumulative Distribution Frequency analysis graphs for Moran's I p-values less than 0.05 and Modified Moran's I

The spatial efficiency analysis demonstrates that SBS produced a sampling design that is 28 percent more spatially efficient than SRS. The distribution of ER values for all

1000 (Figure 3.3) iterations is cented at 0.72 with a minimum value of 0.644 and a maximum value of 0.75.

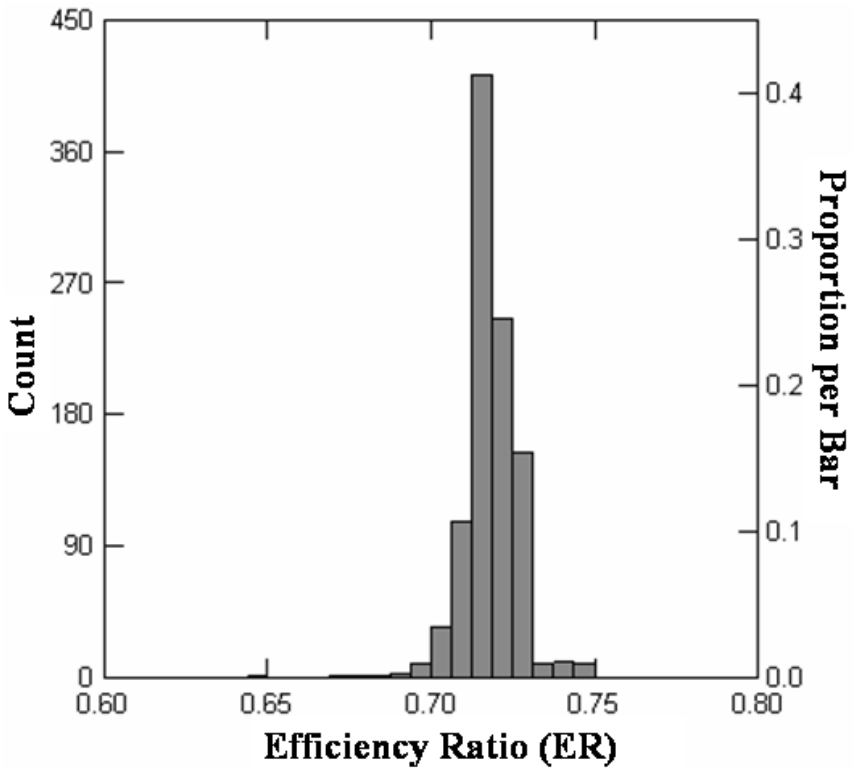


Figure 3.3 - SBSES Efficiency Ratio histogram with a mean of 0.72 and a standard deviation of 0.008

The cost efficiency results show that SBS and SRS sample sites are on average three hours from FEF headquarters with SBS points being 12 minutes closer to FEF headquarters and are located 23 meters lower in elevation than SRS (Table 3.6). The Elevation and Travel time cumulative distributions (Figure 3.4) differed significantly (two-sample Kolmogorov-Smirnov test $P<0.001$) with a maximum percent differences of 25 percent (D= 0.25) and 36 percent (D= 0.36) respectively (Table 3.7). The cost to variance ratio between SBS and SRS is 19.2 (>1.0) indicating that SBS captures more variability with less travel time.

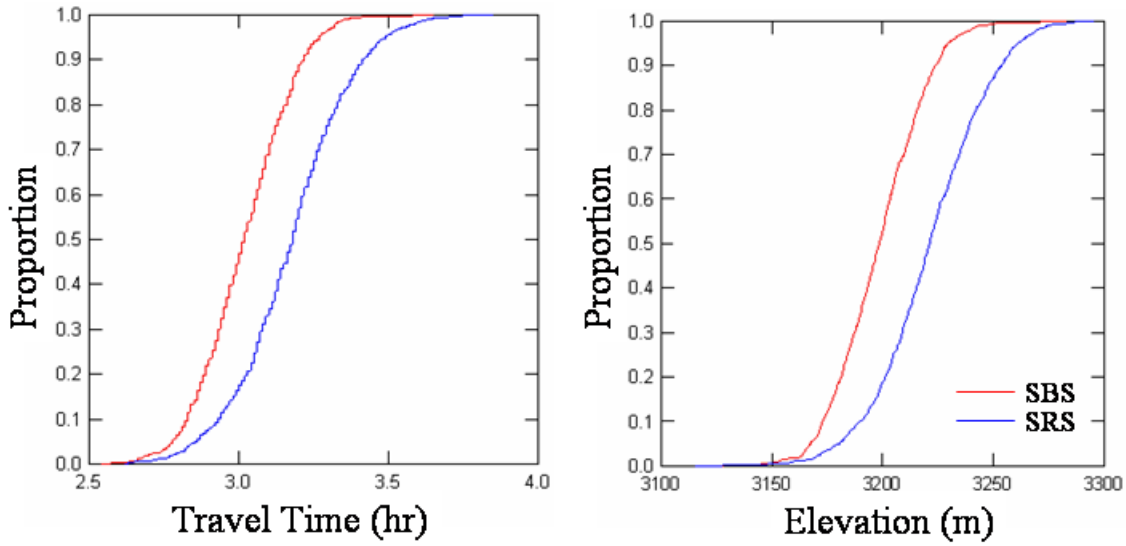


Figure 3.4 - SBSES Cost efficiency metrics Cumulative Distribution Frequency graphs for average one way travel time form FEF headquarters and elevation.

Table 3.7 – SBSES two-sample Kolmogorov-Smirnov test results on efficiency metrics

Performance Metric	D	p-value
Adjusted R^2	0.33	<0.001
MSE	0.98	<0.001
Modified Moan's I	0.42	<0.001
Elevation	0.25	<0.001
Travel Time	0.36	<0.001

3.3 - Soil property spatial models

The regression tree analysis on the Generalized Linear Model (GLM) residuals was investigated but not utilized, due to the strong performance of the GLMs and the added model complexity with very little added model performance. The multi-model methodology is a useful method in generating robust spatial models for a wide range of applications, but was not necessary for this project.

The GLM equations for the soil property models, namely total soil depth, A and O horizon thickness derived from stepwise regression procedures, are shown in Table 3.8. Spectral indices explained the majority of variability across all models with the Tassel Cap Transformation indices (Brightness, Greeness, and Wetness) being incorporated into all three models. Of the six terrain variables developed, three (curvature, TWI, slope) were selected as significant predictors of total soil depth and horizon thickness models. These soil attribute models (Table 3.8) were developed from the 110 selected training sample sites in Iron Creek and Lexen Creek watersheds and were extrapolated to the entire geographical boundaries of FEF.

Table 3.8 - Soil GLM covariates with coefficient and p-value

Spatial Model	Variable	Coefficient	p-value
Soil Depth Model	Intercept	31.4125	< 0.001
	Brightness	0.9288	< 0.001
	Curvature	-2.55842	< 0.001
	Greenness	0.4894	0.002
	TM Band 7	0.69227	<0.001
	MSVI	1.40585	< 0.001
A Horizon Depth Model	Intercept	-36.32	> 0.05
	Wetness	12.33	0.001
	Greenness	2.592	0.002
	NDVI	9.664	<0.001
	TWI	8.930	<0.001
O Horizon Depth Model	Intercept	-12.35	<0.001
	Curvature	-25.91	<0.001
	NDVI	18.53	<0.001
	Greenness	-0.3262	<0.001
	TWI	1.61	<0.001
	Wetness	0.243	<0.001
	Slope	25.621	<0.001

The total soil depth spatial model (Figure 3.5) is made up of spectral indices that are positively correlated and curvature which is negatively correlated with soil depth (Table 3.8). The positive correlation of the spectral indices indicates that areas with higher above ground

biomass have deeper more developed soils. The negatively correlation of curvature indicates that areas that are concave (curvature < 0) have deeper soils and convex areas (curvature > 0) have shallower soils. The performance of the soil depth model is the most robust of the three models based on the following: 1) a high R^2 (Table 3.9); 2) no spatial dependence within the residuals (Table 3.9); and 3) normally distributed residuals (Figure 3.2 A) that are random compared with the prediction values (Figure 3.2 B). The plot of predicted verses observed (Figure 3.4 C) shows a strong linear trend indicating that the form of the model is valid and that model covariance is constant.

