

2015

Spatializing the Soil-Ecological Factorial: Data Driven Integrated Land Management Tools

Travis Nauman

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Nauman, Travis, "Spatializing the Soil-Ecological Factorial: Data Driven Integrated Land Management Tools" (2015). *Graduate Theses, Dissertations, and Problem Reports*. 6299.
<https://researchrepository.wvu.edu/etd/6299>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Spatializing the Soil-Ecological Factorial: Data Driven Integrated Land Management Tools

Travis Nauman

**Dissertation submitted
to the Davis College of Agriculture, Natural Resource and Design
at West Virginia University**

**in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in
Plant and Soil Sciences**

Committee:

**James A Thompson, Ph.D., Chair
Louis McDonald, Ph.D.
James Rentch, Ph.D.
Brenden McNeil, Ph.D.
Timothy Warner, Ph.D.**

Division of Plant and Soil Sciences

**Morgantown, West Virginia
2015**

**Keywords: soil survey, disaggregation, ecological site, digital soil mapping, podzolization,
red spruce, O-horizons, soil organic carbon**

Copyright 2015 Travis Nauman

ABSTRACT

Spatializing the Soil-Ecological Factorial: Data Driven Integrated Land Management Tools

Travis Nauman

Soils form the dynamic interface of many processes key to the function of terrestrial ecosystems. Many soil properties both influence and are influenced by activity of flora and fauna. Interactions between soils, biota, and climate determine the potential ecosystem services that a given unique ecological site (ES) can support, and how resilient a site is to various pressures and disturbances. Soil data are needed to fully understand how these factors interact, but because this data is difficult to obtain, existing soil maps are sometimes not detailed enough to fully explore relationships. Environmental raster GIS data layers were used to increase the detail of maps by representing soil forming factors and associated ecological pedomemory legacies important to understanding ecological potential. This dissertation presents methods and tools to help create these new soil maps at appropriate resolution and theme for field scale assessment of ecological sites that enable land managers to plan and implement appropriate management decisions.

USDA-NRCS soil surveys were disaggregated to higher resolution maps using a semi-automated expert training routine to implement a random forest classification model. This transformed soil map polygons of variable thematic and spatial resolution (soil map unit concepts) to a consistent 30-meter raster grid of unified theme (soil taxa). Disaggregated maps (DM) showed highly variable accuracy (25-75% overall validation accuracy) that mirrored that of the original soil surveys evaluated in Arizona (AZ) and West Virginia (WV). However, disaggregated maps expressed the soil data at a much more detailed spatial scale with a more interpretable legend. The WV surveys exhibited much lower accuracy than the AZ survey evaluated. This lower accuracy in WV is likely due to the forested setting and highly dissected landscape, two factors that create more intrinsic soil variability that is harder to explain with spatial covariates.

Ecological site descriptions (ESD) document soil-ecosystem groups that produce unique amounts and types of biological constituents and respond similarly to disturbance and environmental variation. ESD are linked to soil map unit components in USDA-NRCS soil surveys and are used as the basis for land management planning on rangelands and forestlands. The component level connection makes DM a good way to spatialize ESD because both are spatially represented at the same thematic level, whereas conventional soil maps have polygons that often have multiple components linked to a delineation.

However, in the evaluation of mapping ESD via DM, the DM turned out not to document the key difference in spodic soil properties that distinguished the important ecotone between northern hardwood and alpine red spruce conifer ESDs in Pocahontas and Randolph counties,

WV. So, to adjust, spodic soil properties were mapped directly using digital soil mapping approaches. A strong spatial model of spodic soil morphology presence was developed from a random forest probability model and showed correspondence to red spruce and hemlock occurrences in local historic land deed witness trees from records between 1752 and 1899. From this result, areas with spodic soil properties were assumed to be associated with historic red spruce communities, although 68% of those areas in the WV study area are currently under hardwood cover. This would seem to indicate that hardwoods have encroached on the historic extent of spruce, which is consistent with other recent studies. O-horizon thickness was also observed to be one cm thicker for every 10% greater importance value of red spruce or hemlock versus that of hardwood species at field sites. From these observations, it was calculated conservatively that at least 3.74-6.62 Tg of C have likely been lost from red spruce influenced ecological sites in WV due to historic disturbance related conversions of forest to hardwood composition. These results highlight the value of working within a soil-ecological factorial framework (e.g. an ESD) to contextualize land management options and potential derived services or negative consequences of each available action.

Dedication

I dedicate this work to those who seek knowledge in its base form. It is rare to see scientists that can go outside of their comfort zone and political sphere in the pursuit of knowledge. I feel like I have been privileged to work with individuals that have those qualities in West Virginia. Specifically, I dedicate this to the backwoods soil scientists whose curiosity about our living soils are the foundation for figuring out how we can live in harmony with our surroundings.

I also dedicate this work to my closest friends and family. The adventures we took to the amazing wild places of the western US inspired my curiosity in the natural world and have motivated my work. My parents instilled in me an intellectual independence and willingness to question and evaluate the status quo that has allowed me to contribute some small bit to science. It is this willingness to question and be questioned that allows progress to be made in the expansion of knowledge and wisdom.

I'm compelled to note my inspirational figures in natural science; those who have strived to conserve, preserve and protect wild spaces. Those individuals, the likes of John Muir, Henry Thoreau, Edward Abbey, and Aldo Leopold, recognized that we cannot disconnect ourselves from our wild pasts. These are visionaries who recognized that our destinies are tied to our wild places, and that our collective soul is not whole without the opportunity to connect with the natural world around us.

Acknowledgements

The time I've spent in West Virginia has been punctuated with people who have contributed to this work and the growth necessary for me to complete this story. My advisor, Dr. James Thompson, has been a mentor and supporter all along the way. His ability to give me intellectual space while still helping to guide me through the tedious processes of graduate school, conferences, publishing, and managing professional relationships has been a foundation of my success. Our many long conversations about the soil science and how to best tell the story of landscapes have helped mold me into a 'soil story-teller'. I cannot express my gratitude to him for his support through the variety of ups and downs that both of us have experienced.

My committee has been an absolute pleasure to work with and I thank them for being flexible and supportive through the evolution of this project. It has been great to have the forestry input of Dr. Rentch, the spatial ecology perspective of Dr. McNeil, the soil chemistry foundation of Dr. McDonald, and the remote sensing knowledge of Dr. Warner to draw from. This multi-disciplinary group never failed to inspire good conversation and ideas when gathered into one room!

I have also been surrounded by a cadre of excellent individuals who have provided input, support and sweat and blood to help produce this work. The field data that made this work possible was collected by a variety of individuals over the years. The WV soil partnership, led by the Natural Resource Conservation Service (NRCS) soil survey staff, produced an amazing amount of data during our annual field week events. This data proved invaluable in testing our hypotheses about red spruce ecological sites. Henry Liebermann and Aaron Burkholder were troopers during the summer 2013 field data collection efforts that resulted in our conclusions about forest floor controls. Working through swarms of biting flies, dense thickets of red spruce saplings, and sometimes suffocating heat, Henry and Aaron were dedicated and tough.

I'd also like to specifically acknowledge some key individuals on the Morgantown NRCS staff. A special thanks goes out to Jason Teets, Skip Bell, Tim Dilliplane, and Katey Yoast for their knowledge and leadership in telling the 'red spruce story', and for their help in classifying soil descriptions. Without their help intellectually and in the field, this work would not have been possible. I'd also like to recognize the leadership of David Kingsbury at the NRCS regional office in Morgantown. He does an excellent job galvanizing the soils community here.

The NRCS has also supported my work through the National Soil Survey Center (NSSC) and Geospatial Research Unit (GRU). These groups have financially and intellectually supported my efforts. I especially would like to recognize the support of NSSC director Jon Hempel and NSSC Soil Business Systems leader, David Hoover, for their guidance and help procuring research funds. Dr. Zamir Libohova of the NSSC was also a consistently helpful collaborator. Sharon Waltman and Tom D'Avello of the GRU were also invaluable resources when questions about soils data, database issues, and modeling arose. I am extremely fortunate to have colleagues with such extensive experience in soil mapping.

The Monongahela National Forest (MNF) has also been extremely supportive logistically, financially, and administratively in our work. Specifically Stephanie Connolly, the MNF soil scientist, has been an instrumental figure in building the research collaborative and financial support necessary to make this work happen. Her tenacity and support at all levels have produced a momentum behind this work that has made the implications ring loud and clear to the broader scientific and forestry communities. I'd also like to acknowledge the support and coordination provided by the Greenbrier Ranger District of the MNF. Jack Tribble and Shane Jones were instrumental in helping with field efforts during the summer of 2013. The MNF administration was also very supportive; and a special thanks goes out to Kent Karriker and Clyde Thompson for seeing the value of our efforts. We also received field and editing support from Dr. Greg Nowacki, regional forest service ecologist, whose expertise I am grateful for.

As we know from Heisenberg's uncertainty principal, we often disturb the phenomenon we are observing, and soil science is as guilty as any field in that respect. One of our challenges working in the MNF was respecting the rare and endangered species that make their home there. We had a particular challenge in not disturbing the endangered and endemic Cheat Mountain Salamander (CMS). Luckily, we were able to work with Dr. Thomas Pauley, the foremost expert in CMS habitat to minimize our impacts. I am deeply thankful to him for helping us to do our work in such a way as to maintain this precious resource.

Last, but not least, I'd like to thank the group of various student workers who have helped in the laboratory or field along the way with various, and often quite tedious tasks. This pool of talented young individuals includes Jimmy Leonard, Emily Wells, Adrienne Nottingham, and Emily Romano among others. I know this group will be the next cohort of dynamic soil scientists and do great things!

TABLE OF CONTENTS

1	Table and Figure Indices	x
1.1	List of Tables by Chapter	x
1.2	List of Figures by Chapter	xiii
1.3	List of Symbols, Abbreviations, or Nomenclature	xvii
2	Introduction	1
2.1	Document Summary	1
2.2	Conceptual Rational	1
2.3	Field scale conservation: integrating updated soil survey and forest ecology	3
2.3.1	Overview	3
2.3.2	Background	4
2.4	Summary of Research Objectives	7
2.4.1	Modernizing Soil Information	7
2.5	References	9
3	Semi-Automated Disaggregation of Conventional Soil Maps using Knowledge Driven Data Mining and Classification Trees	14
3.1	Citation	14
3.2	Highlights	14
3.3	Abstract	14
3.4	Introduction	17
3.4.1	Soil Survey Spatial Disaggregation	19
3.5	Materials and Methods	21
3.5.1	Study Area	22
3.5.2	SSURGO Training Areas	23
3.5.3	Terrace Height Raster Construction	24
3.5.4	Rule Matching for Training Area Identification	25
3.5.5	Model Implementation	27
3.5.6	Model Performance and Validation	29
3.6	Results	30
3.6.1	Training Set Agreement	30
3.6.2	SSURGO versus Disaggregation	31
3.6.3	Validation using independent pedons	32
3.6.4	Uncertainty in Predictions	34
3.6.5	Environmental Covariate Influence on Decision Tree Models	35
3.7	Discussion	36
3.7.1	Uncertainty	38
3.7.2	Future Soil Survey Applications	39
3.8	Conclusions	40
3.9	References	42
3.10	Tables and Figures	50

4	Semi-Automated Disaggregation of a Conventional Soil Map using Knowledge Driven Data Mining and Random Forests in the Sonoran Desert, USA	70
4.1	Citation	70
4.2	Abstract	70
4.3	Introduction	71
4.3.1	Soil Survey Spatial Disaggregation.....	73
4.4	Methods	75
4.4.1	Study Area.....	75
4.4.2	Training Set Creation	77
4.4.3	Model Building.....	80
4.4.4	Validation of Disaggregated Maps.....	86
4.5	Results	88
4.5.1	Uncertainty Map.....	90
4.5.2	Important Variables for Models	94
4.6	Discussion and Conclusions	98
4.6.1	Thematic Issues in Disaggregation.....	99
4.6.2	Uncertainty Assessment	100
4.6.3	Covariate Comparisons	102
4.6.4	Future Efforts	104
4.7	Acknowledgements	105
4.8	References	106
5	Ghosts of the forest: mapping pedomemory to guide forest restoration	116
	Citation	116
5.1	116	
5.2	Highlights	116
5.3	Abstract	116
5.4	Introduction	118
5.4.1	Soil pathways and pedomemory.....	118
5.4.2	Importance of Red Spruce Forests in the Central Appalachians.....	125
5.4.3	Digital soil mapping of podzolization.....	127
5.5	Materials and Methods	128
5.5.1	Study area.....	128
5.5.2	Data collection and analysis.....	129
5.5.3	Spatial modeling using DSM	133
5.6	Results	135
5.6.1	Soil profile data	135
5.6.2	Spatial models of spodic probability.....	135
5.6.3	Environmental controls on spodic probabilities.....	136
5.6.4	Witness tree comparison	136
5.6.5	Conifer importance and soil morphology.....	137
5.7	Discussion	138
5.7.1	Understanding historic red-spruce community distribution and spodic soil properties	140
5.7.2	Future implications.....	143
5.8	Conclusions	144
5.9	Acknowledgements	145

5.10	Literature Cited.....	146
5.11	Tables	161
5.12	Figures	165
6	Pedoecological modeling to guide forest restoration using ecological site descriptions	173
6.1	Citation	173
6.2	Abstract	173
6.3	Introduction	175
6.3.1	Podzolization Pathway and Soil Organic Carbon	176
6.3.2	Forest History and Ecological Change in the Allegheny Mountains of West Virginia.....	179
6.4	Materials and Methods	182
6.4.1	Study Area.....	182
6.4.2	Data Collection and Analysis	183
6.4.3	Pedoecological Mapping and Restoration Carbon Sequestration Estimates..	185
6.5	Results	189
6.5.1	Plot Data.....	189
6.5.2	Conifer Importance and O horizons	190
6.5.3	Pedoecological Mapping and O horizon Carbon Accumulation.....	191
6.6	Discussion	192
6.6.1	Timing of O horizon Accumulation	193
6.6.2	Implications of Red Spruce Restoration for Wildlife and Climate Change ...	194
6.7	Conclusions	197
6.8	Acknowledgements	198
6.9	Literature Cited.....	199
6.10	Tables	205
6.11	Figures.....	208
7	Chapter 6: Conclusions	214
8	Appendix A: Plot Diagram and related Data Collection.....	216
9	Appendix B Chapter 3 Data	218
10	Appendix C Chapter 4 Data	221
11	Appendix D Chapters 5 and 6 Data.....	223

1 TABLE AND FIGURE INDICES

1.1 *List of Tables by Chapter*

Semi-Automated Disaggregation of Conventional Soil Maps using Knowledge Driven Data

Mining and Classification Trees

Table 1. Two multi-component map units recorded in the Webster County soil survey, West Virginia (Delp, 1998).....	50
Table 2. Tables used from SSURGO for training set rule matching. An example of the language used to match with digital terrain raster values is shown.....	50
Table 3. Hillslope profile and curvature rule matches derived from the SSURGO database. Slope position is a 0-100 index of how far up a slope a location is with 100 being a summit and 0 being valley floor.....	51
Table 4: Component general landform descriptions from SSURGO and the interpreted environmental raster value rules for the landform descriptions. All unique combinations of the SSURGO geomorphic description queries are shown with on the left along with the translated rules for the terrain rasters used to select training areas.....	53
Table 5. List of environmental variables used for classification tree ensembles.....	57
Table 6. Validation agreement rates for different spatial neighborhoods matching criteria. ‘Series’ matches indicate exact class matches in the final ensemble prediction, ‘series or like’ means the percent of predictions that either matched the validation series or were a similar soil as defined in Section 2.4, ‘any tree’ refers to at least one of the trees in the ensemble predicting the correct series, and both the 5+ and 33+ refer to at least that many trees predicting the correct	

series. SSURGO matches compared validation pedons to the original SSURGO map units. If the validation pedon matched any of the named components in a map unit, it was deemed a match.....58

Table 7. Importance values showing how often all covariates were used in each disaggregation model. All models are averaged in right column. Standard deviation values indicate how much variation there was within the 100 tree ensembles. Variables are ordered from highest (top) to lowest by the averaged column (right).....59

Semi-Automated Disaggregation of a Conventional Soil Map using Knowledge Driven Data Mining and Random Forests in the Sonoran Desert, USA

Table 1. Component level rule matching for training set creation. Some components have more than one geomorphic description.....81

Table 2. List of raster covariate layers used for building tree models from selected training sets.....83

Table 3. Summarizes the distribution of validation points in original SSURGO map units (left), and the Updated disaggregation model (right).....87

Table 4. Layers chosen by PCA data reduction to most efficiently represent soil-landscape variability for soil mapping (from Nauman, 2009).....97

GHOSTS OF THE FOREST: MAPPING PEDOMEMORY TO GUIDE FOREST RESTORATION

Table 1. Description of spodic intensity (SI) classes based on observable field morphology.....161

Table 2. Spatial variables used to model spodic probability.....	162
Table 3. Confusion matrix from withheld 1/3 validation set for spodic probability predictions using a 0.57 threshold for classification as ‘spodic’	164

**PEDOECOLOGICAL MODELING TO GUIDE FOREST RESTORATION USING
ECOLOGICAL SITE DESCRIPTIONS**

Table 1. Description of spodic intensity (SI) classes based on observable field morphology.....	205
Table 2. Organic C calculations for O horizons from laboratory analysis used in restoration predictions of C sequestration. Standard deviations of averaged values are in parentheses.....	206
Table 3. Organic C accumulation calculations for SCF restoration scenarios. Estimates of total organic C stored in O horizons above 880 m, assuming the same areal proportions of states across those areas and same proportion of spodic areas (31%).....	207

Appendix B Chapter 3 Data

Table B.1 Soil classes at validation sites used to evaluate the disaggregation models in Chapter 3.....	218
---	-----

Appendix C Chapter 4 Data

Table C.1 Validation site data from Chapter 4 validation of disaggregated soil survey.....	221
--	-----

Appendix D Chapters 5 and 6 Data

Table D.1 Field transect observations used to train spodic probability model.....	223
Table D.2 Forest plot data.....	231
Table D.3 Laboratory data documenting Al and Fe acid oxalate extractions for different spodic intensity classes.....	232
Table D.4 O-horizon carbon stock data.....	234

1.2 *List of Figures by Chapter*

Introduction

Figure 1. Photos of Spodosols in the Allegheny Mountains of West Virginia. Photo on right is from the Pocahontas Soil Survey report (Flegel, 1999).....	5
---	---

Semi-Automated Disaggregation of Conventional Soil Maps using Knowledge Driven Data Mining and Classification Trees

Figure 1. Study area in West Virginia, USA. Points show pedon locations used in the validation of disaggregation results. Black outlines delineate the Pocahontas and Webster County soil surveys used to train the model. Results were extrapolated to nearby areas with consistent covariate data to incorporate more validation points. (Inset: Location of Pocahontas and Webster Counties in West Virginia.).....	60
Figure 2. The ruleset used to create the terrace raster from relative elevation layers of various neighborhoods. This was used for training area selection and in decision tree models.....	61
Figure 3. User’s versus producer’s accuracy summary figures. Values are expressed in fractions. Note the generally lower values of producer’s accuracy that require a larger x-axis scale in plots. Each class is represented by its parent material type as indicated in the legend.....	62

Figure 4. Bivariate plots comparing training accuracies versus the size of class training samples for different sampling schemes. Each class is symbolized by general parent material type.....63

Figure 5. Maps created from the models created from original proportions, square root, and base-ten logarithm sampling ensembles. Soil types are symbolized by unique colors and are consistent across all three maps.....64

Figure 6. Maps comparing the SqRt model and the original soil survey map units for Webster and Pocahontas County Soil surveys. Soil types (SqRt model) and map units (surveys) are symbolized by unique colors.....65

Figure 7. Map with SSURGO map units overlaying disaggregation predictions for the SqRt (square rootsampled) model for part of the study area.....66

Figure 8. Distributions of prediction frequencies for pixels that correctly predicted (matches) soil series and misclassified (misses) soil series for the nearest neighbor validation of the square root disaggregation model: (a) boxplots of all data; (b) distribution of prediction frequencies for misclassified points; (c) distribution of prediction frequencies for correctly classified points; (d) plot of ground truth probabilities for different ranges of prediction frequencies.....67

Figure 9. Maps showing translation of prediction frequencies to ground truth probabilities using the step function from Fig. 8. This extent covers some of the same data (SqRt model) extent as Figure 7.....68

Semi-Automated Disaggregation of a Conventional Soil Map using Knowledge Driven Data Mining and Random Forests in the Sonoran Desert, USA

Figure 1. ASTER satellite image of Organ Pipe Cactus National Monument and validation locations.....75

Plate 1. (a) Map of original soil survey map units. Percentages indicate slope gradient ranges attributed to map units. (b) Updated disaggregated map of soil series components.....89

Figure 2. Validation site probabilities compared for Updated model showing the higher tendency of prediction probabilities at matches or correctly predicted sites (a), histograms of prediction probabilities at missed sites (b) and matched sites (c), and the empirical relationship relating prediction probabilities and field data probabilities (d).....91

Figure 3. Field data probability representation of uncertainty for the Updated model based on step function in Fig. 2d.....92

Figure 4. Mean field data probabilities with standard deviation bars summarized by Updated model disaggregation components (left) and SSURGO map units (right).....93

Figure 5a. Average frequency of use of covariate layers (See Table 2 for definitions) used on average in more than 2% of tree nodes in the 500 trees in the Updated model. Black brackets give standard deviations of the frequencies to show how these varied over all ensemble trees.....96

Figure 5b. Same as 7a, but for the variables used from Nauman (2009).....97

GHOSTS OF THE FOREST: MAPPING PEDOMEMORY TO GUIDE FOREST RESTORATION

Figure 1. Well expressed podzol soil morphology in a red spruce forest in WV.....165

Figure 2. Study area and data collection locations overlaid on ArcGIS 10 National Geographic mapping baselayer of local features (roads, shaded relief, cities, landmarks).....166

Figure 3. Examples of site conditions, soil profiles, and acid oxalate data of the non-spodic hardwood ecological site (SI=0), spodic integrademixed forest (SI=1), and spodic conifer forest

(SI=2). Green line within graphs represent pictured soil profile. Pictures are of current vegetation at the pictured profile.....167

Figure 4. Spodic morphology probability map with witness tree points overlaid.....168

Figure 5. Classification tree showing how GIS variable splits can isolate more and less spodic groups of soil observations. Correct predictions over total node set size are shown under classification labels (e.g. spodic, 153/195 on upper right leaf). The confusion matrix of the fitted data is shown under the tree.....169

Figure 6. Spodic model probabilities at witness tree sites where no spruce or hemlock were recorded (top), and where spruce or hemlock were observed (bottom).....170

Figure 7. Relationships between conifer importance (CN) with spodic intensity (top) and O horizon.....171

Figure 8. Conceptual diagram of how climatic and topographic controls of red spruce appear to change over elevation.....172

**PEDOECOLOGICAL MODELING TO GUIDE FOREST RESTORATION USING
ECOLOGICAL SITE DESCRIPTIONS**

Figure 1. Study area soil map units (MU), areas of the Spodic Shale Uplands Conifer Forest (SCF) ecological site, and data collection locations overlaid on ArcGIS 10 National Geographic mapping baselayer.....208

Figure 2. State and transition model from spodic shale upland conifer forest (SCF) (Teets, 2013).....209

Figure 3. Flow chart of data analysis to create pedoecological areal estimate of SCF extent (Point data) and alternative states (Byers et al., 2013) as well as O-horizon accumulation (Pedon data and Plot Data).....210

Figure 4. Graph of fixed area plot SI values (x-axis) versus probabilities from the random forest spodic probability model (y-axis). The outlined area at the top of the plot delineates the SCF ecological site, and letters represent dominant tree composition groups (C = conifer, M = mixed, H = hardwood).....211

Figure 5. Relationships between conifer importance (CNIMP) with O horizon depth in at plots with SI of 1-2.....212

Figure 6. Pedoecological map of the different vegetative states within the spodic shale uplands conifer forest (SCF) ecological site. States are denoted by the different colors.....213

APPENDIX A: PLOT DIAGRAM AND RELATED DATA

COLLECTION.....216

1.3 List of Symbols, Abbreviations, or Nomenclature

Symbols, abbreviations and nomenclature are defined in text in the various articles as they are standalone articles.

2 INTRODUCTION

2.1 *Document Summary*

This dissertation is organized around four peer reviewed journal manuscripts. Preceding these articles is an introduction (Ch. 2) that outlines the broader proposed project. Chapters 3-6 are unformatted journal articles, three of which were published and one submitted for review at the time this dissertation was put together. The first two papers document disaggregating soil surveys in WV and AZ. The third and fourth paper document i) the spatial modeling of spodic soils and their connection with red spruce, and ii) the carbon implications of the current status of red spruce ecological sites in WV. All are in press except the carbon implications paper which was submitted to Soil Sci. Soc. Am. Journal on 2/10/15.

2.2 *Conceptual Rational*

Effective management of scarce natural resources demands data that is both accurate and precise. In most ecosystems, soil is a crucial interface for water and nutrient cycling, which we rely on for clean water and vegetation production. Understanding both the current spatial distribution of soils and how they function in the most accurate and detailed fashion is important to making decisions about the implications of various management options for land. Soil inventory surveys have been completed for many parts of the world, but have received criticism for their scope and spatial structure as users include soils data in more technical modeling (Burrough, 1989; Burrough et al., 1997; McBratney et al., 2003; Grunwald, 2009; Grunwald et al., 2011). The main criticism is that these conventional soil maps (CSM) use polygon delineations that enforce crisp boundaries that often do not reflect the more gradual nature of soil

spatial distributions. CSM also often lump more than one soil type into one mapping unit leaving users the task of ‘disaggregating’ the soil map within those polygon delineations (McBratney, 1998; Bui et al., 1999; de Bruin et al., 1999; Wielemaker et al., 2001; Bui, 2004; Thompson et al., 2010). The first portion of this dissertation will aim to help build methods to disaggregate CSM into more realistic and continuous raster maps of soil distribution that require less interpretation by users. Researchers have already laid a foundation for these methods (Zhu et al., 1996, 2010; Zhu, 1997; Bui et al., 1999; Bui and Moran, 2001; Qi et al., 2006; Smith et al., 2010; Thompson et al., 2010; Wei et al., 2010; Goovaerts, 2011; Haring et al., 2012; Kerry et al., 2012; Nauman et al., 2012, 2014; Nauman and Thompson, 2014), but there is still a need to provide more standardized and pragmatic procedures accessible to broader soil science professionals (e.g. CH 3: Nauman and Thompson, 2014). The first section of the proposed dissertation will focus on developing disaggregation methods using pilot studies in West Virginia and in Arizona to show applicability across a wide physiographic range.

The second section of the dissertation will focus on how to use disaggregated higher resolution soil maps to create detailed maps of soil-ecosystem states and communities for use in conservation planning. Ecological Site Descriptions (ESD) (Grazing Lands Technology Institute, 2003; USDA-NRCS, 2014) have been used by the USDA Natural Resources Conservation Service (NRCS) for use in conservation planning across the western U.S., and are being expanded to the eastern US. ESD are linked to soils mapped in CSM for purposes of providing a framework for conservation planning, but because of the problems with sometimes coarse and aggregated soil map units, conservation planners often have to remap lands when building a conservation plan (e.g., for grazing rotations or vegetation management). So, to spatially combine high resolution disaggregated soil maps (DM) maps with ESD data could provide a

much more efficient product to planners (Bestelmeyer et al., 2011). A pilot study to evaluate the feasibility and coherency of producing DM-ESD maps for conservation planning purposes was carried out in the Monongahela National Forest. The demands on the resources in this area for timber, mining, biodiversity conservation, and recreation provide an opportunity for use and trial of ESD. Co-located soil descriptions, forestry metrics, vegetative composition, and digital soil mapping covariates needed to carry out this production will also offer opportunity for insightful analysis of classic pedology questions in an ecological context. The data and ESD structure should help reveal systematic information about how the soils and vegetation have been co-evolving over time through massive area disturbances (Hopkins, 1899; Pielke, 1981) and climate change (NCADAC, 2013) in the last two centuries.

2.3 *Field scale conservation: integrating updated soil survey and forest ecology*

2.3.1 *Overview*

Linking more detailed DM to ESD (Grazing Lands Institute, 2003, Chapter 3; USDA-NRCS, 2014) could provide an appropriate spatial platform for better use in conservation planning. DM were compared to point based soil predictive models and ESD groupings in forest vegetation communities near Cheat Mountain, and surrounding areas representative of the Allegheny Highlands, WV. Analysis focused on areas thought to be historically dominated by red spruce (*Picea rubens*) communities and the associated organic-rich Spodosol and Histosol soils that are thought to be typical of historic climax communities in the area (Byers et al., 2010). Signatures in observed soils common in Spodosols will be used to help determine how red spruce were distributed prior to timber exploitation and associated fires in the last 200 years

(Pielke, 1981; Hopkins, 1899). Fixed area plots collecting forest stand production, stand basal area by species, stand age, stand regeneration, canopy structure, vegetative composition, surface cover and debris, and soil profile data were co-located in field work (See Appendix A). These data were used to build ES descriptions to help determine trends and management frameworks as well as provide insight into ecosystem services such as wildlife habitat, water quality, and carbon sequestration. These data were modeled spatially by correlating field data with digital terrain data, remotely sensed imagery, and other environmental spatial data. The data produced helped support area needs for habitat management of sensitive species including the northern flying squirrel (*Glaucomys sabrinus fuscus*) (Odom et al., 2001; Ford et al., 2004; Menzel et al., 2004; Menzel et al., 2006), Cheat Mountain Salamander (*Plethodon nettingi*) (Dillard et al., 2008a; Dillard et al., 2008b; Pauley, 2008), and general restoration efforts for red spruce communities (Byers et al., 2010; Rentch et al., 2007; Rentch et al., 2010; Schuler et al., 2002).

2.3.2 Background

The connections between the soil, flora, and fauna in ecosystems are often poorly documented by researchers who sometimes study each of these components separately due to a high degree of specialization amongst scientists. In the case of the Central Appalachians, we also have to consider the implications of widespread disturbance due to high intensity timber harvest and wildfire that often followed. Researchers have estimated that red spruce dominated forests once covered ~200,000 - 500,000 ha in West Virginia and Virginia before the middle 19th century, but have been reduced to current extents of roughly 20,000 ha in scattered patches (Hopkins, 1899; Pielke, 1981; Adams and Stephenson, 1989; Byers, 2010; Rollins et al., 2010). The almost complete cut-over and associated fires in the high elevation conifer forests in West Virginia and Virginia make it hard to determine how the historic forest functioned with respect

to wildlife habitat, carbon stocks, nutrient cycling, and sustainable forest productivity. This leaves great uncertainty in attempts to create management plans for these forests with regards to ecological services or resource harvest in the future.



Figure 1. Photos of Spodosols in the Allegheny Mountains of West Virginia. Photo on right is from the Pocahontas Soil Survey report (Flegel, 1999).

2.3.2.1 Vegetation links to podzolization

The small areas of forest in West Virginia that still have red spruce (*Picea rubens*), eastern hemlock (*Tsuga Canadensis*), and/or balsam fir (*Abies balsamea*) give us some clues into the structure and function of those past communities. Generally, these tree species, along with ericaceous shrubs like Rhododendron, and mountain laurel (*Kalmia latifolia*) favor a process of

reactions in the soil called podzolization that result in soils called Spodosols in U.S. Soil Taxonomy (see Figure 1; Soil Survey Staff, 1999). Spodosols are usually found in subalpine and boreal zones and are common in Canada and across New England, but also can be found in sandier soils of warmer humid regions like Florida, and in alpine regions of lower latitudes (Schaetzl and Isard, 1996; Lundström et al., 2000; Sauer et al., 2007). These soils generally form where low-base substrate favors vegetation types that produce nutrient poor litter (e.g., spruce). With these conditions, when ample moisture and cooler temperatures persist, these factors cause organic carbon build-up in the soil resulting in acidification that mobilizes Al, Fe, and sometimes Si in organometallic complexes and other forms. The mobilization of these compounds creates unique horizons or layers in the soil that persist for various lengths of time (Stanley and Ciolkosz, 1981; Lundström et al., 2000). We are proposing to use the more persistent and long-lived parts of these horizons as markers that can help to indicate the past influence of red spruce alpine conifer communities before the region was disturbed and often replaced by hardwood forest or grasslands.

Research has suggested that the process of podzolization is reversible, but that different parts of the resultant horizons degrade faster than others. Specifically, organic carbon and organometallic components of podzol horizons were observed to be lower in podzol soils that had lost white pine/northern hardwood forests in the northern Michigan peninsula as compared to similar areas that had regenerated that forest type (Barrett and Schaetzl, 1998). This region was logged and burned at a similar time to the Central Appalachians, and we think that a similar process of ‘depodzolization’ may be an influence in both locations.

Although depodzolization is thought to occur in areas that have lost the conifer/ericaceous carbon inputs, the large quantities of sesquioxides in B horizons (Barrett and

Schaetzl, 1998) and acid mycorrhizal weathering of sand particle pores (Jongmans et al., 1997; van Breemen et al., 2000; Schöll et al., 2008) are signatures that appear to be much longer lived than the organic molecules lost in Barrett and Schaetzl's (1998) findings. These properties could act as tracers for areas of relic Spodosol soils that formed under the pre-harvest conifer forests.

2.4 *Summary of Research Objectives*

2.4.1 *Modernizing Soil Information*

The overarching goal with this work is to help modernize soil mapping data for better integration into land management and modeling applications. Modern computing abilities and the proliferation of environmental raster spatial data have opened up new possibilities for precision land management. The difficulty and expense of sampling soils, and the complicated spatial distribution of soils, has historically limited map detail in CSM, and thus any analysis that needs soil data. A goal of this project is to promote a vision of efficient use of scarce field data to update soil maps via robust digital soil mapping techniques (e.g., Kempen et al., 2009; Yang et al., 2011; Kempen et al., 2012; Nauman et al., 2012; Nauman and Thompson, 2014), and to make updates in a pragmatic and multidisciplinary fashion that can be used by the widest possible audience. Updated soil maps of the project study area will be produced as a deliverable for this research and will be documented with appropriate reports to help others repeat this process. These maps will be independent validated with field data, and have associated measures of uncertainty in mapping predictions produced by these methods.

Another goal of this research is to improve soils data for conservation planning using Ecological Site Descriptions (ESD). ESD integrate rangeland management, forestry, ecology, wildlife, soils, hydrology, geology, and land management practices into descriptive frameworks

to guide decision making. This kind of integrated approach to natural resource management is a powerful tool that can not only help us make good decisions locally for the land, but help us to better understand how global trends in climate, societal growth, and resource demand might affect the diverse lands we have claimed stewardship over.

Local concerns regarding historic spruce-conifer community dynamics were clarified based on current spodic soil spatial models for comparison with other articles documenting Central Appalachian red spruce habitat (e.g., Pielke, 1981; Byers et al., 2010; Thomas-Van Gundy and Strager, 2012). The spatial extent and expression of spodic sesquioxide horizons were used to corroborate other research on historic red spruce community spatial distribution. It was hypothesized that timber harvest, fires and other factors facilitated loss of large areas of alpine conifer and resulted in significant carbon dioxide release into the atmosphere from the associated organic soils and burnt woody remains.

2.5 References

- Adams, H.S. and Stephenson, S.L., 1989. Old-Growth Red Spruce Communities in the Mid-Appalachians. *Vegetatio*, 85(1-2): 45-56.
- Barrett, L.R. and Schaetzl, R.J., 1998. Regressive Pedogenesis Following a Century of Deforestation: Evidence for Depodzolization. *Soil Science*, 163(6): 482-497.
- Bestelmeyer, B.T., D.P. Goolsby, and S.R. Archer. 2011. Spatial perspectives in state-and-transition models: a missing link to land management? *Journal of Applied Ecology* 48:746-757.
- Bui, E.N., 2004. Soil survey as a knowledge system. *Geoderma*, 120(1/2): 17.
- Bui, E.N., Loughhead, A. and Corner, R., 1999. Extracting soil-landscape rules from previous soil surveys. *Australian Journal of Soil Research*, 37(3): 495-508.
- Bui, E.N. and Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma*, 103(1-2): 79-94.
- Burrough, P.A., 1989. Fuzzy mathematical methods for soil survey and land evaluation. *European Journal of Soil Science*, 40(3): 477-492.
- Burrough, P.A., van Gaans, P.F.M. and Hootsmans, R., 1997. Continuous classification in soil survey: spatial correlation, confusion and boundaries. *Geoderma*, 77(2-4): 115-135.
- Byers, E.A., Vanderhorst, J.P. and Streets, B.P., 2010. Classification and Conservation Assessment of Upland Red Spruce Communities in West Virginia. In: W.V.D.o.N.R. West Virginia Natural Heritage Program (Editor). Wildlife Resources Section.
- de Bruin, S., Wielemaker, W.G. and Molenaar, M., 1999. Formalisation of soil-landscape knowledge through interactive hierarchical disaggregation. *Geoderma*, 91(1-2): 151-172.
- Dillard, L.O., Russell, K.R. and Ford, W.M., 2008a. Macrohabitat models of occurrence for the threatened Cheat Mountain salamander, *Plethodon nettingi*. *Applied Herpetology*, 5(3): 201-224.
- Dillard, L.O., Russell, K.R. and Ford, W.M., 2008b. Site-level habitat models for the endemic, threatened Cheat Mountain salamander (*Plethodon nettingi*): the importance of geophysical and biotic attributes for predicting occurrence. *Biodiversity and Conservation*, 17(6): 1475-1492.

- Ford, W.M., Stephenson, S.L., Menzel, J.M., Black, D.R. and Edwards, J.W., 2004. Habitat characteristics of the endangered Virginia northern flying squirrel (*Glaucomys sabrinus fuscus*) in the central Appalachian mountains. *American Midland Naturalist*, 152(2): 430-438.
- Goovaerts, P., 2011. A coherent geostatistical approach for combining choropleth map and field data in the spatial interpolation of soil properties. *European Journal of Soil Science*, 62(3): 371-380.
- Grazing Lands Technology Institute, 2003. National range and pasture handbook [computer file]. [Fort Worth, Tex.] : U.S. Dept. of Agriculture, Natural Resources Conservation Service, Grazing Lands Technology Institute, [1997].
- Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152(3-4): 195-207.
- Grunwald, S., Thompson, J.A. and Boettinger, J.L., 2011. Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues. *Soil Science Society of America Journal*, 75(4): 1201-1213.
- Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T. and Schröder, B., 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma*, 185–186(0): 37-47.
- Hopkins, A.D., 1899. Report on Investigations to Determine the Cause of Unhealthy Conditions of the Spruce and Pine from 1880-1893. In: W.V.A.E. Station (Editor). Fairmont Index Steam Print, Morgantown, WV.
- Jongmans, A.G. et al., 1997. Rock-eating fungi. *Nature*, 389(6652): 682-683.
- Kempen, B., Brus, D.J., Stoorvogel, J.J., Heuvelink, G.B.M. and de Vries, F., 2012. Efficiency Comparison of Conventional and Digital Soil Mapping for Updating Soil Maps. *Soil Sci. Soc. Am. J.*, 76(6): 2097-2115.
- Kempen, B., D.J. Brus, G.B.M. Heuvelink, and J.J. Stoorvogel. 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma* 151:311-326.
- Kerry, R., Goovaerts, P., Rawlins, B.G. and Marchant, B.P., 2012. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma*, 170: 347-358.
- Lundström, U.S., van Breemen, N. and Bain, D., 2000. The podzolization process. A review. *Geoderma*, 94(2–4): 91-107.
- McBratney, A.B., 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. *Nutrient Cycling in Agroecosystems*, 50(1-3): 51-62.

- McBratney, A.B., Santos, M.L.M. and Minasny, B., 2003. On digital soil mapping. *Geoderma*, 117(1-2): 3-52.
- Menzel, J.M., Ford, W.M., Edwards, J.W. and Menzel, M.A., 2004. Nest tree use by the endangered Virginia northern flying squirrel in the central Appalachian mountains. *American Midland Naturalist*, 151(2): 355-368.
- Menzel, J.M., Ford, W.M., Edwards, J.W. and Terry, T.M., 2006. Home range and habitat use of the Vulnerable Virginia northern flying squirrel *Glaucomys sabrinus fuscus* in the Central Appalachian Mountains, USA. *Oryx*, 40(2): 204-210.
- National Climate Assessment and Development Advisory Committee. 2013. National Climate Assessment Report (Draft). <http://www.globalchange.gov/what-we-do/assessment>
- Nauman, T.W., Thompson, J.A., 2014. Semi-Automated Disaggregation of Conventional Soil Maps using Knowledge Driven Data Mining and Classification Trees. *Geoderma*.
- Nauman, T., Thompson, J.A., Odgers, N. and Libohova, Z., 2012. Fuzzy Disaggregation of Conventional Soil Maps using Database Knowledge Extraction to Produce Soil Property Maps. In: B. Minasny, B. Malone and A. McBratney (Editors), *Digital Soil Assessments and Beyond: 5th Global Workshop on Digital Soil Mapping*, Sydney, Australia.
- Nauman, T.W., J.A. Thompson, and C. Rasmussen. 2014. Semi-Automated Disaggregation of a Conventional Soil Map Using Knowledge Driven Data Mining and Random Forests in the Sonoran Desert, USA. *Photogrammetric Engineering & Remote Sensing* 80:353-366.
- Odom, R.H., Ford, W.M., Edwards, J.W., Stihler, C.W. and Menzel, J.M., 2001. Developing a habitat model for the endangered Virginia northern flying squirrel (*Glaucomys sabrinus fuscus*) in the Allegheny Mountains of West Virginia. *Biological Conservation*, 99(2): 245-252.
- Pauley, T.K., 2008. The Appalachian Inferno: Historical Causes for the Disjunct Distribution of *Plethodon nettingi* (Cheat Mountain Salamander). *Northeastern Naturalist*, 15(4): 595-606.
- Pielke, R.A., 1981. The Distribution of Spruce in West-Central Virginia before Lumbering. *Castanea*, 46(3): 201-216.
- Qi, F., Zhu, A.X., Harrower, M. and Burt, J.E., 2006. Fuzzy soil mapping based on prototype category theory. *Geoderma*, 136(3-4): 774-787.

- Rentch, J.S., Schuler, T.M., Ford, W.M. and Nowacki, G.J., 2007. Red spruce stand dynamics, simulations, and restoration opportunities in the central Appalachians. *Restoration Ecology*, 15(3): 440-452.
- Rentch, J.S., Schuler, T.M., Nowacki, G.J., Beane, N.R. and Ford, W.M., 2010. Canopy gap dynamics of second-growth red spruce-northern hardwood stands in West Virginia. *Forest Ecology and Management*, 260(10): 1921-1929.
- Rollins, A.W., Adams, H.S. and Stephenson, S.L., 2010. Changes in Forest Composition and Structure across the Red Spruce-Hardwood Ecotone in the Central Appalachians. *Castanea*, 75(3): 303-314.
- Sauer, D. et al., 2007. Podzol: Soil of the Year 2007. A review on its genesis, occurrence, and functions. *Journal of Plant Nutrition and Soil Science*, 170(5): 581-597.
- Schaetzl, R.J. and Isard, S.A., 1996. Regional-scale relationships between climate and strength of podzolization in the Great Lakes Region, North America. *Catena*, 28(1-2): 47-69.
- Schöll, L. et al., 2008. Rock-eating mycorrhizas: their role in plant nutrition and biogeochemical cycles. *Plant and Soil*, 303(1-2): 35-47.
- Schuler, T.M., Ford, W.M. and Collins, R.J., 2002. Successional dynamics and restoration implications of a montane coniferous forest in the central Appalachians, USA. *Natural Areas Journal*, 22(2): 88-98.
- Smith, S., Bulmer, C., Flager, E., Frank, G. and Filatow, D., 2010. Digital Soil Mapping and multiple scales in British Columbia, Canada, 4th Global Workshop on Digital Soil Mapping, Rome, Italy.
- Soil Survey Staff, 1999. Soil Taxonomy A Basis System of Soil Classification for Making and Interpreting Soil Surveys. In: U.S.D.o.A.N.R.C. Service (Editor). U.S. Government Printing Office, Washington.
- Stanley, S.R. and Ciolkosz, E.J., 1981. Classification and Genesis of Spodosols in the Central Appalachians¹. *Soil Sci. Soc. Am. J.*, 45(5): 912-917.
- Thomas-Van Gundy, M., M. Strager, and J. Rentch. 2012. Site characteristics of red spruce witness tree locations in the uplands of West Virginia, USA. *The Journal of the Torrey Botanical Society* 139:391-405.
- Thompson, J.A., Nauman, T., Odgers, N., Libohova, Z. and Hempel, J., 2012. Harmonization of Legacy Soil Maps in North America: Status, Trends, and Implications for Digital Soil Mapping Efforts. In: A. McBratney, B. Minasny and B. Malone (Editors), *The 5th Global Workshop on Digital Soil Mapping. Digital Soil Assessments and Beyond*, Sydney, Australia.

- Thompson, J.A. et al., 2010. Regional Approach to Soil Property Mapping using Legacy Data and Spatial Disaggregation Techniques, 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia.
- USDA Natural Resources Conservation Service, 2014. National Ecological Site Handbook.
- van Breemen, N., Lundström, U.S. and Jongmans, A.G., 2000. Do plants drive podzolization via rock-eating mycorrhizal fungi? *Geoderma*, 94(2-4): 163-171.
- Wei, S. et al., 2010. Digital Harmonisation of Adjacent Soil Survey areas - 4 Iowa Counties, 19th World Congress of Soil Science, Soils Solutions for a Changing World, Brisbane, Australia.
- Wielemaker, W.G., de Bruin, S., Epema, G.F. and Veldkamp, A., 2001. Significance and application of the multi-hierarchical landsystem in soil mapping. *Catena*, 43(1): 15-34.
- Yang, L. et al., 2011. Updating Conventional Soil Maps through Digital Soil Mapping. *Soil Science Society of America Journal*, 75(3): 1044-1053.
- Zhu, A.X., 1997. A similarity model for representing soil spatial information. *Geoderma*, 77(2-4): 217-242.
- Zhu, A.X., L.E. Band, B. Dutton, and T.J. Nimlos. 1996. Automated soil inference under fuzzy logic. *Ecological Modelling* 90:123-145.
- Zhu, A.X., L. Yang, B. Li, C. Qin, T. Pei, and B. Liu. 2010. Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma* 155:164-174.

3 SEMI-AUTOMATED DISAGGREGATION OF CONVENTIONAL SOIL MAPS USING KNOWLEDGE DRIVEN DATA MINING AND CLASSIFICATION TREES

3.1 *Citation*

Nauman, T.W., and J.A. Thompson. 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. *Geoderma* 213:385-399. Reprinted from *Geoderma* under license #3603870298207 with permission from Elsevier.

3.2 *Highlights*

- Multiple conventional soil maps were disaggregated to a digital soil series map with no new field observations.
- Disaggregation and original soil maps had similar accuracies with validation pedon soil series matching at 39%-44% of locations within 60-meter neighborhoods.
- Uncertainty was characterized spatially for updated map.
- We present a repeatable methodology for updating and harmonizing conventional soil maps.

3.3 *Abstract*

Disaggregation of conventional soil surveys has been identified as a potential source for much of the next generation of model-ready digital soil spatial data. This process aims to apportion vector soil surveys into raster (gridded) representations of the component soils that are often aggregated together in map unit designs. Most soil surveys are published with some description of the soil-landscape relationships that distinguish component soils within map units. We used these descriptions found in the Soil Survey Geographic (SSURGO) database of Webster and Pocahontas Counties in West Virginia, USA, to build a set of representative training

areas for all soil components by using 1-arc second digital elevation data and derived geomorphic indices. These training areas were then used in classification tree ensembles with a more extensive environmental database to transform the original SSURGO map into a gridded soil series map. We created underlying prediction frequency surfaces from the models that can be used for creating continuous representations of soil class and property distributions.

Disaggregation models matched training sets in 71%-74% of pixels and matched components in original SSURGO map units in 56%-65% of the study area. We evaluated both the original SSURGO data and our models using 87 independent pedons not used in model building. Validation pedons matched components in SSURGO map units at 39% of sites, but in map units that only included one named component (as opposed to multiple soils that could be matched to validation pedons) only 27% of sites matched. Disaggregation predictions matched validation pedon classes 22-24% of the time using nearest neighbor spatial matches, and these rates increased to 39-44% for correct predictions within a 60-meter radius of the pedon. To characterize uncertainty, we compared relative ensemble prediction frequency (probability) of final hardened model classes at validation sites. Sites with correct predictions had generally higher prediction frequencies; which lead us to use them to create an uncertainty model. Uncertainty was calculated by determining the rate of correct predictions at validation sites within different intervals of prediction frequencies using nearest neighbor validation results. We were able to discern four uncertainty classes with values of 7%, 18%, 20% and 43%, which we called “ground truth probabilities”. We present the methods to create these models with the aim of making them more accessible to soil science professionals, and we think there is much potential to use them to aid in updating national soil survey inventories.

Keywords: Polygon disaggregation; Spatial disaggregation; Soil–landscape models;
Classification trees; Uncertainty; Variable importance

3.4 Introduction

Soil properties and soil functions influence many of the problems facing society today. Soil is a primary storage mechanism for carbon and nutrients that control how vegetation grows and how our climate is changing. However, our knowledge of soils is imprecise, with estimates of global soil carbon stocks in the top meter of soil that range from 1400 to 3250 petagrams (Grunwald et al., 2011). In light of the projected challenges of global warming and maintaining natural resource services like crops and clean water (IPCC, 2007), high quality soils information is key to making sustainable decisions. Although many soil inventories have been carried out around the world, the scope and spatial structure of these have been criticized (Burrough, 1989; Burrough et al., 1997; McBratney et al., 2003; Grunwald, 2009; Grunwald et al., 2011) as more and more researchers use soils data in environmental, agricultural, and engineering related models. Many studies try to improve on past soil inventories using digital soil mapping and related methods (Cook et al., 1996; Zhu, 1997; Zhu et al., 1997, 2001; Bui et al., 1999, 2006, 2009; de Bruin et al., 1999; Bui and Moran, 2001; Moran and Bui, 2002; Hansen et al., 2009; Kempen et al., 2009; Thompson et al., 2010; Yang et al., 2011; Häring et al., 2012; Kerry et al., 2012; Nauman et al., 2012). The *GlobalSoilMap* project (www.globalsoilmap.net) is a recent effort to help produce standard basic soil property maps for the whole world that can be used in more modern contexts (Sanchez et al., 2009; Hartemink et al., 2010). The *GlobalSoilMap* consortium has recognized that methods to best utilize old maps for production of new digital models is one of the best ways to begin creating new and more detailed soil maps (Minasny and McBratney, 2010).

One of the main challenges to providing appropriate data is that the classic paradigm of soil survey is management based, and properties attributed to soils are most often estimates based on sparse data at representative locations, not quantifications based on statistics. A large part of the goals of the original design of these maps was to determine suitability or hazards to human activities. These interpretations provide pragmatic initial guidance to developers, farmers, and other land management institutions for issues like road building, septic tank evaluations, and many other uses (Soil Survey Staff, 1993). The soil survey was supposed to be a starting point in planning and general management, but more current users have stretched far beyond these original concepts (Soil Survey Staff, 1993; Bouma, 1989).

Many studies have used soil survey spatial data with property estimates as inputs into models (e.g., Wilson et al., 1993; Lineback Gritzner et al., 2001; Bandaragoda et al., 2004; Causarano et al., 2008; Gatzke et al., 2011; Zhang et al., 2011). In the U.S., both the U.S. General Soil Map (STATSGO2: 1:250,000 to 1:1,000,000 scale) and the Soil Survey Geographic (SSURGO: usually close to 1:24,000 scale, but varies between roughly 1:1,000 and 1:250,000 depending on land use) databases often aggregate multiple soil classes into spatial polygon delineations used in maps (Soil Survey Staff, 1993; Thompson et al., 2012). The data model for SSURGO, which is the primary high resolution soil inventory for the US, includes polygonal map units with generally one to four named soil series (soil taxonomic class) per map unit, plus minor inclusions of soils or non-soil areas, which are sometimes but not always fully documented. This aggregation, and the inherently crisp breaks that choropleth style mapping imposes on spatial data, make spatial representation of estimated soil properties (e.g., soil texture, organic matter, pH) somewhat convoluted and predisposed to artifacts. For example, there are often distinct changes in property values between polygons or at survey project

boundaries that do not make logical sense (Loerch, 2012; Thompson et al., 2012). The problem that now emerges is how to use the wealth of information in legacy soil surveys in an appropriate way. Part of the answer might be to restructure the data to more appropriately address current applications, and one basic step to doing that is to spatially disaggregate the information into its component parts in a manner that better represents how soils truly occur in the field. This paper illustrates a technique to use widely available elevation, lithology, and remote sensing data to disaggregate two existing adjacent soil surveys in West Virginia, USA, into one continuous soil series class map using no new soil field data. This process potentially reveals much more information about spatial soil distribution and spatially harmonizes somewhat disjoint mapping projects that have artifacts along their boundaries (Thompson et al., 2010, 2012; Nauman et al., 2012).

3.4.1 *Soil Survey Spatial Disaggregation*

The primary focus of soil survey disaggregation is to express the spatial distribution of soil individuals in cases where older soil maps have lumped them into one spatial unit (Table 1). Another way to describe it would be the enhancement of a prior generalized soil map to produce a more detailed map that spatially distinguishes soil properties or types at a greater level of detail. Generally these techniques also tend to translate the data from object-based polygon maps to grid-based raster formats by using new point or environmental maps (e.g. DEM or Satellite Imagery) as predictors to map within polygons. Disaggregation has been identified as a conceptual approach to translate current data into formats compatible with modern needs and with pedologic concepts of soil formation (McBratney, 1998; Bui et al., 1999; de Bruin et al., 1999; Bui and Moran, 2001; Wielemaker et al., 2001; Bui, 2004). Generally, approaches use new pedon data and/or environmental covariate data to determine how soils within polygon map units

vary spatially. Approaches tend to draw from digital soil mapping frameworks (McBratney et al., 2003; Scull et al., 2003; Grunwald, 2009; Grunwald et al., 2011) that employ a state-factor theory of soil formation summarized by Jenny (1941).

Spatial disaggregation of multi-component soil map polygons into individual component soil classes has been demonstrated in attempts to universally update soil maps (Bui and Moran, 2001; Hansen et al., 2009; Wei et al., 2010; Smith et al., 2012), and to create class distinctions within the bounds of original survey map units (e.g. Bui and Moran, 2001; Thompson et al., 2010; Häring et al., 2012). Other studies have looked at disaggregating polygons for specific soil properties using conventional soil survey. Goovaerts (2011) evaluated geostatistical methods that can combine point data with choropleth data to look at intra-polygon variation in a specific variable, and Kerry et al. (2012) applied parts of these methods to soil organic carbon mapping in northern Ireland. Fuzzy logic has been used in disaggregation through applications like SoLIM (Zhu et al., 1996, 2010; Zhu, 1997; Qi et al., 2006) to help organize and implement soil-landscape relationships for mapping soils. SoLIM has been used in coordination with both expert knowledge (Smith et al., 2010) and statistical approaches (Yang et al., 2011) to implement discovered soil-landscape relationships used in updating and disaggregating soil maps. Other fuzzy knowledge systems have leveraged landform element classifications to better disaggregate landscapes into units with similar soils (MacMillan et al., 2000). Landform elements have also been combined with other ecological mapping and environmental maps using expert fuzzy logic rules to create ecosystem maps that incorporate soils information (MacMillan et al., 2007). Classification and regression trees have also been a prominent technique used in disaggregation. Bui et al. (2001) and Wei et al. (2010) both used ensembles of decision trees and Haring et al. (2012) used random forests to refine soil and surficial geology classes. Tree-based models have

also been used extensively in general digital soil mapping applications and seem to have the greatest flexibility of common modeling methods (McKenzie and Ryan, 1999; Moran and Bui, 2002; Scull et al., 2005; Saunders and Boettinger, 2007; Schmidt et al., 2008; Bui et al., 2009; Behrens et al., 2010a,b; Lemercier et al., 2011).

The objective of this research was to identify a pragmatic and repeatable method for systematic disaggregation of legacy soil maps. This technique addresses the common situation where an older soil map is available, but more detailed soil spatial data is needed, and too few new soil observations are available to use in geostatistical approaches or for building empirical models. We utilize soil-landscape rules that are usually present in soil survey database map unit descriptions in combination with a classification tree ensemble with different randomization schemes to universally disaggregate two adjacent soil survey projects into one harmonized soil series map. This approach captures both implicit and explicit expert knowledge about soil-landscape relationships in SSURGO and pairs that with available elevation, imagery, and geology data in a classification tree ensemble model. We selected methods and data sources based on repeatability, transparency, and manageability in an effort to make them accessible to soil science professionals in government and consulting.

3.5 *Materials and Methods*

We demonstrate and evaluate disaggregation methods in two adjacent county soil surveys in West Virginia. Soil-landscape rules were first extracted from the SSURGO database from every component (soil) of every map unit in the surveys. The landscape rules were then matched to value ranges of different DEM-based topographic metrics (e.g., slope position, landform element) to identify representative areas in each map unit for each component. These representative areas were then used as training sets for randomized classification tree ensembles.

We created several different randomized sampling techniques for balancing the relative sizes of training sets of different soil classes so that the training set size distribution was not too skewed, which facilitated detection of less extensive soil types. We evaluated three different ensemble models using an independent set of 87 geo-referenced pedons gathered from other local projects. Prediction frequencies from the most balanced ensemble model were also evaluated for use as an empirical proxy of prediction uncertainty..

3.5.1 *Study Area*

The study area is located in the Appalachian mountains of West Virginia, USA, and covers approximately 3,877 km² (Fig. 1). It includes data from two separate soil surveys completed by the USDA-NRCS for Webster (Delp, 1998) and Pocahontas (Flegel, 1998) counties. It includes parts of two U.S. Major Land Resource Areas: the Eastern Allegheny Plateau and Mountains and the Southern Appalachian Ridges and Valleys (Flegel, 1998). The Eastern Allegheny Plateau and Mountains makes up roughly the northwestern two thirds of the area, and is a highly dissected, level-bedded sedimentary plateau that includes the highest mountains in West Virginia. The Ridge and Valley province is a sequence of trellis ridges and valleys that run south-southwest to north-northeast consisting of more altered and folded sedimentary rocks. Soils in the study area generally form in residual sedimentary rocks and colluvial deposits along slope sequences (Delp, 1998; Flegel, 1998). The area is generally a rolling to steep terrain with only smaller flat areas along drainages and in limited areas where ridge tops or mountain tops are flat. There are alluvial soils along drainages, but most alluvial valleys are less than a mile wide. Most drainages in these areas are steep and narrow valleys that do not promote much alluvial deposition.

Soil profile descriptions at 87 locations classified according to U.S. Soil Taxonomy (Soil Survey Staff, 2010) were used to independently validate model results. Of these, 62 came from a

previous study (S. Roecker, unpublished data), and 25 came from the USDA-NRCS national pedon database (National Cooperative Soil Survey, 2012). The NRCS pedons come from a variety of collection dates and project and thus have somewhat variable spatial accuracy (some points were digitized from old manual topographic map notes). The pedon locations used from Roecker's work were likely more accurate as they were collected with a Trimble GeoXT, but were also still collected in a forest setting where accuracies are probably at least two meters. These observations were used solely for validation and not in any part of the model building process.

3.5.2 *SSURGO Training Areas*

The SSURGO dataset consists of a polygon format vector map attributed with a map unit label and a relational database that connects the map units to information about the soils and survey area. There is usually an associated hardcopy survey manuscript that was published for survey project areas (usually counties). The mapping infrastructure in SSURGO includes multiple types of map units that generally have one to four named soil series components as well as 'inclusions' of other soils or non-soil areas. Each of these component soil series can have different property distributions that must be generalized or aggregated somehow if a user wants to display a soil property using SSURGO polygons (e.g., Bliss et al., 1995; Thompson and Kolka, 2005).

In SSURGO, each component of every map unit has information regarding soil properties and environmental context attributed to it (Tables 2, 3, and 4; rule-matching examples shown in Section 2.2.2). By querying the geomorphic and landform related attributes in SSURGO, soil-landscape relationship descriptions were extracted from the database to help determine where within a map unit a component is expected to occur (e.g., Gilpin series exists on the upper third of mountain flanks). The language in these queried descriptions was then matched to values in

environmental rasters that represent hillslope position (0-100 index; Hatfield, 1996), landforms (Schmidt and Hewitt, 2004), terrace height (relative elevation with reference to local minimum in alluvial map-units; described below), percent slope gradient, slope aspect, and catchment areas (Tarboton, 1997) within each map unit. As an example, the descriptor ‘upper third of mountain flanks’ was associated with a hillslope position index (Table 4) raster by specifying that the soil exists on hillslope index values between 66 and 95. This essentially translates soil-landscape relationship records in the database to environmental raster values. This rule translation streamlines the approach Thompson et al. (2010) used to create soil-landscape rulesets. All rules identified typical landscapes for respective component soils within each map unit. These areas were added to a training set that was compiled for all soil series and other named survey components (e.g., rock outcrops or higher taxa such as Fluvaquents) in the study area. Areas from all map units that were typical of a given soil series were combined into one training class. All environmental rasters used in rule-matching were derived from the 1-arc second USGS national elevation dataset (NED) (Gesch et al., 2002; Gesch, 2007). Final maps and other raster data used in later steps were co-registered to the NED grid in a North American Datum of 1983 Universal Transverse Mercator projection in Zone 17-North.

3.5.3 *Terrace Height Raster Construction*

For the SSURGO rulesets described in section 2.2, a terrace raster was built from the NED data for distinguishing alluvial soils by using a combination of neighborhoods to look at the difference of each pixel from local minimum elevations depending on slope position, slope gradient, and Morphometric Protection Index (PI) using a 2000-meter radius (Yokoyama et al., 2002; Conrad and Wichmann, 2011). This approach uses differing neighborhood sizes in similar ways to other soil mapping studies (Moran and Bui, 2002; Hansen et al., 2009; Behrens et al.,

2010, 2010b). Different neighborhoods were used to distinguish terraces in wider alluvial systems versus terraces in narrow high gradient drainages (Fig. 2). The ‘terrace’ raster created is actually a stratified relative elevation index that determines how high a pixel is in relation to its neighbors within the context of its slope position, slope gradient, and PI (e.g., for a headwater stream, a smaller neighborhood radius is used for calculating relative elevation than a wide floodplain). This terrace height raster was created mainly to distinguish levels of terraces in alluvial areas during training area selection for alluvial map units, but also was used for decision tree modeling because it seemed to capture variations in landforms well in non-alluvial areas. The relative elevation calculations used in the terrace served to distinguish finer scale high spots or benches in upland locations making it useful beyond the alluvial units.

3.5.4 *Rule Matching for Training Area Identification*

Four main tables in the SSURGO database can be queried to develop geomorphic and hillslope profile descriptors. These were used to create two logic strings per geomorphic description for rule creation (Tables 2, 3, and 4). Table 2 shows an example of how these descriptors are queried from these tables using joins for the Dekalb soil series in the Gilpin-Dekalb complex map unit. In this case the geomorphic descriptors from the CoGeomorphDesc and CoSurfMorphGC tables indicate that mountain tops on ridges are typical locations for Dekalb. The curvature and hillslope descriptors from the CoSurfMorphSS and CoSurfMorphPP similarly indicate that Dekalb is found on summits with linear curvatures. All unique combinations of records from the *cogeomorphdesc* and *cosurfmorphgc* tables were linked together to create general landform element descriptions that were used for all mapped components (Table 4). Tables *cosurfmorphss* and *cosurfmorphpp* were also combined in unique combinations to create hillslope position and curvature descriptions (Table 3). One logic

string of environmental raster value ranges was created for each of these two descriptions. Value ranges for all rasters that pertained to each description were strung together in an ‘AND’ statement that required any grid cell to meet all the rules for each raster to become a training cell. Then the logic statements from each description were linked together in an ‘OR’ logic string allowing inclusion of pixels that met either of the ‘AND’ strings. Many soil series had more than one ruleset for both the general landform and/or hillslope position and curvature descriptions either due to presence of the soil as a component in multiple map units or because a component of that soil type had more than one set of landform or hillslope descriptions linked in the database. The logic strings that came out of these descriptions within linked records were combined with a logical ‘OR’ so that a cell could meet one set of rules (e.g., general landform) or the other (e.g., hillslope position and curvature) to become a potential training cell for the component linked to those descriptions in the database. The multiple strings of logic were put into single statements by soil series (same component name) by constraining each ‘AND’ statement to the original map unit of the respective component and then stringing common soil series rules together by logical ‘OR’ connectors. The translation of these rules to raster values was done by creating a list of unique instances of the descriptions that occurred for all the components through the study area. Fifty eight unique rules were created for geomorphic landforms (Table 4) and thirty two from the hillslope profile and curvature (Table 3) logic sets.

Once a list of rules was created for all the components in the survey area, training areas were created for all uniquely named components. Training areas were created for 50 soil series, 4 higher taxa classes, and two non-soil areas (water and rock outcrops) for Pocahontas and Webster counties in West Virginia, USA. There were three soils (Sees, Lodi, and Medihemists) that were mapped as single-component map units in limited areas that did not produce enough

training pixels to be detectable in preliminary single tree models. For these cases, the full extent of all of the map unit delineations for each of these soils was used for training. Because each soil series training set was built independently of the others, much of the training areas in multi-component map units overlapped, and this was addressed with the sampling design used for the decision tree ensemble method in the modeling stage of the project.

3.5.5 *Model Implementation*

Series training areas were randomly sampled with replacement to train 100 decision tree models to produce an ensemble model. To address a range of training class sizes and overlap in some of the training areas between soil series, it was deemed necessary to adjust the number of pixels selected from the training set for tree building for each soil series or class to be proportional to the original relative area of each series computed from the SSURGO component percentages similar to Moran and Bui (2002). We sampled at 1% of the SSURGO derived proportional areas for tree building. Class sample sizes averaged 24,595 pixels with a large standard deviation of 50,372 due to a wide range of soil class area extents.

Based on experimental trials with data and the use of plurality in decision tree algorithms (Breiman, 1984), we suspected that proportionality would influence detectability of classes. For this reason we tested three different sampling schemes to determine how scaling the relative training area proportions of the classes to be predicted would affect results. We transformed the original SSURGO area estimations of soil class extents to a square root (SqRt: ave. class size = 21454 pixels; st. dev. = 21638) and base-ten logarithm (Log10: ave. class size = 20079 pixels; st. dev. = 2027) to allocate proportions of training class sizes to create two more ensemble models to compare results with the ensemble using original class proportions (Orig). The SqRt and

Log10 transformed proportions were multiplied into the total study area size and then divided by 100 to produce new sample size for each class.

A more exhaustive set of environmental rasters were used for the classification (Table 5). The imagery chosen for use, Landsat Geocover, is a mosaic Landsat product offered by the USGS for 1990 (MDA, 2004b) and 2000 (MDA, 2004a) with Band 7 (mid-infrared), Band 4 (near-infrared), and Band 2 (visible green) spectra. These mosaics were summarized using principal components analysis in Erdas Imagine (Erdas, 2010) into two components per image that represented almost all of the variance in each scene. All terrain-based rasters were derived from the 1 arc-second USGS National Elevation Dataset (Gesch et al., 2002; Gesch, 2007).

3.5.5.1 Decision Tree Classification

Tree-based machine learning techniques have shown great potential in the modeling of ecology and soil systems (Bui and Moran, 2001; Moran and Bui, 2002; Henderson et al., 2005; Bui et al., 2006; Minasny and McBratney, 2007; Schmidt et al., 2008; Hansen et al., 2009; Behrens et al., 2010b). Generally, these algorithms recursively split a dataset by picking breaks in covariate data that help to purify or increase the information content of the model nodes (branches) (Breiman, 1984; Pedregosa et al., 2011). The Scikit Machine Learning Tree module was used in Python for decision tree implementation and follows a CART implementation (Pedregosa et al., 2011). The algorithm as we implemented in the Tree module uses Gini's impurity to measure the quality of splits for tree building. Gini impurity is a measure of the heterogeneity of classes at a node and is minimized in the tree building process to try and create leaf nodes with just one class, or a Gini impurity value of zero (Breiman, 1984). We conducted an informal sensitivity analysis with the parameters controlling maximum tree depth and minimum node split sample size to try and balance model agreement with training data without over-fitting the tree (with too many

branches). With consideration for the large number of training classes and a complicated and geologically stratified study area, a maximum tree depth of 20, a minimum number of samples to attempt a split of 20, and a minimum leaf size of 5 were chosen for tree building.

3.5.6 *Model Performance and Validation*

Evaluation of the model was done with (i) training set agreement, user's accuracies, and producer's accuracies (Congalton, 1991) of individual trees, (ii) overall comparison of predictions to SSURGO, and (iii) an independent validation dataset of spatially referenced soil pedon data. Our predictions and SSURGO were both compared to the independent pedons to gauge accuracy. Simple point sampling (nearest neighbor) and 60-meter radius neighborhood spatial supports were used in comparisons to allow for some error in the spatial referencing of covariates, SSURGO, and pedons.

Validation in these spatial supports also included determining if predictions were of morphologically similar soils. A similar soil was defined by the following criteria: same parent material type (i.e., alluvium colluvium, residuum, or mine fill), same soil depth class (or within 10 cm), same texture class in control section or within 15% for all fractions (including rock fragments), same or similar drainage class (within one class), and similar horizonation. All criteria were based on U.S. Soil Taxonomy definitions (Soil Survey Staff, 2010).

Predictions of the individual grid cell coregistered with each validation pedon were evaluated for agreement and also for confidence based on the number of trees (out of 100) that predicted the majority class reported for final classification. The agreement between pedon series and the final ensemble predicted series is a strict evaluation of overall classification accuracy. Evaluating the number of correct underlying tree predictions was used to try to estimate uncertainty. We expected prediction counts referred to henceforth as prediction frequencies, to

be higher for correctly predicted pixels. We tested to see if correct ensemble predictions had higher prediction frequencies, a scenario we thought would support using the counts to create ground-truth probabilities that represent confidence or uncertainty in each grid cell ensemble prediction. This use of prediction frequencies essentially amounted to creating an empirical uncertainty model.

3.6 *Results*

3.6.1 *Training Set Agreement*

Overall tree model agreements with training data averaged 74% for the original (Orig) sampling design, 71% for the square root (SqRt) sampling design, and 72% for the base ten logarithm (Log10) sampling design. These represent the average overall training accuracy for all 100 trees in each design. These agreements were consistent among trees in each ensemble, with all three having standard deviations of the agreements under 0.1%. These consistent accuracies indicate that the approach taken was able to distinguish a considerable amount of pattern in the covariates from the training areas. User's accuracies ($[\# \text{ correctly predicted class } x] / [\text{total } \# \text{ predicted of class } x]$) were more consistently high than producer's accuracies ($[\# \text{ correctly predicted class } x] / [\text{total } \# \text{ actual instances of class } x]$), which tended to have a few lower classes (Fig. 3). The transformed designs tended to predict more of the classes well, although those schemes did not increase overall accuracy. The author's thought that the ability to predict more of the classes well at similar accuracy made the SqRt and Log10 models more useful. Upon visual analysis of output maps, SqRt was chosen over Log10 as the optimal model due to the occurrence of inflated areal extent of minor soils not thought to exist in such extents (Fig 5).

For all three models, the classes with the highest combined user's and producer's accuracies were residual soils which tended to have accuracies near 90%, although there were some residual classes that were predicted with less success (Fig. 3). However, there were some groupings in accuracy based on sample size and parent material (Fig. 4). Residual soils with a larger training area tend to have the most consistently high user's and producer's accuracy. A training area size disparity in producer's accuracies is primarily seen in the Orig model. The smaller classes show a much larger range of producer's accuracies in this model; whereas larger classes all have producer's accuracies around 80% (Fig. 4, top right). This pattern is somewhat lost in the SqRt model and almost reverses in the Log10 model, but in all cases residual classes seem to perform the best in general. Colluvial and mine spoil classes were predicted the next best, and alluvial classes seemed to have consistently lower accuracy than other parent materials. The three models also show some visual differences with more heterogeneity of classes in the Log10 and SqRt models, whereas the Orig model appeared more dominated by a few larger classes (Fig. 5).

3.6.2 *SSURGO versus Disaggregation*

Visual comparison of SSURGO and disaggregation results shows that disaggregation predictions follow SSURGO lines much of the time, but also show some differences from delineations, likely map unit inclusions, and county line harmonization (Fig. 6). This is illustrated for a small portion of the study area (Fig. 7) where two multi-component map units, Pineville-Gilpin-Guyandotte association (PLF) and Gilpin-Dekalb complex (GdE), were mapped dominantly. In these map units the disaggregated map highlights all the named components and also shows inclusions, such as areas of Dekalb, Laidig, and Craigsville in PLF. Although these are not listed in the SSURGO database to be present in PLF, they are listed in the hardcopy

manuscript as inclusions (Delp, 1998). In SSURGO, these would fall into “Other soils” components and would not be referenced. This result demonstrates that the decision tree ensemble detected these subtle inclusion areas based on the actual soil-landscape properties of these soils in other areas of the survey without having specific information on included soils in the SSURGO database.

When we compared SSURGO and disaggregation maps directly we recorded agreements of 65.1% for the Orig model, 61.9% for the SqRt model, and 56.4% for the Log10 model. Conceptually, these values are probably low estimates because all map units contain 10%-25% “other soils”, or inclusions, to which we cannot match predictions unless all inclusions from all map units are added to the database from old soil survey manuscripts that vary with age in format and content. Therefore, we saw these as quite high amounts of correspondence between the original survey and the disaggregated soil-landscape patterns. We also noted that the disaggregated map is much more harmonized across the county line because it is actually mapping soil series, the common unit to both surveys as opposed to map units (Fig 6). This consistency within the study area is an advantage of training a model across both surveys.

3.6.3 *Validation using independent pedons*

Model predictions agreed with independent validation pedons 22%-24% of the time when compared using nearest neighbor spatial sampling. Model predictions agreed with validation 39%-44% of the time when comparing using a 60-meter radius sampling to check for matches (Table 6). The SqRt and Log10 models tended to have slightly higher accuracies than the Orig model through most of the measures. Validation pedons matched any of the named components in SSURGO map units 39% of the time for the nearest neighbor sampling and 41% of the time for the 60-meter radius sampling (Table 6). However, if we constrain validation of SSURGO to

just map units with one named component to make it more conceptually comparable to disaggregation predictions, the agreement rate drops to 27%, much closer to that of the nearest neighbor validation of disaggregation models (22-24%).

From the results (Table 6), it appears that the ensemble match rates nearly double when we expand from a nearest neighbor match to a 60-meter radius match, whereas the SSURGO matching rates were relatively unaffected by the matching approach. Although we expected slightly higher agreement rates by expanding the search radius, this large increase for the prediction models seems to indicate that there might have been spatial mismatches in the georeferencing of validation pedons to that of the model spatial data. Thus, the predictions on the covariate data likely represent the validation soil, but did not always line up exactly with validation sites due to spatial error. In general, the 60-meter validations were very similar between original SSURGO and the disaggregation models with even a slight improvement in the Log10 model performance. Even for the nearest neighbor evaluation (Table 6), we see that underlying tree models (any tree or 5+ tree) are detecting the correct soil with similar accuracy to that of SSURGO just not consistently enough for it to make the plurality required to be represented in the final hardened ensemble.

Again, it should be noted that 54% of the SSURGO matches to the validation pedons occurred in multi-component map units where the validation pedon could match any one of the multiple components and be counted as a match. If the validation pedons are evaluated just for the sites located in single component map units where there is only one soil to match, the accuracy for SSURGO falls to 27% using the nearest neighbor sampling. This is much more comparable to the disaggregation results. The validation pedons must also be used with caution because many of them were classified to taxadjunct or family equivalent level by NRCS staff.

Only 37 of the 87 pedons were fully matched to a soil series, the rest were used as the closest soil series if they fit the general concepts for that soil.

3.6.4 *Uncertainty in Predictions*

An advantage to the randomized sampling in the 100-tree ensemble models is that the prediction frequencies, or number of trees that predict a given grid cell outcome can be used as an estimator of confidence in that prediction. So, to see if these frequencies might reflect the likelihood of a correct prediction we compared their values at correctly predicted validation sites to those at sites that were incorrectly classified to see if higher values were associated with correct predictions. We evaluated the SqRt prediction model, which we deemed most the consistent performer over all evaluation metrics (Table 6), for the nearest neighborhood spatial validation. We found that correctly predicted sites did indeed have higher prediction frequencies in general (Wilcoxon rank sum test, $W = 883$, $p\text{-value} = 0.016$, one-sided, Fig. 8). The estimated shift in prediction frequencies from incorrect to correctly predicted validation sites was computed as 10.99 from sample estimates (95% C.I.: 2.0 to infinity) (R Core Development Team, 2008). This 11% probability shift is not of a large magnitude, but it does provide evidence that you can be less uncertain of predictions with higher prediction frequencies.

In further examination, 12 of 67 missed predictions had frequencies between 90 and 100, while a much higher proportioned 9 of 20 correct predictions had frequencies between 90 and 100. Reshuffled, 9 of 21 sites with prediction frequencies between 90 and 100 were correctly predicted, which translates into a ground truth probability of 43% for making a correct prediction at sites in that interval (Fig. 8d). This same approach was applied to the rest of the data in prediction frequency ranges of 0-0.5, 0.5-0.7, 0.7-0.9 and 0.9-1.0 to create a ground truth probability step function to represent uncertainty (Fig. 8). If more validation data were available

this could be modeled more continuously with an empirical function. However, our simple function enabled mapping of estimated ground truth probabilities because the prediction frequencies are available for every grid cell.

Although the ground truth probabilities we calculated are rather coarse, it can still give us an idea of where we are making better predictions without needing complicated calculations. Such a map of uncertainty is illustrated for a small portion of the study area (Fig. 9). For context, the PLF is a sideslope map unit, the GdE and DrF are summit and ridgetop map units, and the LdE is a footslope and small drainageway map unit. In this uncertainty map, we see that concave higher positions of PLF tend to have higher uncertainty (lower probabilities), while the more linear to convex and lower positions of PLF tend to have less uncertainty in predictions (higher probabilities) (Fig. 9).

3.6.5 *Environmental Covariate Influence on Decision Tree Models*

Variable importance values help to determine the most influential environmental covariates used in the models, thus providing insight into the original soil survey paradigm. Bedrock geology and elevation were consistently the most used variables in decision tree breaks (Table 7). The importance of geology and elevation were expected as the study area has strong topographic gradients (Fig. 1) and a variety of contrasting geologic strata (West Virginia Geologic and Economic Survey, 1968). In initial research, local soil scientists stressed the importance of geology and how the wide range of local geologic grain size and base cation content tended to produce distinctly different groupings of soil types. This area also includes a large range in elevation (254 to 1466 m). Slope gradient and slope position were consistently the third and fourth most used variables, whereas the rest of the variables had slightly greater differences among relative order of importance between models. In the less used variables, the

2000 Geocover PC1 and profile curvature showed the most difference in usage between models suggesting that the effects of training class sample sizes effected how these co-varied with soil classes.

Standard deviations of importance values in the ensembles were all generally low, with a maximum of 1.35%. This indicates that the individual trees within ensembles were relatively similar and stable. However, when comparing the different ensemble sampling approaches, the Orig model shows markedly higher deviation in some variables than the SqRt and Log10 models. These higher deviations are seen almost entirely in the top four most important variables (geology, elevation, slope gradient, and slope position). So, even with the overall stability in all the ensemble models, the original model showed less stability than the others. This instability might have been due to the much larger disparity between the sizes of training classes, making patterns in covariates harder to detect.

3.7 Discussion

The disaggregation approach presented in this study integrates the conceptual themes of legacy soil survey into a coherent method to quantitatively refine documented pedologic patterns. This is done while maintaining a similar accuracy to the original product, but doing so with increased spatial and thematic resolution, and in a more continuous and field-oriented raster format. Our goal was to do this in a repeatable fashion with data and software that is widely available; a goal we felt was met. The disaggregation maps we produced may serve as a first step in soil survey update or distributed as an additional digital soil map product that could be useful in projects like *GlobalSoilMap*. The underlying prediction frequencies can be used in cell-by-cell weighted averaging of soil class properties to create fuzzy soil property maps similar to other recent studies (Zhu et al., 2010; Nauman et al., 2012). The underlying tree ensembles used also help us

understand where predictions are better and where new data may be needed in efforts to create more accurate maps.

We attempted to leverage prior mapping work by directly matching environmental rasters like slope position, slope gradient, and other geomorphic metrics to actual published descriptions of soil-landscape relationships. By propagating the expert knowledge contained in the original mapping through to new updated digital mapping products, we can test it and try to model it with modern computing methods. The rule-matching process also provides a direct and understandable way for soil scientists to help tweak these models by refining training areas. In this case, decision trees performed well in interpreting these rules, supporting theories describing a hierarchal nature of many spatial-environmental soil patterns (de Bruin et al., 1999; Bui and Moran, 2001; Wielemaker et al., 2001).

We also observed that adjusting the proportionality of soil class sampling for model training can influence how consistent individual classes are predicted. With a highly skewed distribution of class sizes, smaller classes are predicted poorly, or not detected at all. Using the original sample proportions (Orig model), there were three orders of magnitude of difference between the large and small class sizes, which allowed the larger, more generalized soils to dominate classifications. However, the transformed classifications tended to promote better predictions of smaller classes, and the square root sampling (SqRt model) in particular seemed to maintain the dominance of soils that were mapped extensively while still including less dominant soils. Validation results seemed to confirm that the SqRt and Log10 models showed slightly better accuracies across most of the metrics, although these results only showed slight and not entirely conclusive differences.

3.7.1 *Uncertainty*

Uncertainty and validation of digital soil maps has been addressed by various researchers (e.g., Lark and Bolam, 1997; Brus et al., 2011; Malone et al., 2011; Bishop et al., 2001, 2006), and has been identified as a critical evaluation tool to provide for soil map end-users by the *GlobalSoilMap* consortium (Hartemink et al., 2010; Minasny and McBratney, 2010; GlobalSoilMap, 2012). Fully characterizing the error and uncertainty in predictive models is challenging because there are many potential sources of error. Any comprehensive assessment of predictive models must have some independent data to test against that is properly sampled (Brus et al., 2011), some way to keep track of how error of input variables propagates through a model (e.g., Lagacherie and Holmes, 1997; Hengl et al., 2004; Bishop et al., 2006), and some way to put the errors together into an uncertainty representation (e.g., Malone et al., 2011). This process is quite tedious, and our simple uncertainty result lacks a truly representative sampling scheme and does not comprehensively address all potential error in covariates used. However, it is an informative and understandable approach to help determine where predictions worked better or worse.

Our findings relating model prediction frequencies to ground truth accuracies showed the power of randomized ensemble sampling in estimating uncertainty in predictions. We looked only at the nearest neighbor based validation agreements in our uncertainty calculations (see Table 6), which had lower validation match rates than using a local neighborhood around validation sites to look for matches (which acknowledges that there is spatial error in all data being used). Incorporating the neighborhood validation data to create a ground truth might result in overall higher ground truth probabilities of one finding the predicted soil within a given neighborhood radius of a prediction. This will be an object of future research as it requires more

detailed calculations and assumptions beyond the scope of this investigation. The situation of having limited validation sites and modeling data with varying degrees of spatial resolution and error is a common problem (e.g. Yang et al. 2010; Smith et al., 2012) and difficult to address in a manner that fully integrates error at all steps of modeling. This is especially true with the large number of spatial referencing and raster calculation steps done in GIS when preparing covariates.

3.7.2 *Future Soil Survey Applications*

From a practical perspective, if we can successfully disaggregate legacy soil maps to field scale continuous representations, then they can be better used for management and understanding ecological processes and associated dynamics. Disaggregation also offers a way to help aid in the process of harmonizing the large number of soil survey projects into more contiguous and consistent products. Just having more consistent soil series distribution maps across the U.S. with a disaggregated SSURGO product would aid in better understanding and interpretation of soils in the environment. With SSURGO currently, all that has to be represented on a map unit basis that does not actually spatially represent the underlying soil series directly (See Fig. 6). So just in the translation of the spatial symbology we can help better our geographic understanding of soil populations.

In the United States, ecological site descriptions (ESD) are often built to describe how soil components are linked to sets of ecological communities (Grazing Lands Technology, 2003). ESD are built into conceptual frameworks that describe how potential soil-vegetation-wildlife communities respond to different natural and anthropogenic pressures. Currently, ESD are linked to SSURGO components in soil maps in the western U.S. and being expanded east, but are difficult to use on the ground because much of the soil mapping is coarser than many

conservation plans. So, if disaggregation can produce maps on a scale more appropriate for conservation plans that use ESD, then these efforts can be streamlined quicker.

There is also the potential to start implementing dynamic representations of soil health within soil maps if ESD state and transition models can be integrated directly into soil map databases. This would move soil mapping into the temporal region where soils can be looked at as a series of properties changing at different rates as a result of real-time environmental inputs, a framework that would help meet calls from others for more dynamic soil information in the digital soil mapping community (Grunwald et al., 2011). This could be represented well in a disaggregated soil survey-ESD mapping framework. With better links between management, vegetation, soil, and wildlife documented in ESD, soil change can be modeled as resultant to disturbance or change in other parts of the system, and that could all be mapped spatially as an assessment of interdisciplinary ecological health.

3.8 *Conclusions*

This work demonstrates a method that combines soil-landscape knowledge, data-mining, and machine learning to disaggregate legacy soil surveys into soil component level maps. Although the technique uses the original survey spatial data to help determine typical areas for each soil to use in training, the original mapping polygon lines are not used for the final modeling. This technique also does not require any new field data to create the disaggregated model. However, sparse field data available for validation was used to help determine performance and attempt to determine uncertainty in a spatial representation from classification tree ensemble probabilities.

The geomorphic inputs that help determine training areas represent a direct use of the original expert knowledge used to produce soil surveys. Thus if those geomorphic rule inputs are

updated, this offers a route to be able to iteratively refine the disaggregation product. Other modeling techniques (e.g., random forest, boosted classification trees) could also be tested on these training sets. In the surveys used for this study, we noticed that the rules in the SSURGO database tend to be less specific than those published in the original hard copy manuscript as was used by Thompson et al. (2010). This is especially important as countries like the U.S. try to harmonize and update soil surveys (Loerch, 2012; Thompson et al., 2012) because new refinements can then be rerun into updated disaggregation products using the approach presented.

Our results also offer a small insight into the true accuracy of legacy soil data. Both disaggregation results and original survey data showed approximately 40% agreement with an independent validation when some spatial error is allowed in matching validation sites to predictions. These results were from a spatially limited validation set from multiple sources and dates, and as such must be interpreted with caution. However, these results leave much to be desired, and give us insight into the future work required to update soils data to standards deemed acceptable for modern applications. The uncertainty maps produced from these efforts are likely to be valuable in helping establish targeted field collection of new samples to help increase accuracies in the next generation of digital soil maps that might use disaggregation results as inputs into new models.

3.9 References

- Bandaragoda, C., Tarboton, D.G., and Woods, R., 2004. Application of TOPNET in the distributed model intercomparison project. *Journal of Hydrology*, 298(1-4): 178-201.
- Behrens, T., Schmidt, K., Zhu, A.X., and Scholten, T., 2010a. The ConMap approach for terrain-based digital soil mapping. *European Journal of Soil Science*, 61(1): 133-143.
- Behrens, T., Zhu, A.X., Schmidt, K., and Scholten, T., 2010b. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, 155(3/4): 175-185.
- Bishop, T.F.A., McBratney, A.B., and Whelan, B.M., 2001. Measuring the quality of digital soil maps using information criteria. *Geoderma*, 103(1-2): 95-111.
- Bishop, T.F.A., Minasny, B., and McBratney, A.B., 2006. Uncertainty analysis for soil-terrain models. *International Journal of Geographical Information Science*, 20(2): 117-134.
- Bliss, N.B., S.W. Waltman, and G.W. Petersen. 1995. Preparing and soil carbon inventory for the United States using geographic information systems. p. 275-295. In R. Lal et al. (ed.) *Soils and global change*. *Adv Soil Science*. CRC Press, Boca Raton, FL.
- Bouma, J. 1989. Using Soil Survey Data for Quantitative Land Evaluation, p. 177-213, In B. A. Stewart, ed. *Advances in Soil Science*, Vol. 9. Springer US.
- Breiman, L., 1984. *Classification and regression trees*. Wadsworth International Group, 358 pp.
- Brus, D.J., Kempen, B., and Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science*, 62(3): 394-407.
- Bui, E., Henderson, B., and Viergever, K., 2009. Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. *Global Biogeochemical Cycles*, 23.
- Bui, E.N., 2004. Soil survey as a knowledge system. *Geoderma*, 120(1/2): 17.
- Bui, E.N., Henderson, B.L., and Viergever, K., 2006. Knowledge discovery from models of soil properties developed through data mining. *Ecological Modelling*, 191(3-4): 431-446.
- Bui, E.N., Loughhead, A., and Corner, R., 1999. Extracting soil-landscape rules from previous soil surveys. *Australian Journal of Soil Research*, 37(3): 495-508.