Table 3.9 - Soil spatial model performance summary statistics

Spatial Model	R^2 Model	p-value Model	R^2 Validation	Moran's I	p-value Moran's I	Modified Moran's I
Solum Thickness	0.88	< 0.001	0.91	-.001	0.67	0.06
A-Horizon Thickness	0.86	<0.001	0.82	-0.051	0.091	0.18
O-horizon Thickness	0.93	<0.001	0.85	-0.0098	0.91	0.10

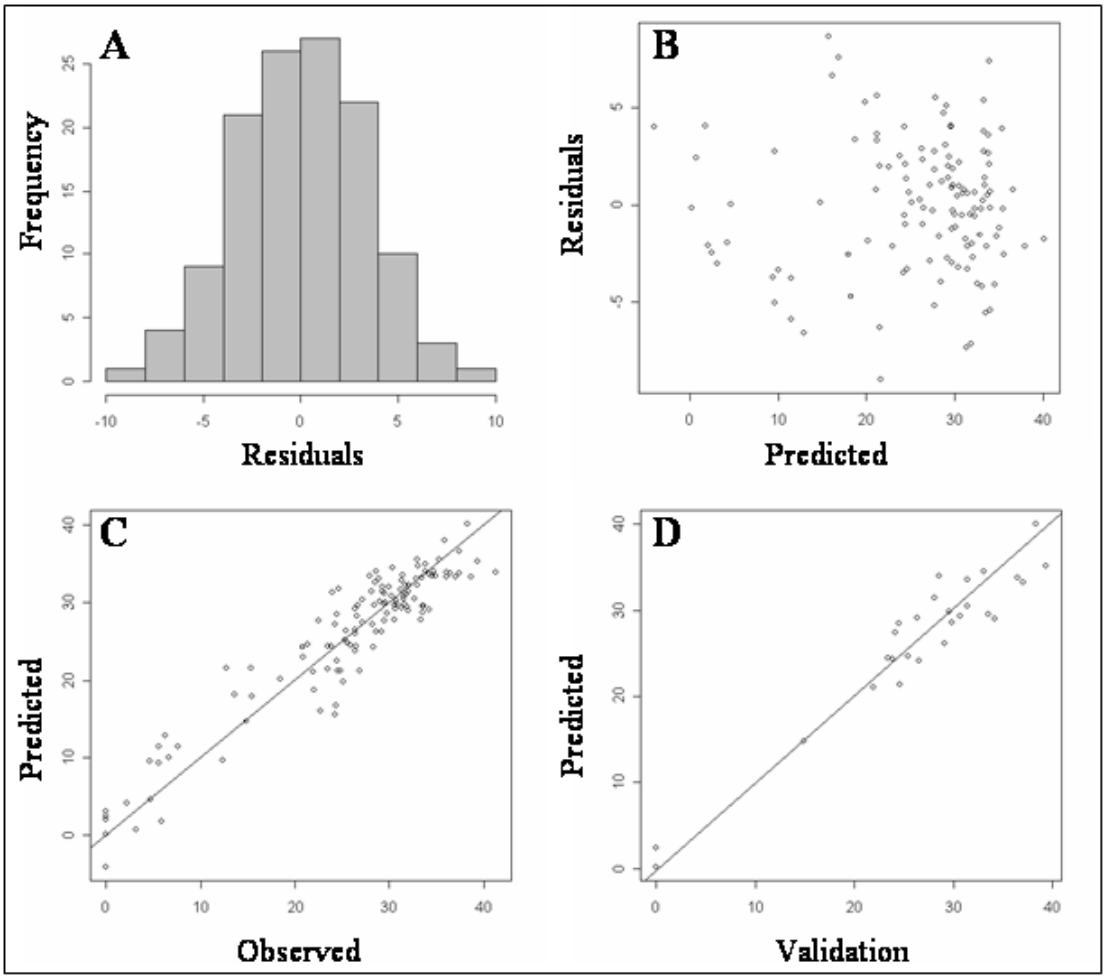


Figure 3.5 – Total soil depth spatial model performance plots: A) Histogram of model residuals; B) Scatter plot between model residuals and predicted values; C) Scatter plot with trend line of model predicted and observed values; D) Scatter plot with trend line between predicted values and validation values

As with total soil depth, spectral indices explain the majority of the variability in A-horizon thickness spatial model (Figure 3.6). The variable TWI (Topographic Wetness Index) has a significantly positive correlation with correlated with A-horizon thickness suggesting that wet areas have thicker A-horizons. The performance of the A-horizon depth model is not as strong as the soil depth or O horizon thickness models; however, it still explains 86 percent of the variability (Table 3.9) for the training data. The residuals are normally distributed (Figure 3.3 A) and the predicted vs. observed scatter plot shows very little trend (Figure 3.3 B). The spatial dependence of the residuals are weakly correlated

(Table 3.9) with a modified Moran's I of 0.2 and a Moran's I p-value > 0.005. As with the total soil depth model, the form of the model is valid with constant covariance (Figure 3.3 C)

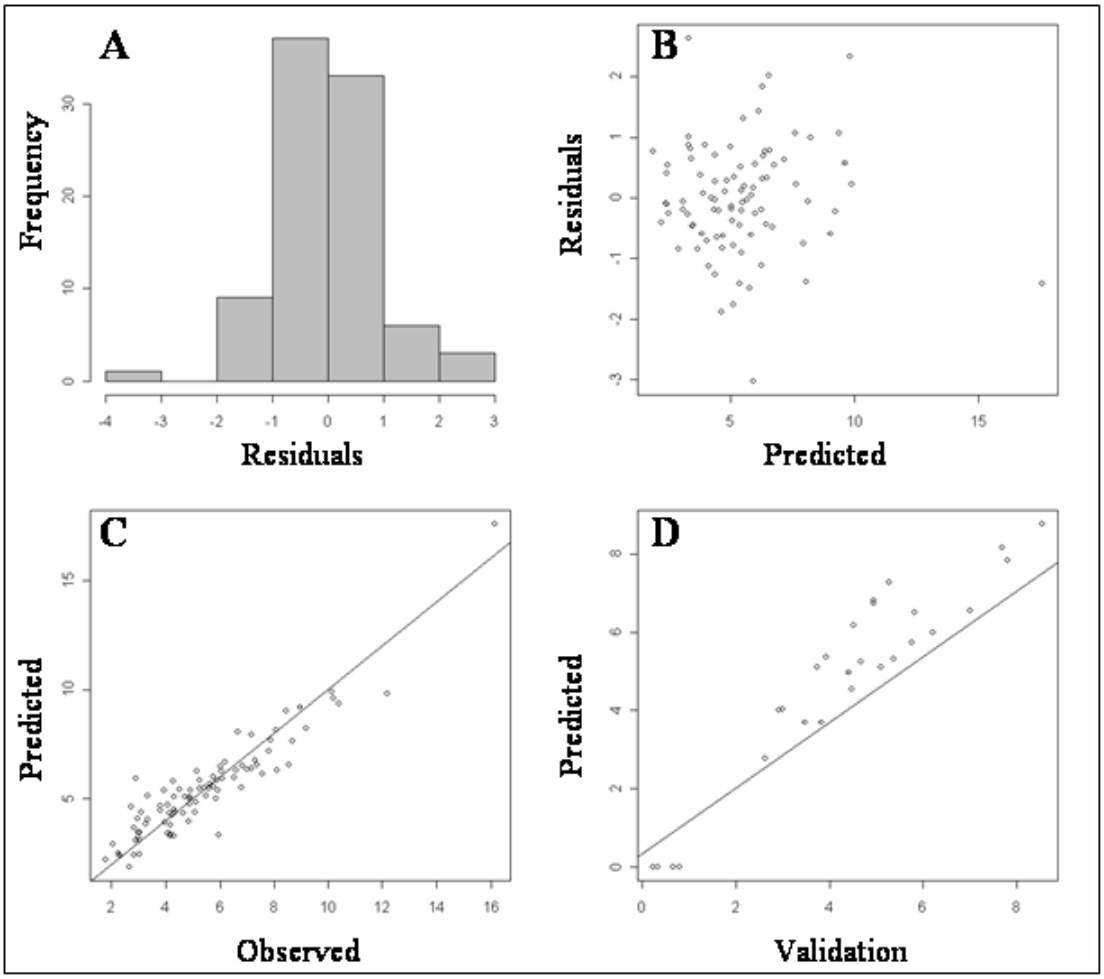


Figure 3.6 – A-horizon thickness spatial model performance plots: A) Histogram of model residuals; B) Scatter plot between model residuals and predicted values; C) Scatter plot with trend line of model predicted and observed values; D) Scatter plot with trend line between predicted values and validation values

The O-horizon thickness spatial model (Figure 3.7) is a strong fitting model with all six covariates including the intercept being significantly correlated. The covariates that make up the model are split between spectral indices that measure vegetation vigor and moisture

content and terrain covariates that capture terrain shape and hydrologic position (Table 3.8). The O-horizon thickness model accounted for 93 percent of the variability with the residuals being tightly clustered between 1 and -1 (Figure 3.4 A and B) with a mean value of 0.17. They are also not significantly spatially correlated with a Moran's I p-value of 0.91 (Table 3.9) and a Modified Moran's I value of 0.1.

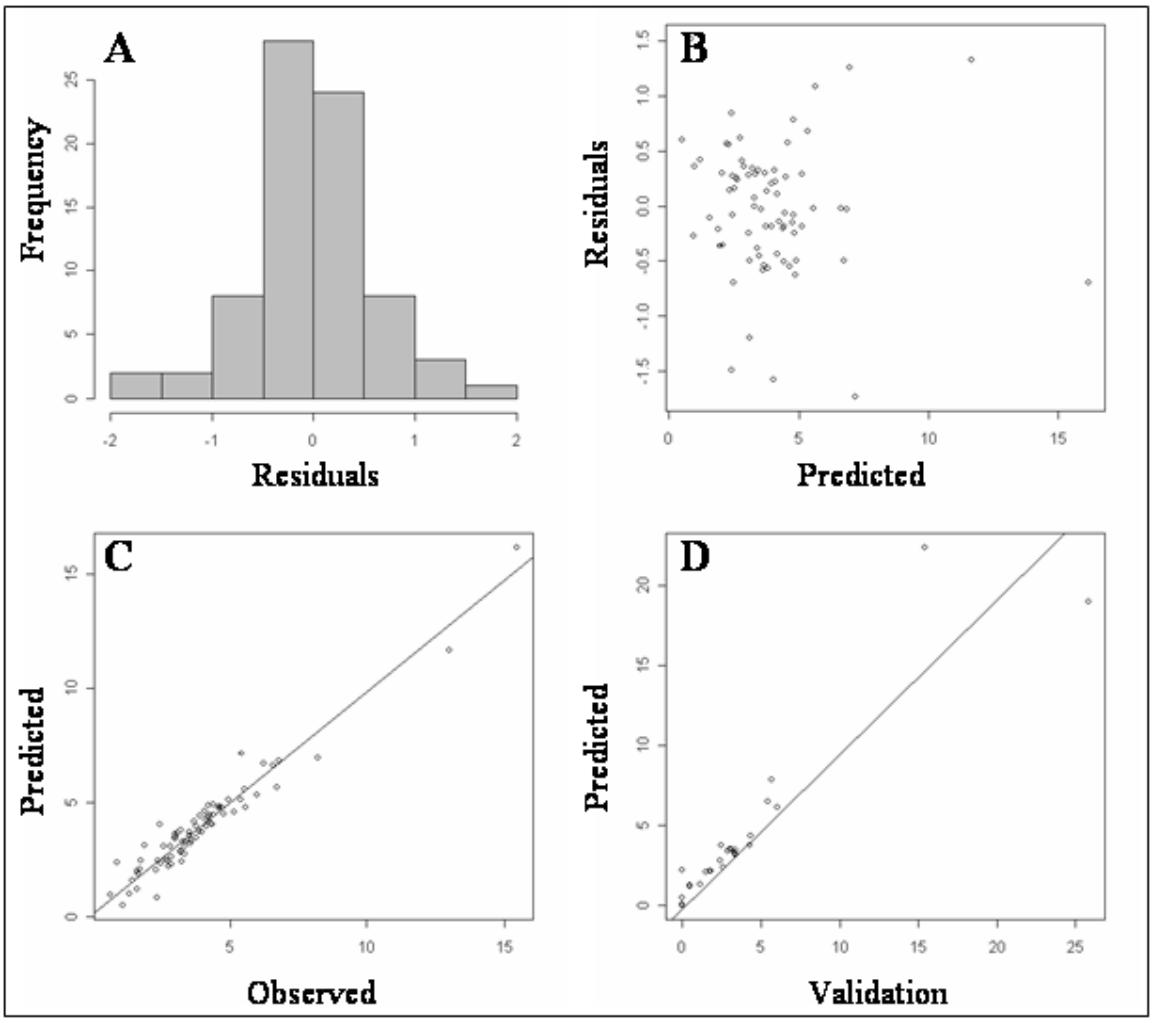


Figure 3.7 – O-horizon thickness spatial model performance plots: A) Histogram of model residuals; B) Scatter plot between model residuals and predicted values; C) Scatter plot with trend line of model predicted and observed values; D) Scatter plot with trend line between predicted values and validation values

3.4 - Soil attribute model validation

Validation of the soil attribute models utilizes three analyses to evaluate statistical and spatial performance. The statistical performance analysis is based on the validation dataset withheld from the original sampled dataset in Iron and Lexen creek watersheds. This analysis focuses how well the soil attribute prediction values correlate to the validation (measured) values. The spatial performance analysis evaluates how well the spatial soil attribute model's predictions compare with the measured (sampled) and soil survey map unit values. This analysis involves calculating total soil depth, A-horizon and O-horizon thickness summary statistics for each land cover type and soil survey map unit in FEF (Table 3.10, Appendix 5, Appendix 6 and Appendix 7). This is useful in determining where the model is robust and where its estimates are not reliable.

The soil depth model has the best statistical fit of the three models by capturing 90 percent of measured variability (Table 3.9 and Figure 3.2 D). Soil depth summaries for the land cover types in FEF (Table 3.10) highlights Alpine Meadow, Lodgepole pine and the Spruce-fir as areas where the model performed strongly. By producing mean soil depth estimates within 2 inches of the soil survey and measured datasets. These land cover types make up the majority of the sampling area (Figure 2.4) and account for 87 percent of FEF. The land cover types with the largest differential (>10 inches) between the model and soil survey are Aspen, Riparian and Rock (Table 3.10). For Rock and Riparian areas the measured and modeled values are very similar (~1 inch).

The soil survey map unit comparison between the model and individual map units (Appendix 5) is weak, with the majority of map units (70%) being greater than 10 inches from the model. Of these map units with sample sites (Alluvial lands (Aa and Ac),

Ptarmigan (Pa, Pb, Pc, Pd) and Vasquez (Va)), have measured values that are within one inch of the model estimates. The 30 percent of map units that are within 10 inches of the soil depth model make up 56 percent of FEF and are associated with forested, alpine organic, alpine meadow soils (Appendix 2).

Table 3.10 – Soil property summary statistics for land cover types in FEF based on prediction model (Spatial Model), sample points (Measured) and Fraser Alpine Area soil survey values. The summary statistics calculated by intersection of soil survey map units and spatial model cell that fall inside a land cover type. The mean and standard deviation values for soil survey values are based on area weights. The measured mean and standard deviations are calculated based on sampled point values weighted by the inclusion probability value.

		Spatial Model		Measured		Soil survey	
Soil Attribute	**Land cover**	Mean	STD	Mean	STD	Mean	STD
Soil Depth	Alpine Meadow	15.1	8.2	18.7	8.7	17.9	15.3
	Aspen	33.6	7.2	NA	NA	23.0	17.0
	Lodgepole Pine	29.8	4.3	32.1	4.1	27.2	16.0
	Riparian	29.5	7.8	29.6	10.2	18.9	15.9
	Rock	2.5	7.0	1.5	2.1	11.2	15.8
	Spruce-fir	27.6	5.3	29.6	4.3	27.0	16.4
A-Horizon Thickness	Alpine Meadow	4.5	2.9	4.9	3.1	2.9	2.6
	Aspen	4.9	2.1	NA	NA	2.4	1.7
	Lodgepole Pine	4.2	1.8	3.9	2.7	2.8	1.6
	Riparian	4.2	1.6	3.9	2.1	2.1	1.7
	Rock	1.6	2.8	1.2	4.2	1.5	2.1
	Spruce-fir	1.5	1.8	3.8	2.4	2.5	1.2
O-Horizon Thickness	Alpine Meadow	0.8	9.7	0.8	1.5	0.5	1.9
	Aspen	4.4	4.6	NA	NA	3.6	5.3
	Lodgepole Pine	2.4	3.7	2.3	1.7	3.0	3.6
	Riparian	5.1	4.4	5.1	6.0	3.3	2.5
	Rock	0.5	0.2	0.8	1.8	0.7	1.8
	Spruce-fir	3.2	2.4	2.9	2.1	3.1	2.0

The A-horizon spatial model performed the poorest against the validation dataset capturing 82 percent of the measured variability (Table 3.9). The scatter plot between predicted and measured (validation) (Figure 3.3 D) shows that the predicted A-horizon values are systematically higher against the measured values.

The land cover summaries support the results from the validation analysis with the A-horizon model over estimating thickness compared with the measured and soil survey thicknesses values (Table 3.10). The difference between the measured data and the model are slight with spruce-fir having the largest error. Comparing the soil survey value with the model indicates that the model in most cases is over estimating A-horizon thickness by 1.75 inches.