- Bui, E.N., and Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma*, 103(1-2): 79-94.
- Burrough, P.A., 1989. Fuzzy mathematical methods for soil survey and land evaluation. *European Journal of Soil Science*, 40(3): 477-492.
- Burrough, P.A., van Gaans, P.F.M., and Hootsmans, R., 1997. Continuous classification in soil survey: spatial correlation, confusion and boundaries. *Geoderma*, 77(2-4): 115-135.
- Causarano, H.J., Doraiswarny, P.C., McCarty, G.W., Hatfield, J.L., Milak, S., and Stern, A.J., 2008. EPIC Modeling of soil organic carbon sequestration in croplands of Iowa. *Journal of Environmental Quality*, 37(4): 1345-1353.
- Conrad, O., and Wichmann, V., 2011. SAGA GIS (www.saga-gis.org), Hamburg, Germany.
- Cook, S.E., Corner, R.J., Grealish, G., Gessler, P.E., and Chartres, C.J., 1996. A rule-based system to map soil properties. *Soil Science Society of America Journal*, 60(6): 1893-1900.
- de Bruin, S., Wielemaker, W.G., and Molenaar, M., 1999. Formalisation of soil-landscape knowledge through interactive hierarchical disaggregation. *Geoderma*, 91(1-2): 151-172.
- Delp, C.H., 1998. Soil Survey of Webster County, West Virginia (SSURGO). In: N.R.C.S. U.S. Department of Agriculture, W.V.A.F.E. Station and F.S. U.S. Department of Agriculture (Editors).
- ESRI, 2011. ArcGIS Desktop: Release 10. Environmental Systems Research Institute, Redlands, CA.
- Flegel, D.G., 1998. Soil Survey of Pocahontas County, West Virginia (SSURGO). In: N.R.C.S. U.S. Department of Agriculture, W.V.D.A.F.E. Station and F.S. U.S. Department of Agriculture (Editors).
- Gatzke, S.E., Beaudette, D.E., Ficklin, D.L., Yuzhou, L., O'Geen, A.T., Minghua, Z., 2011. Aggregation Strategies for SSURGO Data: Effects on SWAT Soil Inputs and Hydrologic Outputs. *Soil Science Society of America Journal*, 75(5): 1908-1921.
- Hansen, M.K., Brown, D.J., Dennison, P.E., Graves, S.A. and Brickleymer, R.S., 2009. Inductively mapping expert-derived soil-landscape units within dambo wetland catenae using multispectral and topographic data. *Geoderma*, 150(1-2): 72-84.
- Erdas, 2010. Erdas Field Guide December 2010. Norcross, Georgia. pp. 492-496.

- Gesch, D.B., 2007. The National Elevation Dataset. In: D. Maune (Editor), *Digital Elevation Model Technologies and Applications: The DEM Users Manual*. American Society for Photogrammetry and Remote Sensing, pp. 99-118.
- Gesch, D.B., Oimoen, M., Greenless, S., Nelson, C., Steuck, M., Tyler, D., 2002. The National Elevation Dataset. *Photogrammetric Engineering and Remote Sensing*, 68(1): 5-11.
- GlobalSoilMap, 2012. Specifications, Version 1 GlobalSoilMap.net Products. Release 2.2. Available online at <http://www.globalsoilmap.net/specifications>.
- Goovaerts, P., 2011. A coherent geostatistical approach for combining choropleth map and field data in the spatial interpolation of soil properties. *European Journal of Soil Science*, 62(3): 371-380.
- Grazing Lands Technology Institute, 2003. *National range and pasture handbook*. [Fort Worth, Tex.] : U.S. Dept. of Agriculture, Natural Resources Conservation Service.
- Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152(3-4): 195-207.
- Grunwald, S., Thompson, J.A., and Boettinger, J.L., 2011. Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues. *Soil Science Society of America Journal*, 75(4): 1201-1213.
- Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., and Schröder, B., 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma*, 185–186(0): 37-47.
- Hartemink, A.E., Hempel, J., Lagacherie, P., McBratney, A., McKenzie, N., MacMillan, R.A., Minasny, B., Montanarella, L., Mendonça Santos, M.L., Sanchez, P., Walsh, M., and Zhang, G.-L., 2010. GlobalSoilMap.net—A New Digital Soil Map of the World. In: J. Boettinger, D. Howell, A. Moore, A. Hartemink and S. Kienast-Brown (Editors), *Digital Soil Mapping*. Progress in Soil Science. Springer Netherlands, pp. 423-427.
- Hatfield, D.C., 1996. *TopoTools - A Collection of Topographic Modeling Tools for ArcINFO: SLOPEPOSITION*. US Forest Service. <http://proceedings.esri.com/library/userconf/proc00/professional/papers/pap560/p560.htm>
- Henderson, B.L., Bui, E.N., Moran, C.J., and Simon, D.A.P., 2005. Australia-wide predictions of soil properties using decision trees. *Geoderma*, 124(3-4): 383-398.
- Hengl, T., Gruber, S., and Shrestha, D.P., 2004. Reduction of errors in digital terrain parameters used in soil-landscape modelling. *International Journal of Applied Earth Observation and Geoinformation*, 5(2): 97-112.

- IPCC, 2007. Climate change 2007: Synthesis report. Contribution of Working Groups I, II, and III to the Fourth Assessment Report., Geneva, Switzerland.
- Jenny, H., 1941. Factors of Soil Formation. McGraw-Hill, New York, New York.
- Kempen, B., Brus, D.J., Heuvelink, G.B.M., and Stoorvogel, J.J., 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, 151(3/4): 311-326.
- Kerry, R., Goovaerts, P., Rawlins, B.G., and Marchant, B.P., 2012. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma*, 170: 347-358.
- Lagacherie, P., and Holmes, S., 1997. Addressing geographical errors in a classification tree for soil unit prediction. *International Journal of Geographical Information Science*, 11(2): 183-198.
- Lark, R.M., and Bolam, H.C., 1997. Uncertainty in prediction and interpretation of spatially variable data on soils. *Geoderma*, 77(2-4): 263-282.
- Lemercier, B., Lacoste, M., Loum, M., and Walter, C., 2011. Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach. *Geoderma*(0).
- Lineback Gritzner, M., Marcus, W.A., Aspinall, R., and Custer, S.G., 2001. Assessing landslide potential using GIS, soil wetness modeling and topographic attributes, Payette River, Idaho. *Geomorphology*, 37(1-2): 149-165.
- Loerch, C., 2012. Soil Data Join Recorrelation Initiative, West Regional National Cooperative Soil Survey Meeting, Davis, CA. Available online at <ftp://ftp-fc.sc.egov.usda.gov/NSSC/NCSS/Conferences/regional/2012/west/loerch.pdf>
- Malone, B.P., McBratney, A.B., and Minasny, B., 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma*, 160(3/4): 614-626.
- McBratney, A.B., 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. *Nutrient Cycling in Agroecosystems*, 50(1-3): 51-62.
- McBratney, A.B., Santos, M.L.M., and Minasny, B., 2003. On digital soil mapping. *Geoderma*, 117(1-2): 3-52.
- McKenzie, N.J., and Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89(1-2): 67-94.
- MDA, Federal, 2004a. Landsat GeoCover ETM+ 2000 Edition Mosaics Tile N-17-35.ETM-EarthSat-MrSID. USGS, Sioux Falls, South Dakota.

- MDA, Federal, 2004b. Landsat Geocover TM 1990 Edition Mosaics Tile N-17-35 TM-EarthSat-MrSID. USGS, Sioux Falls, South Dakota.
- Minasny, B., and McBratney, A.B., 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma*, 142(3-4): 285-293.
- Minasny, B., and McBratney, A.B., 2010. Methodologies for Global Soil Mapping. In: J. Boettinger, D. Howell, A. Moore, A. Hartemink and S. Kienast-Brown (Editors), *Digital Soil Mapping*. Progress in Soil Science. Springer Netherlands, pp. 429-436.
- Moran, C. J., & Bui, E. N. 2002. Spatial data mining for enhanced soil map modeling. *International Journal of Geographical Information Science*, 16(6), 533-549.
- National Cooperative Soil Survey., 2012. National Cooperative Soil Characterization Database. Available online at <http://ncsslabdatamart.sc.egov.usda.gov>.
- Nauman, T., Thompson, J.A., Odgers, N., and Libohova, Z., 2012. Fuzzy Disaggregation of Conventional Soil Maps using Database Knowledge Extraction to Produce Soil Property Maps. In: B. Minasny, B. Malone and A. McBratney (Editors), *Digital Soil Assessments and Beyond: 5th Global Workshop on Digital Soil Mapping*, Sydney, Australia.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python (also see: <http://scikit-learn.org/stable/modules/tree.html#classification-criteria>). *Journal of Machine Learning Research*, 12: 2825-2830.
- Qi, F., Zhu, A.X., Harrower, M., and Burt, J.E., 2006. Fuzzy soil mapping based on prototype category theory. *Geoderma*, 136(3-4): 774-787.
- R Core Development Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Sanchez, P.A., Ahamed, S., Carre, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, Alex B., McKenzie, N.J., Mendonca-Santos, M., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vagen, Tor-Gunnar, Vanlauwe, B., Walsh, M.G., Winowiecki, LA., Zhang, Gan-Lin, 2009. Digital soil map of the world. *Science*, 325(5941): 680-681.

- Saunders, A.M., and Boettinger, J.L., 2007. Incorporating Classification Trees into a Pedogenic Understanding Raster Classification Methodology, Green River Basin, Wyoming, USA. In: P. Lagacherie, A.B. McBratney and M. Voltz (Editors), *Digital Soil Mapping An Introductory Perspective*. Elsevier, Amsterdam, The Netherlands.
- Schmidt, J., and Hewitt, A., 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. *Geoderma*, 121(3-4): 243-256.
- Schmidt, K., Behrens, T., and Scholten, T., 2008. Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. *Geoderma*, 146(1-2): 138-146.
- Scull, P., Franklin, J., and Chadwick, O.A., 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modelling*, 181(1): 1-15.
- Scull, P., Franklin, J., Chadwick, O.A., and McArthur, D., 2003. Predictive soil mapping: a review. *Progress in Physical Geography*, 27(2): 171-197.
- Smith, S., Bulmer, C., Flager, E., Frank, G., and Filatow, D., 2010. *Digital Soil Mapping and multiple scales in British Columbia, Canada*, 4th Global Workshop on Digital Soil Mapping, Rome, Italy.
- Smith, S., Daneshfar, B., Frank, G., Flager, E., and Bulmer, C., 2012. Use of weights of evidence statistics to define inference rules to disaggregate soil survey maps, 5th Global Workshop on Digital Soil Mapping, Sydney, Australia.
- Soil Survey Staff, 2010. *Keys to Soil Taxonomy*, 11th ed. USDA-Natural Resources Conservation Service, Washington, DC.
- Soil Survey Division Staff, 1993. *Soil Survey Manual*. In: Soil Conservation Service U.S.D.A. (Editor), *Handbook* 18.
- Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, 33(2): 309-319.
- Thompson, J.A. and Kolka, R.K., 2005. Soil carbon storage estimation in a forested watershed using quantitative soil-landscape modeling. *Soil Science Society of America Journal*, 69(4): 1086-1093.
- Thompson, J.A., Nauman, T., Odgers, N., Libohova, Z., and Hempel, J., 2012. Harmonization of Legacy Soil Maps in North America: Status, Trends, and Implications for Digital Soil Mapping Efforts. In: A. McBratney, B. Minasny and B. Malone (Editors), *The 5th Global Workshop on Digital Soil Mapping. Digital Soil Assessments and Beyond*, Sydney, Australia.

- Thompson, J.A., Prescott, T., Moore, A., Bell, J., Kautz, D., Hempel, J., Waltman, S., Perry C.H., 2010. Regional Approach to Soil Property Mapping using Legacy Data and Spatial Disaggregation Techniques, 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia.
- Wei, S., McBratney, A., Hempel, J., Minasny, B., Malone, B., D'Avella, T., Burras, L., Thompson, J.A., 2010. Digital Harmonisation of Adjacent Soil Survey areas - 4 Iowa Counties, 19th World Congress of Soil Science, Soils Solutions for a Changing World, Brisbane, Australia.
- West Virginia Geologic and Economic Survey, 1968. Surface Geology - Rock Units. In: T.U. West Virginia DEP (Editor). William and Heintz Map Corporation, pp. Online Digitized Map: Available at <http://wvgis.wvu.edu/>.
- Wielemaker, W.G., de Bruin, S., Epema, G.F., and Veldkamp, A., 2001. Significance and application of the multi-hierarchical landsystem in soil mapping. *Catena*, 43(1): 15-34.
- Wilson, J.P., Inskeep, W.P., Rubright, P.R., Cooksey, D., Jacobsen, J.S., Snyder, R.D., 1993. Coupling Geographic Information-Systems and Models for Weed-Control and Groundwater Protection. *Weed Technology*, 7(1): 255-264.
- Yang, L., Jiao, Y., Fahmy, S., Zhu, A.X., Hann, S., Burt, J.E., Qi, F., 2011. Updating Conventional Soil Maps through Digital Soil Mapping. *Soil Science Society of America Journal*, 75(3): 1044-1053.
- Yokoyama, R., Shirasawa, M., and Pike, R.J., 2002. Visualizing Topography by Openness: A New Application of Image Processing to Digital Elevation Models. *Photogrammetric Engineering & Remote Sensing*, 68(3): 257-265.
- Zhang, Y., Zhang, Z., Reed, S., and Koren, V., 2011. An enhanced and automated approach for deriving a priori SAC-SMA parameters from the soil survey geographic database. *Computers & Geosciences*, 37(2): 219-231.
- Zhu, A.X., 1997. A similarity model for representing soil spatial information. *Geoderma*, 77(2-4): 217-242.
- Zhu, A.X., Band, L.E., Dutton, B., and Nimlos, T.J., 1996. Automated soil inference under fuzzy logic. *Ecological Modelling*, 90(2): 123-145.
- Zhu, A.X., Band, L., Vertessy, R., and Dutton, B., 1997. Derivation of soil properties using a soil land inference model (SoLIM). *Soil Science Society of America Journal*, 61(2): 523-533.
- Zhu, A.X., Hudson, B., Burt, J., Lubich, K., and Simonson, D., 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal*, 65(5): 1463-1472.

Zhu, A.X., Yang, L., Li, B., Qin, C., Pei, T., Liu, B., 2010. Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma*, 155(3-4): 164-174.

3.10 Tables and Figures

Table 1. Two multi-component map units recorded in the Webster County soil survey, West Virginia (Delp, 1998).

Map unit (MU) name	MU kind	Components	Parent material	% of MU
Gilpin-Laidig association, very steep, extremely stony	Association	Gilpin	Residuum	45
		Laidig	Colluvium	35
		Included soils	n/a	20
Pineville-Gilpin-Guyandotte association, very steep, extremely stony	Association	Pineville	Colluvium	35
		Gilpin	Residuum	25
		Guyandotte	Colluvium	15
		Included soils	n/a	25

Table 2. Tables used from SSURGO for training set rule matching. An example of the language used to match with digital terrain raster values is shown.

SSURGO table	Upwards SSURGO link table	Table concept	Example	Rulematch
Mapunit	Legend	Spatial unit attribute (polygon types)	Gilpin-Dekalb complex, 15 to 35 percent slopes, extremely stony	n/a
Component	Map Unit	Soil series and other components	Dekalb Series, 35% of map unit	
CoGeomorDesc	Component	Geomorphic description: landform	Ridges	Geomorphic element Rules: Slope
CoSurfMorphGC	CoGeomorDesc	Geomorphic component	Mountaintop	Position >= 95
CoSurfMorphSS	CoGeomorDesc	Surface shape	Linear linear	Hillslope context Rules: Slope
CoSurfMorphPP	CoGeomorDesc	Hillslope profile	Summit	Position >= 75 and a 'Plain'

Table 3. Hillslope profile and curvature rule matches derived from the SSURGO database. Slope position is a 0-100 index of how far up a slope a location is with 100 being a summit and 0 being valley floor.

SSURGO descriptions				Raster rules
Slope shape (up/down)	Slope shape (across)	Hillslope profile	Slope position	Landform (Schmidt and Hewitt, 2004)
Concave	Concave	Backslope		Hollow foot
Concave	Concave	Footslope	<50	Hollow foot or pit
Concave	Concave	Shoulder	50-95	Hollow foot or pit
Concave	Concave	Summit	>=75	Pit
Concave	Convex	Backslope		Spur foot
Concave	Convex	Footslope	<50	Spur foot or saddle
Concave	Convex	Shoulder	>=50	Spur foot or saddle
Concave	Convex	Summit		Saddle
Concave	Linear	Backslope		Footslope
Concave	Linear	Footslope	<50	Footslope or channel
Concave	Linear	Shoulder	>50	Footslope or channel
Concave	Linear	Summit	>=75	Channel
Convex	Concave	Backslope		Hollow shoulder
Convex	Concave	Footslope	<50	Hollow shoulder
Convex	Convex	Backslope		Nose
Convex	Convex	Footslope	<50	Nose
Convex	Convex	Shoulder	>=50	Nose
Convex	Convex	Summit		Peak
Convex	Linear	Backslope		Shoulder slope
Convex	Linear	Footslope	<50	Shoulder slope
Convex	Linear	Shoulder	50-95	Shoulder slope
Convex	Linear	Summit		Ridge
Convex	Linear	Toeslope	<10	Ridge
Linear	Concave	Backslope		Hollow
Linear	Concave	Footslope	<50	Hollow or channel
Linear	Concave	Summit	>=75	Channel
Linear	Convex	Backslope		Spur
Linear	Linear	Backslope		Planar slope
Linear	Linear	Footslope	<50	Planar slope or plain
Linear	Linear	Shoulder	>=50	Planar slope or plain

Linear	Linear	Summit	≥ 75	Plain
Linear	Linear	Toeslope	< 10	Plain

Table 4: Component general landform descriptions from SSURGO and the interpreted environmental raster value rules for the landform descriptions. All unique combinations of the SSURGO geomorphic description queries are shown with on the left along with the translated rules for the terrain rasters used to select training areas.

SSURGO Descriptions				Raster Rules						
Feature name	Feature modifier	Geomorphic component			Hillslope position	Slope gradient (%)	Slope aspect (deg)	Contrib. area (m ²)	Terrace (m)	Landform†
		Mountains	Hills	Terraces						
reclaimed lands	v. steep; mostly coal & high - carbon shale	Mountainflank			10<=x<=95					
alluvial fans	mouth of hollows	Tread			1<=x<=5	>1	101, 111, 100, 120, or 10			
coves		Upper third of mountainflank			66<=x<=95					
coves	lower elevations	Center third of mountainflank			33<=x<=66					
coves	lower elevations	Lower third of mountainflank			10<=x<=33					
coves	northern facing	Mountainbase			1<=x<=10	270<=y<=36 0 or 0<=y<=90			21, 1, 11, 101	
coves	northern facing	Mountainflank			10<=x<=95					
depressions	upland depressions on mountains	Mountaintop			>95					
drainageways	along drainageways and on head slopes	Mountaintop			>95			or >75000		101
drainageways	along drainageways and on head slopes	Upper third of mountainflank			66<=x<=95			or >100000		101
flats		Mountaintop			>95					
flood plains		Tread			<5					
flood plains	high bottom land	Tread			0<=x<=5			>2		100 or 101
flood plains	nearly level	Tread			<5			< 3		100 or 101

flood plains	nearly level & poorly drained		Tread	<5	< 3		100 or 101
flood plains	nearly level & well drained		Tread	<5	< 3	>0.5	100 or 101
hillslopes			Side Slope	10<=x<=95			22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes		Lower third of mountainflank		10<=x<=33			22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes		Mountainflank		10<=x<=95			22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes		Upper third of mountainflank		66<=x<=95			
mountain slopes	benches	Center third of mountainflank		33<=x<=66	<25		
mountain slopes	benches	Mountainflank		10<=x<=95	<25		
mountain slopes	benches	Upper third of mountainflank		66<=x<=95	<25		
mountain slopes	benches on low elevation mountains and hills	Center third of mountainflank		33<=x<=66	<25		
mountain slopes	benches on low elevation mountains and hills	Upper third of mountainflank		66<=x<=95	<25		
mountain slopes	benches; > 3400 ft. elevation	Upper third of mountainflank		66<=x<=95	<25		
mountain slopes	concave sideslopes of uplands	Mountainflank		10<=x<=95			21, 1, 11, 10, 12
mountain slopes	disected uplands - convex	Mountainflank		10<=x<=95			20, 22, 2, 12, 120, 122
mountain slopes	disected uplands - convex and benches	Mountainflank		10<=x<=95			20, 22, 2, 12, 100, 120, 122
mountain slopes	disected uplands - convex; & benches	Mountainflank		10<=x<=95			20, 22, 2, 12, 100, 120, 122
mountain slopes	drainageways & footslopes	Mountainbase		1<=x<=10		or >350000	101, 10, 11

mountain slopes	drainageways & footslopes	Mountainflank	$10 \leq x \leq 95$	or >150000	10, 11, 101
mountain slopes	elevations > 3400 ft.	Upper third of mountainflank	$66 \leq x \leq 95$		22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes	lower sideslopes & footslopes	Center third of mountainflank	$33 \leq x \leq 66$		22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes	lower sideslopes & footslopes	Lower third of mountainflank	$10 \leq x \leq 33$		22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes	northern facing	Mountainbase	$1 \leq x \leq 10$	$270 \leq y \leq 360$ or $0 \leq y \leq 90$	22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes	northern facing	Mountainflank	$10 \leq x \leq 95$	$270 \leq y \leq 360$ or $0 \leq y \leq 90$	22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes	side slopes & 10s	Mountainflank	$10 \leq x \leq 95$		2, 0, 1, 11, 10, or 12
mountain slopes	steep & very steep slopes and benches; very stony	Mountainflank	$10 \leq x \leq 95$		
mountain slopes	steep and very steep slopes and benches	Mountainflank	$10 \leq x \leq 95$		
mountain slopes	upper	Center third of mountainflank	$33 \leq x \leq 66$		22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes	upper	Mountainflank	$33 \leq x \leq 95$		22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes	upper	Upper third of mountainflank	$66 \leq x \leq 95$		22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes	upper; > 3400 ft. elevation	Mountainflank	$33 \leq x \leq 95$		22, 20, 21, 2, 0, 1, 12, 10, or 11
mountain slopes	very steep uplands; extremely stony	Mountainflank	$10 \leq x \leq 95$		
ridges		Mountaintop	>95		

ridges	braod; > 3400 ft. elevation	Mountaintop	>95	100
ridges	broad	Mountaintop	>95	100
ridges	broad; > 3400 ft elevation	Mountaintop	>95	100
ridges	broad; > 3400 ft. elevations	Mountaintop	>95	100
ridges	narrow - low elevation mountains and hills	Mountaintop	>80	120 or 122 or 121
ridges	narrow ridgetops on low elevation mountains and hills	Mountaintop	>80	120 or 122 or 121
ridges	narrow; low elevation mountains & hills	Mountaintop	>80	120 or 122 or 121
sinkholes		Lower third of mountainflank	10<=x<=33	111
stream terraces		Tread	0<=x<=50	1<=z<=5 100 or 101
streams	headwaters of	Lower third of mountainflank	10<=x<=33	or >150000 101 or 1 or 11
streams	headwaters of	Mountainbase	1<=x<=10	or >350000 101 or 1 or 11
terraces	low stream	Tread	0<=x<=10	>1 100 or 101

† Landform key: 0 = backslope, 1 = hollow, 2 = spur, 10 = footslope, 11 = hollow foot, 12 = spur foot, 20 = shoulder, 21 = hollow shoulder, 22 = nose, 100 = plain, 101 = channel, 111 = pit, 120 = ridge, 121 = saddle, 122 = peak (Schmidt and Hewitt, 2004)

Table 5. List of environmental variables used for classification tree ensembles.

Layer name	Description	Reference
Slope position	0 to 100 index of hillslope position	(Hatfield, 1996)
% slope gradient	slope gradient of pixel in percent	(ESRI, 2011)
Southness	north/south aspect; -1 is north; 1 is south	cosine(aspect - 180); (ESRI, 2011)
Eastness	east/west aspect: -1 is west; 1 is east	cosine(aspect - 90); (ESRI, 2011)
Planar curvature	curvature parallel to contour line	(ESRI, 2011)
Profile curvature	curvature parallel to slope fall-line	(ESRI, 2011)
log10(catchment area)	upstream catchment area log10 transformed	(Tarboton, 1997)
Elevation	elevation in meters	(Gesch, 2007; Gesch et al., 2002)
Terrace/relative height	stratified relative elevation index	Figure 2; Section 2.2.1
1990 Landsat Geocover PC1	1990 Landsat scene principal component 1	(MDA, 2004b); (Erdas, 2010)
1990 Landsat Geocover PC2	1990 Landsat scene principal component 2	(MDA, 2004b); (Erdas, 2010)
2000 Landsat Geocover PC1	2000 Landsat scene principal component 1	(MDA, 2004a); (Erdas, 2010)
2000 Landsat Geocover PC2	2000 Landsat scene principal component 2	(MDA, 2004a); (Erdas, 2010)
Bedrock formations	Different bedrock lithology formations	(West Virginia Geologic and Economic Survey, 1968)

Table 6. Validation agreement rates for different spatial neighborhoods matching criteria.

‘Series’ matches indicate exact class matches in the final ensemble prediction, ‘series or like’ means the percent of predictions that either matched the validation series or were a similar soil as defined in Section 2.4, ‘any tree’ refers to at least one of the trees in the ensemble predicting the correct series, and both the 5+ and 33+ refer to at least that many trees predicting the correct series. SSURGO matches compared validation pedons to the original SSURGO map units. If the validation pedon matched any of the named components in a map unit, it was deemed a match.

		Data Model			
		Original	Square root	Log10	SSURGO
Nearest neighbor	Series	22%	24%	24%	39%
	Series or like	38%	36%	32%	52%
	Any tree	47%	51%	49%	--
	5+ trees	33%	39%	37%	--
	33+ trees	24%	25%	24%	--
60-meter radius	Series	39%	41%	44%	41%
	Series or like	61%	64%	62%	56%
	Any tree	59%	56%	56%	--
	5+ trees	44%	51%	51%	--
	33+ trees	38%	41%	44%	--

Table 7. Importance values showing how often all covariates were used in each disaggregation model. All models are averaged in right column. Standard deviation values indicate how much variation there was within the 100 tree ensembles. Variables are ordered from highest (top) to lowest by the averaged column (right).

Variables	Original model		Square root model		Log10 model		All
	Importance	St. dev	Importance	St. dev.	Importance	St. dev.	Average
Bedrock geology	30.70%	1.35%	25.27%	0.08%	21.47%	0.09%	25.81%
National Elev. Dataset	23.41%	1.13%	20.03%	0.07%	25.07%	0.09%	22.84%
Slope position	12.91%	0.42%	14.64%	0.10%	15.27%	0.05%	14.28%
% slope	6.57%	0.18%	8.32%	0.13%	8.40%	0.12%	7.76%
2000 Geocover PC2	2.79%	0.04%	7.68%	0.03%	4.45%	0.05%	4.98%
log(catchment area)	4.74%	0.34%	5.11%	0.04%	4.80%	0.06%	4.88%
Profile curvature	6.15%	0.03%	2.50%	0.04%	2.54%	0.06%	3.73%
1990 Geocover PC2	2.56%	0.04%	3.03%	0.04%	3.20%	0.05%	2.93%
2000 Geocover PC1	1.81%	0.03%	2.83%	0.04%	3.27%	0.05%	2.64%
Planar curvature	2.57%	0.05%	2.56%	0.03%	2.37%	0.04%	2.50%
Terrace	1.57%	0.06%	2.56%	0.06%	3.35%	0.09%	2.49%
Eastness	1.76%	0.03%	2.26%	0.03%	2.42%	0.04%	2.15%
Southness	1.49%	0.03%	2.03%	0.03%	2.18%	0.04%	1.90%
1990 Geocover PC1	0.97%	0.02%	1.16%	0.03%	1.21%	0.03%	1.11%

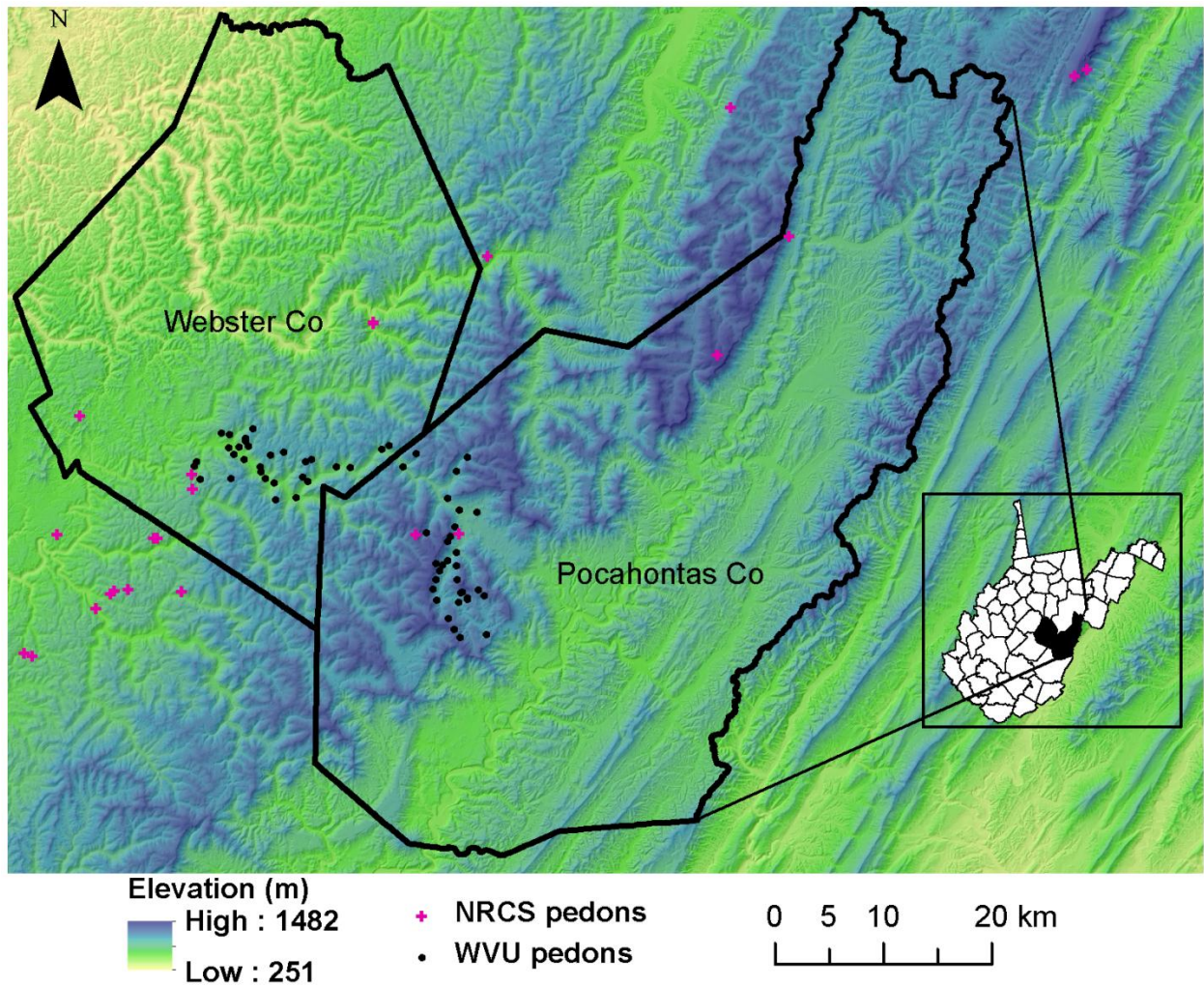


Figure 1. Study area in West Virginia, USA. Points show pedon locations used in the validation of disaggregation results. Black outlines delineate the Pocahontas and Webster County soil surveys used to train the model. Results were extrapolated to nearby areas with consistent covariate data to incorporate more validation points. (Inset: Location of Pocahontas and Webster Counties in West Virginia.)

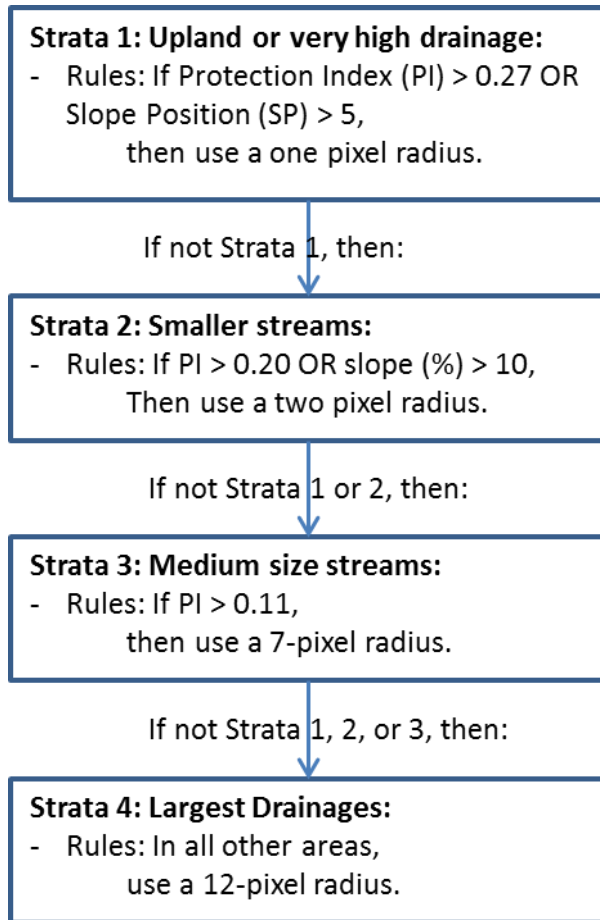


Figure 2. The ruleset used to create the terrace raster from relative elevation layers of various neighborhoods. This was used for training area selection and in decision tree models.

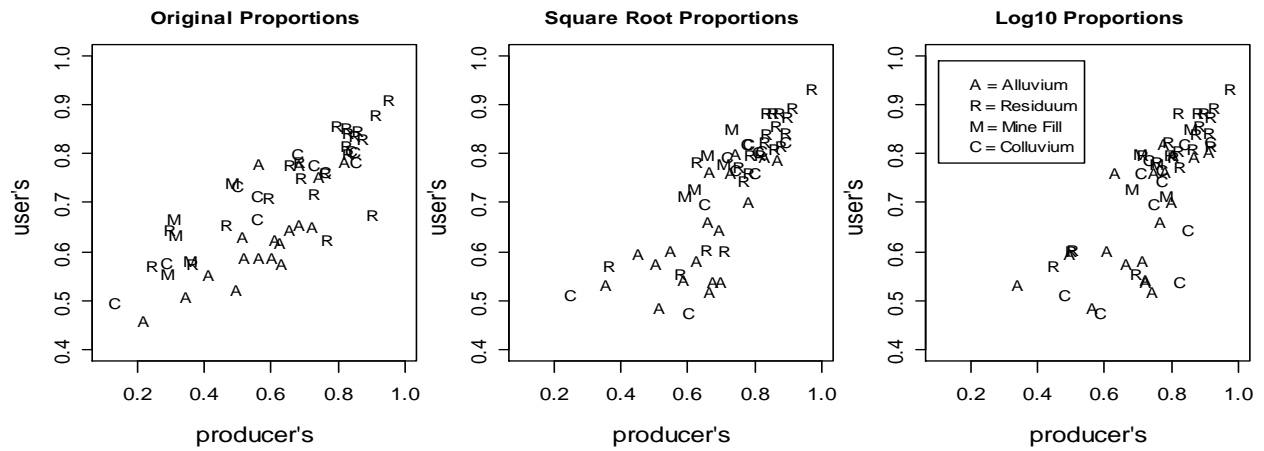


Figure 3. User's versus producer's accuracy summary figures. Values are expressed in fractions.

Note the generally lower values of producer's accuracy that require a larger x-axis scale in plots.

Each class is represented by its parent material type as indicated in the legend.

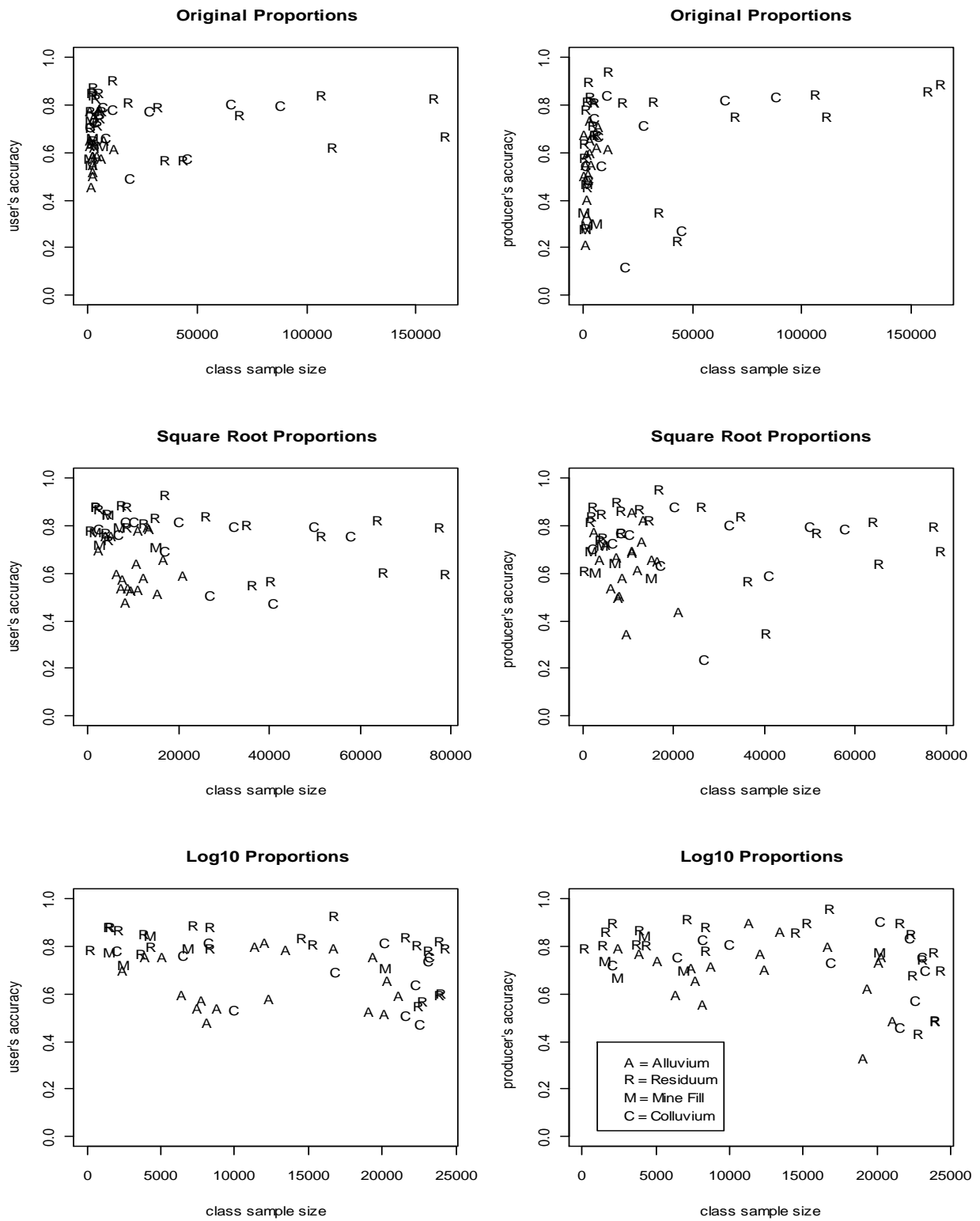


Figure 4. Bivariate plots comparing training accuracies versus the size of class training samples for different sampling schemes. Each class is symbolized by general parent material type.

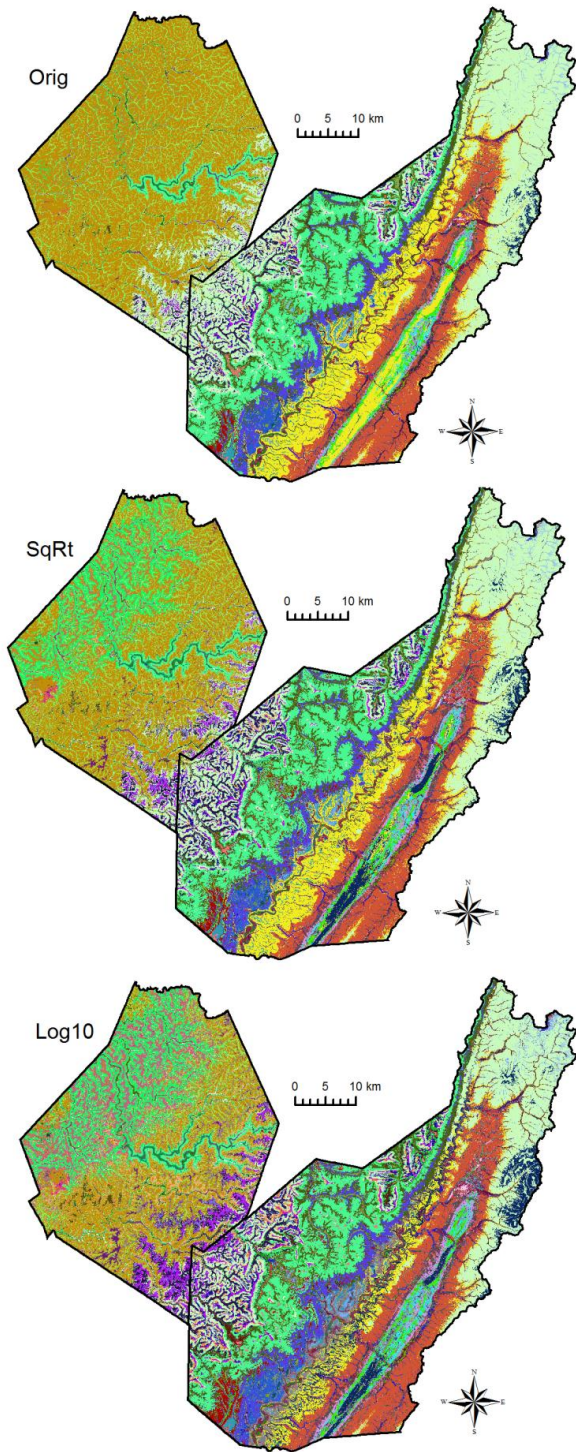


Figure 5. Maps created from the models created from original proportions, square root, and base-ten logarithm sampling ensembles. Soil types are symbolized by unique colors and are consistent across all three maps.

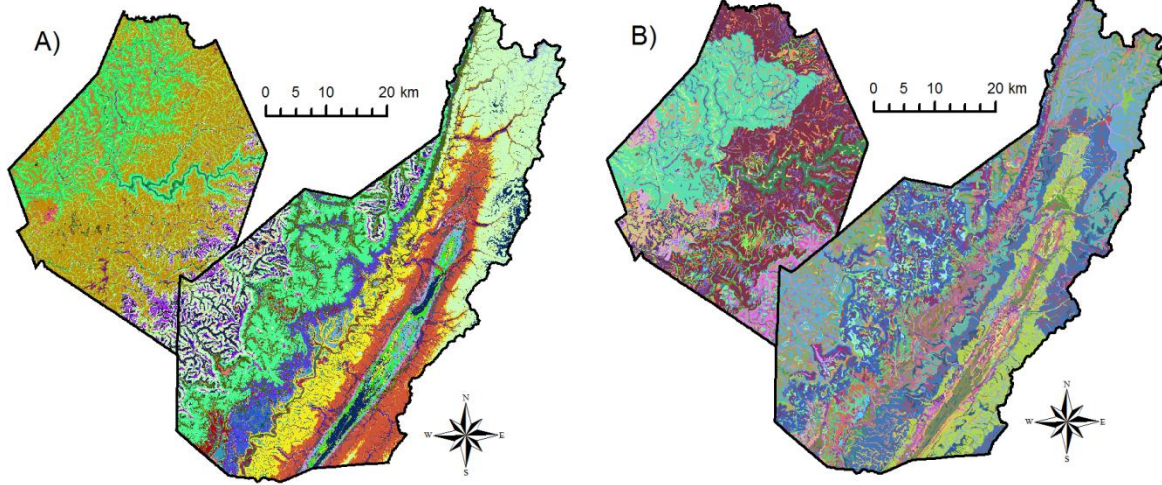


Figure 6. Maps comparing the SqRt model and the original soil survey map units for Webster and Pocahontas County Soil surveys. Soil types (SqRt model) and map units (surveys) are symbolized by unique colors.

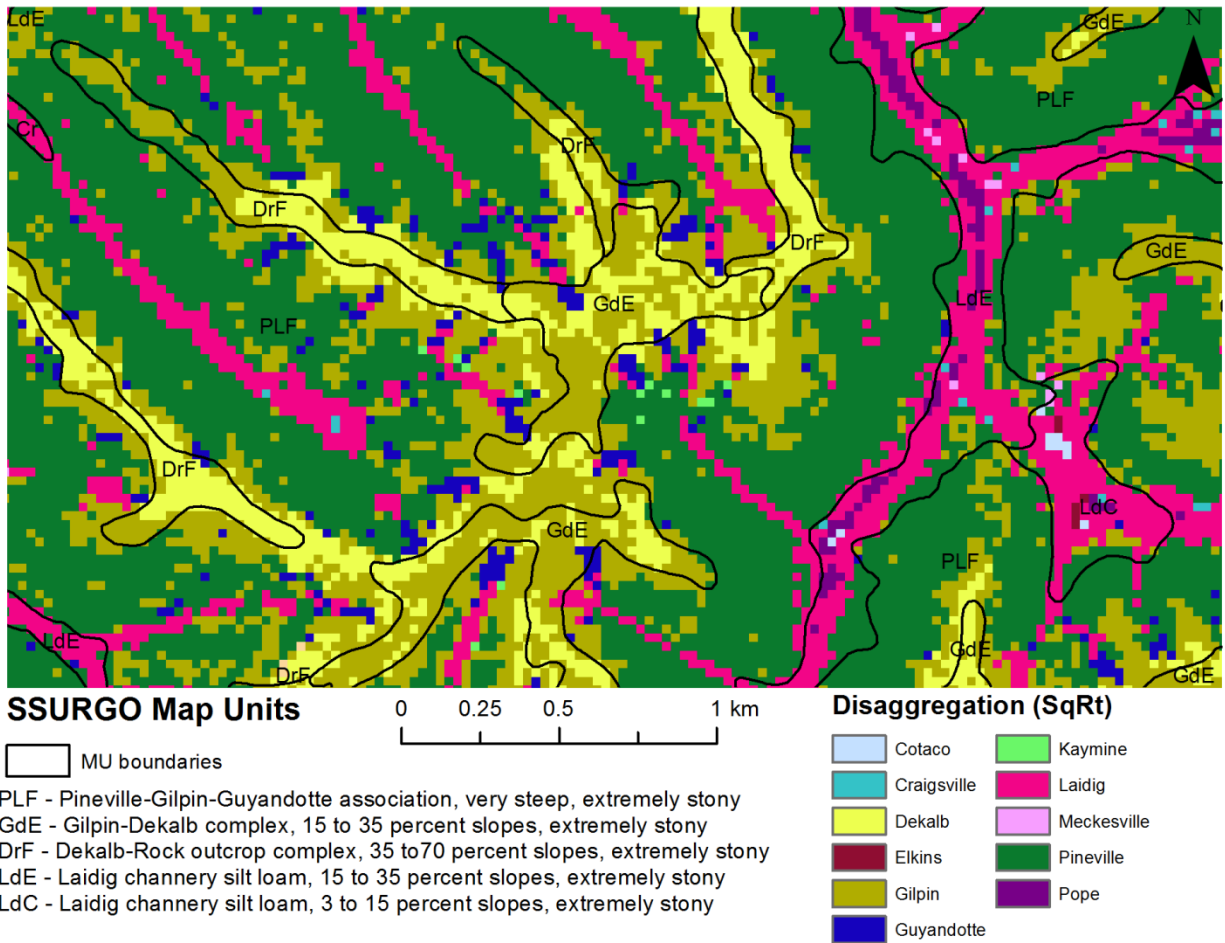


Figure 7. Map with SSURGO map units overlaying disaggregation predictions for the SqRt (square rootsampled) model for part of the study area.

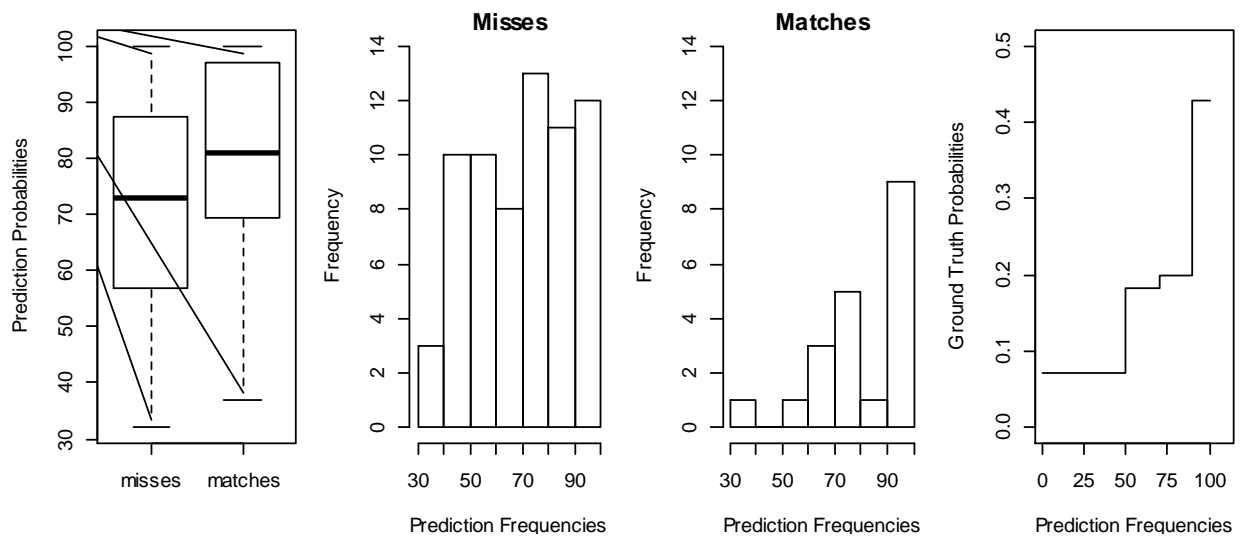


Figure 8. Distributions of prediction frequencies for pixels that correctly predicted (matches) soil series and misclassified (misses) soil series for the nearest neighbor validation of the square root disaggregation model: (a) boxplots of all data; (b) distribution of prediction frequencies for misclassified points; (c) distribution of prediction frequencies for correctly classified points; (d) plot of ground truth probabilities for different ranges of prediction frequencies.

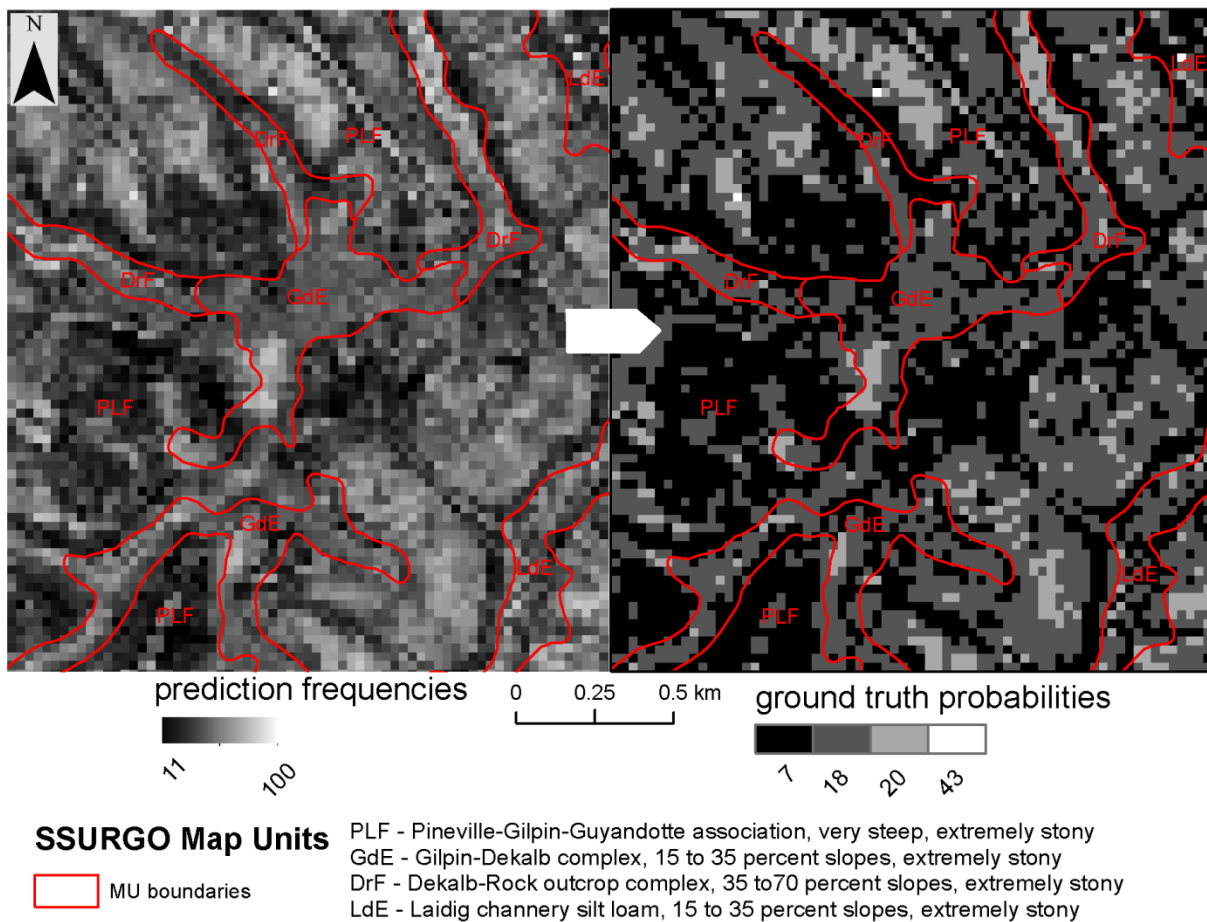


Figure 9. Maps showing translation of prediction frequencies to ground truth probabilities using the step function from Fig. 8. This extent covers some of the same data (SqRt model) extent as Figure 7.

4 SEMI-AUTOMATED DISAGGREGATION OF A CONVENTIONAL SOIL MAP USING KNOWLEDGE DRIVEN DATA MINING AND RANDOM FORESTS IN THE SONORAN DESERT, USA

4.1 *Citation*

Nauman, T.W., J.A. Thompson, and C. Rasmussen. 2014. Semi-Automated Disaggregation of a Conventional Soil Map Using Knowledge Driven Data Mining and Random Forests in the Sonoran Desert, USA. *Photogrammetric Engineering & Remote Sensing* 80:353-366. Reproduced with permission from the American Society for Photogrammetry and Remote Sensing, Bethesda, MD, www.asprs.org.

4.2 *Abstract*

Conventional soil maps (CSM) have provided baseline soil information for land use planning for over 100 years. Although CSM have been widely used, they are not suitable to meet growing demands for high resolution soil information at field scale. We present a repeatable method to disaggregate CSM data into ~30-meter resolution rasterized soil class maps that include continuous representation of probabilistic map uncertainty. Methods include training set creation for original CSM component soil classes from soil-landscape descriptions within the original survey database. Training sets are used to build a random forest predictive model utilizing environmental covariate maps derived from ASTER satellite imagery and the USGS National Elevation Dataset. Results showed agreement at 70% of independent field validation sites and equivalent accuracy between original CSM map units and the disaggregated map. Uncertainty of predictions was mapped by relating prediction frequencies of the random forest model and success rates at validation sites.

4.3 *Introduction*

The increased availability of both digital elevation data and remote sensing data have prompted many studies that use these data to improve soil property prediction and inventory in a field that has been coined ‘digital soil mapping’ (DSM) (Grunwald et al., 2011; Grunwald, 2009; McBratney et al., 2003; Scull et al., 2003). Many of these studies use elevation data and remotely sensed imagery to represent one or more soil forming factors that include climate, organisms, relief, parent material, and time (Jenny, 1941). In this form, soil classes or properties are predicted from topographic or spectral indices derived from elevation and imagery.

Soil properties and functions influence many societal challenges particularly the response of ecosystem services such as carbon and nutrient cycling, water storage, purification, and cycling, pollutant transport, and vegetation growth to climate change (Brady and Weil, 2008). However, our knowledge of soils is imprecise as demonstrated by estimates of global soil carbon stocks in the top meter of soil that range from 1400 to 3250 petagrams (Grunwald et al., 2011). In light of the projected challenges that climate change presents to ecosystem services (IPCC, 2007), high quality soil information is central to natural resource management and land use planning.

Although many soil inventories in the form of CSM have been carried out around the world, the scope and coarse spatial resolution of many soil databases have been criticized as limitations to effective incorporation of soil information into models of ecosystem services and other earth surface processes (Burrough, 1989; Burrough et al., 1997; Grunwald, 2009; Grunwald et al., 2011; McBratney et al., 2003). The field of DSM has responded to this challenge with concerted efforts to quantitatively improve CSM soil information using a wide array of statistical, spatial, and information technology (Bui et al., 2009; Bui et al., 2006; Bui et al., 1999; Bui and Moran, 2001; Cook et al., 1996a; Cook et al., 1996b; de Bruin et al., 1999; Häring et al., 2012; Kempen

et al., 2009; Kerry et al., 2012; McBratney, 1998; Minasny and McBratney, 2010; Nauman and Thompson, 2014; Nauman et al., 2012; Thompson et al., 2010; Yang et al., 2011; Zhu, 1997; Zhu et al., 1997, 2001).

One of the main challenges to improving CSM data representation is that the original intent of CSM was management oriented, and properties attributed to soils were often estimates based on sparse data at representative locations and not quantifications based on rigorous statistical sampling and interpolation (USDA-NRCS, 2013). A large part of the goals of the original design of CSM was to provide somewhat qualitative interpretations intended to provide pragmatic initial guidance to developers, farmers, and other land management institutions (Soil Survey Staff, 1993). However, many current users of soil information, particularly those not familiar with CSM history and evolution, have attempted to use CSM data beyond their original purposes leading to the potential for spurious relationships and possible incorrect data and process interpretation.

Various models and analyses have been developed using spatial soil information from CSM (e.g., Gatzke et al., 2011; Lineback Gritzner et al., 2001; Thomas-Van Gundy et al., 2012; Thomas-Van Gundy and Strager, 2012). In the U.S., both the U.S. General Soil Map (STATSGO2) and the Soil Survey Geographic (SSURGO) database aggregate multiple soil classes into spatial map unit polygon delineations (Soil Survey Staff, 1993; Thompson et al., 2012). The data model for SSURGO includes polygon map units with generally one to four named soil series (soil taxonomic class) per map unit, plus minor inclusions of soils or non-soil areas. This aggregation, and the inherently crisp breaks that choropleth style mapping impose on spatial data, make spatial representation and analysis of soils somewhat convoluted and predisposed to improper interpretation. For example, there are often abrupt changes in property

values between polygons or at survey project boundaries that do not follow natural breaks in the landscape (Loerch, 2012; Thompson et al., 2012). The emerging challenge is how to use the wealth of information in CSM in an appropriate way. We believe that part of the answer is to restructure the data by spatially disaggregating the information in soil map units into component parts in a manner that better represents the continuous nature of soil data in the field.

4.3.1 *Soil Survey Spatial Disaggregation*

The primary focus of soil survey spatial disaggregation is to express a realistic spatial distribution of soil individuals in cases where CSM aggregate multiple soils or create unrealistic crisp boundaries between soil classes. This can be considered an enhancement of a prior generalized soil map to produce a more consistent and detailed map that spatially distinguishes soil types and/or properties at a finer resolution. These techniques also tend to translate the data from polygonal maps to gridded raster formats. Disaggregation has been identified as a conceptual approach to translate current data into new higher resolution products better suited for modern applications (Bui, 2004; Bui et al., 1999; Bui and Moran, 2001; de Bruin et al., 1999; McBratney, 1998; Wielemaker et al., 2001). Generally, approaches use new pedon data and/or environmental covariate data in a DSM framework to determine how soils within polygon map units vary spatially.

Spatial disaggregation of CSM has been demonstrated in attempts to recreate soil maps without the original polygons, which we call universal soil map updates (Bui and Moran, 2001; Hansen et al., 2009; Moran and Bui, 2002; Nauman and Thompson, 2014; Smith et al., 2012; Wei et al., 2010; Yang et al., 2011). Others have updated CSM within the bounds of original survey map units (Bui and Moran, 2001; Thompson et al., 2010; Häring et al., 2012). Other studies have looked at disaggregating CSM for specific soil properties (Goovaerts, 2011; Kerry

et al., 2012; Nauman et al., 2012). Goovaerts (2011) evaluated geostatistical methods that can combine point data with choropleth data to look at intra-polygon variation of a specific variable, and Kerry et al. (2012) applied parts of these methods to soil organic carbon mapping in northern Ireland. Fuzzy logic has been used in disaggregation through applications like SoLIM (Qi et al., 2006; Zhu, 1997; Zhu et al., 1996) to help organize and implement soil-landscape relationships for mapping soils. SoLIM has been used in coordination with both expert knowledge (Smith et al., 2010) and statistical approaches (Yang et al., 2011) to implement discovered soil-landscape relationships for updating and disaggregating soil maps. Other fuzzy knowledge systems have leveraged landform element classifications to better disaggregate landscapes into units with similar soils (MacMillan et al., 2000). Classification and regression trees have also been a prominent technique used in disaggregation. Bui et al. (2001) and Wei et al. (2010) both used ensembles of decision trees and Haring et. al. (2012) used random forests to refine soil and surficial geology classes. Tree-based models have also been used extensively in general DSM applications and seem to have the greatest flexibility of common modeling methods (Behrens et al., 2005; Behrens et al., 2010a; Behrens et al., 2010b; Bui et al., 2009; Lemerrier et al., 2012; McKenzie and Ryan, 1999; Moran and Bui, 2002; Saunders and Boettinger, 2007; Schmidt et al., 2008; Scull et al., 2005).

The objective of this research was to address the common situation where an older CSM is available, but more detailed soil spatial data is needed and few soil observations are available. We also compare the usefulness of a variety of ASTER satellite imagery and USGS 1 arc-second National Elevation Dataset (NED) derived data layers for use in CSM disaggregation. We utilize soil-landscape rules that are usually present in soil survey database map unit descriptions in combination with a random forest to universally disaggregate a CSM to a ~30-meter resolution

raster soil class map without collecting new field data. This approach leverages both the implicit information of the SSURGO spatial data (the standard CSM data product), and explicit expert knowledge about soil-geomorphology relationships attributed in the SSURGO database to create a training set. It pairs the training set with elevation and imagery in a random forest classification tree ensemble model. We selected methods and data sources based on repeatability, transparency, and manageability in an effort to make them more accessible to soil science professionals in government and consulting.

4.4 *Methods*

4.4.1 *Study Area*

Organ Pipe Cactus National Monument (ORPI) is located in the Basin and Range physiographic province of southern Arizona, USA. The area is characterized by alternating mountain ranges of diverse lithology and broad alluvial valleys with bajada and basin floor systems (Hendricks, 1985). The mountain ranges within ORPI include both intrusive and extrusive igneous rocks as well as sedimentary and meta-sedimentary materials with a wide variety of mineral assemblages (Fig. 1) (Bezy et al., 2000; Eddy et al., 1991). The geologic history of the area includes four distinct periods of volcanism starting 1.6 billion years ago and ending with the last episode 26 to 14 million years before present (Bezy et al., 2000; Eddy et al., 1991). Tectonic uplift and erosion have worked and reworked the landscape during this time to produce a complicated arrangement of mountains rising up to 1,465 meters in elevation with an intricate assemblage of associated alluvial outwash landforms (Bezy et al., 2000). Current area geomorphology is a result of Pleistocene and Holocene aggradations and entrenchment cycles leaving complex arrangements of deposits with varying dissection and escarpment patterns that

differ between lithologic source materials (Bezy et al., 2000; McAuliffe, 1994; Parker, 1991; Parker, 1995; Simpson, 1991).

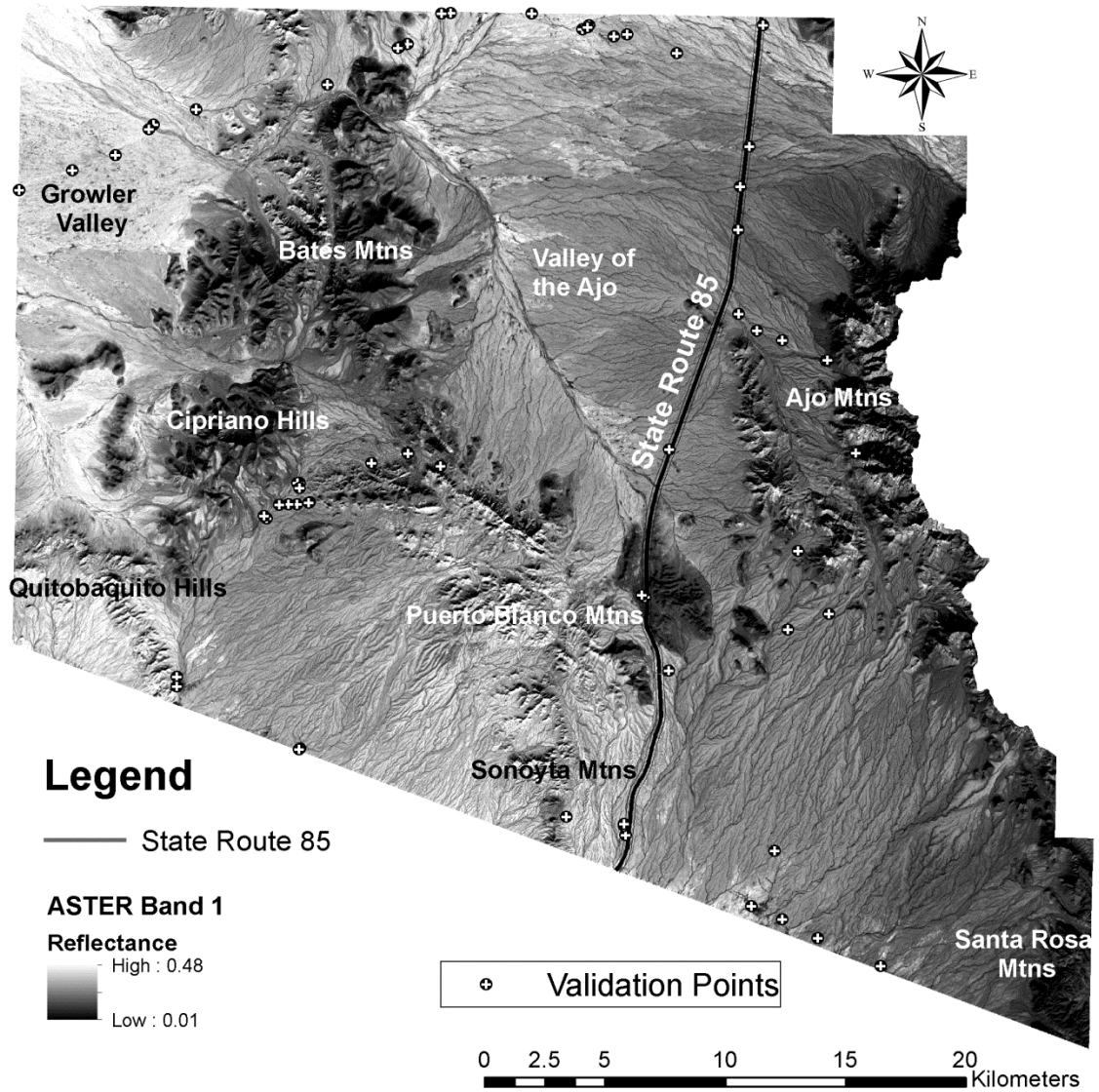


Figure 1. ASTER satellite image of Organ Pipe Cactus National Monument and validation locations.

The area spans the transition from the Arizona Upland to Lower Colorado River Valley subdivision of the Sonoran Desert. This includes a variety of vegetation communities including juniper woodland in the high Ajo Mountains, desert scrub columnar cacti communities on

Bajadas, and sparsely vegetated creosote flats in the Growler Valley (Fig. 1) (Parker, 1991). Average annual precipitation at the ORPI headquarters is reported to be 251 mm with a strong bimodal distribution characterized by summer monsoon precipitation and a moderate winter rainy season (NOAA, 2004). The mean annual temperature from 1971 to 2000 was 21.6°C with a range of -10 to 47.8°C (NOAA, 2004). A precipitation gradient exists within ORPI decreasing from 342 mm in the high elevation Ajo Mountains to 190 mm in the western areas of the monument (Parker, 1991). Soils in ORPI were all classified as having an aridic soil moisture regime and hyperthermic soil temperature regime (USDA-NRCS, 1972). However, high elevation areas characterized by juniper woodland may include ustic soil moisture and thermic soil temperature regimes.

4.4.2 *Training Set Creation*

The CSM dataset used here includes a SSURGO dataset that consists of a polygon format vector map attributed with a map unit label and a relational database that attaches soil information to the map units. SSURGO includes multiple types of map units that generally have one to four named soil series components as well as ‘inclusions’ of other soils or non-soil areas. Each of these component soil series can have different property distributions that must be generalized or aggregated somehow if a user wants to display a soil property using SSURGO polygons (e.g. Bliss et al., 1995; Thompson and Kolka, 2005).

In SSURGO, each component of every map unit has information regarding soil properties, as well as geologic and geomorphic characteristics, attributed to it (Table 1). After reviewing an extensive set of environmental rasters and SSURGO attributes for potential use in training models, we determined that a simple scheme that matches DEM derived layers to geomorphology attributes would be effective. By querying the geomorphic landform tables in

SSURGO (*cogeomorphdesc* and *cosurfmorphgc*), soil-landscape relationship descriptions were extracted from the database to help determine where within a map unit a component is expected to occur, e.g., Growler series exists on the convex portions of valley floors. The geomorphic landform descriptions in these queries were then matched to values in environmental rasters that represent topographic wetness index (TWI) (Yang et al., 2007), created using Tarboton's (1997) surface flow routing algorithm, and to relative elevation metrics within different neighborhoods to help distinguish components within map units (Table 1).

Rule sets were developed to match the descriptive language from the geomorphology queries as well as delineate small washes present in map units as inclusions. In most single component map units this only involved trying to eliminate small washes that were inclusions by setting a TWI cutoff of 17. This threshold was chosen by draping the TWI layer onto high resolution USGS 1-meter DOQQ imagery and matching a TWI value to delineate visible washes. Single component map units with younger soils that still flood, e.g. Gilman very fine sandy loam, were left alone for training selection because they were deemed acceptable as is.

For multi-component map units, each component geomorphic description was examined, and DEM variables were selected on the basis of which variable best distinguished the labeled differences. This was often difficult because the language in SSURGO can seem contradictory. For example, the Growler series in the Growler-Antho complex is described on 'valley floors in convex portions', and Antho is described on 'flood plains in dips' or on 'alluvial fans' in areas with 'terrace tread' (Table 1). This seems to indicate that there are areas of flood plain, alluvial fan and valley floor in the map unit. However, when these map units were examined in the field and in aerial photography, they appeared to be alluvial fan shaped delineations that also have very subtle topography more like a valley floor with slightly lower areas that still flood, and

other slightly higher areas that do not receive much overland flow. Based on this observation, we decided to use the TWI raster to split these areas apart into lower wetter areas and higher convex locations. We also checked the original hardcopy soil survey manuscript for clarification which indicated that “Growler soils lie on slightly elevated convex areas” and are “easily recognized” by “varnished desert pavement and sparse vegetation”. Similarly, the report states that “Antho soils lie between the Growler areas and along shallow drainage ways” (USDA-NRCS, 1972). Based on this we found that the TWI raster distinguished drainage patterns in Growler-Antho complex in such a way as to delineate subtle washes and varnished surfaces based on comparing TWI breaks with visual inspection of 1-meter USGS DOQQ aerial photography. Following this process, we translated soil-landscape relationship records in the SSURGO database to environmental raster values. This rule translation attempted to emulate and streamline the approach Thompson et al. (2010) used to create spatial soil-landscape rule-sets. We simply employ these rule-sets as a means to create a model training set.

All rules were presumed to identify typical landscapes for respective component soils within each map unit. The areas selected by each component rule were added to a training set that was compiled for all soil series and other named components (e.g., rock outcrops or higher taxa such as Torrifluvents). In other words, areas from all map units that were ‘typical’ of a given soil series were combined into one training set for each respective soil series. All environmental rasters used in rule-matching were derived from the 1-arc second USGS national elevation dataset (NED) (Gesch, 2007; Gesch et al., 2002). Final maps and other raster data used in later modeling steps were co-registered to the NED grid with the North American Datum of 1983 Universal Transverse Mercator projection in Zone 12-North.

4.4.3 *Model Building*

Training areas for each soil series/component were randomly sampled proportionally to component areal extent in the original CSM (following Moran and Bui, 2002) to produce two ensemble models. Random forest classifications were built from the training sets using a more exhaustive set of environmental covariate rasters than were used in the original training rule-matching (Table 2). Variables used by Nauman (2009) for unsupervised soil-landscape classifications were used for one model ('Original' model) to compare with that study, and a set of new variables that highlight subtle topographical differences were added to that dataset to build an 'Updated' model. The Updated model added a suite of relative elevation metrics to help with classification based on results from recent studies that show that varying neighborhood sizes of terrain indices can improve spatial prediction of soils (Behrens et al., 2010a; Behrens et al., 2010b).

Table 1. Component level rule matching for training set creation. Some components have more than one geomorphic description.

Map Unit Name	Component (% of map unit)	SSURGO Component Geomorphic Descriptions*	Raster rules		
			TWI**	H3**	R20**
Ajo very cobbly sandy loam, 2 to 5 percent slopes	Ajo (90%)	terrace tread on alluvial fans	twi < 17		
Ajo very gravelly loam, 1 to 5 percent slopes	Ajo (90%)	terrace tread on alluvial fans	twi < 17		
Antho fine sandy loam	Antho (95%)	terrace tread on alluvial fans	no rules***		
Antho fine sandy loam	Antho (95%)	dips in flood plains	no rules***		
Antho soils, very gravelly variants	Antho (85%)	terrace tread on alluvial fans	twi < 17		
Cherioni gravelly very fine sandy loam, 0 to 8 percent slopes	Cherioni (95%)	low beveled side slopes of hills	twi < 17		
Cipriano gravelly loam	Cipriano (90%)	terrace tread on bajadas	twi < 17		
Cipriano gravelly loam	Cipriano (90%)	terrace tread on alluvial fans	twi < 17		
Gachado extremely cobbly loam, 2 to 8 percent slopes	Gachado (75%)	toe slopes of hills	twi < 17		
Gachado extremely cobbly loam, 2 to 8 percent slopes	Gachado (75%)	mountainflank toe slopes	twi < 17		
Gilman very fine sandy loam	Gilman (90%)	alluvial fans, lower, terrace tread	no rules***		
Gilman very fine sandy loam	Gilman (90%)	dips in flood plains	no rules***		
Gilman very fine sandy loam, saline	Gilmansaline (90%)	terrace tread on lower alluvial fans	no rules***		
Gilman very fine sandy loam, saline	Gilmansaline (90%)	dips in flood plains	no rules***		
Growler-Antho complex	Antho (45%)	dips in flood plains	twi >= 14.5		
Growler-Antho complex	Antho (45%)	terrace tread on alluvial fans	twi >= 14.5		
Growler-Antho complex	Growler (35%)	convex portions of valley floors	twi < 14.5		
Gunsight very gravelly loam, 0 to 2 percent slopes	Gunsight (75%)	terrace tread on lower portions of alluvial fans	twi < 17		
Gunsight very gravelly loam, 2 to 15 percent slopes	Gunsight (80%)	terrace tread on lower portions of alluvial fans	twi < 17		
Harqua very cobbly loam, 0 to 8 percent slopes	Harqua (90%)	terrace tread of degrading surface on plains	twi < 17		
Harqua very gravelly loam, 0 to 3 percent slopes	Harqua (90%)	terrace tread of degrading surfaces on plains	twi < 17		

Map Unit Name	Component (% of map unit)	SSURGO Component Geomorphic Descriptions*	Raster rules		
Harqua-Gunsight complex	Gunsight (40%)	terrace tread on lower portions of alluvial fans	twi < 17		r20 > 0
Harqua-Gunsight complex	Harqua (40%)	terrace tread on degrading surfaces of plains	twi < 17		r20 <= 0
Laveen loam	Laveen (85%)	tread of old terraces	twi < 17		
Lomitas extremely stony loam, 8 to 40 percent slopes	Lomitas (75%)	hills, Side Slope	twi < 17		
Lomitas extremely stony loam, 8 to 40 percent slopes	Lomitas (75%)	mountainflanks	twi < 17		
Perryville very cobbly fine sandy loam, 0 to 8 percent slopes	Perryville (80%)	terrace tread of old alluvial fans	twi < 17		
Rillito gravelly sandy loam	Rillito (75%)	terrace tread on alluvial fans	twi < 17		
Rock land	Rock land (90%)	mountain slopes	twi < 17		
Rock outcrop	Rock outcrop (90%)	mountain slopes and peaks	twi < 17		
Stony land-Rock outcrop association	Rock outcrop (30%)	mountain peaks	twi < 17	h3 > 45	
Stony land-Rock outcrop association	Stony land (65%)	hill side slopes	twi < 17	h3 <= 45	
Stony land-Rock outcrop association	Stony land (65%)	mountainflanks	twi < 17	h3 < 45	
Torrifluvents	Torrifluvents (90%)	dips with eroded overflow stream channels	original map units or twi > 17 (any map unit)		

* These are interpretations that take the original nouns in the SSURGO database and link them using prepositions to create meaningful context.
** twi = topographic wetness index,
h3 = relative elevation above local 3-pixel minimum
r20 = relative elevation w/ respect to 20 pixel local average
*** No rules were included for these soils because they exist on floodplains effectively eliminating the need for exclusion of washes by twi.

Table 2. List of raster covariate layers used for building tree models from selected training sets.

Variable Symbol	Name	Original or Updated*	Explanation
<i>b1</i>	ASTER Band 1	Original	VNIR Reflectance (0.52-0.6 μm)
<i>b2</i>	ASTER Band 2	Original	VNIR Reflectance (0.63-0.69 μm)
<i>b3</i>	ASTER Band 3	Original	VNIR Reflectance (0.76-0.86 μm)
<i>b4</i>	ASTER Band 4	Original	SWIR Reflectance (1.6-1.7 μm)
<i>b5</i>	ASTER Band 5	Original	SWIR Reflectance (2.145-2.185 μm)
<i>b6</i>	ASTER Band 6	Original	SWIR Reflectance (2.185-2.225 μm)
<i>b7</i>	ASTER Band 7	Original	SWIR Reflectance (2.235-2.285 μm)
<i>b8</i>	ASTER Band 8	Original	SWIR Reflectance (2.295-2.365 μm)
<i>b9</i>	ASTER Band 9	Original	SWIR Reflectance (2.360-2.430 μm)
<i>b2b1</i>	ASTER Ratio 2/1	Original	Reflectance Ratio Bands 2/1
<i>b2b4</i>	ASTER Ratio 2/4	Original	Reflectance Ratio Bands 2/4
<i>b2b5</i>	ASTER Ratio 2/5	Original	Reflectance Ratio Bands 2/5
<i>b2b6</i>	ASTER Ratio 2/6	Original	Reflectance Ratio Bands 2/6
<i>b2b7</i>	ASTER Ratio 2/7	Original	Reflectance Ratio Bands 2/7
<i>b2b8</i>	ASTER Ratio 2/8	Original	Reflectance Ratio Bands 2/8
<i>b2b9</i>	ASTER Ratio 2/9	Original	Reflectance Ratio Bands 2/9
<i>b2sd14</i>	Band 2 Std. Dev - 14-pixel	Original	Std Dev of Band 2 in a 14-pixel radius
<i>b2sd3</i>	Band 2 Std. Dev - 3-pixel	Original	Std Dev of Band 2 in a 3-pixel radius
<i>b2sd5</i>	Band 2 Std. Dev - 5-pixel	Original	Strd Dev of Band 2 in a 5-pixel radius
<i>b3b2</i>	ASTER Ratio 3/2	Original	Reflectance Ratio Bands 3/2
<i>b4b3</i>	ASTER Ratio 4/3	Original	Reflectance Ratio Bands 4/3
<i>b4b5</i>	ASTER Ratio 4/5	Original	Reflectance Ratio Bands 4/5
<i>b4b6</i>	ASTER Ratio 4/6	Original	Reflectance Ratio Bands 4/6
<i>b4b7</i>	ASTER Ratio 4/7	Original	Reflectance Ratio Bands 4/7
<i>b4b8</i>	ASTER Ratio 4/8	Original	Reflectance Ratio Bands 4/8
<i>b4b9</i>	ASTER Ratio 4/9	Original	Reflectance Ratio Bands 4/9
<i>b4sd14</i>	Band 4 Std. Dev - 14-pixel	Original	Std Dev of Band 4 in a 14-pixel radius
<i>b4sd3</i>	Band 4 Std. Dev - 3-pixel	Original	Std Dev of Band 4 in a 3-pixel radius

Variable Symbol	Name	Original or Updated*	Explanation
<i>b4sd5</i>	Band 4 Std. Dev - 5-pixel	Original	Strd Dev of Band 4 in a 5-pixel radius
<i>dem</i>	Elevation - meters	Original	1-arc sec. Nat'l Elev. Dataset (Gesch et al., 2002)
<i>ca</i>	Contributing Area	Original	Upstream surface area contributing flow to pixel
<i>twi</i>	Topographic Wetness Index	Original	Calc: $\ln(ca / \tan(\text{slope}))$
<i>wetness_tn</i>	Modified twi	Original	Calc: $\ln(ca / (\text{meandiff25} / \text{range of meandiff25}))$
<i>curvx</i>	Horizontal Curvature	Original	2nd-derivative across slope contour
<i>diff25</i>	Difference from Max - 25-pixel	Original	Max elevation in 25-pixel radius minus the cell value
<i>meandiff25</i>	Mean difference - 25-pixel	Original	Mean elevation in 25-pixel radius minus the cell value
<i>slppos25</i>	Slope Position - 25-pixel	Original	$(\text{Max elevation in 25-pixel radius minus the cell value})/(\text{range})$
<i>swness</i>	Southwestness	Original	A -1 to 1 index of how southwest a slope faces: $\cos(\text{aspect} - 225^\circ)$
<i>plen</i>	Longest Upslope Length	Original	Length of longest flow path above each cell
<i>tlen</i>	Total Upslope Length	Original	Additive length of all upslope flowpaths for each cell
<i>rel_ht_3</i>	Local Height - 3-pixel	Updated	Height of cell above the local minimum elevation in 3-pixel radius
<i>rel_ht_5</i>	Local Height - 5-pixel	Updated	Height of cell above the local minimum elevation in 5-pixel radius
<i>rel_ht_10</i>	Local Height - 10-pixel	Updated	Height of cell above the local minimum elevation in 10-pixel radius
<i>rel_ht_20</i>	Local Height - 20-pixel	Updated	Height of cell above the local minimum elevation in 20-pixel radius
<i>rel_ht_30</i>	Local Height - 30-pixel	Updated	Height of cell above the local minimum elevation in 30-pixel radius
<i>rel_ht_50</i>	Local Height - 50-pixel	Updated	Height of cell above the local minimum elevation in 50-pixel radius
<i>rel_ht_70</i>	Local Height - 70-pixel	Updated	Height of cell above the local minimum elevation in 70-pixel radius
<i>rel_meanht3</i>	Local Relief - 10-pixel	Updated	Height of cell relative to local mean elevation in 3-pixel radius
<i>rel_meanht5</i>	Local Relief - 10-pixel	Updated	Height of cell relative to local mean elevation in 5-pixel radius
<i>rel_meanht10</i>	Local Relief - 10-pixel	Updated	Height of cell relative to local mean elevation in 10-pixel radius
<i>rel_meanht20</i>	Local Relief - 20-pixel	Updated	Height of cell relative to local mean elevation in 20-pixel radius
<i>rel_meanht30</i>	Local Relief - 30-pixel	Updated	Height of cell relative to local mean elevation in 30-pixel radius
<i>rel_meanht50</i>	Local Relief - 50-pixel	Updated	Height of cell relative to local mean elevation in 50-pixel radius
<i>rel_meanht70</i>	Local Relief - 70-pixel	Updated	Height of cell relative to local mean elevation in 70-pixel radius

*Original refers to variables that were used directly from Nauman (2009), and Updated refers to variables that were added to the 'Updated' model beyond those from the 2009 study.

4.4.3.1 Covariate Data Sources

ASTER satellite imagery (Abrams, 2000) and the 1 arc-second USGS National Elevation Dataset (NED) (Gesch, 2007; Gesch et al., 2002) were used for covariate layers. All terrain-based rasters were derived from NED after re-projection to NAD83-UTM12N using a bilinear interpolation (Table 2). An ASTER scene from December 18, 2001, was chosen for clear conditions and spatial coverage of ORPI. The ASTER On-Demand L3 Orthorectified imagery was acquired from LP-DAAC (<http://edcimswww.cr.usgs.gov/pub/imswelcome/>). Radiance at the sensor was calculated from the original imagery scaled radiance (DN values) based upon ASTER coefficients published at the LP-DAAC website (Abrams et al., 2001). These radiance values were subsequently modified using a ground based correction coefficient supplied by the University of Arizona Optical Sciences Remote Sensing Group (Buchanan, 2007). Radiance images were then converted to reflectance values using the COST method (Chavez, 1996; instructions in Appendix A of Nauman, 2009). Inputs for this conversion included an average of two commonly used solar irradiation models for ASTER bands, World Radiation Center (WRC) and 'MODTRAN4' (Thome et al., 2001, p. 264). Earth-sun distances were obtained online from the NASA horizons web-server (Giorgini et al., 1996). ASTER reflectance bands were used for all imagery variables (Table 2).

4.4.3.2 Decision Tree Classification

Tree-based machine learning techniques have shown great potential in the modeling of ecology and soil systems (Bui and Moran, 2001; Henderson et al., 2005; Bui et al., 2006; Minasny and McBratney, 2007; Schmidt et al., 2008; Behrens et al., 2010b). Generally, these algorithms recursively split a dataset by picking breaks in covariate data that help to purify or increase the

information content of the model branches (Breiman, 1984; Pedregosa et al., 2011). The Scikit Machine Learning Tree module was used in Python for decision tree implementation and follows a Classification and Regression Tree (CART) implementation which allows for numerical and categorical variables to be used as inputs as well as for a target variable (Pedregosa et al., 2011; Scikit-learn.org, 2013). The algorithm as we implemented in the Tree module uses Gini's impurity to measure the quality of splits for tree building and randomizes variable selection at each node to implement a Random Forest (Breiman, 2001). We conducted an informal sensitivity analysis with the parameters controlling maximum tree depth and minimum node split sample size to try and balance fit with tree complexity. Due to the large number of training classes and a complicated and geologically stratified study area, we felt that trees needed to be allowed to grow relatively large. A maximum tree depth of 20 splits and a 20-pixel minimum sample size to attempt a split were selected for simple tree pruning parameters. At each split in all trees, 18 variables were randomly chosen from the greater suite for possible use in rule creation. Fifty percent of the training set was randomly sampled with replacement for use in each tree. In each tree training sample, individual component class sizes were trimmed down so that all components retained the same relative proportions as in the original SSURGO survey. A 500-tree ensemble was generated for both models.