The A-horizon thickness model estimated mean A-horizon thicknesses that are greater than the majority (60%) of the soil map unit estimates (Appendix 3). This is especially true for map units with zero A-horizon thickness (Alluvial lands (Aa), Alpine lands (Ab, and Ac), Leal Series (La and Lb) and Rock outcrops (Ra and Rb)) of these map units the model has a large standard deviations and have measured estimates that are within 2.5 inches of the model estimates (Appendix 3). The other map units that the model estimated a higher mean A-horizon thickness are within 2.5 inches of the soil survey. The 40 percent of map units that model underestimated A-horizon thickness have a difference of about 2.5 inches with the measured values being within 1.5 inches.

The O-horizon model performed well against the validation dataset by capturing 85 percent of measured variability (Table 3.9). The scatter plot between predicted and measured (validation) (Figure 3.4 D) shows that at deeper O-horizon thickness (>15 inches) the model error increases.

The land cover O-horizon summaries indicate that there is a close agreement between the model, measured and soil survey map units across all land cover types (Table 3.10). The lodgepole pine and Spruce-fir land cover types have the strongest agreement between all three dataset with less than 0.6 inches separating them and standard deviations the over lap.

The Alpine Meadow land cover has a close agreement (mean <0.3 inches) between all three datasets, but the model has a very high standard deviation. The Riparian land cover shows a very close agreement between the model and measured with the soil survey map units having a mean depth 1.8 inches less than the modeled and measured datasets.

The soil survey map unit comparison (Appendix 7) shows a strong agreement with the majority of map units (60%) having modeled, measured, and soil survey thickness that are with one inch of each other. The map units with the greatest difference between the model and soil survey are the Alluvial land (Aa), Lunch (Ld), Nystrom (Na) and Tabernash series (Appendix 2). These map units have very large differences between the model and soil survey with the Lunch and Nystrom series exceeding 10 inches. The Alluvial land and Tabernash series differences between the model are five inches with standard devotions equal to the mean indicating the there is a lot of variability within them.