4.4.4 *Validation of Disaggregated Maps*

Disaggregated maps were validated with independent field data from 10 USDA-NRCS pedon database locations (National Cooperative Soil Survey, 2012), and 53 field checks in 2006 and 2007 (Fig. 1). Access to ORPI is very difficult due to the remoteness of the area and active smuggling issues along the international border with Mexico. Due to these logistical challenges, field checks were only allowed in limited areas along certain roads. As such,

validation points were not randomly allocated and were located to best represent the units where access was granted. We were able to get points in 17 of 23 SSURGO map units and 16 of 18 disaggregated components (Updated model) given the field limitations (see Table 3). Field checks were determined by small hole and/or auger check of diagnostic soil features (e.g., argillic horizon) and basic soil morphology (i.e., rock content, texture, carbonates, surface rock, and color) to match with the nearest soil series. These field locations were intersected with disaggregation results to estimate overall classification accuracy, and to determine uncertainty based on the underlying ensemble model frequencies.

Table 3. Summarizes the distribution of validation points in original SSURGO map units (left), and the Updated disaggregation model (right).

SSURGO Map Unit (Plate 1a)	# Validation Points	# Correct	% Correct	Predicted Component in Updated Model (Plate 1b)	# Validation Points	# Correct	% Correct
Ajo very gravelly loam, 1% to 5% slopes	6	3	50.0%	Ajo	6	3	50.0%
Antho fine sandy loam	7	6	85.7%	Antho	7	6	85.7%
Cherioni gravelly very fine sandy loam, 0% to 8% slopes	3	3	100.0%	Cherioni	3	3	100.0%
Cipriano gravelly loam	3	3	100.0%	Cipriano	3	3	100.0%
Gachado extremely cobbly loam, 2% to 8% slopes	1	1	100.0%	Gachado	1	1	100.0%
Gilman very fine sandy loam	4	4	100.0%	Gilman	6	5	83.3%
Gilman very fine sandy loam, saline	1	0	0.0%	Gilmansaline	1	0	0.0%
Growler-Antho complex	4	4	100.0%	Growler	2	2	100.0%
Gunsight very gravelly loam, 2% to 15% slopes	9	6	66.7%	Gunsight	18	10	55.6%
Harqua very cobbly loam, 0% to 8% slopes	3	2	66.7%	Harqua	4	2	50.0%
Harqua-Gunsight complex	7	5	71.4%	Laveen	1	1	100.0%
Laveen loam	1	1	100.0%	Lomitas	4	2	50.0%
Lomitas extremely stony loam, 8% to 40% slopes	4	3	75.0%	Perryville	1	1	100.0%
Perryville very cobbly fine sandy loam, 0% to 8% slopes	1	1	100.0%	Rillito	3	3	100.0%
Rillito gravelly sandy loam	3	3	100.0%	Rock land	1	0	0.0%
Rock land	1	0	0.0%	Torrifluvents	2	2	100.0%
Torrifluvents	5	2	40.0%	Total	63	44	69.8%
Total	63	47	74.6%				

4.5 Results

The two disaggregation models performed well with training accuracies of 80% for the Original model and 85% for the Updated model. Corresponding validation accuracies were 66.7% for the Original model and 69.8% for the Updated model (Table 3). At validation sites, the original SSURGO map units listed one of the correct validation soils 74.6% of the time. Although this agreement is higher than the disaggregation models, the multi-component map units inflate the accuracy because they offer more than one possible class that can count for a match in a polygon. In contrast, the disaggregation models always predict one soil for one pixel, so a comparison to the validation of SSURGO map units with multinomial themes is not objective. To better compare performance between the original CSM and disaggregated maps, we looked at validation points that fell into single-component soil consociation SSURGO map units (52 total sites) to see if that affected agreement rates. In consociations, SSURGO matched at 73.1% of sites and the Updated disaggregated map at 75.0% of those same locations indicating very similar accuracies. The resulting disaggregated map presents a single consistent theme (i.e. one soil component per pixel; Plate 1b) whereas the SSURGO map units sometimes aggregate multiple components in map units and also sometimes delineate multiple map units with the same soil by breaking out general slope gradient classes (Plate 1a).

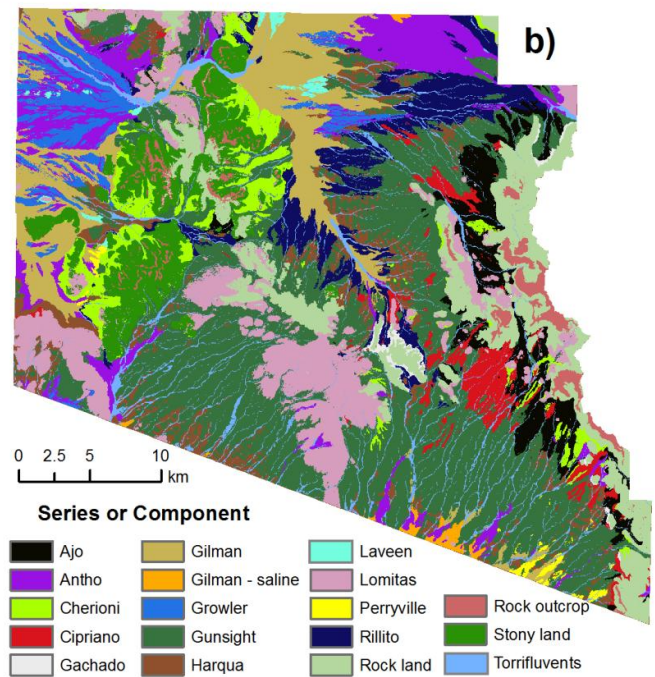
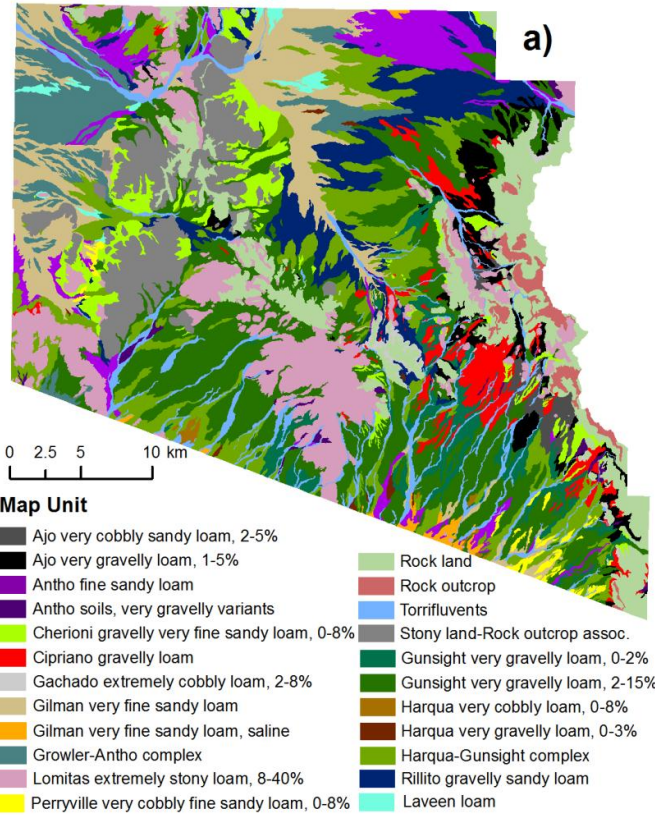


Plate 1. (a) Map of original soil survey map units. Percentages indicate slope gradient ranges attributed to map units. (b) Updated disaggregated map of soil series components.

4.5.1 *Uncertainty Map*

Prediction frequencies of classes in the best performing random forest model (Updated) were compared between pixels that both matched and did not match validation site soil observations to create a simple field data derived uncertainty surface. Figure 2 shows how validation site prediction probabilities (derived from model prediction frequencies) were compared and translated to an uncertainty surface. Prediction probabilities were found to be higher at sites where validation matched predictions (Wilcoxon rank sum test w continuity correction, $W = 571.5$, 1-sided $p = 0.0187$, 2-sided $p = 0.0373$; Fig. 2a). Although match rates seem to increase with prediction probabilities, there is a small drop in field data agreement rates in the highest bin (0.9-1.0) that does not follow the positive trend (Fig. 2d). It is difficult to make any detailed conclusions beyond the overall positive relationship between field data probabilities and prediction probabilities because the actual sample sizes of the step function bins are all less than twenty and are variable in size. These small sample sizes mean that a change of just one validation match would influence any bin by more than $1/20$, or 5%. In this case, the difference in bin sample size between the 0.8-0.9 bin ($n=11$) and 0.9-1.0 bin ($n=19$) makes it difficult to determine if the drop in the 0.9-1.0 probability (Fig. 2d) is due to the difference in bin sample sizes or the predictive ability of the model. However, the overall field data accuracies still are all above 0.56, indicating a good deal of predictive ability across all prediction probabilities. It also appears clear that above a prediction probability of 0.70 the field data probability also goes up to above 0.70. Field data probabilities were mapped by translating prediction probabilities produced from the random forest using the step function shown in Figure 2d in order to create a map that can serve to represent uncertainty in predictions (Fig. 3).

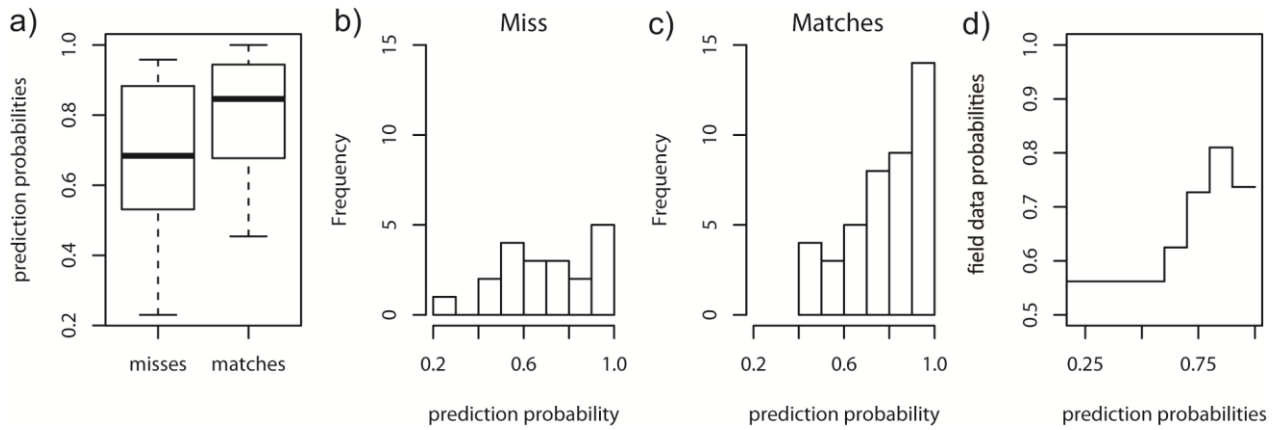


Figure 2. Validation site probabilities compared for Updated model showing the higher tendency of prediction probabilities at matches or correctly predicted sites (a), histograms of prediction probabilities at missed sites (b) and matched sites (c), and the empirical relationship relating prediction probabilities and field data probabilities (d).

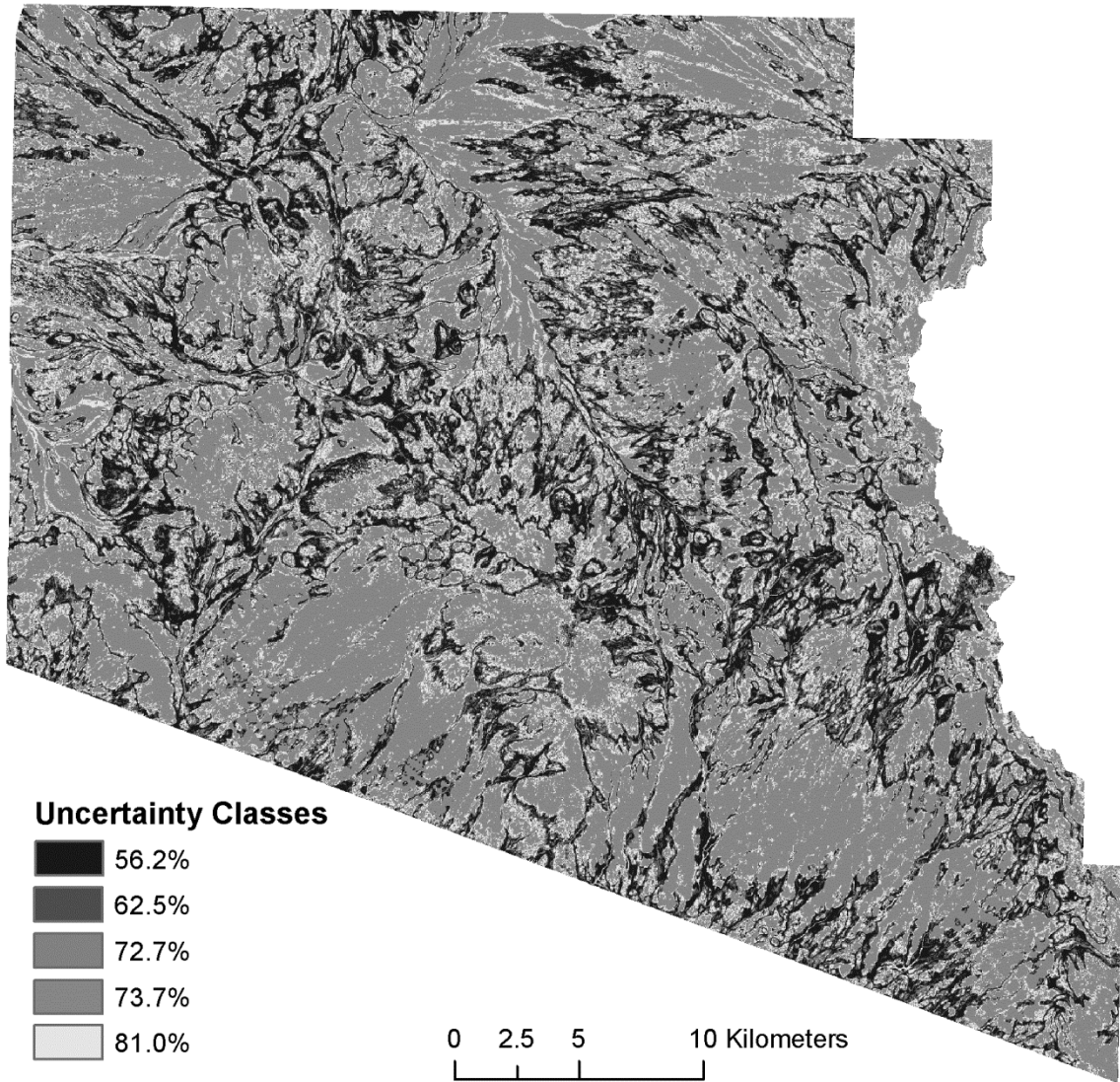


Figure 3. Field data probability representation of uncertainty for the Updated model based on step function in Fig. 2d.

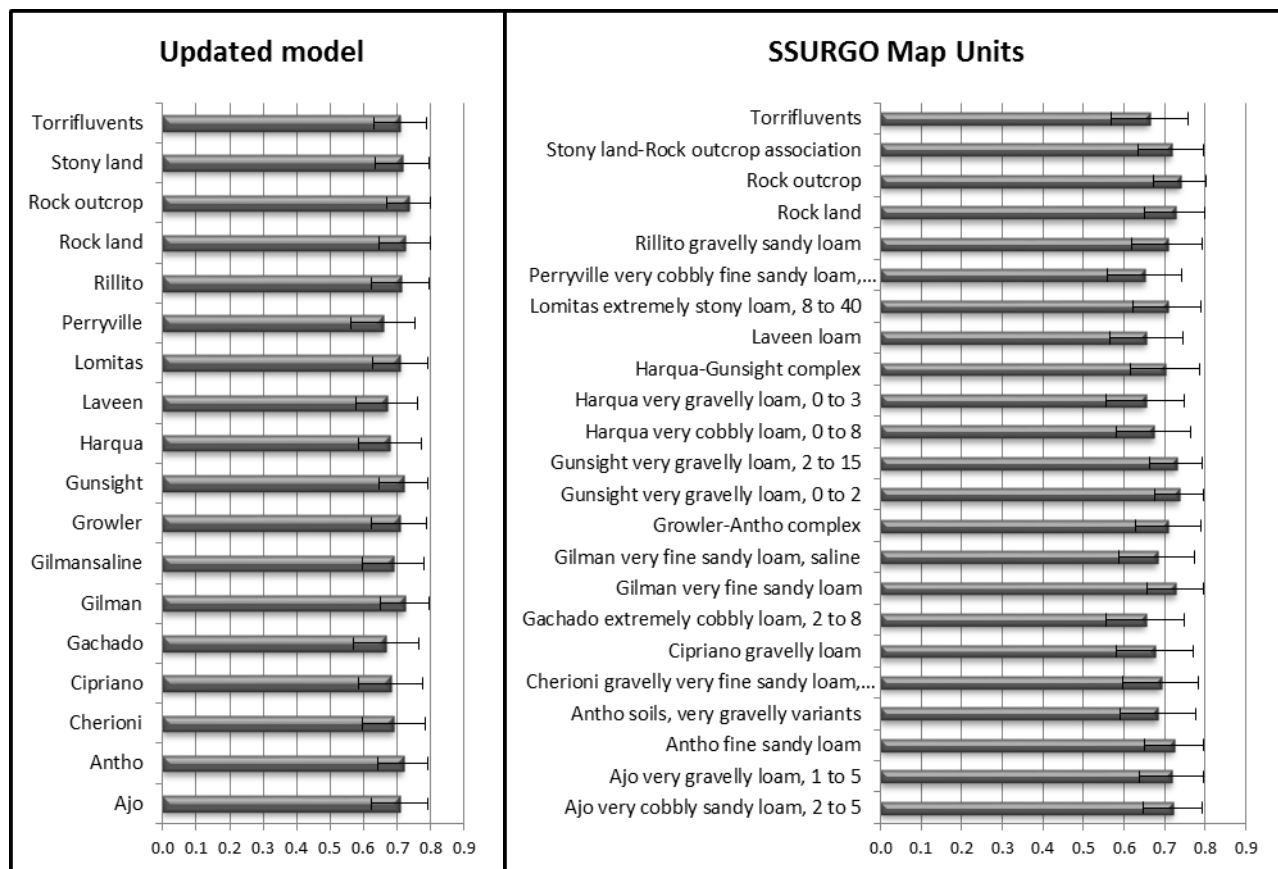


Figure 4. Mean field data probabilities with standard deviation bars summarized by Updated model disaggregation components (left) and SSURGO map units (right).

Examination of field data probability (uncertainty) values among original SSURGO map units and Updated disaggregation model components did show some variability between classes, but all class means were between 0.65 and 0.74 (Fig. 4). Variation is evident within classes when SSURGO and Updated model maps are overlaid on the field data probability map. Visually, there were lower probability areas around some of the SSURGO map unit boundaries where the model appears to transition between classes (unpublished map). Other lower probability areas followed landscape attributes that didn't track any one component type or map unit but tended to occur on certain lower alluvial fan sequences, mountain footslopes, or fan and wash scarps. Delineations of the Growler-Antho complex in the Valley of the Ajo (Fig.1) were a good example of lower alluvial fan units that had lower probability values (Fig. 3).

However, the large area of the Growler-Antho complex in the Growler Valley (Fig. 1) had generally higher probabilities (Fig. 3) suggesting that there might be some kind of difference between Growler-Antho units in the different valleys. Areas of valley floor (e.g. Gilman), Torrifluents (in the Updated model), middle bajadas (e.g. Gunsight), upper alluvial fans (e.g. Ajo), and non-soil components of the larger mountains (e.g. Rock land) tended to have higher field data probabilities. We also observed that in both the Updated model and SSURGO that classes with less than 10,000 pixels (e.g. Gachado, Gilman-saline, Laveen, and Perryville) all had lower field data probability averages while the largest classes (e.g. Antho, Gilman, Gunsight, Lomitas, Rock land, and Stony land) with 100,000 or more pixels had generally higher means.

We are cautious about over-interpretation of this uncertainty data because the field data probability values are means of discretized probability classes created from 63 validation points, and this analysis is generalizing 1.6 million pixels. The large sample sizes (number of pixels) in individual classes would likely result in statistical differences between probability class means in both the SSURGO and Updated model maps. This would mainly be a result of the large class sample sizes that produce extremely low standard errors and hence greater statistical detectability that may not be meaningful. For example, the largest class standard error of field data probability values for all SSURGO maps units or Updated model components was 0.002.

4.5.2 *Important Variables for Models*

Variable usefulness was evaluated based on the relative frequency of each variable's use in the random forests. In both the Original and Updated models, NED derived variables were generally used more than ASTER variables (Fig. 5). The variables *dem* and *diff25* were in the top three used for tree building in both models, suggesting a strong influence from watershed-scale elevation gradients. The Original model, which included fewer NED derived

variables, tended to rely more evenly on ASTER and NED variables, whereas the Updated model uses more NED based layers with 10 of the 14 newly added NED relative elevation variables being used quite frequently. This might suggest that the relative height metrics added to the Updated model provided more consistent predictive power than the ASTER variables they appear to replace in the Original model. However, considering that the Original model was only 3.1% less accurate at validation sites than the Updated model, the differences in variable use may not be that significant. Frequently used ASTER derived layers span a wide scope of types in the Original model including three reflectance layers, one band ratio, and three band neighborhood variation layers. In contrast, the Updated model only includes Band 4 and Band 2 neighborhood variation layers from ASTER. Overall, the *dem* variable and relative elevation surfaces that integrate more than 10 pixels seem to be used most often in trees possibly indicating more predictive power than the other included variables.

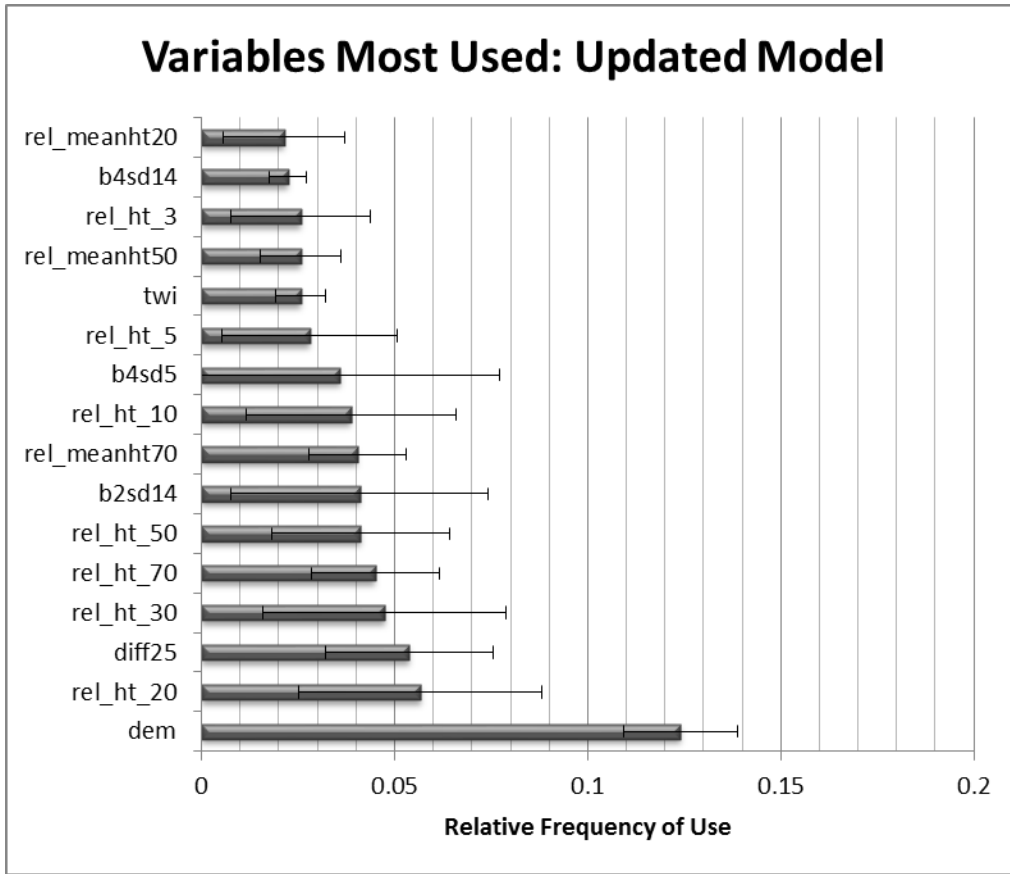


Figure 5a. Average frequency of use of covariate layers (See Table 2 for definitions) used on average in more than 2% of tree nodes in the 500 trees in the Updated model. Black brackets give standard deviations of the frequencies to show how these varied over all ensemble trees.

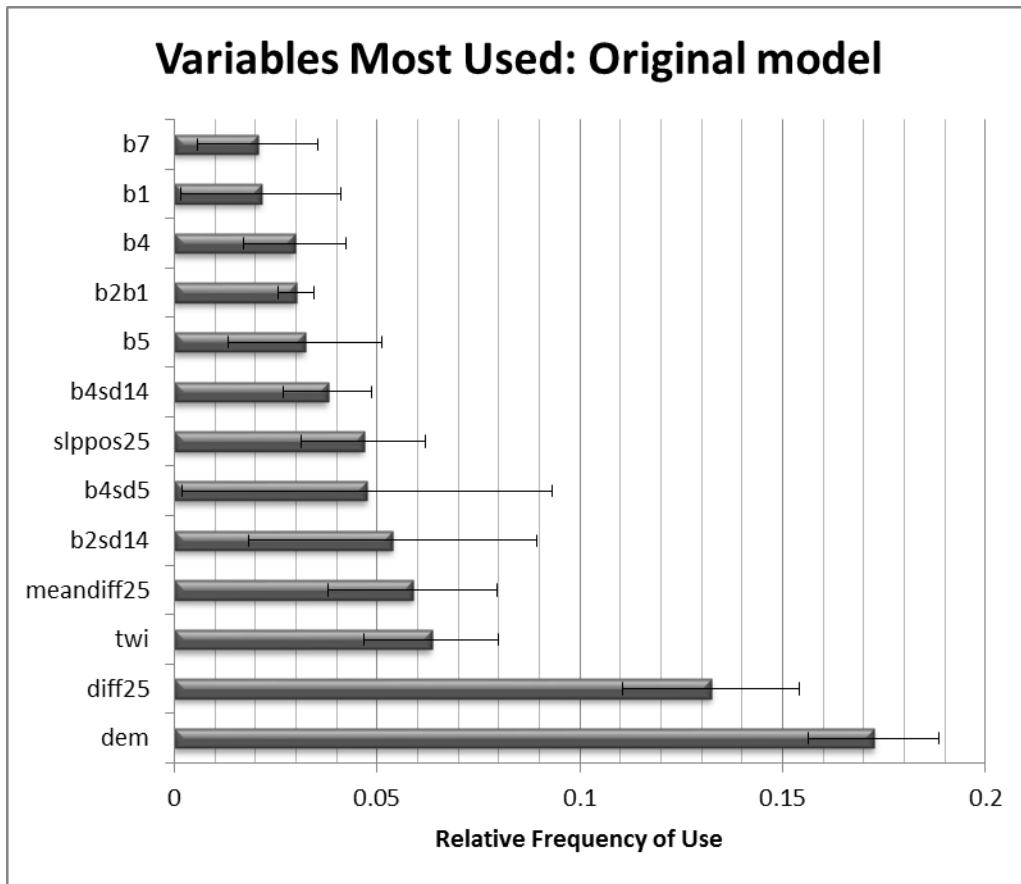


Figure 5b. Same as 7a, but for the variables used from Nauman (2009).

Table 4. Layers chosen by PCA data reduction to most efficiently represent soil-landscape variability for soil mapping (from Nauman, 2009)

Fluvial Strata Layers	Mountain Strata Layers
dem	dem
meandiff25	swness
b2sd3	b4sd14
b2sd14	b4b5
b2b1	b4b8
b1	b1

In comparing the principal components analysis (PCA) unsupervised data reduction variable selection from Nauman (2009) to the random forest use of same data in the Original

disaggregation model, eight of the 10 variables used were selected in both of the studies as important (Fig. 5, Table 4). This commonality drops to four of 10 when comparing with the Updated random forest model. It is interesting that the *diff25* variable was not selected in Nauman (2009) as it is a more important variable in both of the random forest models. Of the ASTER derived variables, neighborhood standard deviation of reflectance layers are the only variables that show up as important in both disaggregation models as well as the PCA unsupervised data reduction (Nauman, 2009).

4.6 Discussion and Conclusions

Our assessment is that both versions of random forest models worked well to disaggregate the CSM of ORPI, and that the near equivalent accuracy of SSURGO and the Updated model indicated that the models were able to reproduce much of the information captured by the survey. The accuracy at validation points in ORPI was higher than implementation of this same general methodology in the dissected Allegheny Plateau and Mountains of southern West Virginia (WV) (Nauman and Thompson, 2014) where classification validation accuracies ranged from 24% to 44% depending on spatial supports used in validation. However, similar to the ORPI results, the reported accuracy of the original CSM used in the WV study was also lower (27%-41%). It is encouraging that in both studies this method seems to produce a higher spatial resolution soil map at accuracies similar to the original soil surveys. The field validation accuracy in ORPI (69.8%) was also similar to results of similar studies in New Brunswick, Canada (64.9% to 67.6%, Yang et al, 2011), the African dambo (75.5%, Hansen et al., 2009), and in the Bavarian forests of Germany (70%, Haring et al., 2012).

The relationship between ensemble model prediction frequency (prediction probabilities) and validation accuracies allowed for a simple representation of classification uncertainty in both

ORPI and WV (Nauman and Thompson, 2014) studies, indicating some degree of consistency across different physiographic regions using these approaches. This relationship might also prove useful in future studies for using prediction probabilities in tree ensembles for creating fuzzy membership classifications. This scenario would involve using the proportion of tree predictions as membership functions where multiple soils are predicted for the same pixel in different trees within the ensemble model. This should be investigated in future disaggregation attempts as it would allow for continuous fuzzy thematic representations of soil classes (e.g. Burrough, 1989; Burrough et al., 1997; De Gruijter et al., 1997; Hodza, 2010; Lagacherie et al., 1997; McBratney and Odeh, 1997; Qi et al., 2006; Zhu, 1997; Zhu et al., 1996, 2001, 2010) that could be translated into soil property maps (e.g. Malone et al., 2011; Nauman et al., 2012; Qi et al., 2006; Zhu et al., 1997, 2001, 2010).

4.6.1 *Thematic Issues in Disaggregation*

Disaggregated maps create a singular and consistent theme of one soil class (or non-soil component, e.g. rock outcrops) per pixel. This is different than SSURGO maps units which can have multiple soil classes, slope gradient modifiers, and soil taxonomic variants. Soil series variants can be dealt with in these disaggregation approaches if they are included as classes at the training stage, but care should be taken because they might be so closely related to the non-variant soil series that the environmental covariates used in modeling might not be able to discern the two. Slope gradient modifiers have also been used to help with management interpretations in SSURGO (Soil Survey Division Staff, 1993). However, because slope gradient maps can be made at such high resolutions with modern digital elevation models, perhaps it is a better option to overlay a true soils themed map with a slope map for such purposes in modern contexts.

In ORPI, variants of soil series were mapped for the Ajo, Antho, Harqua, and Gilman soil series. Variants of Ajo and Harqua were split out based on the presence of cobble sized rock fragments at the surface and slightly different surface texture (both only in the top 5 cm), and Antho variants were distinguished based on variations of rock content at depth. A saline variant of the Gilman series was also mapped. Among these variants, we chose to only split out the Gilman saline variant for disaggregation because it was the only series where we had validation points in both the original series and the variant. Saline variants of the Gilman series also have a very unique ecology with Atriplex (saltbrush) dominated vegetation communities and a particular susceptibility to erosion (USDA-NRCS, 1972). We do feel that all of these variants could likely be identified in disaggregation models based on how well our models were able to match the original survey concepts, but due to our lack of ability to validate this and concerns about how different these soils truly are from a soil genesis perspective, we chose not to separate variants of the Ajo, Antho, and Harqua soil series. Thematic choices at the training stage of these models are difficult because soil series variants are generally only locally defined; so if results from ORPI were compared to other CSM, concepts will be more consistent at the soil series taxonomic level. Even among soils series there is often inconsistent degrees of detail because formal classification within U.S. *Soil Taxonomy* ends at the family level (Soil Survey Staff, 2010, 1999).

4.6.2 *Uncertainty Assessment*

Map accuracy and spatial representation of uncertainty are not offered with the SSURGO product. We aimed to produce these with a simple approach using pragmatic means with a limited validation set. Our results demonstrated that areas where the ensemble trees predicted classes the most consistently also tended to have higher validation agreements. Using this

relationship, we were able to represent this spatially to display where accuracies were higher or lower. In analyzing the uncertainty data we observed a few basic patterns in how estimated uncertainty was distributed. Specifically, small zones of higher uncertainty (lower field data probability) ran along SSURGO map unit boundaries in some areas, possibly representing transitions between soils along those boundaries. We also observed relatively higher uncertainties that seem to follow geomorphic patterns that sometimes, but not always, follow certain soil types as mapped in SSURGO and the disaggregated map (Updated model). These geomorphic differences might indicate that there are certain groups of soils or landforms that either have variability not represented in the maps, or that the differences between soils in those areas may be so subtle that they were not easily distinguished in classifications. Further, we observed distinctly higher uncertainty values (lower probabilities) among individual delineations of certain map units (e.g. Growler-Antho complex), possibly suggesting that the more uncertain delineations were incorrectly correlated in mapping. In these cases, a field update of the more uncertain delineations would be prudent. As the Growler-Antho example shows, the uncertainty map, in addition to being a gauge of accuracy, can help to identify potential inconsistencies in soil maps and aid in the development of priorities for additional field work.

In these results there was also a small, but systematic difference between smaller classes with higher uncertainty and large classes with lower uncertainty. This result might be related to the effects of class proportionality on tree model results as has been reported by previous studies (Moran and Bui, 2002; Nauman and Thompson, 2014). Moran and Bui (2002) showed that sampling classes proportionally for tree building improved results. Nauman and Thompson (2014) showed that sampling proportionally can degrade accuracy of smaller classes if there is a wide disparity between the size of small and large classes. They showed that this can be

somewhat mitigated by transforming the proportionality of sample sizes to have less disparity between the larger and smaller class sample sizes during tree building. This would appear to be a limitation of tree-based methods because the results indicate a need to experiment with class sample proportionalities to find an optimum for sampling training classes.

4.6.3 *Covariate Comparisons*

Overall results indicate that both NED elevation data and ASTER provide predictive power for soil survey disaggregation modeling. Variable importance values showed that the layers derived from the elevation data were likely more useful in soil classifications in this environment. Based on the dominance of the top four variables in both random forest models by the *dem* and other topographic variables, it would seem ASTER might not perform well without NED data to supplement predictions. The presence of the *dem* variable as dominant in both the fluvial and mountain areas in the PCA analysis (Table 4) (Nauman, 2009) also seems to support this. However, the PCA variable selection also selected more ASTER layers overall, which does support ASTER as a viable predictor. It would be useful in the future to generate a model just using ASTER or similar imagery and comparing that to models using only DEM derived covariates to better test the predictive power of both data sources based on classification success as opposed to variable importance values within models as presented here.

The dominance of elevation and relative height metrics seem to relate to the relief-driven topographic sequence of landforms in ORPI. These landforms include relatively young mountains with rugged outcrops at summits and limited soil development on side-slopes. Lower in the topographic sequence sets of alluvial fans are arranged in step-wise patterns moving away from the mountains, with basin floor deposits at the valley bottom. Soils tend to follow these step patterns because each riser between alluvial fans represents erosion cutting into

an older aged deposit with the oldest soils (i.e. Ajo - Argic Petrocalcids) being the closest to the mountains and representing past basin base levels (USDA-NRCS, 1972; Parker, 1995). These fans seem to have relatively distinct drainage patterns with some having mainly deep gullies whereas others include reticulating washes with depositional areas, which were likely detected by the relative height metrics (e.g. *rel_ht_20* in Fig. 5, Table 2) based on the height of fan treads above the drainage cuts.

The only ASTER layers used on average in more the 2% of tree nodes in both models were neighborhood standard deviation of reflectance layers of ASTER band 4 (lower wavelength short-wave infrared) and band 2 (red visible). Generally, these layers highlight areas that have more active geomorphic or hydrologic dynamics. They distinguish areas with higher densities of washes; especially where drainages reticulate or have greater vegetation cover. These layers also highlight areas where mountain and alluvial landforms adjoin. The lack of influential band ratios and reflectance bands in the Updated model was unexpected. We were expecting more of the ASTER covariates to possibly distinguish mineralogy or albedo differences based on the diverse lithology sources in ORPI. It was also perplexing that both band ratio and reflectance layers were highlighted in the Original disaggregation model and the PCA data reduction done by Nauman (2009), but not in the Updated model where more topography variables were introduced. However, examination of the original SSURGO map units reveals that soil series in ORPI were often mapped across areas sourced from multiple types of lithology, possibly indicating that mineralogy was not influential in distinguishing map units. This is supported by the SSURGO mineralogy classes, which is mixed for all soils in ORPI except for Perryville, which was attributed as carbonatic (USDA-NRCS, 1972).

We found that our model covariate importance values shared considerable similarities to the PCA-driven unsupervised soil-landscape classification in ORPI done previously (Nauman, 2009). Our Original disaggregation model used the same variables in the random forest classification as used by Nauman (2009) to see if similar variables would be used more frequently in the random forest. The previous study used a PCA-based approach to try to identify the most useful variables for soil mapping without any *a priori* knowledge of an area from a large suite of possible DEM and ASTER variables. The similar selection of variables by our Original random forest disaggregation model and the PCA-based method used by Nauman (2009) seems to confirm that the PCA-based approach can help select covariates from large datasets effectively for initial soil mapping at a site.

4.6.4 *Future Efforts*

The success with these general methods for disaggregating CSM at ORPI and in WV (Nauman and Thompson, 2014) seem to demonstrate that a consistent general approach can be taken to updating CSM around the United States. The key to this method is finding suitable initial variables in raster format to match with soil-landscape descriptions published in soil surveys to properly train a model. We would point out that this does not need to be limited to the terrain metrics used in ORPI and WV. There are vegetation and geologic attributes in SSURGO that could also be matched to imagery or other data sources for training. The main differences between the WV and ORPI studies included: (i) using different initial rule matching variables, (ii) inclusion of hillslope position descriptions (e.g. footslope, backslope, shoulder, etc.) in addition to the geomorphic table in the WV study, (iii) implementation of a full random forest algorithm in ORPI rather than just a classification tree ensemble, and (iv) inclusion of a larger set

of covariate rasters in ORPI. The use of the random forest model in ORPI was more appropriate given the larger number of covariates being used (Breiman, 2001).

Based on the higher accuracies at ORPI, updating the work in WV to incorporate more variables and a random forest framework might help results there. However, both disaggregation studies showed similar accuracies to the original CSM from which they were derived; which might indicate that the slight differences in methods were not as important as the original CSM accuracy in the reported differences in disaggregation accuracy. There are many factors that might influence these original CSM accuracies, but the scale of soil variation and the actual mapping scale are likely responsible for this in large part. The dominant soils in ORPI follow more contiguous patterns of alluvial sediments that might have less intrinsic variability than the forested and highly dissected plateaus and mountains in the WV study. We think the general workflow presented here and in Nauman and Thompson (2014) offers an opportunity to both improve and harmonize large CSM databases into more useful modern data products.

4.7 Acknowledgements

We wish to acknowledge Sue Rutman and the rest of the ORPI staff for their insight and help, as well as Chris Bertrand and Corrie Hannah for dedicated assistance in the field. Portions of this research were supported by USDA-NRCS Cooperative Agreements #68-9457-8-466 and #68-7482-12-535 with Dr. Craig Rasmussen, as well as Arizona Agricultural Experiment Station ARZT-1367190-H21-155. Funding supporting Travis Nauman at the University of Arizona came through USDA-NRCS Cooperative Agreement #68-3A75-4-100. Additional funding was provided by USDA-NRCS Cooperative Agreements #68-7482-11-527 and #68-7482-9-511 with Dr. James Thompson. Scientific contribution no. 3193 from the West Virginia Agricultural and Forestry Experiment Station, Morgantown, WV.

4.8 References

- Abrams, M., 2000. The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER): data products for the high spatial resolution imager on NASA's Terra platform. *International Journal of Remote Sensing*, 21(5): 847-859.
- Abrams, M., Hook, S. and Ramachandran, B., 2001. ASTER User Handbook Version 2. NASA, Pasadena, CA.
- Behrens, T. et al., 2005. Digital soil mapping using artificial neural networks. *Journal of Plant Nutrition and Soil Science-Zeitschrift Fur Pflanzenernahrung Und Bodenkunde*, 168(1): 21-33.
- Behrens, T., Schmidt, K., Zhu, A.X. and Scholten, T., 2010a. The ConMap approach for terrain-based digital soil mapping. *European Journal of Soil Science*, 61(1): 133-143.
- Behrens, T., Zhu, A.X., Schmidt, K. and Scholten, T., 2010b. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, 155(3/4): 175-185.
- Bezy, J.V., Gutmann, J.T. and Haxel, G.B., 2000. A Guide to the Geology of Organ Pipe Cactus National Monument and the Pinacate Biosphere Reserve. Arizona Geologic Survey, Tucson, AZ, 63 pp.
- Bliss, N., Waltman, S. and Petersen, G.W., 1995. Preparing a soil carbon inventory for the United States using geographic information systems. In: R. Lal (Editor), Soils and global change. *Adv Soil Science*. . CRC Press, Boca Raton, FL.
- Brady, N.C., and R.R. Weil. 2008. The Nature and Properties of Soils, 14th edition Prentice Hall/Pearson Education.
- Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1): 5-32.

- Buchanan, J., 2007. ASTER Radiance Scalars. In: T. Nauman (Editor). University of Arizona Remote Sensing Group, Tucson, AZ.
- Bui, E., Henderson, B. and Viergever, K., 2009. Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. *Global Biogeochemical Cycles*, 23.
- Bui, E.N., 2004. Soil survey as a knowledge system. *Geoderma*, 120(1/2): 17.
- Bui, E.N., Henderson, B.L. and Viergever, K., 2006. Knowledge discovery from models of soil properties developed through data mining. *Ecological Modelling*, 191(3-4): 431-446.
- Bui, E.N., Loughhead, A. and Corner, R., 1999. Extracting soil-landscape rules from previous soil surveys. *Australian Journal of Soil Research*, 37(3): 495-508.
- Bui, E.N. and Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma*, 103(1-2): 79-94.
- Burrough, P.A., 1989. Fuzzy mathematical methods for soil survey and land evaluation. *European Journal of Soil Science*, 40(3): 477-492.
- Burrough, P.A., van Gaans, P.F.M. and Hootsmans, R., 1997. Continuous classification in soil survey: spatial correlation, confusion and boundaries. *Geoderma*, 77(2-4): 115-135.
- Chavez, P.S., 1996. Image-Based Atmospheric Corrections Revisited and Improved. *Photogrammetric Engineering and Remote Sensing*, 62(9): 1025-1036.
- Cook, S.E., Corner, R.J., Grealish, G., Gessler, P.E. and Chartres, C.J., 1996a. A rule-based system to map soil properties. *Soil Science Society of America Journal*, 60(6): 1893-1900.
- Cook, S.E., Corner, R.J., Groves, P.R. and Grealish, G.J., 1996b. Use of airborne gamma radiometric data for soil mapping. *Australian Journal of Soil Research*, 34(1): 183-194.

- de Bruin, S., Wielemaker, W.G. and Molenaar, M., 1999. Formalisation of soil-landscape knowledge through interactive hierarchical disaggregation. *Geoderma*, 91(1-2): 151-172.
- De Gruijter, J.J., Walvoort, D.J.J. and van Gams, P.F.M., 1997. Continuous soil maps -- a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. *Geoderma*, 77(2-4): 169-195.
- Eddy, A. et al., 1991. Geological Reconnaissance at Organ Pipe Cactus National Monument, Arizona. National Park Service, Tucson, AZ.
- Gatzke, S.E. et al., 2011. Aggregation Strategies for SSURGO Data: Effects on SWAT Soil Inputs and Hydrologic Outputs. *Soil Science Society of America Journal*, 75(5): 1908-1921.
- Gesch, D.B., 2007. The National Elevation Dataset. In: D. Maune (Editor), Digital Elevation Model Technologies and Applications: The DEM Users Manual. American Society for Photogrammetry and Remote Sensing, pp. 99-118.
- Gesch, D.B. et al., 2002. The National Elevation Dataset. *Photogrammetric Engineering and Remote Sensing*, 68(1): 5-11.
- Giorgini, J.D. et al., 1996. JPL's On-Line Solar System Data Service. Bulletin of the American Astronomical Society.
- Goovaerts, P., 2011. A coherent geostatistical approach for combining choropleth map and field data in the spatial interpolation of soil properties. *European Journal of Soil Science*, 62(3): 371-380.
- Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152(3-4): 195-207.

- Grunwald, S., Thompson, J.A. and Boettinger, J.L., 2011. Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues. *Soil Science Society of America Journal*, 75(4): 1201-1213.
- Hansen, M.K., Brown, D.J., Dennison, P.E., Graves, S.A. and Brickleyer, R.S., 2009. Inductively mapping expert-derived soil-landscape units within dambo wetland catenae using multispectral and topographic data. *Geoderma*, 150(1–2): 72-84.
- Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T. and Schröder, B., 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma*, 185–186(0): 37-47.
- Henderson, B.L., Bui, E.N., Moran, C.J. and Simon, D.A.P., 2005. Australia-wide predictions of soil properties using decision trees. *Geoderma*, 124(3-4): 383-398.
- Hendricks, D.M., 1985. Airzona Soils. University of Arizona, Tucson, AZ, 243 pp.
- Hodza, P., 2010. Fuzzy logic and differences between interpretive soil maps. *Geoderma*, 156(3–4): 189-199.
- IPCC, 2007. Climate change 2007: Synthesis report. Contribution of Working Groups I, II, and III to the Fourth Assessment Report., Geneva, Switzerland.
- Jenny, H., 1941. Factors of Soil Formation. McGraw-Hill, New York, New York.
- Kempen, B., Brus, D.J., Heuvelink, G.B.M. and Stoorvogel, J.J., 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, 151(3/4): 311-326.
- Kerry, R., Goovaerts, P., Rawlins, B.G. and Marchant, B.P., 2012. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma*, 170: 347-358.

- Lagacherie, P., Cazemier, D.R., van Gaans, P.F. and Burrough, P.A., 1997. Fuzzy k-means clustering of fields in an elementary catchment and extrapolation to a larger area. *Geoderma*, 77(2-4): 197-216.
- Lemercier, B., Lacoste, M., Loum, M. and Walter, C., 2012. Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach. *Geoderma*, 171–172(0): 75-84.
- Lineback Gritzner, M., Marcus, W.A., Aspinall, R. and Custer, S.G., 2001. Assessing landslide potential using GIS, soil wetness modeling and topographic attributes, Payette River, Idaho. *Geomorphology*, 37(1-2): 149-165.
- Loerch, C., 2012. Soil Data Join Recorrelation Initiative, West Regional National Cooperative Soil Survey Meeting, Davis, CA.
- MacMillan, R.A., Pettapiece, W.W., Nolan, S.C. and Goddard, T.W., 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy Sets and Systems*, 113(1): 81-109.
- Malone, B.P., McBratney, A.B. and Minasny, B., 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma*, 160(3/4): 614-626.
- McAuliffe, J.R., 1994. Landscape Evolution, Soil Formation, and Ecological Patterns in Sonoran Desert Bajada. *Ecological Monographs*, 64(2): 111-148.
- McBratney, A.B., 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. *Nutrient Cycling in Agroecosystems*, 50(1-3): 51-62.
- McBratney, A.B. and Odeh, I.O.A., 1997. Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma*, 77(2-4): 85-113.

- McBratney, A.B., Santos, M.L.M. and Minasny, B., 2003. On digital soil mapping. *Geoderma*, 117(1-2): 3-52.
- McKenzie, N.J. and Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89(1-2): 67-94.
- Minasny, B. and McBratney, A.B., 2010. Methodologies for Global Soil Mapping. In: J. Boettinger, D. Howell, A. Moore, A. Hartemink and S. Kienast-Brown (Editors), *Digital Soil Mapping. Progress in Soil Science*. Springer Netherlands, pp. 429-436.
- Moran, C.J. and Bui, E.N., 2002. Spatial data mining for enhanced soil map modelling. *International Journal of Geographical Information Science*, 16(6): 533-549.
- National Cooperative Soil Survey., 2012. National Cooperative Soil Characterization Database. Available online at <http://ncsslabsdatamart.sc.egov.usda.gov>.
- Nauman, T.W. and Thompson, J.A., 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. *Geoderma*, 213(0): 385-399.
- Nauman, T.W., Thompson, J.A., Odgers, N. and Libohova, Z., 2012. Fuzzy Disaggregation of Conventional Soil Maps using Database Knowledge Extraction to Produce Soil Property Maps. In: B. Minasny, B. Malone and A. McBratney (Editors), *Digital Soil Assessments and Beyond: 5th Global Workshop on Digital Soil Mapping*, Sydney, Australia.
- Nauman, T.W., 2009. Digital Soil-Landscape Classification for Soil Survey using ASTER Satellite and Digital Elevation Data in Organ Pipe Cactus National Monument, Arizona, M.S. Thesis. The University of Arizona, Tucson, Arizona.
- NOAA, 2004. Climatography of the United States No. 20 1971-2000 Organ Pipe Cactus NM, AZ Station, U.S. Department of Commerce, Asheville, NC.

- Parker, K.C., 1991. Topography, Substrate, and Vegetation Patterns in the Northern Sonoran Desert. *Journal of Biogeography*, 18(2): 151-163.
- Parker, K.C., 1995. Effects of Complex Geomorphic History on Soil and Vegetation Patterns on and Alluvial Fans. *Journal of Arid Environments*, 30(1): 19-39.
- Qi, F., Zhu, A.X., Harrower, M. and Burt, J.E., 2006. Fuzzy soil mapping based on prototype category theory. *Geoderma*, 136(3-4): 774-787.
- Saunders, A.M. and Boettinger, J.L., 2007. Incorporating Classification Trees into a Pedogenic Understanding Raster Classification Methodology, Green River Basin, Wyoming, USA. In: P. Lagacherie, A.B. McBratney and M. Voltz (Editors), *Digital Soil Mapping An Introductory Perspective*. Elsevier, Amsterdam, The Netherlands.
- Schmidt, K., Behrens, T. and Scholten, T., 2008. Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. *Geoderma*, 146(1-2): 138-146.
- Scikit-learn.org. 2013. User Guide: Decision Trees. Online: <http://scikit-learn.org/stable/modules/tree.html>
- Scull, P., Franklin, J. and Chadwick, O.A., 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modelling*, 181(1): 1-15.
- Scull, P., Franklin, J., Chadwick, O.A. and McArthur, D., 2003. Predictive soil mapping: a review. *Progress in Physical Geography*, 27(2): 171-197.
- Simpson, D.T., 1991. Soils and geomorphology of the Quaternary alluvial sequence on the western Piedmont of the Ajo Mountains, Organ Pipe Cactus National Monument, Pima County, Arizona. M.S. Thesis, University of New Mexico, Albuquerque, New Mexico, 160 pp.

- Smith, S., Bulmer, C., Flager, E., Frank, G. and Filatow, D., 2010. Digital Soil Mapping and multiple scales in British Columbia, Canada, 4th Global Workshop on Digital Soil Mapping, Rome, Italy.
- Smith, S., Daneshfar, B., Frank, G., Flager, E. and Bulmer, C., 2012. Use of weights of evidence statistics to define inference rules to disaggregate soil survey maps, 5th Global Workshop on Digital Soil Mapping, Sydney, Australia.
- Soil Survey Division Staff., 1993. Soil Survey Manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.
- Soil Survey Staff. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. 2nd edition. Natural Resources Conservation Service. U.S. Department of Agriculture Handbook 436.
- Soil Survey Staff. 2010. Keys to Soil Taxonomy, 11th ed. USDA-Natural Resources Conservation Service, Washington, DC.
- USDA-NRCS, 2013. National soil survey handbook, title 430-VI. Available online at <http://soils.usda.gov/technical/handbook/>. Accessed [04/15/2013].
- Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*, 33(2): 309-319.
- Thomas-Van Gundy, M., Strager, M. and Rentch, J., 2012. Site characteristics of red spruce witness tree locations in the uplands of West Virginia, USA. *The Journal of the Torrey Botanical Society*, 139(4): 391-405.
- Thomas-Van Gundy, M. and Strager, M.P., 2012. European settlement-era vegetation of the Monongahela National Forest, West Virginia. In: F.S. U.S. Department of Agriculture (Editor). Northern Research Station, Newtown Square, PA.

- Thome, K., Biggar, S. and Slater, P., 2001. Effects of assumed solar spectral irradiance on intercomparisons of earth-observing sensors. In: H. Fujisada, J.B. Lurie and K. Weber (Editors), *Sensors, Systems and Next-Generation Satellites V*. SPIE.
- Thompson, J.A. and Kolka, R.K., 2005. Soil carbon storage estimation in a forested watershed using quantitative soil-landscape modeling. *Soil Science Society of America Journal*, 69(4): 1086-1093.
- Thompson, J.A., Nauman, T.W., Odgers, N., Libohova, Z. and Hempel, J., 2012. Harmonization of Legacy Soil Maps in North America: Status, Trends, and Implications for Digital Soil Mapping Efforts. In: A. McBratney, B. Minasny and B. Malone (Editors), *The 5th Global Workshop on Digital Soil Mapping. Digital Soil Assessments and Beyond*, Sydney, Australia.
- Thompson, J.A., Prescott, T., Moore, A., Bell, J., Kautz, D., Hempel, J., Waltman, S., & Perry, C.H. 2010. Regional Approach to Soil Property Mapping using Legacy Data and Spatial Disaggregation Techniques. In, *19th World Congress of Soil Science, Soil Solutions for a Changing World*. Brisbane, Australia
- USDA-NRCS, 1972. Soil Survey Organ Pipe Cactus National Monument Pima County Arizona a special report. United States Department of Agriculture Soil Conservation Service, United States Department of the Interior, National Park Service Arizona Agricultural Experiment Station.
- Wei, S., McBratney, A., Hempel, J., Minasny, B., Malone, B., D'Avello, T., Burras, L., & Thompson, J.A. 2010. Digital Harmonisation of Adjacent Soil Survey areas - 4 Iowa Counties. In, *19th World Congress of Soil Science, Soils Solutions for a Changing World*. Brisbane, Australia

- Wielemaker, W.G., de Bruin, S., Epema, G.F. and Veldkamp, A., 2001. Significance and application of the multi-hierarchical landsystem in soil mapping. *Catena*, 43(1): 15-34.
- Yang, L. et al., 2011. Updating Conventional Soil Maps through Digital Soil Mapping. *Soil Science Society of America Journal*, 75(3): 1044-1053.
- Yang, X., Chapman, G.A., Gray, J.M. and Young, M.A., 2007. Delineating soil landscape facets from digital elevation models using compound topographic index in a geographic information system. *Australian Journal of Soil Research*, 45(8): 569-576.
- Zhu, A.X., 1997. A similarity model for representing soil spatial information. *Geoderma*, 77(2-4): 217-242.
- Zhu, A.X., Band, L., Vertessy, R. and Dutton, B., 1997. Derivation of soil properties using a soil land inference model (SoLIM). *Soil Science Society of America Journal*, 61(2): 523-533.
- Zhu, A.X., Band, L.E., Dutton, B. and Nimlos, T.J., 1996. Automated soil inference under fuzzy logic. *Ecological Modelling*, 90(2): 123-145.
- Zhu, A.X., Hudson, B., Burt, J., Lubich, K. and Simonson, D., 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal*, 65(5): 1463-1472.
- Zhu, A.X. et al., 2010. Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma*, 155(3-4): 164-174.

5 GHOSTS OF THE FOREST: MAPPING PEDOMEMORY TO GUIDE FOREST RESTORATION

5.1 Citation

Nauman, T.W., J.A. Thompson, S.J. Teets, T.A. Dilliplane, J.W. Bell, S.J. Connolly, H.J. Liebermann, and K.M. Yoast. 2015. Ghosts of the forest: Mapping pedomemory to guide forest restoration. *Geoderma* 247:51-64. Reprinted from *Geoderma* under license #3603861273405 with permission from Elsevier.

5.2 Highlights

- Pedomemory mapping demonstrated as a restoration tool.
- Spodic soil morphology linked to historic red spruce (*Picea rubens*) forests.
- Mapped spodic morphology occurrence with 70-78% accuracy.
- Maps provide direct guidance for red spruce restoration efforts.
- Conifer composition showed strong relationship to O-horizon thickness.

5.3 Abstract

Soil morphology can provide insight into how ecosystems change following periods of extensive disturbance. Soils properties can often be linked to historic environmental influences (e.g. vegetation or climate) to provide a record of pedomemory. Identification and mapping of soil pedomemory properties shows promise in providing context for ecological restoration. We have developed a novel use of digital soil mapping of spodic morphology to estimate historical forest composition in the high-elevation forests of the Central Appalachians. This region was extensively disturbed by clear-cut harvests and related fires during the 1880's-1930's. Hardwood forest species recovered much better than local conifers and generally encroached into historic populations of red spruce (*Picea rubens*) and eastern hemlock (*Tsuga canadensis*). Spodic soil morphology, which is often associated with subalpine and boreal conifer forests, was mapped

using a random forest probability model and showed correspondence to red spruce – eastern hemlock distribution, as derived from local historic property deed witness tree records from 1752-1899. These data and resulting models indicate a greater spatial extent of spodic soil properties than documented in previous soil maps, which is more consistent with general theories of much more extensive historic spruce populations. The resulting maps and models provide guidance for field scale restoration planning for historically disturbed spruce-hemlock forests. Our results suggest that historic Euro-American disturbance probably induced conifer-to-hardwood state transitions at mid to high elevation coniferous ecological sites within the Appalachians. Where transitions have occurred, there appears to have been dramatic losses in forest floor thickness (O-horizons) and associated soil organic carbon stocks into atmospheric carbon pools. Spatial modeling of similar pedomemory properties and other soil-ecology linkages is likely to be a powerful tool to guide restoration in other regions as well.

Key words: podzolization, pedomemory, digital soil mapping, soil organic carbon, forest restoration, red spruce

5.4 Introduction

5.4.1 Soil pathways and pedomemory

Soil properties can help reveal the history of interactions between abiotic and biotic drivers at the Earth's surface. In soil science, this has been conceptualized as a state factor model where the state or properties of a soil are a result of interactions between climate, organisms, relief, and parent material over time (clorpt) (Dokuchaev, 1899; Jenny, 1941). The state factor model evolved to an ecosystem level model where soils and organisms have some parallel drivers, but also interact strongly (Amundson and Jenny, 1997; Jenny, 1961, 1980). Equation 1 reformats Jenny's (1941) 'clorpt' model into an ecological equation where different groups of the original soil forming factors interact over time to result in a set of ecosystem properties (including soil) at a given point in time.

$$l, s, v, a = f(L_0, P_x, t) \text{ [Eq. 1: Ecological factorial; Jenny, 1961]}$$

The dependent factors in this case include ecosystem properties (l), soil properties (s), vegetation (v), and animals (a). The related state factors in an ecosystem based approach include the initial state (L_0) and external potentials (P_x), and time (t). Initial state L_0 includes the parent material (bedrock or substrate), initial relief, and water table. Climate and organism changes are grouped as the P_x variable, which represent the primary energy sources (sun), receptors (plants), and catalysts (e.g. water) that drive processes (Jenny, 1961). Amundson and Jenny (1991, 1997) have introduced these conceptual models into ecological sciences, with humans included in the factorial equation. In an ecosystem, soils bear the imprint and help record the history of organisms—including humans—as well as the climate. For conceptual and measurement

purposes, we define an ecosystem as the living organisms and physical environment of a defined unit space or a plot (e.g. 20x20 meters) that we can sample in the field.

Climatic and biological factors drive processes in soils that involve additions, removals, translocations, and transformations (Simonson, 1959) of materials in the soil column that have associated energies (Nikiforoff, 1959; Runge, 1973). When environmental drivers remain relatively constant over a period of time they can direct a soil down a developmental pathway toward expressions of specific horizonation (Johnson and Watson-Stegner, 1987). Changes in climate and/or organisms can alter the balance of processes and thus the pathway of a soil. At any one time, many processes are occurring in a soil, which can create complicated superimposed distributions of soil properties within a soil profile (Burrough, 1983).

The properties observed in soils reflect a record of information, often called soil memory or pedomemory, where the specific patterns of reorganization and transformation of the original soil parent material into new physical and chemical distributions in the soil profile can often be attributed to how historic climate and vegetation promote soil processes that result in a specific morphology (Hole, 1975; Lin, 2011; Targulian and Goryachkin, 2004). Related studies have linked mottling, iron chemistry, and other morphology to historic soil-water-landscape models (Coventry et al., 1983; Coventry and Williams, 1984; Fritsch and Fitzpatrick; 1994 ; Schwertmann, 1988). Others have found that vegetation communities interact with the soil over time to create soil property signatures recorded in the pedomemory useful in determining a site history (Hole, 1975; Phillips and Marion, 2004; Willis et al., 1997). Thus, a soil property like spodic materials can potentially provide a time-space record that can help decipher historic ecosystem vegetative reference conditions, which are an accepted basis for ecological restoration to a certain target community type and condition (Higgs et al., 2014; SER, 2004;

<http://www.ser.org/resources/resources-detail-view/ser-international-primer-on-ecological-restoration>). Linking soil types with historic reference communities has become the basis for land management frameworks such as ecological site descriptions (ESD) (Caudle et al., 2013; NRCS, 2014). We aim to show how mapping key pedomemory properties linked to vegetative communities can inform restoration at a field ecosystem scale. We demonstrate this using an example along the ecologically important transition between northern hardwood and spruce-hemlock forest types in the Central Appalachian mountains of the eastern US (Byers et al., 2010).

For distinguishing the historic transition between northern hardwood and spruce-hemlock, we chose the podzolization pathway (Lundström et al., 2000a, 2000b; Sauer et al., 2007; Schaetzl and Harris, 2011) as our pedomemory indicator because of its association with similar moist conifer forest and heathland species composition globally (Hole, 1975; Miles, 1985; Willis et al., 1997; Lundström et al., 2000a; Sauer et al., 2007). In a typical cool, moist conifer site where Spodosols form as a result of podzolization, the soil morphology generally is a sequence of Oi-Oe-Oa surface horizons forming a mor forest floor, then a leached E horizon, and a sequence of Bh-Bhs-Bs-BC subsurface horizons (Fig. 1) (Soil Survey Staff, 1999; Soil Survey Staff, 2010). The podzolization pathway includes multiple soil processes that promote aluminum, iron, and organic matter mobilization and translocation to deeper soil depths in acidic, permeable parent materials. Thick surface O horizons also frequently form at the soil surface in these typically moist conifer systems (Hix and Barnes, 1984; Lietzke and McGuire, 1987; Lundström et al., 2000a). Leaching is usually associated with soluble organic acids from the forest floor and actively mining ectomycorrhizal communities causing mineral weathering and the ultimate transport of aluminum, iron, and organic matter from near surface soil horizons (O, A, E) into

subsurface (B) soil horizons (Blum et al., 2002; Giesler et al., 2000; Hoffland et al., 2004; Jongmans et al., 1997; Lundström et al., 2000b; Schaetzl and Harris, 2011; Schöll et al., 2008; Van Breemen et al., 2000).

Much of the organic carbon distribution in Spodosols can be lost in 30-100 years just by converting cool, moist acidic conifer forest stands to differing species compositions (prairie or hardwood) that favor more decomposition (Barrett and Schaetzl, 1998; Hix and Barnes, 1984; Hole, 1975; Miles, 1985). The most pronounced losses in organic carbon occur in the forest floor O horizons, which generally get thinner in conversions. Conversely, studies have also demonstrated that conversion from mesic hardwood forests (mostly *Quercus spp.*, *Betula spp.*, and *Fagus spp.*) to Norway spruce (*Picea abies*) and/or scots pine (*Pinus sylvestris*) initiates O horizon buildup and podzolization within a century (Herbaults and Buyl, 1981; Miles, 1985; Ranger and Nys, 1994; Sohet et al., 1988). Common garden experiments studying replanted monoculture plots of various tree species have also documented tree species gradients of influence on soil organic matter accumulation and acidity. On the two extremes, *Acer spp.* and *Tilia spp.* promote increased base cation activity which favors heterotrophic organic matter decomposition, whereas *Pinus spp.* and *Larix decidua* enhances acidic Al and Fe activity which limit decomposition of soil organic matter (Hobbie et al., 2007). Garden experiments also showed higher tree litter calcium content appeared to increase pH, decomposition, and earthworm activity that resulted in less forest floor mass (Reich et al., 2005; Hobbie et al., 2006). Hobbie et al., (2006) also recorded that plots with spruce and fir species had lower mean annual soil temperatures and less litter decomposition. Although general differences in litter chemistry exist between angiosperms (basic) and gymnosperms (acidic), these studies showed that there is significant variation within these tree groups. Another recent common garden study in New York

documented a similar influence of worms under northern red oak (*Quercus rubra.*) and sugar maple (*Acer saccharum*), but not under Norway spruce, which had a thicker forest floor (Melvin and Goodale, 2013). Although Ca^{2+} content was similar under all three species, pH was lower under the spruce, suggesting that base cation activity might not be the only factor to examine. Other studies of tree species interactions with soil have recorded similar trends (Finzi et al., 1998; Van Breeman and Finzi, 1998). Overall, these studies tell a story where heterotrophic forest litter decomposition and O horizon accumulation are intricately linked to dominant tree species at a site.

Autotrophic mycorrhizal partnerships are another important consideration in understanding carbon and nutrient cycling in soils (Högberg and Read, 2006). Studies have demonstrated intensive ectomycorrhizal (ECM) colonization of E horizons that appear to be a significant nutrient acquisition adaptation strategy of conifers in acidic Al-dominated soil environments, thereby overcoming conditions that might otherwise be toxic (Blum et al., 2002; Giesler et al., 2000; Hoffland et al., 2004; Högberg and Read, 2006; Jongmans et al., 1997; Lundström et al., 2000b; Van Breemen et al., 2000). Giesler et al., (2000) were able to show that the expansion of mineral-boring ECM hyphae looking for other nutrients is a likely mechanism for Al, Fe and Si transport to, and subsequent flux out of, O horizons. The buildup of autotrophic root hypha in the forest floor and associated host carbon allocation seem to be much more dominant processes than the classic heterotrophic model of litter and fine root decomposition and respiration in acid conifer systems (Högberg and Read, 2006). The development of deep O horizons under acidic conifer must, by definition, mean that heterotrophic communities are either suppressed or very inefficient in cycling carbon in these systems, which is also consistent with the results of garden studies (Reich et al., 2005; Hobbie et al., 2006; Hobbie et al., 2007).

Red spruce is one of the most acidophilic conifers, producing nutrient-poor litter (especially low in Ca^{2+}) relative to other North American trees (compare from: Berg and McClaugherty, 2008; Côté and Fyles, 1994; Friedland et al., 1988; Rustad and Fernandez, 1998). This implies that red spruce should promote podzolization and O horizon accumulation (Herbaults and Buyl, 1981; Lundström et al., 2000a; Miles, 1985; Ranger and Nys, 1994; Sauer et al., 2007; Sohet et al., 1988). Conversely, we expect that where spruce was converted to base-promoting hardwoods, like red maple (*Acer rubrum*), black cherry (*Prunus serotina*), and American beech (*Fagus grandifolia*), organic material loss has probably occurred from O and B horizons (Hix and Barnes, 1984; Miles, 1985; Hole, 1975). O horizon loss was probably initially exacerbated by the large-scale fires documented in these parts of West Virginia (WV) after mass clearcutting between 1860 and 1920 (Clarkson, 1964; Hopkins, 1899; Pauley, 2008). Well-developed Spodosols often take 1000-6000 years to form in areas similar to red spruce ecosystems (Lundström et al., 2000a; Schaetzl and Harris, 2011). Loss of Spodosol morphology is not as well documented, but was reported to disappear from a watershed in Hungary in 1000 years after a change in climate triggered a sequence of fires that likely converted forest stands from conifer to hardwood (Willis et al. 1997). However, the Fe and Al sesquioxide accumulations (spodic soil materials in US soil taxonomy; Soil Survey Staff, 1999) in the subsurface soil should still be observable as these are more stable and persistent in soils within the 150-250 year timeframe in this study (Barrett and Schaetzl, 1998; Lundström et al., 2000b; Parfitt, 2009). Indeed, Al-protomogolite, the major diagnostic sesquioxide solid compound in Spodosols, is relatively stable in soils for many millennia when soils maintain a pH greater than four (Parfitt, 2009). We hypothesized that Fe and especially Al sesquioxide accumulation found

in Bhs and Bs (spodic) soil horizons should be good pedomemory evidence for pre-Euro-American spruce-hemlock influence.

Recent work related to ESD development in the Monongahela National Forest (MNF) in WV for the purpose of linking management strategies to pre-settlement vegetation and site potential has suggested that spodic soil properties are linked to past red spruce and eastern hemlock distributions (Nowacki and Wendt, 2010; Teets, 2013). In the most impacted sites where O horizons were probably lost and E horizons were likely transformed or lost due to hardwood conversion, erosion, and/or fires, we think remnant Bs horizons could be a good indicator of past spruce influence. Although we think historic podzolization of these areas was due in large part to the red spruce acidic foliar chemistry, shallow root distribution, and acid producing mycorrhizal activity (Blum et al., 2002; Glenn et al., 1991), there are also climatic parallels between red spruce and eastern hemlock physiological requirements and podzolization. Both require cold and moist environments and are favored by longer winter snowpacks and thus should follow analogous topographic patterns (Lietzke and McGuire, 1987; Schaetzel and Isard, 1996; Nowacki and Wendt, 2010; Nowacki et al., 2010; Stanley and Ciolkosz, 1981). Published modern soil surveys for counties of the MNF only delineate Spodosols on the highest sandstone ridges where red spruce has more successfully regenerated from past disturbance (Delp, 1998; Flegel, 1998; USDA-SCS and USDA-FS, 1982), but not down into siltstone and shale parent materials at slightly lower elevations that are still within the local range of red spruce based on current inventories and related models (Beane et al., 2013; Byers et al., 2010; Nowacki and Wendt, 2010) as well as historic witness tree species related species distribution models from historic county property boundary records (Thomas-Van Gundy et al., 2012). However, an older soil survey (Williams and Fridley, 1931) supports existence of a much larger area of podzol

soils, which we believe corresponds to the more extensive historical distribution of red spruce forest communities prior to the regional harvest and fire disturbance of the late 19th and early 20th centuries. The vast majority of the harvest and fires occurred between 1880 and 1930, but site specific dates are hard to find. It is thought that very few places were not harvested in this period, and that fires also affected the vast majority of the landscape, but historic records are somewhat general in descriptions (Hopkins, 1899; Clarkson, 1964; Pauley, 2008).

5.4.2 *Importance of Red Spruce Forests in the Central Appalachians*

Vast forests of red spruce (*Picea rubens*), either singly or in association with northern hardwoods, once covered the higher elevations of the central Appalachians (Hopkins, 1899). This assemblage is thought to have spanned the last 4-5 millenia (Watts, 1979), and strong associations developed between these forests and various animals, with sensitive species becoming somewhat reliant on red spruce habitat, such as the Virginia northern flying squirrel (*Glaucomys sabrinus fuscus*) and Cheat Mountain salamander (*Plethodon nettingi*) (Dillard et al., 2008a, 2008b; Menzel et al., 2004, Menzel et al., 2006a, 2006b; Pauley, 2008). Wind and ice storms were the principal disturbance agents in presettlement times as the prevailing cool, moist climate greatly retarded fire (Rentch et al., 2010). As such, the natural disturbance regime was probably driven by periodic light-to-moderate severity storms rather than by catastrophic blowdowns and old-growth conditions were abundant. The Euro-American disturbances of the late-1800s to early 1900s were in stark contrast to this naturally low-disturbance environment. As a valuable timber species, red spruce was quickly liquidated by industrial clear-cut logging once railroad technologies afforded access to mountainous areas (Clarkson, 1964, Lewis, 1998, Nowacki and Wendt, 2010). Thereafter, uncontrolled wildfires burned through the remaining slash, largely consuming red spruce regeneration in the process. The rapidity and voracity of

these disturbances completely devastated red spruce, causing significant contraction to its population and range.

Due to its ecological and economic importance, red spruce restoration has received much attention in the central Appalachian region (e.g., Central Appalachian Spruce Restoration Initiative; <http://www.restoreredspruce.org/>). Unfortunately, efforts to restore red spruce are thwarted by the fact that its former range is so poorly documented at the field scale—although recent attempts through modeling (Beane et al., 2013; Byers et al., 2010; Nowacki and Wendt, 2010) and witness-tree analyses (Thomas-Van Gundy et al., 2012) have provided greater clarity on its original distribution.

In West Virginia, historical accounts indicate that the current extent (~20,000 ha) of alpine red spruce forest communities is greatly reduced from estimates prior to railroad era disturbance (~200,000 ha) (Hopkins, 1899; Pauley, 2008; Pielke, 1981; Nowacki and Wendt, 2010). Local studies, along with regional analysis of red spruce distribution (Nowacki et al., 2010), show that the main restriction on red spruce is warmer temperatures (with elevation as a surrogate) and lower precipitation. However, recent work in compiling and analyzing witness-tree databases from the MNF indicate a lower minimum elevation historically (lowest recorded red spruce at 509 meters) than previous models, and more specificity to topographic controls in respect to slope steepness, slope position, slope aspect, and landforms (Thomas-Van Gundy et al., 2012). These subtleties in the pre-settlement distribution of red spruce might indicate historic affinity for topographically-driven cool and moist microclimates that included the highest ridgelines, cooler aspects not in rain shadows, and narrow valleys that foster cold air drainage and foggy inversions.

Human disturbance and pollution have drastically impacted red spruce populations, but climate change and warming temperatures may have also affected populations—and these phenomena are hard to distinguish (Hamburg and Cogbill, 1988). Theoretically, global warming will drive boreal conifer ecosystem species like red spruce higher in elevation and further north, putting large pools of soil organic carbon at risk for further atmospheric release (Lal, 2005; Tarnocai et al., 2009). It is also hard to account for climate-vegetation feedbacks as well, and restoring to more historic communities could mitigate these potential feedbacks. Studies have shown that convectively driven precipitation patterns and radiative dynamics are influenced by changing vegetation type and structure which is likely to mean warmer and drier soil conditions for former spruce sites (Pielke, 1981, 2001; Pielke et al., 2002). Other concerns about acid deposition on red spruce health have been studied (Johnson, 1983; Hornbeck and Smith, 1985; Adams and Eagar, 1992), but might be difficult to discern from the impact of historic disturbance and climate change (Hamburg and Cogbill, 1988). Indeed, red spruce is projected by different climate change scenarios to disappear from West Virginia by the end of the century (Butler et al., 2014; Byers et al., 2010; Iverson et al., 2008; Prasad et al., 2007). However, there are signs that red spruce is recovering from historic disturbance and could be further restored despite climate change (Nowacki et al., 2010; Rentch et al., 2007; Rentch et al., 2010; Rollins et al., 2010). At this time, its future remains uncertain, which has prompted this effort to try to better understand its historic distribution and dynamics.

5.4.3 *Digital soil mapping of podzolization*

Digital soil mapping (DSM) of soil properties often utilizes digital elevation model (DEM) derivatives, remotely sensed imagery, and climate surfaces as predictive soil forming factor surrogates using geographic information systems (GIS) and computer-based statistical

modeling (Grunwald, 2009; Grunwald et al., 2011; McBratney et al., 2003; Scull et al., 2003). Although many DSM studies are aimed at predicting certain soil classes or soil properties at specified depths (e.g., Behrens et al., 2014; Yang et al., 2011), the same general structure can be applied to predicting a soil pathway such as podzolization because the active soil formation factors being represented by topography and imagery (climate and organisms) drive the processes that produce spodic soil properties. We postulated that an effective spatial model of spodic morphology should spatially correlate to the distribution of red spruce and eastern hemlock in the MNF witness tree database (Thomas-Van Gundy et al., 2012). Our aim was to test use of current spodic morphology as a pedomemory proxy to portray the extent of red spruce and eastern hemlock influence in forests before mass industrial timber harvest and subsequent wildfire. Furthermore, we think these same spatial models of podzolization can be used to connote how red spruce restoration could lead to the buildup of surface O horizons and increased forest carbon stocks and other ecosystem services.

5.5 *Materials and Methods*

5.5.1 *Study area*

We examined sites in the Chemung and Hampshire geologic formations across the regional transition between temperate northern hardwood and subalpine spruce communities within the MNF (Fig. 2). These are acid geologies primarily composed of shale and siltstone parent materials with minor inclusions of sandstone (WVGES, 1968). The area is relatively moist, with mean annual precipitation ranging from 1118-1524 mm (44-60 inches; NOAA-NCDC, 2014), which is likely controlled by elevation and orographic effects. Mean annual temperature ranges from 6.0 to 8.3 degrees Celsius (NOAA-NCDC, 2014), which reflect elevation, slope aspect, and cold air drainage patterns. The elevations of sites examined ranged

from 880-1320 meters, which spans the approximate elevation boundary (~1100 m) between the mesic and frigid soil temperature regimes cited as an important boundary by other regional podzol studies (Lietzke and McGuire, 1987; Stanley and Ciolkosz, 1981). The topography in the area includes flat narrow ridgetops, steep mountainsides, occasional rock outcrops, and deep and narrow river valleys. Within slopes there are benches, hollows, and spurs along with cradle-knoll micro-relief that affect how water, energy, and materials are distributed in the soil system (Schaetzl, 1990).

Current vegetation in the study area in Figure 2 grades from northern hardwoods to spruce-hemlock forests, with mixed conifer-northern hardwood areas between. Common tree species observed in the study area include red maple, sugar maple, mountain maple (*Acer spicatum*), striped maple (*Acer pennsylvanicum*), red spruce, eastern hemlock, yellow birch (*Betula alleghaniensis*), sweet birch (*Betula lenta*), American basswood (*Tilia americana*), white ash (*Fraxinus americana*), northern red oak, black cherry, American beech, mountain magnolia (*Magnolia fraseri*), and cucumber magnolia (*Magnolia acuminata*). Commonly seen shrubs include mountain holly (*Ilex montana*), mountain laurel (*Kalmia latifolia*), and rhododendron (*Rhododendron spp.*), as well as shrubby root sprouts as a result of the beech bark disease complex (Shigo, 1972). Common herbaceous and ground cover species observed include New York fern (*Thelypteris noveboracensis*), intermediate woodfern (*Dryopteris intermedia*), hypnum moss (*Hypnum imponens*), liverwort (*Bazzania trilobata*), three *Lycopodium* species, *Viola* spp., and three *Carex* species.

5.5.2 Data collection and analysis

Three types of soils data were collected as part of this research: (i) extensive point observations of soil morphological properties, (ii) detailed pedon descriptions with

comprehensive laboratory characterization of soil physical and chemical properties at selected sites, and (iii) fixed-area forest vegetation plots with detailed pedon descriptions and limited soil laboratory characterization data. Data collected at all visited locations included detailed field descriptions of the soil morphology at hand-excavated pits with a focus on podzol morphology. We express podzol morphology as a ‘spodic intensity’ (SI; Table 1) based on color, horizon characteristics, and smeariness observations typical of ‘spodic soil materials’ in US *Soil Taxonomy* (Schoeneberger et al., 2002; Soil Survey Staff, 1999). Data were collected by a variety of local soil scientists associated with the USDA-NRCS, USDA-Forest Service (FS), and West Virginia University (WVU). Soil descriptions were made consistent with U.S. national soil survey standards (Schoeneberger et al., 2012). Site locations were selected to evaluate soils derived from Devonian-age shale parent materials on upland landscape positions for the purpose of soil survey update and preliminary ESD reconnaissance. Specific soil map units were associated with three common soil series: Mandy (Loamy-skeletal, mixed, active, frigid Spodic Dystrudepts), Berks (Loamy-skeletal, mixed, active, mesic Typic Dystrudepts), and Dekalb (Loamy-skeletal, siliceous, active, mesic Typic Dystrudepts). Overstory and understory vegetation species lists were also noted at every location.

The extensive point observations were obtained from 2010-2012 at 322 locations throughout the study area. Sampling locations were allocated in small watersheds identified by the FS for examination. Specific sample locations were identified using a stratified random sampling technique in each watershed. From within the specified Mandy, Berks, and Dekalb map units, strata were created based upon vegetation (spruce dominated or other; Lammie, 2009), slope curvature (convex, linear, or concave), and slope gradient (>35% or <35%). Slope curvature and slope gradient were calculated in ArcGIS Spatial Analyst (ESRI, 2011) using a

publicly-available 3-meter resolution DEM

(<http://www.wvgis.wvu.edu/data/dataset.php?ID=261>). These criteria were concatenated to produce individual strata classes (e.g., spruce-convex-<35% slope). Points were randomly located within each stratum using the ArcGIS random points generator. The number of points allocated to each stratum was weighted based on the relative areal amount of each stratum in the watershed. In the watersheds, the soil profiles were examined at an approximate density of one every 25 hectares. A variety of handheld GPS units were used to record actual locations in the field, which makes estimating spatial error of these data difficult.

At seven locations within the study area soil pits were excavated, described, and sampled, and the samples were sent to the NRCS Kellogg Soil Survey Laboratory (KSSL) in Lincoln, NE, for full characterization of soil physical and chemical properties using standard soil laboratory procedures (Soil Survey Staff, 2004) to document the re-classification of the Mandy soil series from Typic Dystrudepts to Spodic Dystrudepts, and the establishment of Wildell, a new soil series classified as Typic Haplorthods. Soil depth profiles of acid oxalate extractions of Al and Fe were compared from Mandy (n = 4) and Wildell (n = 3) as well as three similar, but non-podzolized, soils (analyzed at WVU) from the area thought to be associated with historic hardwood communities. Acid oxalate primarily extracts amorphous to poorly crystalline material including Al (e.g., Al rich allophane and imogolite type materials) and Fe (e.g., ferrihydrite) sesquioxides diagnostic of Spodosols (Soil Survey Staff, 2004). *U.S. Soil Taxonomy* (Soil Survey Staff, 1999) uses the percent weight of aluminum plus half of that of iron (Al+0.5Fe) as one criterion of spodic materials, and we provide depth profiles demonstrating consistency between field spodic intensity (SI, Table 1) observations of color, spodic horizon expression, and soil smeariness (Schoeneberger et al., 2012, page 2-65) with laboratory depth profiles of Al+0.5Fe.

Additionally, in 2013, 24 fixed-area forest plots centered on new soil pit observations were observed to quantitatively compare tree species composition to spodic properties and O horizon thickness. Plots were located near a subset ($n = 15$) of the 322 original locations that were easily accessible and representative of the range of variability recorded. Of the nine remaining new sites, three were located at ridgetop sites that were not represented well in the original sample, and six were randomly located in the study area. Of the 15 revisited sites only two fell within the same pixels as the 2010-2012 observations used for spatial modeling predictions, which makes even these revisited sites pseudo-independent of the original observations for validation purposes. Plot locations were all recorded with a Magellan MobileMapper Pro (v 6.52) GPS unit allowed to record in WASS mode for at least 30 minutes at ground level just upslope of the soil pit face at roughly the center of the plot.

Fixed, 20x20 m area plots were oriented with the slope contour. Diameter at breast height (dbh) was measured on all trees greater than 7 cm dbh. From measured dbh values and species tallies, importance values (IMP) were calculated for red spruce and eastern hemlock (Eq. 2; following Rollins et al., 2010).

$$\text{IMP} = 0.5((\text{species basal area}/\text{plot tree basal area}) + (\text{species count}/\text{plot tree count})) \quad [\text{Eq. 2}]$$

Importance values are proportional measures of relative composition of a specific species that range from zero to one. To compare with IMP values within plots, O horizon thicknesses were observed at the soil profile as well as at the center of each plot quadrant ($n = 5$ per plot). The importance of red spruce and hemlock were added to get a ‘conifer importance’ (CNIMP), which we hypothesized would show strong correlation with O horizon thickness.

We expected that conifer importance would trend positively with both spodic intensity (SI) as well as O horizon thickness. However, because reviewed studies indicate that current

conifer communities are much reduced compared to pre-settlement conditions (e.g., Thomas-Van Gundy et al., 2012), we believed that CNIMP values would have a stronger relationship with O horizon thickness because the Al and Fe accumulations reflected in SI visual cues and smeariness observations are longer lived than organic carbon and O horizons in similar soils (Barrett and Schaetzl, 1998; Hix and Barnes, 1984; Lundström et al., 2000b; Parfitt, 2009). We suspected that O horizons have adjusted much more quickly to forest composition changes, and thus would maintain closer correspondence to the current forest state.

5.5.3 *Spatial modeling using DSM*

A binary random forest probability model (Breiman, 2001; Liaw and Wiener, 2002; Niculescu-Mizil and Caruana, 2005) was implemented to relate a suite of DEM and remotely sensed variables (Table 2) to soils that showed no sign of podzolization ($SI = 0$) versus those that did ($SI > 0$). All DEM variables were computed from the 1-arc second USGS National Elevation Dataset (Gesch et al., 2002; Gesch, 2007) in SAGA GIS (Conrad and Wichmann, 2011). Landsat Geocover imagery from 2000 (MDA, 2004) was also included as a potential predictor source representing current vegetation and land use. Tabulated soil observations and spatial predictor data were intersected using nearest neighbor spatial support and exported from SAGA into the R computing software (R Core Development Team, 2008) for model creation and implementation. Underlying random forest probabilities (relative ensemble votes) were exported as an xyz formatted comma delimited file and imported into SAGA GIS to map spodic morphology probability (probability of $SI > 0$).

Validation of the probability model was evaluated using three approaches. First, the randomForest R package out of bag error (oob) was reported for a model built with the full 322 field point observations. Secondly, a model of a random 2/3 subset of the field points was created

and predicted onto the withheld 1/3 of the points for an independent validation. The classification accuracy and confusion matrix of the withheld data was then reported for the probability threshold that maximized overall accuracy in the validation set by trial and error. Thirdly, the 24 plots examined in 2013 were tested against the predicted surface created by the model created from the full 322 field points. Agreements between predictions and plots were reported for (i) all plots (n = 24), (ii) completely independent new observations (n = 9), and (iii) the pseudo-independent sites that were revisited, but fell into different pixels than the original 2010-2012 GPS points (n = 13).

The spodic probability model created from the full field observation set (n = 322) was then compared to the MNF witness tree database (Thomas-Van Gundy et al., 2012). Points that intersect the predictive model data footprint (n = 1031) were tested to see if witness sites where spruce or hemlock were reported had higher spodic probability values compared to sites with neither species recorded. Both a Welch two-sample t-test and a Wilcoxon rank sum test with continuity correction were used to test this hypothesis against a null of no difference in the R statistical computing program (R Core Development Team, 2008). We expected areas predicted to have spodic morphology (higher probabilities) should correspond with areas that had more spruce and hemlock historically. We then compared our map of spodic properties with a current forest inventory (Byers et al., 2013) to determine how much of the modeled area of spodic expression is currently under hardwood dominated cover congruent with the reported historic conversion of large areas out of spruce cover.

5.6 Results

5.6.1 Soil profile data

Acid oxalate extractable Al and Fe in soil depth profiles clearly distinguished field SI observations representing the gradient of spodic soil morphologies seen in the study area (Fig. 3). Analyzed profiles exhibited distinct depth profiles of $Al + 0.5Fe$ acid oxalate extract, which is one of the criteria for Spodosol classification in U.S. *Soil Taxonomy*. Some variation in depth ranges and intensity of peaks within the classes existed, but overall graphed patterns appeared to separate soils by SI class well. The lack of an increase in $Al+0.5Fe$ in the subsoil of the non-spodic data contrasts strikingly to other sites, which provides evidence supporting our decision to separate these sites from the others in our spatial models of spodic expression presence.