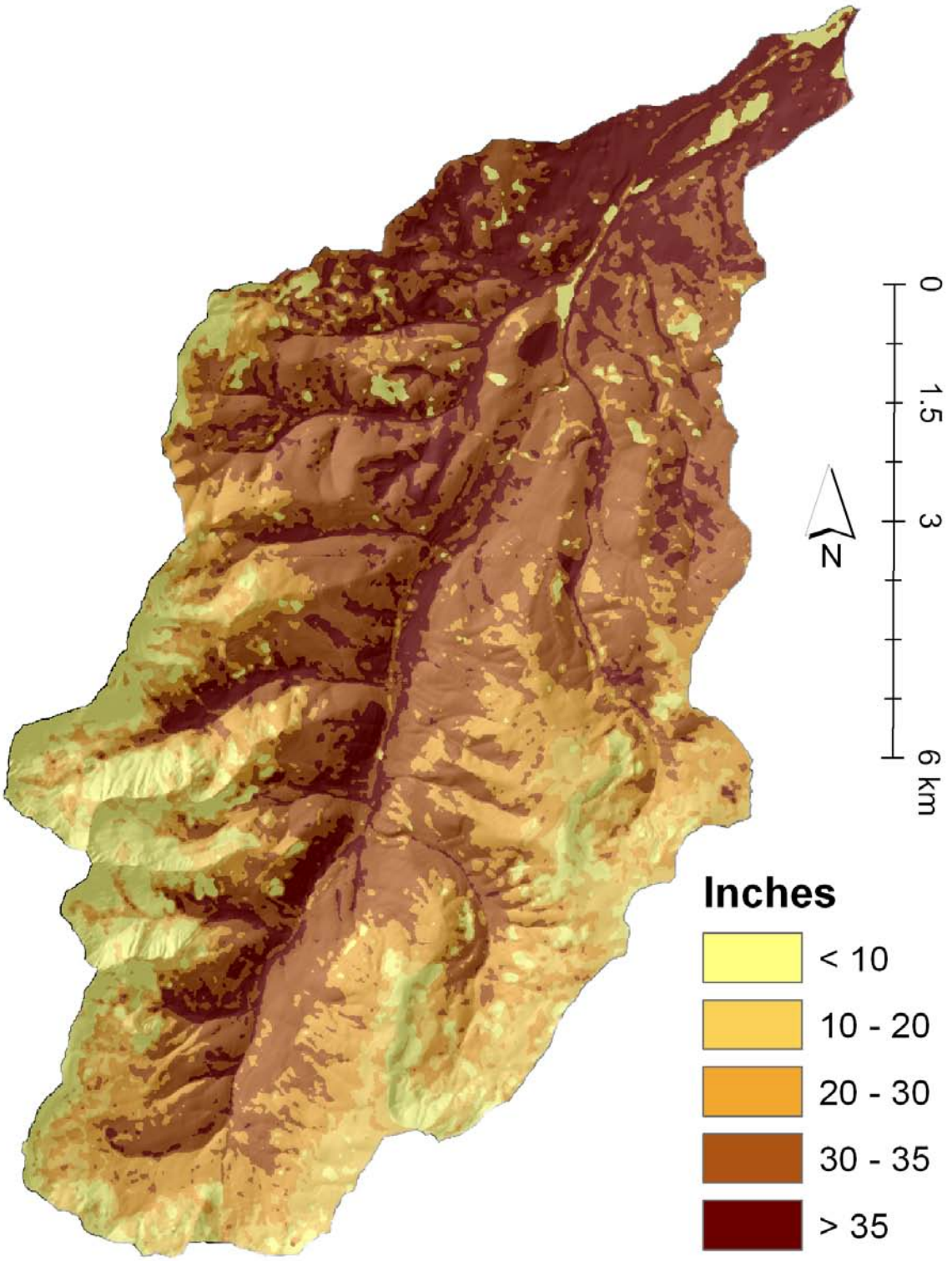


Figure 3.8 – Total soil depth spatial model

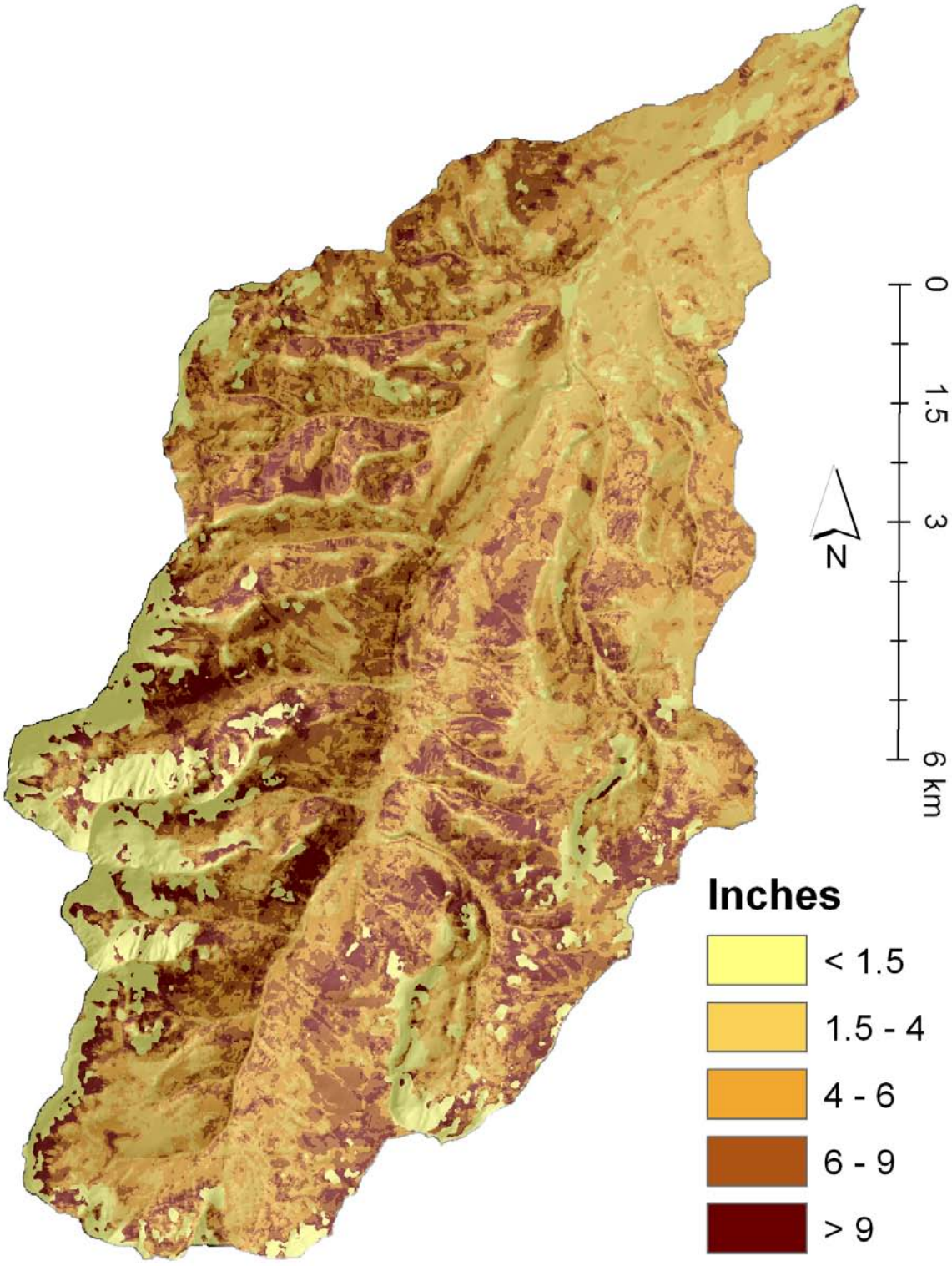


Figure 3.9 - A-horizon thickness spatial model

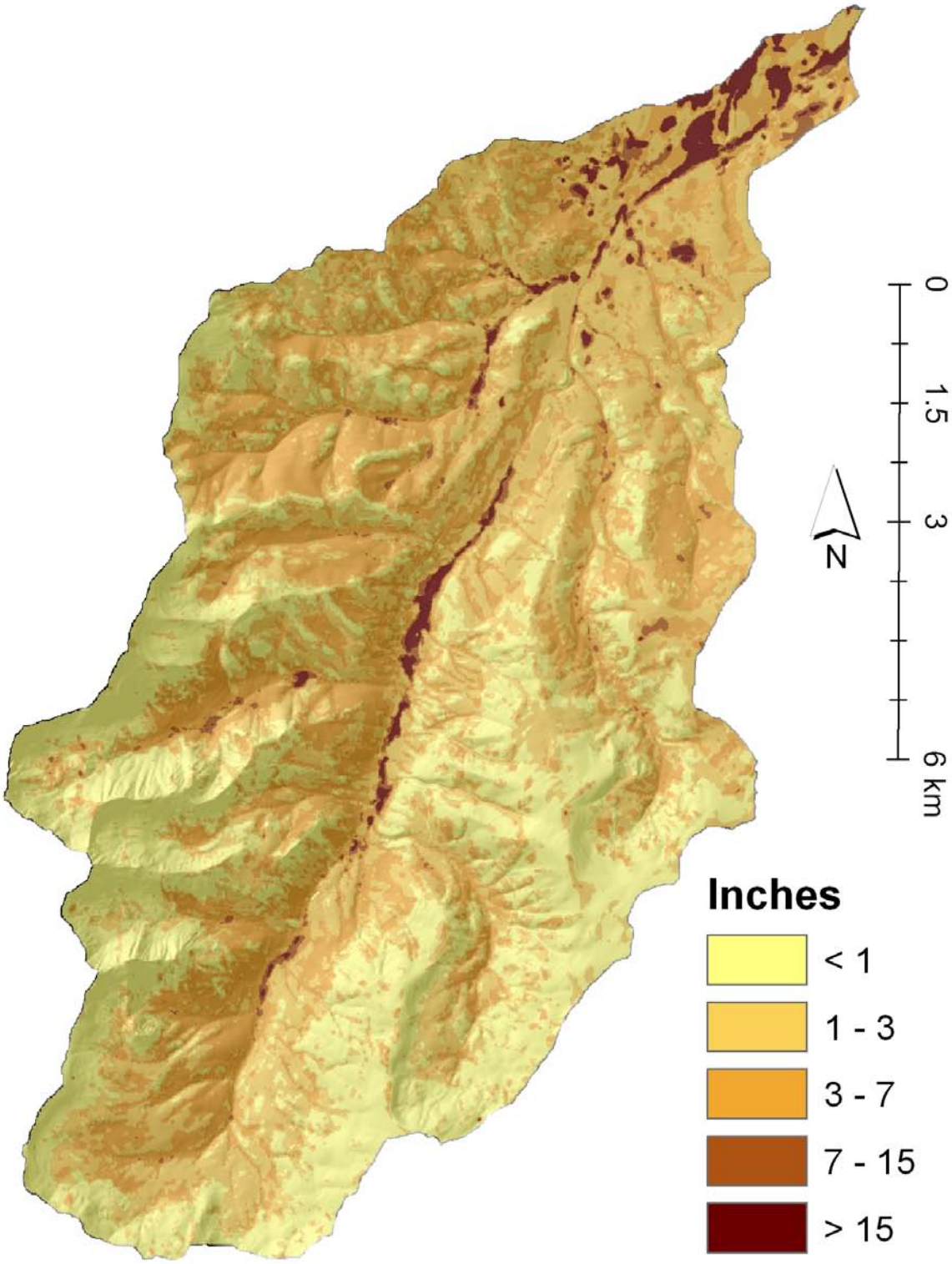


Figure 3.10 - O-horizon thickness spatial model

3.5 - Summary and Conclusions

The overall objective of this thesis is to generate spatially explicate soil attribute surfaces for FEF. This required acquiring geospatial datasets, performing remote sensing and terrain analysis, development of a sampling design, field sampling and geostatistical methods. Due to personal, resource and climatic constraints, the need to streamline this process became evident resulting in an underlying theme of producing viable soil attribute models for FEF in a cost efficient manner. This efficiency theme is most noticeable in the sampling design and the SBSES, but the efficiencies of the terrain and remote sensing analysis, field methods, and geostatistical methods streamlined the modeling effort as well. The efficiencies of the terrain and remote sensing analysis are based on two factors: 1.) Rely on two datasets (elevation and Landsat 7 TM), which for FEF captures its diverse terrain and vegetation dynamics. 2.) Based on widely used, robust across all biomes, and easy to calculate. These metrics along with land cover played key roles in selecting intensive sampling units, developing the sampling design's inclusion probability surface, and generating the geostatistical soil attribute models.

The SBS design, field sampling protocols and accessibility analysis used to intensively sample soils in the Iron and Lexen Creek watersheds produced field base information that captured soil property variability in a short period of time (142 hours). The field based soil properties provided spatial information used to summarize soil depth, A-horizon and O-horizon thickness for land cover types in the intensive sampling units, as well as to generate the soil attribute models.

The methods implemented in the field were streamlined by collecting minimal soil property information and defining sampling loops using cost distance analysis allowing for many samples to be collected per sampling outing. These methods can be implemented within a wide variety of landscapes with many degrees of accessibility. The sampling time statistics provide good estimates of accessibility in FEF for easily accessible watersheds like Lexen Creek (~40%) and also for very inaccessible watersheds like Iron Creek (~60%).

The SBSES analysis demonstrates that the SBS design implemented for this study produced stronger fitting statistical models (R^2 by 2% and MSE by 93%) that are less influenced by spatial autocorrelation (93% Moran's I p-values > 0.005) in a more cost effective manner (CR > 1.0). The simulation results provides evidence that SBS in conjunction with the inclusion probability surface is effective at sampling soil properties to inform statistical algorithms by increasing model fit, reducing spatial bias, and sampling costs.

The soil depth model is the most statistically robust of three models with a high R^2 (0.88), no spatial dependence in the residuals and a strong fit with the validation data (R^2=0.91). Spatially, the soil depth model performed strongly in Lodgepole pine, Spruce-fir,

Alpine Meadow land cover types (Table 3.10) that make up 87 percent of FEF. The model agreement with the soil survey (Appendix 5) is not as strong as the land cover types with 70% of the map units being +/- 10 inches form the model means. In most cases the measured data within the map units is in agreement with the model.

The statistical and spatial performance of the A-horizon depth model shows a systematic inflation of thickness across the validation, land cover type and soil survey datasets. The statistical comparison with the validation data (Figure 3.3 and Table 3.9) indicates that the model accounted for 82 percent of the variability with the residuals being slightly spatially correlated. The spatial comparison for land cover types shows that the model is slightly higher than the measures data and is on average two inches greater than the soil survey. The soil survey comparison shows that the model thickness estimates for soil map units are within two inches of the measured and the soil survey.

The O-horizon thickness model has an R^2 of 0.91 with model errors that are +/- 1.5 and accounts for 85 percent of the validation thickness variability. The spatial comparison on land cover types (Table 3.10) indicates that the model is on average +/- 1 inch form the measured and soil survey estimates. The soil survey comparison (Appendix 7) shows that the model and measured data agree (+/- 0.56 inches) and the majority (60%) of map units are within +/- 1 inch. The map units that are not in agreement are the organic soils (Lunch and Nystrom series) and the forested soil Tabernash series (Appendix 2).

The soil attribute models developed from this study provide a continuous representation of soil properties (Total soil depth, A-horizon and O-horizon thickness) at a fine scale (0.001 ha). These spatial models will provide inputs to hydrological and ecological models, statistical covariates to investigate soil's influence on water chemistry and vegetation

distributions, and provide an initial platform for future soil survey activities in FEF. The high statistical performance of the soil attribute models is only valid for Iron and Lexen Creek watersheds, but the spatial comparisons (model, measured, and soil survey) indicates that the models are robust for large areas in FEF.

The development of the soil attribute models have the potential to provide useful auxiliary information for the soil survey development and updating by quantifying soil property variability, enhancing map unit delineation, and providing insight into where additional samples would capture soil variability. The use of the intensive sampling units within FEF (Iron and Lexen Creek watersheds) was useful in capturing fine scale variability of soil properties within these units, but lacks spatial leverage FEF wide and should be supplemented with more samples to enhance the soil attribute spatial models to increase accuracy and precision model predictions.

REFERENCES

Aguilar, R., Kelly, E.F., Heil, R.D. 1988. Effects of Cultivation on Soils in Northern Great Plains Rangeland, Soil Sci. Soc. Amer. 52:1081-1085

Akaike, H. 1969. Fitting autoregressive models for prediction. Annals of the Institute for Statistics and Mathematics 21, 243-247.

Austin, M.P., Mckenzie, N.J., 1988. Data analysis. In: Gunn, R.H., Beattie, J.A., van der Graff, R.H.M. (eds.), Australian Soil and Land Survey Handbook. Guidelines for Conduction Surveys, Chap. 11. Inkata Press, Melbourne.

Bailey, R. G., 1995 Descriptions of the Ecoregions of the United States (2nd Edition). Miscellaneous Publication No. 1391, Map scale 1:7,500,000, U.S. Department of Agriculture, Forest Service, 108 pp.

Burgess, T.M., Webster, R. and McBratney, A.B. 1981. Optimal and isarithmic mapping of soil properties: IV. Sampling Strategy. Journal of Soil Science 32, 643-659.

Burgess, T.M. and Webster, R. 1980. Optimal interpolation and isarithmic mapping of soil properties II. Block Kriging. Journal of Soil Science 31, 333-341.

Cassel, C.M., Sarndal, C.E. and Wretman, J.H. (1977), Foundations of inference in survey sampling: Wiley, New York, 192 p.

Crist, E.P. and Kauth, R. J. 1986. The tasseled cap de-mystified. Photogrammetric Engineering and Remote Sensing 22, 81-86

Crist, E. P., & Cicone, R. C. 1984. A physically-based transformation of Thematic Mapper data-the TM Tassel Cap. IEEE Transformations on Geoscience and Remote Sensing, 22, 256-263.

Cook, S.E., Corner, R.J., Groves, P.R., Grealish, G.J., (1996). Use of airborne gamma radiometric data for soil mapping. Agust. J. Soil Res. 34, 183-194.

Davis, J.C. 1986. Statistics and data analysis in geology. Second Edition. John Wiley & Sons, New York.

Eppinger, R.G., Theobald, P.K., and Carlson, R.R. (1985). Generalized geologic map of the Vasquez Peak Wilderness Study Area and the Williams Fork and St. Louis Peak Roadless Areas, Clear Creek, Grand and Summit Counties, Colorado. U.S. Geologic Survey, Misc. Field Studies map MF-1588-B, Reston, VA.

ESRI. 2007. ArcGIS Version 9.2. Windows 2000/XP Server Operating System. Build 800. Redlands, CA.

Evans, C.V and Roth, D.C. 1992. Conceptual and statistical modes to characterize soil materials, landforms, and processes. Soil Sci. Soc. Amer. 56, 214-219.

Gerrard, A.J., 1981. Soils and landforms: an integration of geomorphology and pedology. George Allen and Unwin, London.

Guisan, A., Theurillat, J-P., 2000. Equilibrium modeling of alpine plant distribution and climate change: how far can we go? Phytocoenologia 30, 353-384.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 89-100.

Lehmann, A., Leathwick, J.R., Overton, J.McC., 2002. Assessing New Zealand fern diversity from spatial predictions of species assemblages. Biodivers. Conserv. 11, 2217-2238.

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

Hudson, B.D., 1992. The soil survey as a paradigm-based science. Soil Sci. Am. J. 56, 836-841.

Hewitt, A.E., 1993. Predictive modeling in soil survey. Soil and Fertilizers 3, 305-315.

Lillesand, T. and R. Kiefer. 2000. Remote sensing and image interpretation. John Wiley & sons. New York, New York.

Kauth, R.j. and Thomas, G. S. 1976. The Tasseled Cap - a graphic description of the spectral – temporal development of agricultural crops as seen by Landsat. Proceedings second ann. Symp. Machine processing of remotely sensed data. West Lafayette: Purdue University Lab. App. Remote Sensing.

Kelly, E.F., Aguilar, R., Muhaimeed, A.S., Deutsch, P.C. and Heil R.D. (1988). Profile Reconstruction: A method to quantify changes in soil properties resulting from cultivation. Agriculture, Ecosystems, and Environment. 21:153-162

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

Odeh, I.O.A., McBrateny, A.B., Chittleborough, D.J., (1994). Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63, 197-214.

Oliver, M., Webster, R. and Gerrard, J. 1989. Geostatistics in physical geography. Part II: applications. Trans. Inst. Br. Geogr. N.S. 14, 270-286.

Popovich, S.J., Sheppard, W.D., Reichert, D.W., and Cone, M.A. 1993. Flora of the Fraser Experimental Forest, Colorado. Gen. tech. Rep. RMRS-GTR-233. Fort Collins, CO. U.S. Department of Agriculture. Forest Service Rocky Mountain range and Experimental Station. 62 p.

Qi, J., A. Chehbouni, A. R. Huete, and Y.H. Kerr. 1994. Modified Soil Adjusted Vegetation Index (MSAVI). Rem. Sens. Environ. 48, 119-126

R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900045-15-3. [Online] URL http://www.R-progect.org.

Retzer, J.L. 1962. Soil Survey of Fraser alpine area, Colorado. Soil Survey 1956, No. 20. USDA Forest Service and Soil Conservation Service and Colorado Agriculture Experimental Station. U.S. Government Printing Office, Washington, DC. 47 p.

Ripley, B. 2008. An online help: the *MASS* package. [online] URL: http://can.r-project.org/web/packages/VR/VR.pdf

Rondeaux, G., M. Steven, and F. Baret. 1996. Optimization of soil-adjusted vegetation indices. Rem, Sens, Environ. 55, 95-107

Skidmore, A.K., Watford, F., Luckanamurug, P., Ryan, P.J., 1996. An Operational GIS expert system for mapping forest soils. Photogrammetric Engineering and Remote Sensing 62, 501-511.

Sorensen, R., Zinko, U. and J. Seibert. 2006. On the calculation of the topographic wetness index: evaluation of different methods based upon field observations. Hydrology and Earth System Sciences, 10, 101-112.

Stevens, D.L., Olsen, A.R. 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association. 99, 262-278.

Theobald, D.M., Norman, J.B., Peterson, E., Wade, A., and Ferraz, S. 2006. Functional Linkages of Watersheds and streams (FLoWs) User's manual: Network-based ArcGIS tools to analyze freshwater ecosystems. May 2006.

Theobald, D.M., N. Peterson, and W. Romme. 2004. The Colorado Vegetation Model: using National Land Cover Data and Ancillary Spatial Data to Produce a High Resolution, Fine-classification Map of Colorado (v8). 4 February. Unpublished report, Natural Resources Ecology Lab, Colorado State University.

Theobald, D.M., Stevens D.L. Jr., White, D., Urquhart, N.S., Olsen, A.R., Norman, J.B. 2007. Using GIS to generate spatially balanced random survey designs for natural resource applications. Environmental Management. 4, 134-146.

Tobler, W 1961: Map Transformations of Geographic Space, PhD dissertation, Seattle, University of Washington

USDA-NRCS. 1998. Land resource regions and major land resource areas of the United States. USDA-NRCS Agric. Handb. 296. U.S. Gov. Print. Office, Washington, DC.

Webb, T.H. (Ed.), 1994. Soil Landscape Modeling in New Zealand. Landcare Research Science Series No. 5. Manaaki Whenua Press, Lincon.

Webster, R. 1985. Quantitative spatial analysis of soil in the field. Pages 1-70 in B.A. Stewart, editor. Advances in soil science. Volume 3. Springer-verlag, New York, New York, USA.