5.6.2 Spatial models of spodic probability

Spodic probability spatial models (Fig. 4) had overall error rates of 30% for both out-of-bag error and the one-third withholding validation. The validation results using withheld data indicated a maximum classification agreement at a 0.57 probability threshold to separate spodic from non-spodic predictions and indicated that predictions of spodic sites were more reliable than those of non-spodic sites (Table 3). The weaker prediction agreement of non-spodic sites (46.3% user error, 61.3% producer error, Table 3) with a lower user error rate indicates that non-spodic sites were over predicted relative to spodic sites. At fixed area forest plots the error rate was 12.5% for all plots ($n = 24$), 22.2% for strictly independent plots ($n = 9$), and 7.7% for the pseudo-independent site revisits that fell into separate pixel predictions than original soil descriptions. Based on these different metrics, 70% seems to be a consistent conservative estimate of overall prediction accuracy.

5.6.3 *Environmental controls on spodic probabilities*

Slope aspect, mid-infrared (MIR) band of Landsat Geocover, and topographic flow convergence calculated in SAGA GIS were the four most important variables in the randomForest analysis of mean decrease in accuracy when these variables were omitted from model building. Specifically, the EASTNESS and NWNNESS slope aspect variables were the most important followed by MIR, and CONVERGENCE. Visual evaluation of the map output (Fig. 4) indicated that W-NW aspects had higher spodic probability, but other factors were more subtle. A highly pruned classification tree was built in rpart (Therneau et al., 2010) to further help interpretations (Fig. 5). Tree structure shows very similar results to the random forest model, with western aspects most favoring spodic development followed by lower MIR values where imagery picks up conifer canopy (usually in lower slope positions of deep narrow valleys that cut into the mountains). The LS_Factor is a water flow energy term from the Universal Soil Loss Equation that SAGA will calculate from a DEM. It is very similar to the CONVERGENCE variable and both mainly distinguish areas that likely concentrate overland water runoff energy. The LS_Factor split might be indicative of past erosion eliminating some areas of spodic expression that might not represent historic spruce preferences, and only isolates 4.2% of the spodic sites. The confusion matrix of the classification fit shows that these three environmental variable splits correctly classify 75% of the soil descriptions.

5.6.4 *Witness tree comparison*

Comparisons of spodic probabilities at witness tree points showed a positive shift in the distribution of values at sites where hemlock or spruce were listed (Wilcoxon rank sum, $p=0.0052$; Welch 2-sample t-test, $p=0.0077$; Fig. 6). This shift was highly significant statistically, and while the magnitude of the shift is visible in the distribution, it still exhibits considerable distribution overlap. However, this area represents a transitional gradient between

hardwood and conifer that we think produces a concurrent gradient of spodic expression and thus considerable overlap in distribution would be expected logically. Witness tree records are also not exhaustive species listings, and an omission of a species does not indicate that it was not present. We must also account for the imperfect spodic spatial model, which does not account for ~30% of the soil variability.

5.6.5 *Conifer importance and soil morphology*

Conifer importance at forest plots shows positive associations with both the thickness of O horizons and SI values. However, the trends with O horizon thickness are much more consistent indicating support for our hypothesis of a quick O-horizon response to forest change (Fig. 7). Both graphs of O horizon response have a positive trend with conifer importance, with overall responses of 0.96 to 1.1 cm of O horizon thickness increase per 10% of conifer importance increase. It is important to note that conifer importance does not include any calculation of site productivity, it is solely based on the relative composition of tree species. Therefore, this association is somewhat independent of site productivity. Interestingly, for our conifer dominated plots older than 100 years in averaged tree core ring counts ($n = 3$ per plot), O horizon thickness averages 18.8 cm compared to the overall regression average of 15.8 cm, suggesting that over time O horizons may get even thicker similar to the findings of Schaetzl (1994). At those older plots, we observed only one site with no charcoal evidence of past fire, and the average O horizon thickness there was 26.8 cm with a maximum of 37 cm. This might be suggestive of the true old growth condition; however, relatively undisturbed sites are hard to find due to the prolific extent of historic disturbance and thus it is difficult to establish a representative sample.

5.7 Discussion

Our results demonstrate the importance of understanding the ecological soil factorial (Eq. 1; Amundson and Jenny, 1991, 1997; Jenny 1961, 1980) and its relationship to pedomemory. Soil process pathways driven by vegetative influences that manifest themselves in soil morphology can inform our understanding of the ecological history and plausible management responses of a site (Higgs, et al., 2014; Johnson and Watson-Stegner, 1987; Phillips and Marion, 2004; Schaetzl and Anderson, 2005; Schaetzl and Schwenner, 2006; Lin, 2011; Simonson, 1959; Targulian and Goryachkin, 2004). We demonstrate this in the Central Appalachian northern hardwood-red spruce transition using models of spodic morphology tested against historic land deed witness tree data.

We think our findings are also important globally because they bring together independent evidence supporting use of soil properties to map historic reference communities. The concept of carefully selecting pedomemory or pedogenic attributes to help understand vegetation dynamics over time is not limited to these systems. For example, recent studies in Australia have shown geochemical pedogenic linkages to vegetative and hydrological dynamics and diversity that generally relate to pH, mineralogy, and redoximorphic features (Bui et al., 2014; Coventry et al., 1983, 1984; Fritsch and Fitzpatrick, 1994; Laliberté et al., 2014; Mûcher and Coventry, 1993). There are many ecosystems that promote certain soil morphologies that have been converted to other land uses with different influences on soil (Goldewijk, 2001; Hansen, 2013; Johnson and Watson-Stegner, 1987; Karhu, 2011; Miles, 1985). These land use changes include deforestation, forest type conversions, agricultural expansion, and urbanization. Changes are often complex and hard to recreate when detailed historic records don't exist, which makes soils invaluable recordings of site histories (Targulian and Goryachkin, 2004).

Our results suggest that the disturbance in the mountains of WV resulting from extensive past industrial timber harvest and related fire, and resulting forest composition changes, probably caused large losses of soil carbon stocks in the forest floor. However, the fact that O-horizons seem to have already somewhat adjusted to current forest composition seems to indicate that red spruce restoration has the potential for re-accumulation of large amounts of forest floor (and thus organic carbon). Earlier work on the spruce-hardwood ecotone in Vermont also showed a correspondence between more acidic soils with deeper forest floors and red spruce dominated areas, but didn't report as much specificity between spruce and spodic properties (Siccama, 1974; Young, 1934). However, modern studies must account for the possibility that the vast harvest disturbance of forests associated with European colonization has favored hardwood incursion into formerly conifer influenced areas (Nowacki et al., 2010; Pielke, 1981) that might be reflected in spodic soils currently under hardwood cover.

When our spodic probability map was overlaid on a current forest inventory map recently completed by Byers et al. (2013), much of the modeled spodic areas were under hardwood cover (<10% conifer). Of areas of the spodic model with >70% probability (26% of study area), 68% were mapped by Byers et al. (2013) as hardwood. This represents a large area of forest currently dominated by hardwoods that we postulate were dominant or co-dominant spruce or hemlock cover before railroad era disturbance. The 70% threshold was chosen because at that probability level we had even greater confidence in our prediction of spodic property presence (77% using withheld validation set), and the vast majority of fully expressed Spodosols (SI = 2) observed at forest plots (100% of plots with Spodosols) and field validation sites (71% of field transect sites with Spodosols) were also seen at probabilities >70%.

5.7.1 *Understanding historic red-spruce community distribution and spodic soil properties*

Other studies of the red spruce – northern hardwood ecotone have often focused on the elevation of the transition and the associated ecological changes (Siccama, 1974; Beckage et al., 2008). Late twentieth century decreases in the growth of red spruce and upward shifts of the ecotone have largely been attributed to climate warming, but cannot rule out pollution and competition as co-factors (Beckage et al., 2008; McLaughlin et al., 1987). Hamburg and Cogbill (1988) were able to show that climate was probably more influential than air pollution (e.g. acid rain) in red spruce decline since 1800. However, all of these changes in red spruce population are superimposed upon the historic harvest impacts, and make determining pre-industrial population distribution estimates quite complex. This complex history makes a plausible pedomemory proxy attractive.

Although we were able to demonstrate strong statistical evidence of spatial correspondence between modeled spodic soil properties and historic witness tree red spruce and hemlock occurrences, the underlying spatial model covariates also seemed to indicate similar climate-related topographic controls to those of red spruce witness trees. We compared analysis from Thomas-Van Gundy et al. (2012) with our models and found similar topographic relationships. Our field data were taken from the Northern High Allegheny Mountain (NHAM) area, but the spodic model footprint we tested also included areas and witness tree locations from smaller areas of the Southern High Allegheny Mountain (SHAM) and Western Allegheny Mountain (WAM) areas as analyzed by Thomas-Van Gundy et al. (2012). In their analysis of spruce locations in NHAM, SHAM, and WAM, Thomas-Van Gundy and co-authors showed spruce associations with northern slope aspects, with northwest slope aspects being specifically favored more in NHAM and SHAM. They also found that relative elevation and landform preferences were for higher ridgetops in SHAM, more cove-like settings in NHAM, and lower

valley bottoms in WAM. Our models showed that spodic soils were most probable on west-northwest slope aspects, similar to the witness tree database. Spodic morphology was also associated with low MIR pixel values that corresponded with conifer-dominated plots (Wilcoxon rank sum test, $W=89$, $p=0.0324$, alternative of MIR being lower at sites with conifer importance $>50\%$). These same low MIR values were also associated with lower slope positions that typically depict coves and narrow valleys (SLOPEPOS in Table 2; Wilcoxon rank sum test, $W=108$, $p=0.013$, alternative of lower MIR at lower slope positions). These areas with low MIR values seem to be representing remnant spruce populations in coves and at lower elevation narrow valley bottoms analogous to the landform analysis seen at lower elevations by Thomas-Van Gundy et al. (2012). We summarize our postulated topographic-climate relationships in Figure 8. It includes an elevation gradient that starts with dominant spruce on the high ridgelines, and grades into spruce microclimates on cool-wet aspects at mid-elevations, and strongly sheltered cold air drainages at lower elevations.

It is important to recognize that our observations only cover a part of the NHAM area analyzed by Thomas-Van Gundy et al. (2012). Our points cover the more rugged ridges and narrow valleys of the upper Greenbrier River watershed and Middle Mountain that run in a mostly S-SW to N-NE direction. Other parts of NHAM, like Canaan Valley, which sits on top of the Blackwater Falls anticline and weathered limestone, have a variety of ridge orientations and more open topography. We also included eastern hemlock as a red spruce associate in witness-tree comparison, which could also be slightly shifting our model results relative to Thomas-Van Gundy et al. (2012), whose analysis was specific to red spruce.

Overall, we feel that these topographic controls probably indicate cooler and wetter climatic niches. Middle elevations (~1000-1250 m) in the WV historic red spruce range seem to have narrower climate windows that exclude spruce-hemlock conifer dominated stands from southeast-east aspects that are warmer and drier, which is likely a result of orographic rain-shadows and greater solar insolation. We also think that the presence of spruce preferentially in narrow valley bottoms and toeslopes at lower elevations is probably related to cold air drainages where we also observed more persistent fog that probably favors spruce. Recent maximum entropy (MaxEnt) models of red spruce habitat suitability indicate that cooler temperatures (especially summer temperatures) and higher snowfall and precipitation were most important in predicting potential red spruce habitat, but did not identify slope aspect as a major driver (Nowacki and Wendt, 2010; Beane et al., 2013). It is difficult to determine if slope aspect-related climate variability was detected in the climate layers used by these studies because the base data for those spatial layers had 400-meter and 1-km resolution, and was probably too coarse to pick out many fine scale topographic aspect patterns. These MaxEnt models also did not detect the lower valley bottom populations of red spruce found down to below 600 meters in the witness tree database. However, our model did extrapolate spodic predictions into those lower areas, and a significant portion of the witness-tree points we tested against were located below 800 meters in areas near Bowden, WV (upper left corner of Fig. 4), which seems to indicate that our spodic model detected these areas of historic spruce found at lower elevations.

Interestingly, soil variables were included in the MaxEnt models as well as the witness-tree studies. In all studies, USDA-NRCS soil surveys, including the more generalized State Soil Geographic (STATSGO2) database and the more detailed Soil Survey Geographic (SSURGO) database soil maps, were summarized by map unit, which can produce interpretation issues

where multiple soils are grouped into one map unit (Nauman and Thompson, 2014; Nauman et al., 2012; Thompson et al., 2012). However, the Mandy soil series was identified as associated with spruce witness tree locations, and was recently reclassified to include recognition of spodic properties based on data used in this study. Many of the other soils identified by Thomas-Van Gundy et al. (2012) are also likely to be cool-moist variants of Inceptisols and Ultisols that might need to be re-evaluated for re-classification as spodic subgroups or Spodosols. For example, the Shouns soil series was found to be associated with spruce in parts of MNF. We found a Shouns soil profile sampled on the southern side of Spruce Knob and characterized by the NSSC that had a discernable depth peak in acid-oxalate extracted Al and Fe (Pedon ID S03WV-071-001, NCSS, 2014). Notably, the two Spodosols previously mapped in WV (Leetonia and Gauley), which are almost exclusively associated with current red spruce stands (Delp, 1998; Flegel, 1998; Losche and Beverage, 1967; Williams and Fridley, 1931; USDA-SCS and USDA-FS, 1982), were not mentioned in the witness tree paper. Beane et al. (2013) did note associations with STATSGO2 soil map units that included Gauley as well as other similar soils to the witness tree results. Nowacki and Wendt (2010) noted associations with shallower soils and fragipans, which makes intuitive sense because red spruce is shallow rooted and perhaps better adapted to fragipans than other species. Nowacki and Wendt (2010) also discussed the likelihood of spodic soil properties being associated with red spruce, which partially inspired this study, but the SSURGO data available for their analysis at that time did not reflect that relationship.

5.7.2 *Future implications*

More laboratory corroboration and wider spatial sampling would provide greater certainty for our conclusions regarding historic forests and restoration projections in WV. We did not include data describing soil organic carbon dynamics in mineral soil horizons (A, Bh, and

Bhs) at these sites to see how restoration might affect those pools, but we think that they could also represent a significant potential flux after disturbance. Others have shown that mineral horizon organic carbon stocks can be lost via depodzolization after disturbance and vegetation conversion in similar systems (Barret and Schaetzl, 1998; Hole, 1975). Soil pools, along with calculations from forest growth model scenarios (e.g., Krankina et al., 2012; Schulze et al., 2012) could provide a more interdisciplinary illustration of carbon sequestration potential and will likely provide evidence of even greater ability of these forests and soils to mitigate climate change.

We also hope that other researchers will further investigate subalpine/boreal conifer to temperate hardwood ecotones throughout other comparable zones of the world to see if similar scenarios exist where prior disturbance has caused compositional and biogeochemical shifts. We also expect that that future work with quantitative analysis of translocated soil sesquioxides in WV and similar areas, especially Al-rich allophanes and proto-imogolites, could potentially provide a spatially explicit map of quantitative estimates of pre-disturbance forest composition since these compounds have longer residence times in the soil than other spodic properties (Lundström et al., 2000b; Parfitt, 2009).

5.8 *Conclusions*

Soil properties and morphology can reveal pedomemory insights into past vegetative dynamics. The key to this is understanding the time scale and mechanisms associated with different vegetation related soil processes that manifest in soil development. In cool, moist, and acidic conifer forests, persistent subsurface sesquioxide horizons reside in soils for long periods and can serve as indicators of those forest communities. Contrastingly, organic carbon pools can shift quickly when forest composition is changed due to disturbance. Carbon pools that respond

quickly to forest restoration represent an important potential avenue of carbon sequestration and habitat renewal. Although there is uncertainty regarding future effects of climate change on red spruce, there might be a significant mitigation potential in red spruce restoration. Alternatively, if red spruce is lost, similar species that promote podsolization including other selected *Tsuga*, *Larix*, *Picea*, *Pinus*, and *Abies* species could serve as alternatives. Restoration of red spruce and similar carbon-sequestering species represents one of many potential climate and ecological degradation mitigation options that society will need to evaluate in our efforts to balance our global carbon pools and disturbance footprint.

5.9 Acknowledgements

We would like to acknowledge the large group of scientists who contributed to the field soil description efforts for this project. These individuals came from diverse institutions including USDA-NRCS, USDA Forest Service, West Virginia University, West Virginia State University, and Virginia Polytechnic Institute and State University. We also acknowledge the dedicated fieldwork at forest plots by WVU research associate Aaron Burkholder, and the input and assistance from Shane Jones, Monongahela National Forest Biologist. We acknowledge Greg Nowacki for his helpful insights and edits during preparation of this paper. Portions of this research were supported by the United States' Department of Agriculture Natural Resources Conservation Service. Scientific contribution no. XXXX from the West Virginia Agricultural and Forestry Experiment Station, Morgantown, WV.

5.10 *Literature Cited*

- Adams, M. B., and C. Eagar. 1992. Impacts of acidic deposition on high-elevation spruce-fir forests—results from the Spruce-Fir Research Cooperative. *Forest Ecology and Management* 51:195-205.
- Amundson, R., and H. Jenny. 1991. The place of humans in the state factor theory of ecosystems and their soils. *Soil Science* 151:99-109.
- Amundson, R., and H. Jenny. 1997. On a state factor model of ecosystems. *BioScience* 47:536-543.
- Barrett, L. R., and R. J. Schaetzl. 1998. Regressive Pedogenesis Following a Century of Deforestation: Evidence for Depodzolization. *Soil Science* 163(6):482-497.
- Beane, N. R., J. S. Rentch, and T. M. Schuler. 2013. Using Maximum Entropy Modeling to Identify and Prioritize Red Spruce Forest Habitat in West Virginia. USFS Northern Research Station.
- Behrens, T., K. Schmidt, L. Ramirez-Lopez, J. Gallant, A. X. Zhu, and T. Scholten. 2014. Hyper-scale digital soil mapping and soil formation analysis. *Geoderma* 213:578-588.
- Berg, B., and C. McClaugherty. 2008. *Plant litter*. Springer.
- Blum, J. D., A. Klaue, C. A. Nezat, C. T. Driscoll, C. E. Johnson, T. G. Siccama, C. Eagar, T. J. Fahey, and G. E. Likens. 2002. Mycorrhizal weathering of apatite as an important calcium source in base-poor forest ecosystems. *Nature* 417:729-731.
- Breiman, L. 2001. Random Forests. *Machine Learning* 45:5-32.
- Bui, E., C. González-Orozco, and J. Miller. 2014. Acacia, climate, and geochemistry in Australia. *Plant and Soil* 381:161-175.
- Burrough, P. A. 1983. Problems of superimposed effects in the statistical study of the spatial variation of soil. *Agricultural Water Management* 6:123-143.

- Burt, R and Soil Survey Staff. 2004. Soil Survey Laboratory Methods Manual. USDA-NRCS, Lincoln, NE.
- Butler, P. R., L. Iverson, F. R. T. III, L. Brandt, S. Handler, M. Janowiak, P. D. Shannon, C. Swanston, J. Bartig, S. Connelly, W. Dijak, K. Karriker, C. Randall, S. Bearer, S. Blatt, A. Brandon, E. Byers, C. Coon, T. Culbreth, J. Daly, W. Dorsey, D. Ede, C. Euler, N. Gillies, L. Lyte, D. McCarthy, D. Minney, D. I. Murphy, C. O'Dea, D. Hix, C. Johnson, R. Orwan, M. Peters, J. Reed, C. Sandeno, T. Schuler, L. Sneddon, B. Stanley, A. Steele, R. Swaty, S. Stout, J. Teets, T. Tomon, J. Vanderhorst, J. Whatley, and N. Zegre. 2014 (In Review). Central Appalachians ecosystem vulnerability assessment and synthesis: a report from the Central Appalachians Climate Change Response Framework project., Department of Agriculture, Forest Service, Northern Research Station., Newtown Square, PA.
- Byers, E. A., K. C. Love, K. R. Haider, E. J. Burks, and J. E. Rowan. 2013. Red Spruce (*Picea rubens*) Cover in West Virginia, Version 1.0. West Virginia Division of Natural Resources, Central Appalachian Spruce Restoration Initiative, Appalachian Forest Heritage Area Americorps, Monongahela National Forest, and U.S. Fish and Wildlife Service.
- Byers, E. A., J. P. Vanderhorst, and B. P. Streets. 2010. Classification and Conservation Assessment of Upland Red Spruce Communities in West Virginia. WV-DNR West Virginia Natural Heritage Program. Wildlife Resources Section.
- Caudle, D., H. Sanchez, J. DiBenedetto, C. Talbot, and M. Karl. 2013. Interagency Ecological Site Handbook for Rangelands. USDA-NRCS, USDA-FS, & DOI-BLM.

- Clarkson, R. B. 1964. Tumult on the mountains: lumbering in West Virginia, 1770-1920. McClain Printing Company.
- Conrad, O., and V. Wichmann. 2011. SAGA GIS (www.saga-gis.org). Hamburg, Germany.
- Côté, B., and J. W. Fyles. 1994. Nutrient concentration and acid-base status of leaf litter of tree species characteristic of the hardwood forest of southern Quebec. *Canadian Journal of Forest Research* 24:192-196.
- Coventry, R. J., Taylor, R. M., & Fitzpatrick, R. W. (1983). Pedological significance of the gravels in some red and grey earths of central North Queensland. *Australian Journal of Soil Research*, 21(3), 219-240.
- Coventry, R. J., & Williams, J. (1984). Quantitative relationships between morphology and current soil hydrology in some Alfisols in semiarid tropical Australia. *Geoderma*, 33(3), 191-218.
- Delp, C. H. 1998. Soil Survey of Webster County, West Virginia (SSURGO). U.S. Department of Agriculture (USDA) - NRCS, and USDA-FS.
- Dillard, L. O., K. R. Russell, and W. M. Ford. 2008a. Macrohabitat models of occurrence for the threatened Cheat Mountain salamander, *Plethodon nettingi*. *Applied Herpetology* 5:201-224.
- Dillard, L. O., K. R. Russell, and W. M. Ford. 2008b. Site-level habitat models for the endemic, threatened Cheat Mountain salamander (*Plethodon nettingi*): the importance of geophysical and biotic attributes for predicting occurrence. *Biodiversity and Conservation* 17:1475-1492.
- Dokuchaev, V. V. 1999. On the concept of natural zones - St. Petersburg, 1899. *Eurasian Soil Science* 32:726-727.

- ESRI. 2011. ArcGIS Desktop: Release 10. Environmental Systems Research Institute, Redlands, CA.
- Finzi, A. C., N. Van Breemen, and C. D. Canham. 1998. Canopy tree-soil interactions within temperate forests: species effects on soil carbon and nitrogen. *Ecological Applications* 8:440-446.
- Flegel, D. G. 1998. Soil Survey of Pocahontas County, West Virginia (SSURGO). U.S. Department of Agriculture (USDA) - NRCS, and USDA-FS.
- Friedland, A. J., G. J. Hawley, and R. A. Gregory. 1988. Red spruce (*Picea rubens* Sarg.) foliar chemistry in Northern Vermont and New York, USA. *Plant and Soil* 105:189-193.
- Fritsch, E., & Fitzpatrick, R. W. (1994). Interpretation of soil features produced by ancient and modern processes in degraded landscapes. 1. A new method for constructing conceptual soil-water-landscape models. *Australian Journal of Soil Research*, 32(5), 889-907.
- Gesch, D. B. 2007. The National Elevation Dataset. Pages 99-118 *in* D. Maune, editor. Digital Elevation Model Technologies and Applications: The DEM Users Manual. American Society for Photogrammetry and Remote Sensing.
- Gesch, D. B., M. Oimoen, S. Greenless, C. Nelson, M. Steuck, and D. Tyler. 2002. The National Elevation Dataset. *Photogrammetric Engineering and Remote Sensing* 68:5-11.
- Giesler, R., H. Ilvesniemi, L. Nyberg, P. van Hees, M. Starr, K. Bishop, T. Kareinen, and U. S. Lundström. 2000. Distribution and mobilization of Al, Fe and Si in three podzolic soil profiles in relation to the humus layer. *Geoderma* 94:249-263.
- Glenn, M. G., W. S. Wagner, and S. L. Webb. 1991. Mycorrhizal status of mature red spruce (*Picea rubens*) in mesic and wetland sites of northwestern New Jersey. *Canadian Journal of Forest Research* 21:741-749.

- Goldewijk, K.K. 2001. Estimating global land use change over the past 300 years: The HYDE Database. *Global Biogeochemical Cycles* 15:417-433.
- Grunwald, S. 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma* 152:195-207.
- Grunwald, S., J. A. Thompson, and J. L. Boettinger. 2011. Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues. *Soil Science Society of America Journal* 75:1201-1213.
- Hansen, M.C., P.V. Potapov, R. Moore, M. Hancher, S.A. Turubanova, A. Tyukavina, D. Thau, S.V. Stehman, S.J. Goetz, T.R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C.O. Justice, and J.R.G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* 342:850-853.
- Hamburg, S. P., and C. V. Cogbill. 1988. Historical decline of red spruce populations and climatic warming. *Nature* 331:428-431.
- Herbauts, J., and E. Buyl. 1981. The relation between spruce monoculture and incipient podzolisation in ochreous brown earths of the Belgian Ardennes. *Plant and Soil* 59:33-49.
- Higgs, E., D.A. Falk, A. Guerrini, M. Hall, J. Harris, R.J. Hobbs, S.T. Jackson, J.M. Rhemtulla, and W. Throop. 2014. The changing role of history in restoration ecology. *Frontiers in Ecology and the Environment* 12:499-506.
- Hix, D. M., and B. V. Barnes. 1984. Effects of clear-cutting on the vegetation and soil of an eastern hemlock dominated ecosystem, western Upper Michigan. *Canadian Journal of Forest Research* 14:914-923.

- Hobbie, S., M. Ogdahl, J. Chorover, O. Chadwick, J. Oleksyn, R. Zytkowski, and P. Reich. 2007. Tree Species Effects on Soil Organic Matter Dynamics: The Role of Soil Cation Composition. *Ecosystems* 10:999-1018.
- Hobbie, S. E., P. B. Reich, J. Oleksyn, M. Ogdahl, R. Zytkowski, C. Hale, and P. Karolewski. 2006. Tree species effects on decomposition and forest floor dynamics in a common garden. *Ecology* 87:2288-2297.
- Hoffland, E., T. W. Kuyper, H. Wallander, C. Plassard, A. A. Gorbushina, K. Haselwandter, S. Holmström, R. Landeweert, U. S. Lundström, and A. Rosling. 2004. The role of fungi in weathering. *Frontiers in Ecology and the Environment* 2:258-264.
- Högberg, P., and D. J. Read. 2006. Towards a more plant physiological perspective on soil ecology. *Trends in Ecology & Evolution* 21:548-554.
- Hole, F.D. 1975. Some relationships between forest vegetation and Podzol B horizons in soils of Menominee tribal lands, Wisconsin, U.S.A. *Sov. Soil Sci.* 7:714-723.
- Hopkins, A. D. 1899. Report on Investigations to Determine the Cause of Unhealthy Conditions of the Spruce and Pine from 1880-1893. W.V. Agricultural Experimental Station. Fairmont Index Steam Print, Morgantown, WV.
- Hornbeck, J. W., and R. B. Smith. 1985. Documentation of red spruce growth decline. *Canadian Journal of Forest Research* 15:1199-1201.
- Iverson, L. R., A. M. Prasad, S. N. Matthews, and M. Peters. 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecology and Management* 254:390-406.
- Jenny, H. 1941. *Factors of Soil Formation*. McGraw-Hill, New York, New York.

- Jenny, H. 1961. Derivation of state factor equations of soils and ecosystems. *Soil Science Society of America Journal* 25:385-388.
- Jenny, H. 1980. *Ecological Studies: Analysis and Synthesis, Vol. 37. The Soil Resource: Origin and Behavior*. Xxi+377p. Springer-Verlag: New York, N.Y., USA
- Johnson, A. H. 1983. Red Spruce Decline in the Northeastern U.S.: Hypotheses Regarding the Role of Acid Rain. *Journal of the Air Pollution Control Association* 33:1049-1054.
- Johnson, D. L., and D. Watson-Stegner. 1987. Evolution model of Pedogenesis. *Soil Science* 143:349-366.
- Jongmans, A. G., N. van Breemen, U. Lundstrom, P. A. W. van Hees, R. D. Finlay, M. Srinivasan, T. Unestam, R. Giesler, P. A. Melkerud, and M. Olsson. 1997. Rock-eating fungi. *Nature* 389:682-683.
- Karhu, K., A. Wall, P. Vanhala, J. Liski, M. Esala, and K. Regina. 2011. Effects of afforestation and deforestation on boreal soil carbon stocks—Comparison of measured C stocks with Yasso07 model results. *Geoderma* 164:33-45.
- Krankina, O. N., M. E. Harmon, F. Schneckengerger, and C. A. Sierra. 2012. Carbon balance on federal forest lands of Western Oregon and Washington: The impact of the Northwest Forest Plan. *Forest Ecology and Management* 286:171-182.
- Lal, R. 2005. Forest soils and carbon sequestration. *Forest Ecology and Management* 220:242-258.
- Laliberté, E., G. Zemunik, and B. L. Turner. 2014. Environmental filtering explains variation in plant diversity along resource gradients. *Science* 345:1602-1605.
- Lammie, S. 2009. Task Order West Virginia Northern Flying Squirrel (WVNFS) Vegetative Habitat (Map).

- Lewis, R. L. 1998. Transforming the Appalachian countryside: Railroads, deforestation, and social change in West Virginia, 1880-1920. Univ of North Carolina Press.
- Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. R news 2:18-22.
- Lietzke, D. A., and G. A. McGuire. 1987. Characterization and Classification of Soils with Spodic Morphology in the Southern Appalachians. Soil Sci. Soc. Am. J. 51:165-170.
- Lin, H. 2011. Three Principles of Soil Change and Pedogenesis in Time and Space. Soil Sci. Soc. Am. J. 75:2049-2070.
- Losche, C. K., and W. W. Beverage. 1967. Soil survey of Tucker County and part of Northern Randolph County, West Virginia. Soil Survey Reports. United States Department of Agriculture.
- Lundström, U. S., N. van Breemen, and D. Bain. 2000a. The podzolization process. A review. Geoderma 94:91-107.
- Lundström, U. S., N. van Breemen, D. C. Bain, P. A. W. van Hees, R. Giesler, J. P. Gustafsson, H. Ilvesniemi, E. Karlton, P. A. Melkerud, M. Olsson, G. Riise, O. Wahlberg, A. Bergelin, K. Bishop, R. Finlay, A. G. Jongmans, T. Magnusson, H. Mannerkoski, A. Nordgren, L. Nyberg, M. Starr, and L. Tau Strand. 2000b. Advances in understanding the podzolization process resulting from a multidisciplinary study of three coniferous forest soils in the Nordic Countries. Geoderma 94:335-353.
- McBratney, A. B., M. L. M. Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma 117:3-52.
- MDA, F. 2004. Landsat GeoCover ETM+ 2000 Edition Mosaics Tile N-17-35.ETM-EarthSat-MrSID. USGS, Sioux Falls, South Dakota.

- Melvin, A. M., and C. L. Goodale. 2013. Tree species and earthworm effects on soil nutrient distribution and turnover in a northeastern United States common garden. *Canadian Journal of Forest Research* 43:180-187.
- Menzel, J. M., W. M. Ford, J. W. Edwards, and L. J. Ceperley. 2006a. A habitat model for the Virginia northern flying squirrel (*Glaucomys sabrinus fuscus*) in the central Appalachian Mountains. US Department of Agriculture, Forest Service, Northeastern Research Station.
- Menzel, J. M., W. M. Ford, J. W. Edwards, and M. A. Menzel. 2004. Nest tree use by the endangered Virginia northern flying squirrel in the central Appalachian mountains. *American Midland Naturalist* 151:355-368.
- Menzel, J. M., W. M. Ford, J. W. Edwards, and T. M. Terry. 2006b. Home range and habitat use of the Vulnerable Virginia northern flying squirrel *Glaucomys sabrinus fuscus* in the Central Appalachian Mountains, USA. *Oryx* 40:204-210.
- Miles, J. 1985. The pedogenic effects of different species and vegetation types and the implications of succession. *Journal of Soil Science* 36:571-584.
- Mücher, H. J., & Coventry, R. J. (1993). Soil and landscape processes evident in a hydromorphic grey earth (Plinthusalf) in semiarid tropical Australia. *Developments in Soil Science*, 22, 221-231.
- Nauman, T., J. A. Thompson, N. Odgers, and Z. Libohova. 2012. Fuzzy Disaggregation of Conventional Soil Maps using Database Knowledge Extraction to Produce Soil Property Maps. *in* B. Minasny, B. Malone, and A. McBratney, editors. *Digital Soil Assessments and Beyond: 5th Global Workshop on Digital Soil Mapping*, Sydney, Australia.

- Nauman, T. W., and J. A. Thompson. 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. *Geoderma* 213:385-399.
- Niculescu-Mizil, A., and R. Caruana. 2005. Predicting good probabilities with supervised learning. Pages 625-632 *in* Proceedings of the 22nd international conference on Machine learning. ACM.
- Nikiforoff, C. 1959. Reappraisal of the soil. *Science* 129:186-196.
- Nowacki, G., and D. Wendt. 2010. The current distribution, predictive modeling, and restoration potential of red spruce in West Virginia. Pages 163-178 *in* Proceedings from the conference on the ecology and management of high-elevation forests in the central and southern Appalachian Mountains. USDA-FS Northern Research Station, Slatyfork, WV.
- Nowacki, G. J., R. Carr, and M. Van Dyck. 2010. The Current Status of Red Spruce in the Eastern United States: Distribution, Population Trends, and Environmental Drivers. Page 242 *in* Proceedings from the conference on the ecology and management of high-elevation forests in the central and southern Appalachian Mountains. USDA - Forest Service, Northern Research Station, Slaty Fork, WV.
- Parfitt, R. L. 2009. Allophane and imogolite: role in soil biogeochemical processes. *Clay Minerals* 44:135-155.
- Pauley, T. K. 2008. The Appalachian Inferno: Historical Causes for the Disjunct Distribution of *Plethodon nettingi* (Cheat Mountain Salamander). *Northeastern Naturalist* 15:595-606.
- Phillips, J. D., and D. A. Marion. 2004. Pedological memory in forest soil development. *Forest Ecology and Management* 188:363-380.

- Pielke, R. A. 1981. The Distribution of Spruce in West-Central Virginia before Lumbering. *Castanea* 46:201-216.
- Prasad, A., L. Iverson, S. Matthews, and M. Peters. 2007. A climate change atlas for 134 forest tree species of the eastern United States [database]. Northern Research Station, USDA Forest Service, Delaware, Ohio.
- Ranger, J., and C. Nys. 1994. The effect of spruce (*Picea abies* Karst.) on soil development: an analytical and experimental approach. *European Journal of Soil Science* 45:193-204.
- Reich, P. B., J. Oleksyn, J. Modrzyński, P. Mrozinski, S. E. Hobbie, D. M. Eissenstat, J. Chorover, O. A. Chadwick, C. M. Hale, and M. G. Tjoelker. 2005. Linking litter calcium, earthworms and soil properties: a common garden test with 14 tree species. *Ecology Letters* 8:811-818.
- Rentch, J. S., T. M. Schuler, W. M. Ford, and G. J. Nowacki. 2007. Red spruce stand dynamics, simulations, and restoration opportunities in the central Appalachians. *Restoration Ecology* 15:440-452.
- Rentch, J. S., T. M. Schuler, G. J. Nowacki, N. R. Beane, and W. M. Ford. 2010. Canopy gap dynamics of second-growth red spruce-northern hardwood stands in West Virginia. *Forest Ecology and Management* 260:1921-1929.
- Rollins, A. W., H. S. Adams, and S. L. Stephenson. 2010. Changes in Forest Composition and Structure across the Red Spruce-Hardwood Ecotone in the Central Appalachians. *Castanea* 75:303-314.
- Runge, E. C. A. 1973. Soil development sequences and energy models. *Soil Science* 115:183-193.

- Rustad, L. E., and I. J. Fernandez. 1998. Soil Warming: Consequences for Foliar Litter Decay in a Spruce-Fir Forest in Maine, USA. *Soil Sci. Soc. Am. J.* 62:1072-1080.
- Sauer, D., H. Sponagel, M. Sommer, L. Giani, R. Jahn, and K. Stahr. 2007. Podzol: Soil of the Year 2007. A review on its genesis, occurrence, and functions. *Journal of Plant Nutrition and Soil Science* 170:581-597.
- Schaetzl, R.J. 1990. Effects of treethrow microtopography on the characteristics and genesis of Spodosols, Michigan, USA. *Catena* 17:111-126.
- Schaetzl, R.J. 1994. Changes in O horizon mass, thickness and carbon content following fire in northern hardwood forests. *Vegetatio* 115:41-50.
- Schaetzl, R. J., and S. Anderson. 2005. *Soils: Genesis and geomorphology*. Cambridge University Press.
- Schaetzl, R.J. and W. Harris. 2011. Spodosols. In: *Handbook of Soil Sciences*, 2nd ed. P.M. Huang, Y. Li and M.E. Sumner (eds.). CRC Press, New York. pp. 33-113 – 33-127.
- Schaetzl, R. J., and S. A. Isard. 1996. Regional-scale relationships between climate and strength of podzolization in the Great Lakes Region, North America. *Catena* 28:47-69.
- Schaetzl, R. J., and C. Schwenner. 2006. An application of the runge "energy model" of soil development in Michigan's upper peninsula. *Soil Science* 171:152-166.
- Schoeneberger, P. J., E. Wysocki, and S. S. Staff. 2012. *Field book for describing and sampling soils*, Version 3.0. Government Printing Office.
- Schöll, L., T. Kuyper, M. Smits, R. Landeweert, E. Hoffland, and N. Breemen. 2008. Rock-eating mycorrhizas: their role in plant nutrition and biogeochemical cycles. *Plant and Soil* 303:35-47.

- Schulze, E.-D., C. Körner, B. E. Law, H. Haberl, and S. Luysaert. 2012. Large-scale bioenergy from additional harvest of forest biomass is neither sustainable nor greenhouse gas neutral. *GCB Bioenergy* 4:611-616.
- Scull, P., J. Franklin, O. A. Chadwick, and D. McArthur. 2003. Predictive soil mapping: a review. *Progress in Physical Geography* 27:171-197.
- SER. 2004. International Primer on Ecological Restoration. Online:
<http://www.ser.org/resources/resources-detail-view/ser-international-primer-on-ecological-restoration>
- Shigo, A. L. 1972. The beech bark disease today in the northeastern US. *Journal of Forestry* 70:286-289.
- Simonson, R. W. 1959. Outline of a generalized theory of soil genesis. *Soil Sci Soc Amer Proc* 23:152-156.
- Sohet, K., J. Herbauts, and W. Gruber. 1988. Changes Caused by Norway Spruce in an Ochreous Brown Earth, Assessed by the Isoquartz Method. *Journal of Soil Science* 39:549-561.
- Soil Survey Staff. 1999. *Soil Taxonomy A Basis System of Soil Classification for Making and Interpreting Soil Surveys*. in U. S. D. o. A. N. R. C. Service, editor. U.S. Government Printing Office, Washington.
- Soil Survey Staff. 2010. *Keys to Soil Taxonomy*. in U.-N. R. C. Service, editor., Washington, DC.
- Stanley, S. R., and E. J. Ciolkosz. 1981. Classification and Genesis of Spodosols in the Central Appalachians¹. *Soil Sci. Soc. Am. J.* 45:912-917.
- National Cooperative Soil Survey. 2014. National Cooperative Soil Characterization Database. Available online at <http://ncsslabdatamart.sc.egov.usda.gov>

- Targulian, V. O., and S. V. Goryachkin. 2004. Soil memory: Types of record, carriers, hierarchy and diversity. *Revista Mexicana de Ciencias Geológicas* 21.
- Tarnocai, C., J. G. Canadell, E. A. G. Schuur, P. Kuhry, G. Mazhitova, and S. Zimov. 2009. Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochemical Cycles* 23:GB2023.
- R Core Development Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Teets, J. 2013. Ecological Site Description: Spodic Shale Upland Conifer Forest (Draft).in USDA-NRCS, editor., Morgantown, WV.
- Therneau, T. M., B. Atkinson, and B. Ripley. 2010. rpart: Recursive partitioning. R package version 3.
- Thomas-Van Gundy, M., M. Strager, and J. Rentch. 2012. Site characteristics of red spruce witness tree locations in the uplands of West Virginia, USA. *The Journal of the Torrey Botanical Society* 139:391-405.
- Thompson, J. A., T. Nauman, N. Odgers, Z. Libohova, and J. Hempel. 2012. Harmonization of Legacy Soil Maps in North America: Status, Trends, and Implications for Digital Soil Mapping Efforts.in *The 5th Global Workshop on Digital Soil Mapping. Digital Soil Assessments and Beyond*, Sydney, Australia.
- USDA-NRCS. 2014. National Ecological Site Handbook.
- USDA-SCS, and USDA-FS. 1982. Soil Survey of Randolph County Area, Main Part, West Virginia.
- Van Breemen, N., and A. C. Finzi. 1998. Plant-soil interactions: ecological aspects and evolutionary implications. *Biogeochemistry* 42:1-19.

- van Breemen, N., U. S. Lundström, and A. G. Jongmans. 2000. Do plants drive podzolization via rock-eating mycorrhizal fungi? *Geoderma* 94:163-171.
- Watts, W. A. 1979. Late Quaternary Vegetation of Central Appalachia and the New Jersey Coastal Plain. *Ecological Monographs* 49:427-469.
- West Virginia Geological and Economic Survey. 1968. Surface Geology - Rock Units. Online Digitized Map: Available at <http://wvgis.wvu.edu/> in T. U. West Virginia DEP, editor. William and Heintz Map Corporation.
- Williams, B. H., and H. M. Fridley. 1931. Soil Survey of Randolph County, West Virginia.
- Willis, K. J., M. Braun, P. Sumegi, and A. Toth. 1997. Does soil change cause vegetation change or vice versa? A temporal perspective from Hungary. *Ecology* 78:740-750.
- Yang, L., Y. Jiao, S. Fahmy, A. X. Zhu, S. Hann, J. E. Burt, and F. Qi. 2011. Updating Conventional Soil Maps through Digital Soil Mapping. *Soil Science Society of America Journal* 75:1044-1053.

5.11 Tables

Table 1. Description of spodic intensity (SI) classes based on observable field morphology.

SI Class	Description
0.0	No evidence of podzolization.
0.5	Very weak expression of podzolization. There is only slight physical evidence of podzolization. A slightly redder hue and higher value is present at the top of the B horizon, but the hue is less than one Munsell hue redder than an underlying horizon. The soil is non-smearly*.
1.0	Weak expression of podzolization (spodic intergrade, very close to Spodosol). Spodic materials are present, but they might not meet the criteria for a diagnostic spodic horizon. A weakly expressed Bs horizon is present. The Bs horizon is one Munsell hue redder than an underlying horizon. Bhs material is usually absent. An albic E horizon is not present. The spodic materials are sometimes weakly smearly.
1.5	Moderate expression of podzolization (Spodosol). Spodic materials are present as a diagnostic spodic horizon. A moderately expressed Bs horizon is present, often with pockets of Bhs material. An albic E horizon is not present. The spodic materials are often weakly smearly.
2.0	Strong expression of podzolization (well-expressed Spodosol). A diagnostic spodic horizon is present usually underlying an albic E horizon. A Bhs or Bh horizon is continuous across at least 85 percent of the pedon. The spodic materials are often moderately smearly.

* Smeariness (Shoeneberger et al, 2012, page 2-65) is a physical observation about how moistened soil samples fail when they are squeezed and rubbed between the thumb and forefinger. Smeariness can help identify spodic soil materials.

Table 2. Spatial variables used to model spodic probability.