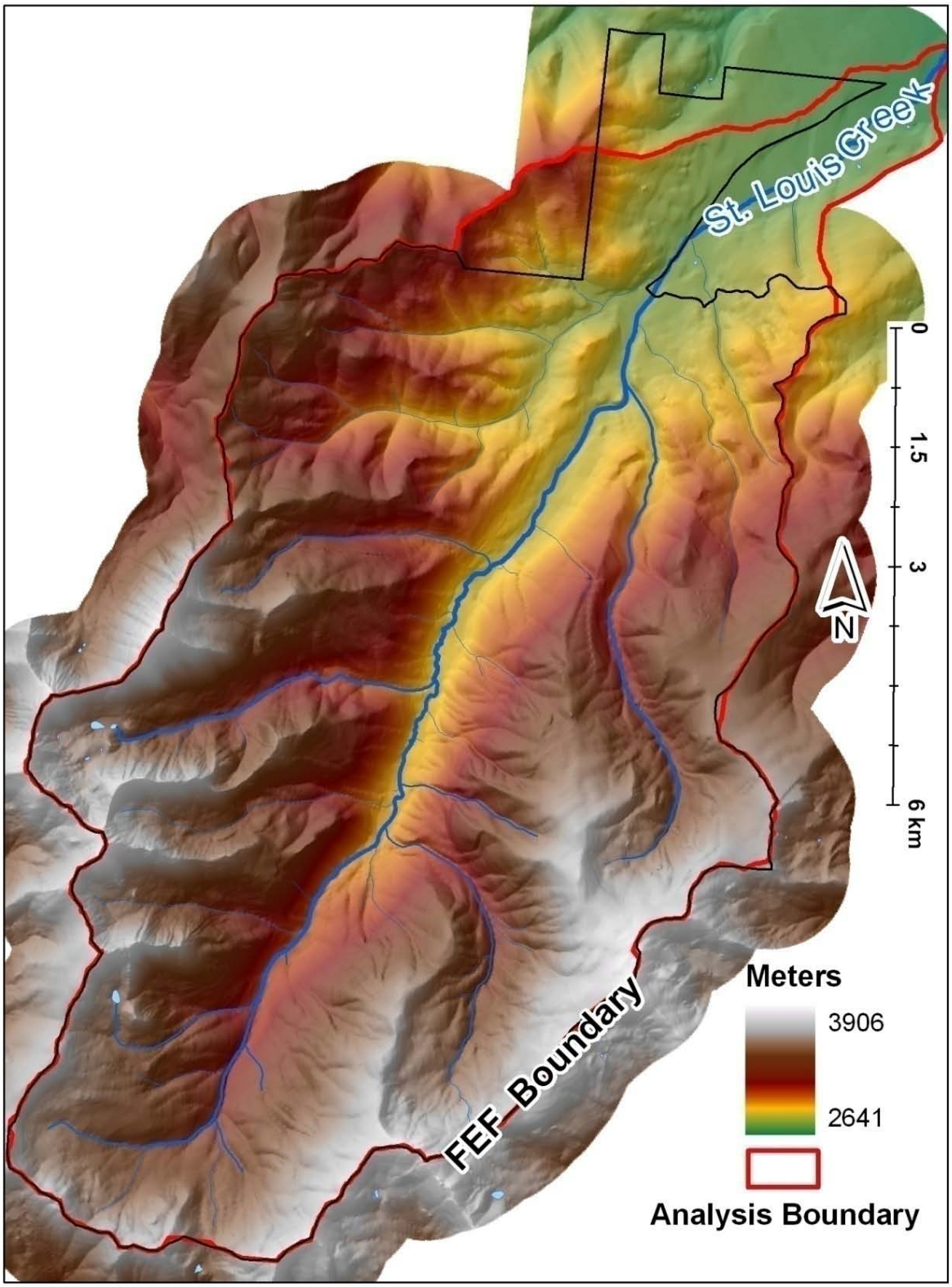
Webster, R., 1997. Soil resources and their assessment. Phil Trans. R. Soc. London, Ser. B 352, 963-973

Weitz, A., Bunte, D., Hersemann, H., 1993. Application of nested sampling technique to determine the scale variation in soil physical and chemical properties. Cantena 20, 207-214.

Xhu, A.X., Band, L., Vertessy, R., Dutton, B., 1997. Derivation of soil properties using a soil land inference model (SoLIM). Soil Sci. Am. J. 61, 523-533.

ADDNEDIX 1

APPENDIX 1. Analysis extent “window” in red used to generate soil attribute surfaces and evaluate SBS design efficiency



APPENDIX 2

APPENDIX 2A. Fraser Alpine Area Soil Survey map unit symbols and component with brief descriptions

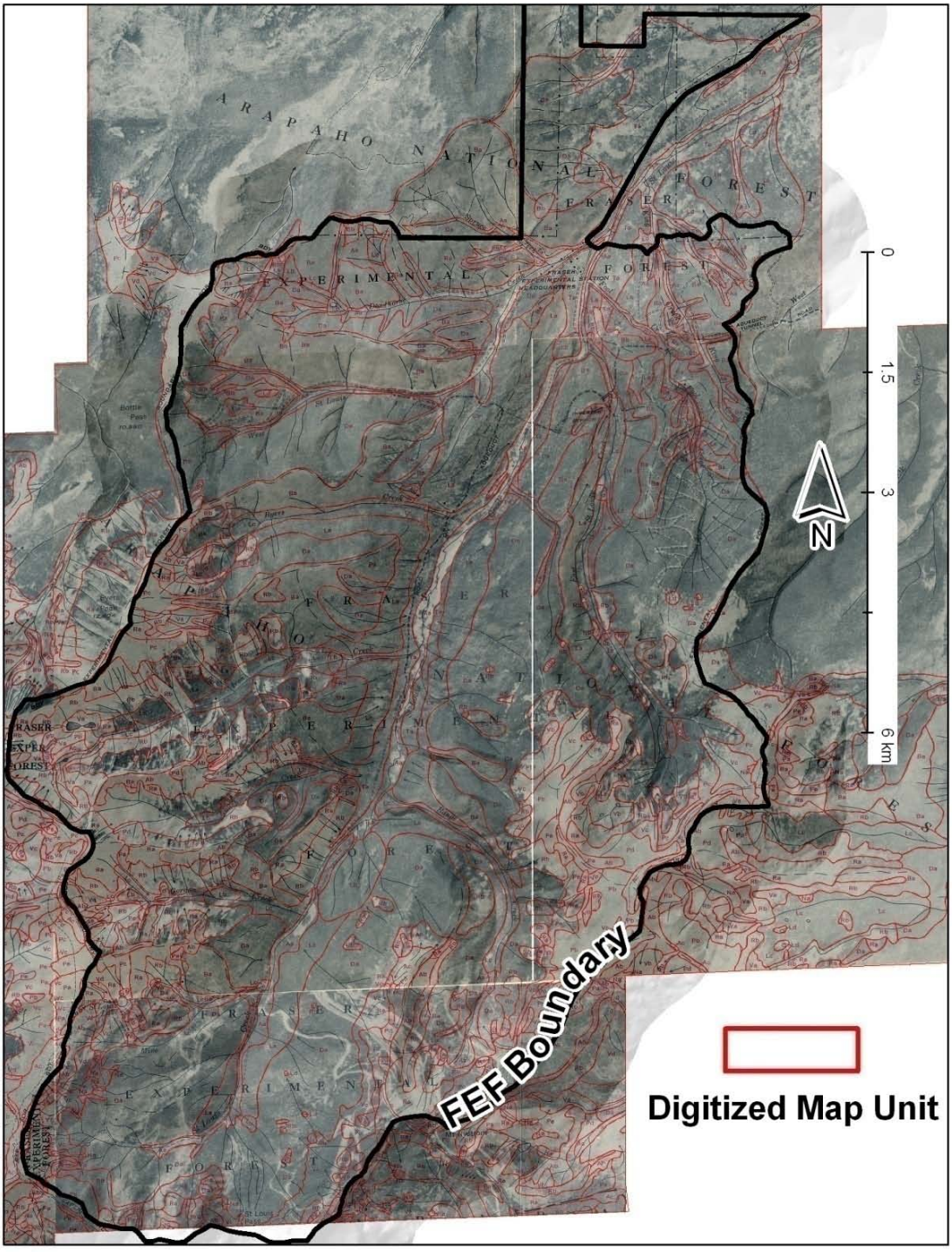
Symbol	Name	Description
Aa	Alluvial land	Coarse-textured, cobbly stratified alluvium making up flood plains associated with the stem of St. Louis Creek.
Ab	Alpine rimland	A sandy, gravelly and rock soil with small amounts of silt and clay supporting sparse vegetation on alpine rimlands.
Ac	Alpine wind-eroded land	Wind swept alpine areas made up of coarse sand, gravel and rock with no vascular vegetation.
Ba	Bobtail series	A gravelly sandy loam lodgepole pine soil with a weak sopdic horizon, but not classified as a Spodosol.
Bb	Bottle series	A fine sandy spodic acidic soil with a strongly developed E-horizon. This soil is associated with sandstone in high-elevations within the West St. Louis Creek watershed.
Da	Darling series	An excessively drained, deep, coarse textured Spodosol in wet shaded areas from the sup-alpine to low elevation areas. Spruce-fir is the dominate vegetation producing a large O-horizon (4 inches) and thick E-horizon (3 inches).
La **Lb** **Lc**	Leal series	A coarse textured acidic Spodosol found along the lateral moraines of St. Louis Creek.
Ld	Lunch series	An organic soil that is one to two feet deep and is associated with steep sub-alpine and alpine slopes.
Na	Nystrom series	An alpine organic soil found in areas where water accumulates and vegetation can grow. The depth of organic material is on average 17 inches deep.
Pa **Pb** **Pc** **Pd**	Ptarmigan series	A well-drained sandy loam soil associated with alpine meadows. This series has four map units which differ by texture and development of the A-horizon.
Ra	Rock outcrop	This map unit is made up of areas consisting of great masses of consolidated rock and large boulder fields.

APPENDIX 2B (Cont.). Fraser Alpine Area Soil Survey map unit symbols and component names with brief descriptions

Symbol	Name	Description
Rb	Rock slides	This map unit is made up of areas consisting of loose rock ranging from large stones to gravel associated with scree fields, colluvium slides and rock glaciers.
Ta	Tabernash series	A medium to fine textured loam wooded soil associated with the north side of an extinct lake with zero slope.
Vb	Vasquez series	A sandy clay loam soil associated with depressed areas in the alpine and surrounded by alpine meadow (Ptarmigan series).

APPENDIX 3

APPENDIX 3. Fraser Alpine Area Soil Survey georectified paper maps with digitized map units (red) and Fraser Experimental Forest boundary (black).



APPENDIX 4

APPENDIX 4. Spatially Balanced Survey Efficiency Simulation procedures per iteration with number corresponding to Figure 2.7

1) Generate a 137 SBS and SRS points for the study area. The SRS points were generated by randomly adding x and y locations that fall within the study area. The SBS points were generated using a RRQRR sequence and inclusion probability raster where a random raster is generated, filtered against the random raster and then sequenced with the sequence raster. The 137 sample points where then extracted from the resulting raster by taking the top 137 values.

2) Voronoi polygons are then calculated at each iteration and ER is calculated between the two point processes.

3) Generate SBS and SRS point data table to be used in the modeling and cost analysis processes. This data table consists of NDVI (dependant variable), Independent variables (Table 4) and travel time, as well as, geographic coordinates (UTM WGS84, zone 13).

4) In R, the LM is created using all independent variables (Table 4) and is then evaluated with a forward stepwise operation that finds the best fitting model. This results in a LM intercept and coefficients, model performance statistics, and prediction residuals.

5) Spatial dependence of the residuals is evaluated based on partial and global Moran's I.

6) The LM model is then converted into a spatial surface via a map algebra string executed in ArcGIS 9.2.

7) The LM prediction surface is subtracted from the true NDVI surface to calculate true spatial error. The error surface is then averaged at each iteration as well as variance and standard deviation of error is calculated.

8) During the simulation two different tables are generated the iteration statistics file focuses on global iteration statistics (R^2, Moran's I, p-value) that represents all points for a given iteration the point statistics data file captures each point involved in the simulation and statistics that pertain to it (partial Moran's I and travel time)

9) During the simulation two different text files are generated:

i. The point iteration file (pts_iteration.txt) records X and Y, dependant, independent and cost values for each SBS/SRS point generated, as well as, the LM residuals partial Moran's I estimates. This file consisted of 137000 records for the full 1000 iterations.

ii. The iteration statistics text file (iteration_stats.txt) records statistical information about each iteration (Table 5) resulting in 1000 records.

APPENDIX 5

APPENDIX 5. Fraser Alpine Area Soil Survey map units with total soil depth model (model) and sample site (measured) means and standard deviations, as well as soil survey estimates. The soil survey estimates don't include standard deviations because there is only one component per map unit and the soil survey supplies one depth value.

	Total Soil Depth (inches)				
Map	**Model**		**Measured**		**Soil**
Symbol	Mean	STD	Mean	STD	**survey**
Aa	31.2	4.6	31.8	0.4	0
Ab	8.5	9.3	0.0	0.0	0
Ac	12.7	9.0	NA	NA	0
Ba	28.5	7.3	29.9	4.0	40
Bb	26.5	8.9	NA	NA	33
Bd	28.5	1.7	NA	NA	33
Da	27.2	5.6	26.0	11.4	33
La	30.4	3.8	28.3	7.4	38
Lb	28.3	7.1	27.4	1.4	2
Lc	28.6	6.7	33.1	3.4	2
Ld	33.1	5.3	NA	NA	2
Na	19.3	6.2	NA	NA	24
Pa	12.7	9.3	NA	NA	14
Pb	18.8	7.2	14.9	2.5	30
Pc	13.2	9.6	24.4	1.5	30
Pd	11.5	9.0	6.6	3.5	30
Pe	15.9	8.3	NA	NA	30
Ra	10.6	11.1	14.2	14.5	0
Rb	15.2	13.1	15.0	11.3	0
Ta	30.5	9.3	NA	NA	36
Va	15.7	9.6	21.6	8.2	31
Vb	13.3	9.8	NA	NA	31
Vc	20.4	6.2	NA	NA	31

APPENDIX 6

APPENDIX 6. Fraser Alpine Area Soil Survey map units with A-horizon thickness model (Model) and sample site means (Measured) and standard deviations, as well as soil survey estimates. The soil survey estimates don't include standard deviations because there is only one component per map unit and the soil survey supplies one depth value.

Map Symbol	A-Horizon Thickness (inches) Model Mean	Model STD	Measured Mean	Measured STD	Soil survey
Aa	4.2	1.5	5.8	3.0	0
Ab	2.6	3.3	0.0	0.0	0
Ac	4.3	3.4	N A	NA	0
Ba	4.5	1.8	3.6	2.6	3
Bb	3.4	1.6	NA	NA	6
Bd	5.0	1.3	NA	NA	3
Da	4.9	1.8	3.5	3.9	2
La	4.4	1.5	4.7	2.5	2
Lb	4.4	1.7	4.0	2.6	2
Lc	4.8	1.7	3.2	2.2	0
Ld	3.8	1.2	NA	NA	0
Na	4.2	1.9	NA	NA	6
Pa	4.1	3.4	NA	NA	6
Pb	5.0	2.2	3.3	0.9	6
Pc	4.0	3.1	8.6	1.7	6
Pd	3.9	3.5	0.0	0.5	6
Pe	5.4	2.8	NA	NA	6
Ra	2.9	3.3	2.3	3.2	0
Rb	3.8	3.4	3.1	4.3	0
Ta	3.5	1.7	NA	NA	5
Va	4.3	3.1	5.1	3.3	4
Vb	4.0	3.2	NA	NA	4
Vc	5.0	2.1	NA	NA	4

APPENDIX 7

APPENDIX 7. Fraser Alpine Area Soil Survey map units with O-horizon thickness model (Model) and sample site (Measured) means and standard deviations, as well as soil survey estimates. The soil survey estimates don't include standard deviations because there is only one component per map unit and the soil survey supplies one depth value.

	O-Horizon Thickness (inches)				
Map	**Model**		**Measured**		**Soil**
Symbol	Mean	STD	Mean	STD	**survey**
Aa	5.3	5.2	4.3	1.2	0
Ab	0.5	0.7	0.0	0.0	0
Ac	1.0	1.2	NA	NA	0
Ba	2.3	1.4	1.8	1.6	2
Bb	2.7	2.3	NA	NA	2
Bd	2.5	1.6	NA	NA	2
Da	2.5	1.7	2.1	1.9	4
La	2.9	1.8	3.1	1.8	2
Lb	3.0	1.9	2.9	2.1	2
Lc	3.7	2.4	5.2	4.0	2
Ld	4.4	3.8	NA	NA	24
Na	1.8	1.2	NA	NA	14
Pa	1.0	1.1	NA	NA	0
Pb	0.9	1.0	2.6	3.2	0
Pc	0.8	0.9	0.0	0.0	0
Pd	0.5	0.8	0.0	0.0	0
Pe	0.7	0.8	NA	NA	0
Ra	0.7	1.0	0.8	1.3	0
Rb	1.4	1.6	1.5	2.3	0
Ta	6.5	6.5	NA	NA	1
Va	1.3	1.3	0.6	1.4	0
Vb	0.9	1.2	NA	NA	0
Vc	1.3	1.1	NA	NA	0

APPENDIX B

FIELD DATA SHEET

Site #: _____ Photo #: _____ Start Time: _____ End Time: _____

Date: _____

Slope: _____

Aspect: _____

PM: _____

Landform:

Horizon Designation:	From (in):	To (in):	Texture class:
1			
2			
3			
4			
5			
6			
7			
8			

Site: H1	Site: H2	Site: H3	Site: H4	Site: H5	Site: H6	Site: H7	Site: H8