Variable Name	Description
<i>National Elevation Dataset (~27.5-meter resolution)</i>	
NWNESS	Index from 1 to -1 of how northwest (1) or southeast (-1) a site faces
EASTNESS	Index from 1 to -1 of how east (1) or west (-1) a site faces
SOUTHNESS	Index from 1 to -1 of how south (1) or north (-1) a site faces
NENESS	Index from 1 to -1 of how northeast (1) or southwest (-1) a site faces
ELEVm	Elevation in meters
PLAN_CURV	Curvature perpendicular to the slope direction
PROF_CURV	Curvature parallel to the slope direction
LS_FACTOR	Slope-length factor from USLE as calculated in SAGA GIS
CONVERGENCE	Overall measure of concavity
SLOPEPOS	Index from 0 (valley floor) to 100 (ridgetop) of slope position (Hatfield, 1996)
SLOPE	Slope gradient (rise/run) in fraction units
MRRTF	Multiple resolution ridgetop flatness index
MRVBF	Multiple resolution valley bottom flatness index
TWI	Topographic wetness index
ALT_OVER_STREAM	Altitude above local stream channel
BASELEVEL	Elevation of nearest channel point to each pixel in its given watershed
CONTRIBAREA	Upstream contributing area
REL_HT_1	Height of cell above the local minimum elevation in 1-pixel radius
REL_HT_2	Height of cell above the local minimum elevation in 2-pixel radius

REL_HT_3	Height of cell above the local minimum elevation in 3-pixel radius
REL_HT_5	Height of cell above the local minimum elevation in 5-pixel radius
REL_HT_10	Height of cell above the local minimum elevation in 10-pixel radius
REL_HT_20	Height of cell above the local minimum elevation in 20-pixel radius
REL_HT_30	Height of cell above the local minimum elevation in 30-pixel radius
REL_HT_50	Height of cell above the local minimum elevation in 50-pixel radius
REL_HT_70	Height of cell above the local minimum elevation in 70-pixel radius

Landsat Geocover 2000 (14.5-meter resolution, resampled to 27.5-m)

NIR	Near Infrared band in 8-bit digital number units
MIR	Middle Infrared band in 8-bit digital number units
GREEN	Green visible band in 8-bit digital number units
MIRNIR	Ratio of MIR/NIR
GREENNIR	Ratio of GREEN/NIR
GREENMIR	Ratio of GREEN/MIR

Table 3. Confusion matrix from withheld 1/3 validation set for spodic probability predictions using a 0.57 threshold for classification as ‘spodic’.

		<u>predicted</u>		
		non spodic	spodic	
<u>observed</u>	non spodic	19	12	61.3%
	spodic	22	61	73.5%
		46.3%	83.6%	70.2%

5.12 *Figures*

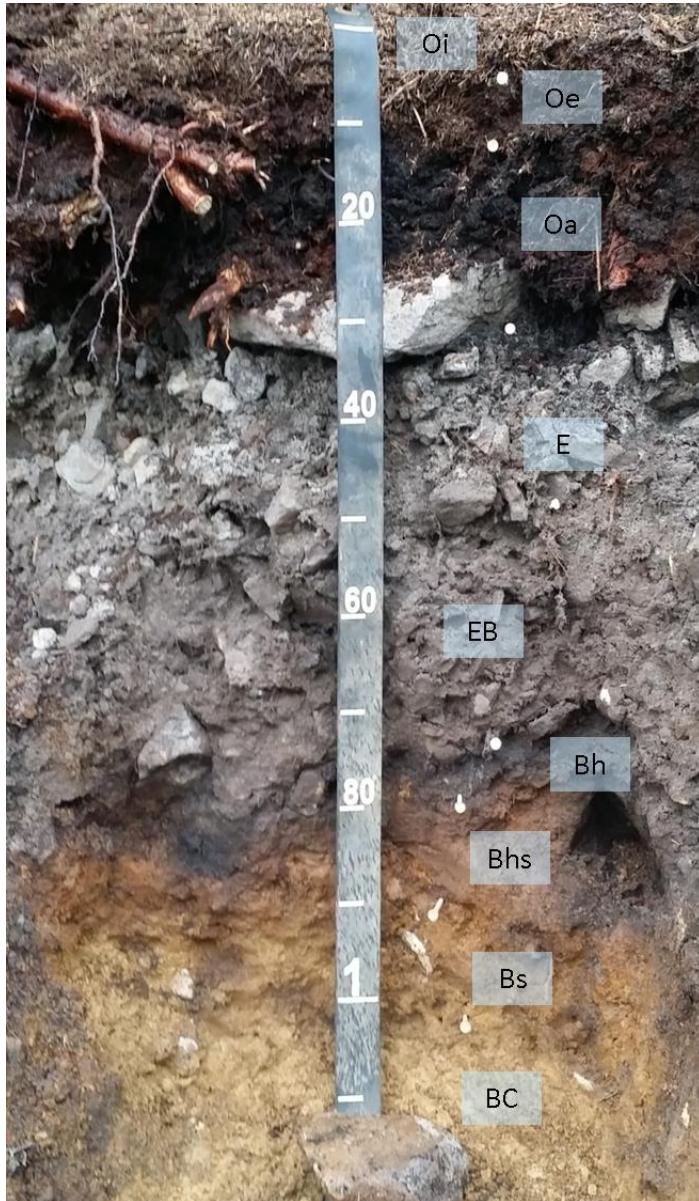


Figure 1. Well expressed podzol soil morphology in a red spruce forest in WV.

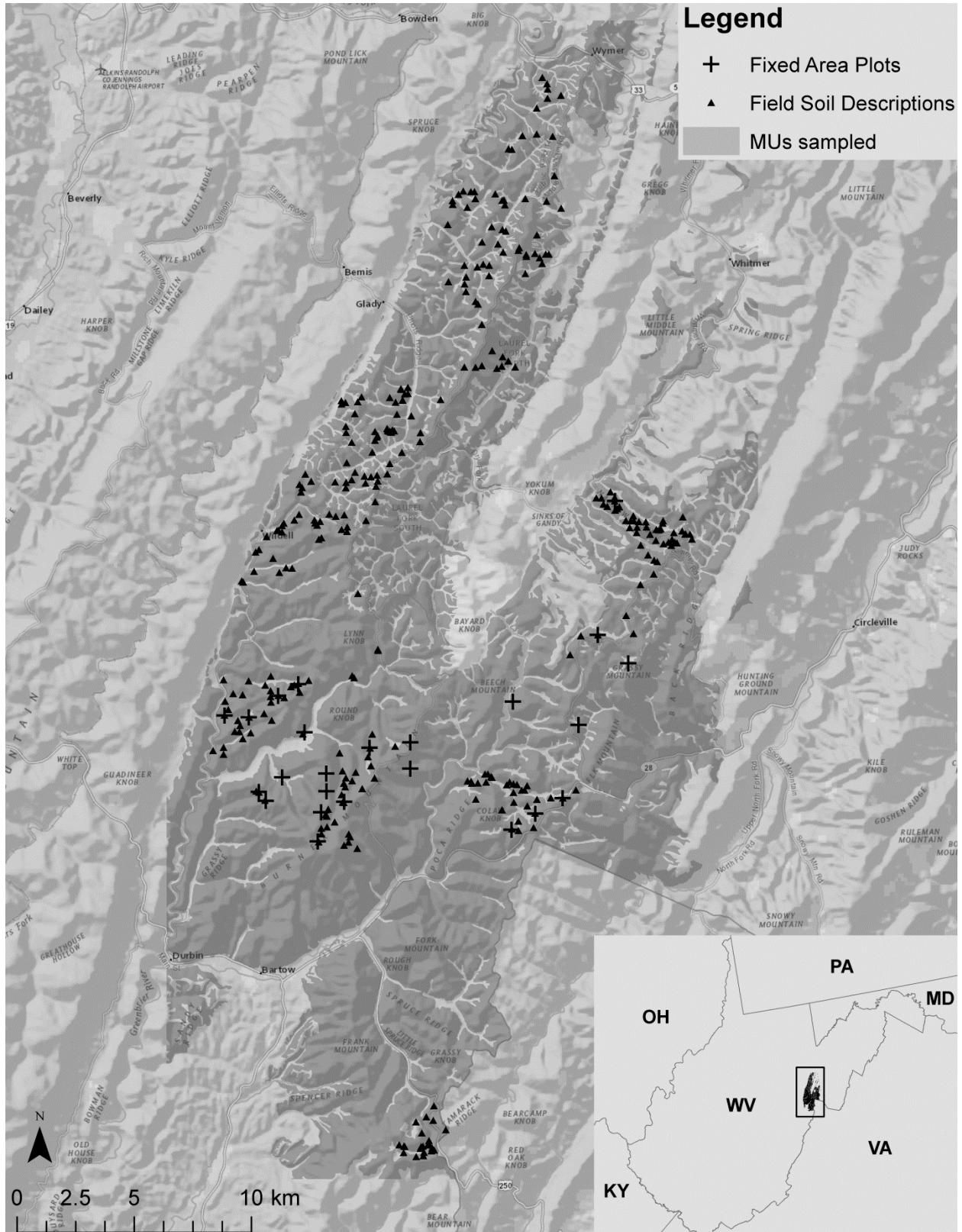


Figure 2. Study area and data collection locations overlaid on ArcGIS 10 National Geographic mapping baselayer of local features (roads, shaded relief, cities, landmarks).

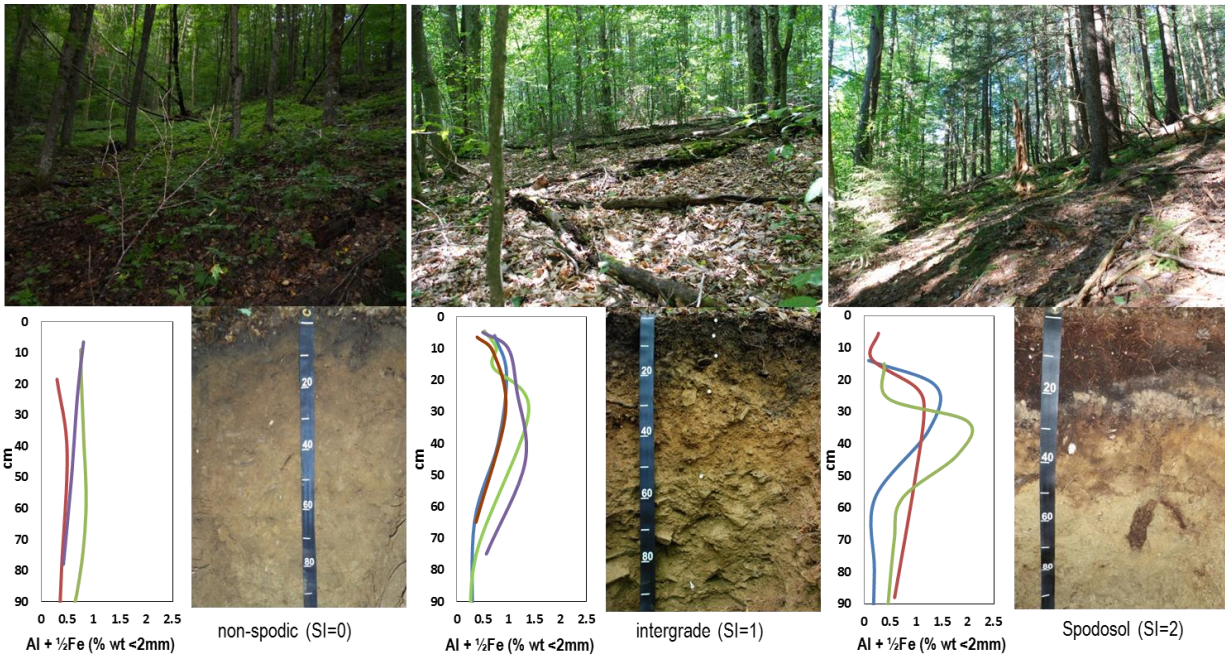


Figure 3. Examples of site conditions, soil profiles, and acid oxalate data of the non-spodic hardwood ecological site (SI=0), spodic integrademixed forest (SI=1), and spodic conifer forest (SI=2). Green line within graphs represent pictured soil profile. Pictures are of current vegetation at the pictured profile.

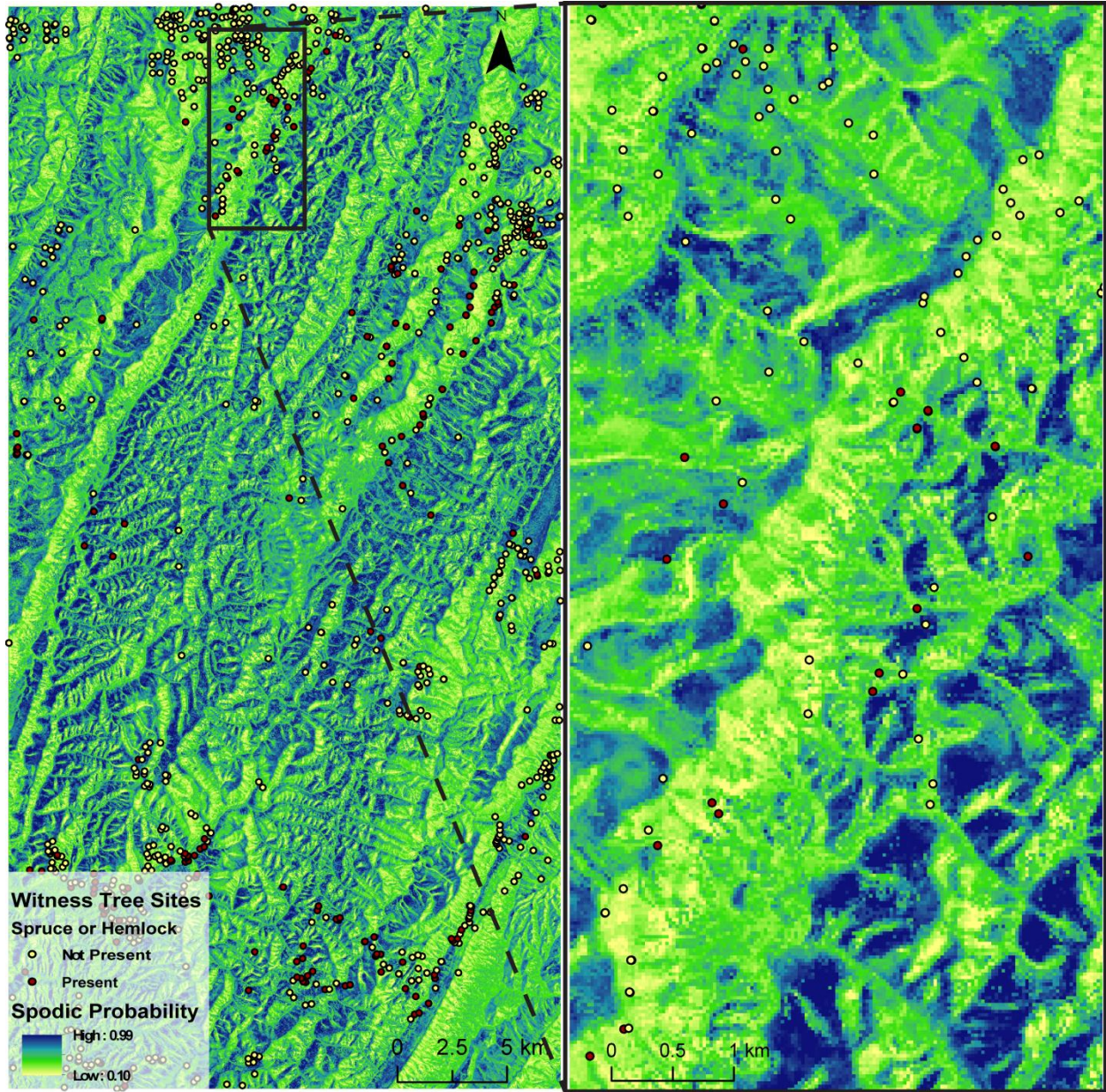


Figure 4. Spodic morphology probability map with witness tree points overlaid.

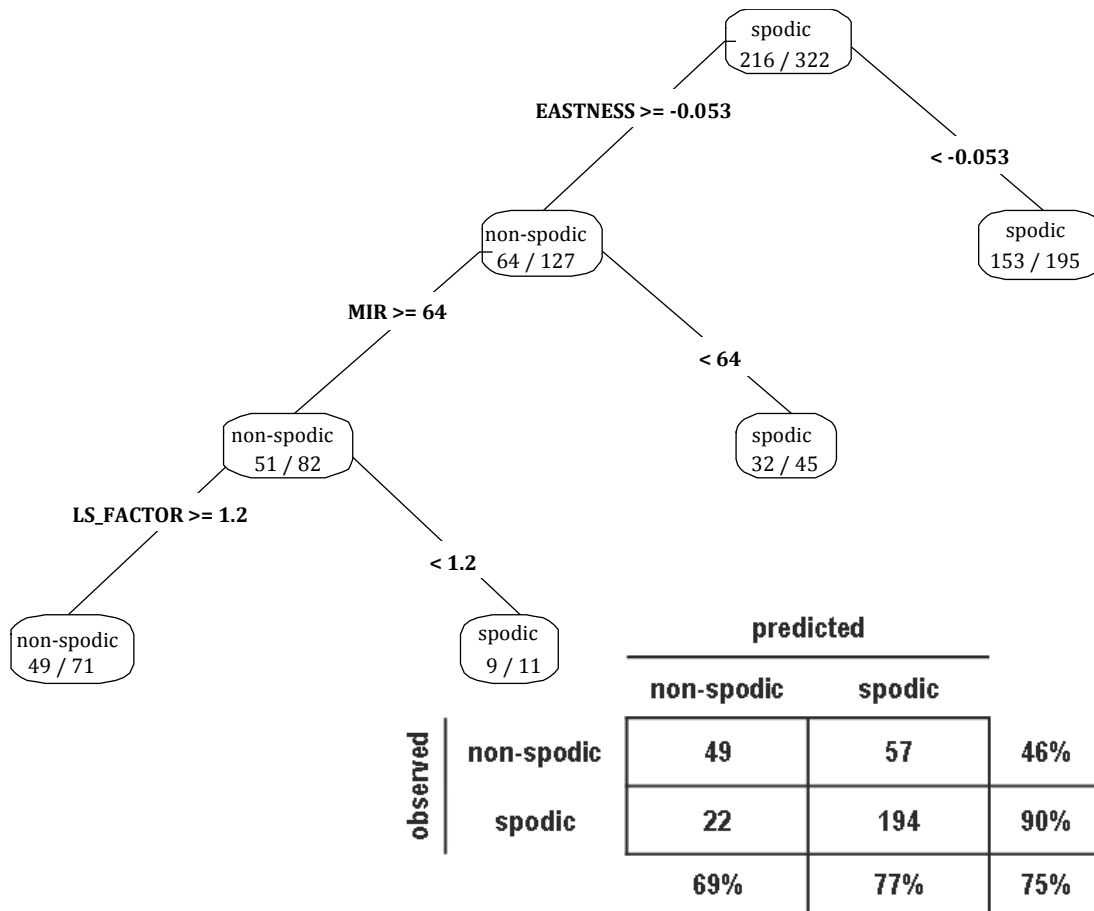


Figure 5. Classification tree showing how GIS variable splits can isolate more and less spodic groups of soil observations. Correct predictions over total node set size are shown under classification labels (e.g. spodic, 153/195 on upper right leaf). The confusion matrix of the fitted data is shown under the tree.

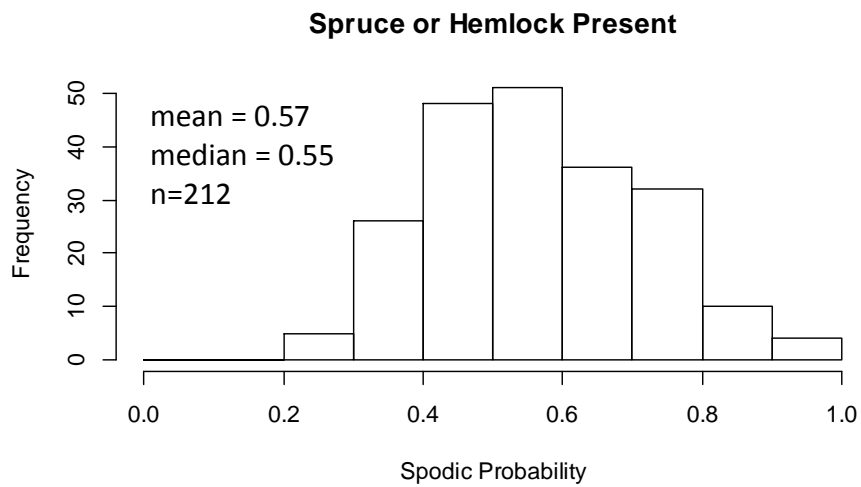
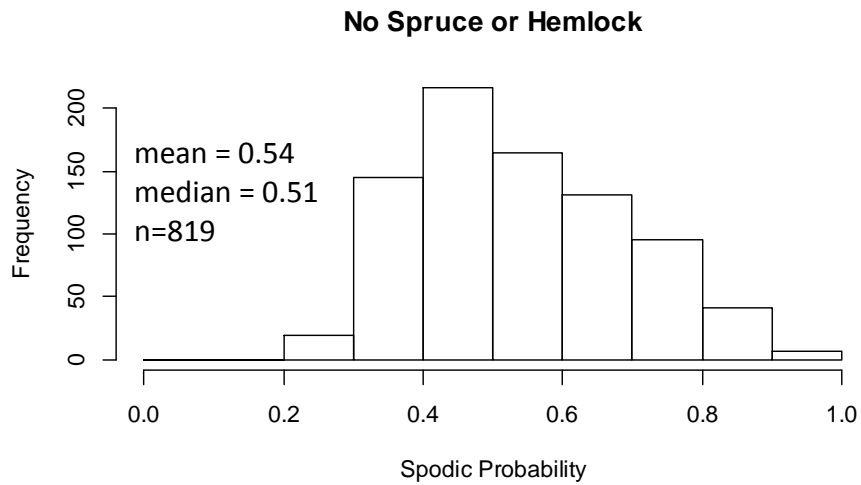


Figure 6. Spodic model probabilities at witness tree sites where no spruce or hemlock were recorded (top), and where spruce or hemlock were observed (bottom).

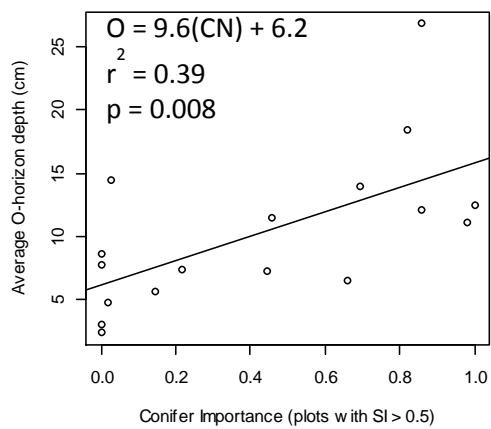
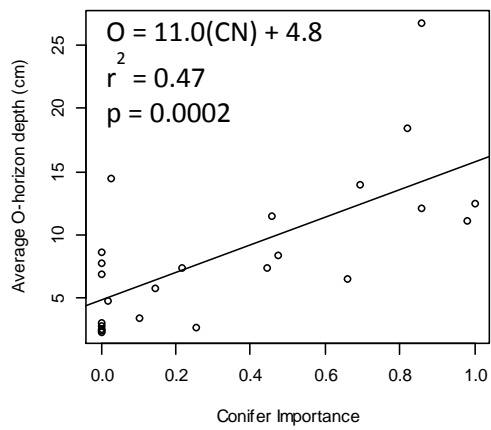
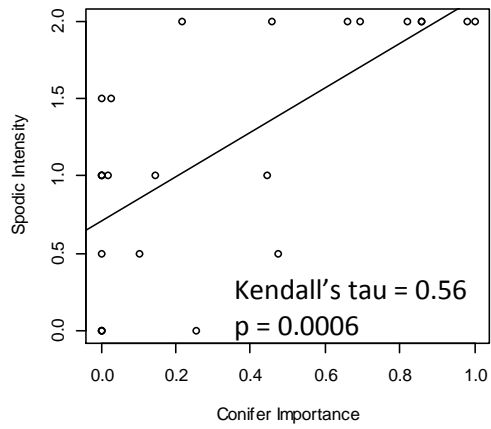


Figure 7. Relationships between conifer importance (CN) with spodic intensity (top) and O horizon

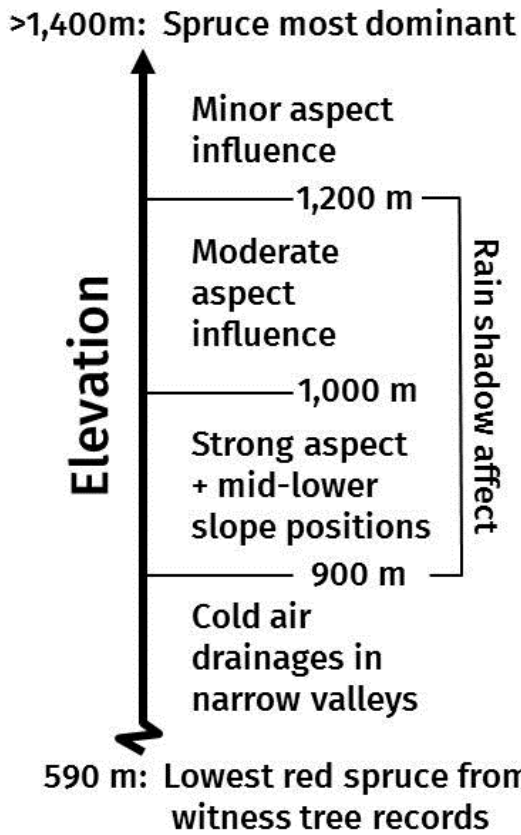


Figure 8. Conceptual diagram of how climatic and topographic controls of red spruce appear to change over elevation.

6 PEDOECOLOGICAL MODELING TO GUIDE FOREST RESTORATION USING ECOLOGICAL SITE DESCRIPTIONS

6.1 *Citation*

Nauman, T.W., J.T. Thompson, S.J. Teets, T.A. Dilliplane, J.W. Bell, S.J. Connolly, H.J. Liebermann, and K.M. Yoast. In revision. Pedoecological modeling to guide forest restoration using ecological site descriptions. *Soil Science Society of America Journal*.

6.2 *Abstract*

The U.S. Department of Agriculture (USDA)-Natural Resources Conservation Service (NRCS) uses ecological site descriptions (ESD) to help incorporate interactions between local soil, climate, flora, fauna, and humans into schema for land management decision-making. We incorporate ESD and digital soil mapping tools to (i) map areas in alternative states that can be targeted for restoration, and (ii) estimate potential forest floor C stock accumulation in the high-elevation forests of the Central Appalachians in West Virginia. This region was extensively disturbed by clear-cut harvests and related fires during the 1880's-1930's. We combined spodic soil property maps, recently linked to historic red spruce – eastern hemlock (*Picea rubens* – *Tsuga canadensis*) forest communities, with current forest inventories to provide guidance for restoration to a historic reference state. This allowed mapping of alternative hardwood states within areas of the Spodic shale uplands conifer forest ESD (SCF) along the regional conifer-hardwood transition of the Appalachian Mountains. Plots examined in these areas suggest that spruce-hemlock dominated stands in West Virginia converted to a hardwood state by historic disturbance have lost at least 10 centimeters of O horizon thickness, and possibly much more. Based on this conservative 10 cm estimate, we calculate that at least 3.74-6.62 Tg of C were lost from areas above 880 meters elevation in West Virginia due to historic disturbance of O horizons, and that much of these stocks and related ecosystem functions could potentially be

restored within 100 years under focused management, but more practical scenarios would likely require closer to 200 years.

Key words: red spruce, podzolization, O horizon, digital soil mapping, ecological sites, ecological site descriptions, soil organic carbon

6.3 *Introduction*

Soils are a dynamic interface between abiotic and biotic drivers and the Earth's crust. In soil science this has been conceptualized as a state factor model where the state or properties of a soil are a result of interactions between climate, organisms, relief, and parent material over time (Dokuchaev, 1899; Jenny, 1941). The state factor model evolved to an ecosystem level model where soils and organisms have some parallel drivers, but also interact strongly (Eq. 1, Amundson and Jenny, 1997; Jenny, 1961; Jenny, 1980).

$$l, s, v, a = f(L_0, P_x, t) \quad [\text{Eq. 1}]$$

The dependent factors in this case include ecosystem properties (l), soil properties (s), vegetation (v), and animals (a). The related state factors in an ecosystem based approach include the initial state (L_0) and external potentials (P_x), and time (t). Initial state L_0 includes the parent material (bedrock or substrate), initial relief, and water table. Climate and organisms are grouped as the P_x variable, which represent the primary energy sources that drive processes (Jenny, 1961). Amundson and Jenny (1991; 1997) have introduced these conceptual models into ecological sciences, with humans included in the factorial equation. Soils bear the imprint and help record the story of organisms—especially humans—and the climate as pedomemory signals in biogeochemical and physical properties that can be valuable in understanding the history of sites (Lin, 2011; Nauman et al., In Press; Phillips and Marion, 2004; Targulian and Goryachkin, 2004).

Within the context of the multi-factorial soil system, understanding the relationship between soils and associated ecosystems has been incorporated into different land management schemes. Ecological site descriptions (ESD) are a framework used by various U.S.

government agencies to help land managers employ sound science in taking appropriate management actions in the rangelands of the western United States (Bestelmeyer et al., 2011; Bestelmeyer et al., 2009; Briske et al., 2005; Caudle et al., 2013; Grazing Lands Technology, 2003; Herrick et al., 2006). Recently the USDA-NRCS, the government agency behind most ESD development, has put more emphasis into applying this framework in the eastern United States and has released a new handbook to help incorporate appropriate methods for inclusion of eastern forested systems (USDA-NRCS, 2014). The conceptual importance of ESD is in recognition of how soils both influence and are influenced by the productive potential of a site by documenting ecological states and transitions associated with different pressures on a site. There is also recognition of connections between groups of floral and faunal species and specific soil properties. Put more specifically by the USDA-NRCS (2014), an ecological site is “a distinctive kind of land based on recurring soil, landform, geological, and climate characteristics that differs from other kinds of land in its ability to produce distinctive kinds and amounts of vegetation and in its ability to respond similarly to management actions and natural disturbances.” This relates well to how a set of ecosystem factors (Eq. 1) driving soil-biota process pathways often result in specific soil morphologic expression that can provide an insightful narrative of a site’s history.

6.3.1 *Podzolization Pathway and Soil Organic Carbon*

In West Virginia, there has been debate over the extent of Spodosols and associated spruce-hemlock forest communities down into shale geologies due to contrasting reporting in soil mapping projects (Losche and Beverage, 1967; USDA-SCS and USDA-FS, 1982; Williams and Fridley, 1931). Spodosols are a result of soil development along the podzolization pathway (Lundström et al., 2000a; Lundström et al., 2000b; Sauer et al., 2007), which has been shown to often relate to forest species composition (Miles, 1985; Willis et al.,

1997). The podzolization pathway includes multiple evolutionary soil processes that promote aluminum, iron, and organic matter mobilization and translocation to deeper soil depths in acidic and permeable and usually (but not always) sandy parent materials. There also is often formation of thick surface O horizons (mor forest floor) at the soil surface, especially in more moist conifer systems (Hix and Barnes, 1984; Lietzke and McGuire, 1987; Lundström et al., 2000a). Ample soil solution leaching along with soluble organic acid inputs from the forest floor and actively mining ectomycorrhizal communities cause mineral weathering and the ultimate transport of aluminum, iron, and organic matter from near surface soil horizons (O, A, E) into subsurface (B) soil horizons (Blum et al., 2002; Giesler et al., 2000; Hoffland et al., 2004; Jongmans et al., 1997; Lundström et al., 2000b; Schöll et al., 2008; van Breemen et al., 2000).

Much of the organic C in Spodosols can be lost in 30-100 years just by converting cool, moist acidic conifer forest stands to differing species compositions (prairie or hardwood) that favor more decomposition (Barrett and Schaetzl, 1998; Hix and Barnes, 1984; Hole, 1975; Miles, 1985). This is most prominent in the forest floor O horizons, which get thinner in the conversion (Barrett and Schaetzl, 1998; Hix and Barnes, 1984; Miles, 1985). Studies have also shown that conversion from mesic hardwood forests (e.g. *Quercus spp.*, *Betula spp.*, and *Fagus spp.*) to Norway spruce (*Picea abies*) and/or scots pine (*Pinus sylvestris*) causes O horizon buildup and increased podzolization (Herbaults and Buyl, 1981; Miles, 1985; Ranger and Nys, 1994; Sohet et al., 1988). Forest common garden plot studies that isolate tree species on individual plots have also shown a gradient among species that promote base cation activity and heterotrophic organic matter decomposition (e.g., *Acer spp.* and *Tilia spp.*), and those that favor acidic Al and Fe activity (e.g. *Pinus spp.* and *Larix decidua*) which were associated with less decomposition of soil organic matter (Hobbie et al., 2007). In these garden plots, higher tree

litter calcium content appeared to control pH, decomposition, and stimulate earthworm activity which resulted in less forest floor mass (Hobbie et al., 2006; Reich et al., 2005). Hobbie et al., (2006) also showed that spruce and fir species were associated with lower mean annual soil temperatures that were associated with less litter decomposition. Although influential general differences in litter chemistry were seen between angiosperms (basic) and gymnosperms (acidic), these studies showed that there is significant variation within these groups of species. Another recent common garden study in New York showed a similar influence of worms under northern red oak (*Quercus rubra.*) and sugar maple (*Acer saccharum*), but not under Norway spruce which formed deeper forest floor (Melvin and Goodale, 2013). Although Ca^{2+} was similar under all three species, pH was lower under the spruce, suggesting that base cation activity might not be the only factor to examine. Overall, these studies tell a story where heterotrophic forest litter decomposition and O horizon accumulation are intricately linked with each tree species present at a site.

Mycorrhizal partnerships are another important consideration in understanding C and nutrient cycling in soils especially in regards to systematic C balance (Högberg and Read, 2006). Intensive ectomycorrhizal (ECM) colonization of E horizons appear to be a significant nutrient acquisition adaptation strategy of conifer systems in acidic Al-dominated soil exchange complexes of Spodosols that might otherwise be toxic to tree roots (Blum et al., 2002; Giesler et al., 2000; Hoffland et al., 2004; Högberg and Read, 2006; Jongmans et al., 1997; Lundström et al., 2000b; van Breemen et al., 2000). Averrill et al. (2014) demonstrated that ECM and ericoid mycorrhizae (ERM) promote SOC accumulation on a global scale, and concluded that this is likely because they can effectively compete for nitrogen in organic matter with saprotrophic bacteria. The buildup of forest floor mycorrhizal fine root hypha and associated host C allocation

now appear to be as important as more established C-cycling models of heterotrophic litter and fine root decomposition and respiration (Högberg and Read, 2006). The development of deep O horizons under acidic ECM-associated conifer must, by definition, mean that heterotrophic communities are either suppressed or very inefficient in cycling C in these systems which is also consistent with the results of garden plot studies (Hobbie et al., 2007; Hobbie et al., 2006; Reich et al., 2005). Species specificity to ECM and ERM should also make it possible to use forest composition as further predictors of SOC stocks (Averill et al., 2014; Binkley and Fisher, 2012; Brundrett, 2009).

6.3.2 *Forest History and Ecological Change in the Allegheny Mountains of West Virginia*

In West Virginia, historical accounts indicate that the current extent (~20,000 ha) of red spruce forest communities is drastically reduced from its range before extensive logging and fires between 1860 and 1920 (~200,000 ha) (Hopkins, 1899; Nowacki and Wendt, 2010; Pauley, 2008; Pielke, 1981). Maximum entropy modeling efforts have similarly shown that the suitable habitat for red spruce in West Virginia is much more extensive than current distributions (Beane et al., 2013; Byers et al., 2010; Nowacki and Wendt, 2010). These studies, along with broader analysis of red spruce habitat (Nowacki et al., 2010) show temperature and precipitation as the main controls on extent. However, recent work in compiling and analyzing witness tree databases from the Monongahela National Forest (MNF) indicate a lower minimum elevation historically (lowest recorded red spruce at 509 meters) than previous models, and more specificity controls linked to topographic position preferences in respect to slope aspect and relative slope positions (Thomas-Van Gundy et al., 2012). Nauman et al., (In Press) found that the spatial distribution of spodic soil properties is associated with the occurrence of red spruce and eastern hemlock witness trees (recorded from 1752-1899), and follow similar topographic

controls to analysis by Thomas-Van Gundy et al. (2012). These results indicate that spodic soil properties are much more widespread in West Virginia than previously thought, and concluded that this also likely represents a much greater extent of conifer historically (Nauman et al., In Press; Thomas-Van Gundy et al., 2012). The subtleties in the pre-disturbance spatial distribution of red spruce might indicate historic affinity for topographically driven cool and moist microclimates that included the highest ridgelines, cooler slope aspects not in rain shadows, and narrow valleys that foster cold air drainage and foggy inversions (Nauman et al., In Press; Thomas-Van Gundy et al., 2012).

Human disturbance, pollution, and climate change are thought to have contracted red spruce populations, but are somewhat hard causations to distinguish (Hamburg and Cogbill, 1988). A warming climate is pushing cooler conifer ecosystem species like red spruce higher in elevation and higher in latitude, putting large pools of soil organic C at risk for further atmospheric release (Lal, 2005). Acid deposition damage to red spruce health has also been studied (Adams and Eagar, 1992; Hornbeck and Smith, 1985; Johnson, 1983), but might be hard to separate from the impact of a changing climate and overall warming (Hamburg and Cogbill, 1988) as well as extensive historic clear cutting and associated fires and pest outbreaks (Clarkson, 1964; Hopkins, 1899; Pauley, 2008; Stephenson and Clovis, 1983). Indeed, red spruce is projected by different climate change scenarios to disappear from West Virginia by the end of the century (Butler et al., 2014; Byers et al., 2010). However, there are signs that red spruce is recovering from historic disturbance and could be further restored despite climate change (Nowacki et al., 2010; Rentch et al., 2007; Rentch et al., 2010; Rollins et al., 2010). At this time, its future remains uncertain.

Red spruce is both one of the most acidophilic conifer species and an ECM associate, suggesting that it should promote SOC accumulation. Red spruce produces quite nutrient-poor litter (especially in Ca^{2+}) relative to other North American tree species (compare from: Berg and McClaugherty, 2008; Côté and Fyles, 1994; Friedland et al., 1988; Rustad and Fernandez, 1998). So we hypothesized that red spruce should promote podzolization and O horizon accumulation based on findings from previously discussed forest composition effects on soils (Herbaults and Buyl, 1981; Lundström et al., 2000a; Miles, 1985; Ranger and Nys, 1994; Sauer et al., 2007; Sohet et al., 1988). We would also expect that the extensive areas of historic red spruce converted to non-ECM/ERM associated and more basic cation promoting species like red maple (*Acer rubrum*) and black cherry (*Prunus serotina*) in West Virginia have probably lost organic material from O horizons and B horizons (Averill et al., 2014; Barrett and Schaetzl, 1998; Brundrett, 2009; Comas and Eissenstat, 2009; Hix and Barnes, 1984; Miles, 1985). This was exacerbated by the large scale fires documented in West Virginia after areas were clear-cut (Hopkins, 1899; Pauley, 2008). However, the Fe and Al sesquioxide accumulations in the subsurface soil are still observable as these are more stable and persistent in soils (Barrett and Schaetzl, 1998; Lundström et al., 2000b; Nauman et al., In Press; Parfitt, 2009).

Recent work related to ESD development in the MNF for the purpose of linking soil management strategies to historic site potential vegetation communities has suggested that spodic soil morphology in the MNF was linked to past red spruce and commonly associated eastern hemlock distribution (Nauman et al., In Press; Nowacki and Wendt, 2010; Teets, 2013). Nauman et al. (In Press) were able to map this using spodic soil properties to help delineate the Spodic shale uplands conifer forest (SCF) ecological site. They also showed that there was a positive linear relationship between the current relative conifer composition and the thickness of

O horizons. We hypothesize that the areas of northern hardwood on the SCF ecological site are alternative states that were converted from a spruce-hemlock dominated state by the railroad era timber harvest disturbance. We think this disturbance and conversion has resulted in large losses of O horizon material and associated C stocks. We aim to build on the analysis of Nauman et al. (In Press) to (i) connect the relationship between O horizon depth and forest composition to ecological site state and transition models, (ii) spatialize the SCF state and transition model using detailed current forest inventories, and (iii) estimate the potential C stocks in West Virginia that could be accumulated by restoring areas in alternative SCF ecological states to a conifer-dominated state according to the prescribed SCF ecological site description (Teets, 2013).

6.4 *Materials and Methods*

6.4.1 *Study Area*

The study extended across the higher elevations of the Chemung and Hampshire geologic formations in parts of the MNF (Fig. 1). These are acid geologies primarily composed of shale and siltstone parent materials with minor inclusions of sandstone (WVGES, 1968). The area is a moist udic to perudic soil moisture regime, with annual precipitation ranging from 1118-1524 mm (44-60 inches; NOAA-NCDC, 2014), which is likely controlled by elevation and topographic rain shadow effects. Mean annual temperature ranges from 6 to 8.3°C (NOAA-NCDC, 2014), which we think follows elevation, slope aspect, and cold air drainage patterns. The elevations of sites examined ranged from 880-1320 m, which spans the approximate elevation boundary (~1100 m) between the mesic and frigid soil temperature regimes cited by other regional podzol studies (Lietzke and McGuire, 1987; Stanley and Ciolkosz, 1981). The topography in the area includes flat narrow ridgetops, steep mountainsides, occasional rock

outcrops, and deep and narrow river valleys. Within slopes there are benches, hollows, and nose slopes along with cradle-knoll micro-relief that mitigate how water, energy, and materials are distributed in the soil system.

Vegetation observed in these areas consists of northern hardwood and spruce-hemlock dominated stands as well as mixed composition stands where hardwood and spruce-hemlock co-dominate. Common tree species observed in the study area include red maple, sugar maple, mountain maple (*Acer spicatum*), striped maple (*Acer pensylvanicum*), red spruce, eastern hemlock, yellow birch (*Betula alleghaniensis*), sweet birch (*Betula lenta*), American basswood (*Tilia americana*), white ash (*Fraxinus americana*), northern red oak, black cherry, American beech (*Fagus Grandifolia*), mountain magnolia (*Magnolia fraseri*), and cucumber magnolia (*Magnolia acuminata*). Commonly seen shrubs include mountain holly (*Ilex montana*), mountain laurel (*Kalmia latifolia*), and rhododendron (*Rhododendron spp.*), as well as shrubby root sprouts as a result of the beech bark disease complex (Shigo, 1972). Common herbaceous and ground cover species include New York fern (*Thelypteris noveboracensis*) intermediate woodfern (*Dryopteris intermedia*), hypnum moss (*Hypnum imponens*), liverwort (*Bazzania trilobata*), three *Lycopodium* species, *Viola* spp., and three *Carex* species.

6.4.2 Data Collection and Analysis

Three types of soils data were collected as part of this research: (i) extensive point observations of soil morphological properties (n = 322), (ii) detailed pedon descriptions at selected sites with associated comprehensive laboratory characterization of soil physical and chemical properties (n = 7), and (iii) fixed-area forest vegetation plots with detailed pedon descriptions and limited soil laboratory characterization data (n = 24). Data collected at all visited locations included detailed field descriptions of the soil morphology at hand-excavated

pits with a focus on spodic morphology expression (i.e. spodic intensity; Table 1). Spodic intensity (SI) was determined on a zero (non-spodic) to two (well-expressed Spodosol) scale by 0.5 increments based on colors, horizon characteristics, and smeariness observations (See Table 1) typical of ‘spodic soil materials’ in US *Soil Taxonomy* (Schoeneberger et al., 2002; Soil Survey Staff, 1999) . Data were collected by a variety of local soil scientists associated mostly with the USDA-NRCS, USDA-Forest Service (FS), and West Virginia University (WVU). Soil descriptions were made consistent with U.S. national soil survey standards (Schoeneberger et al., 2002). Site locations were selected to evaluate soils derived from Devonian shale parent materials on stable upland landscape positions for the purpose of soil survey update and preliminary ESD reconnaissance. Soil map units sampled were associated with three common soil series: Mandy (Loamy-skeletal, mixed, active, frigid Spodic Dystrudepts), Berks (Loamy-skeletal, mixed, active, mesic Typic Dystrudepts), and Dekalb (Loamy-skeletal, siliceous, active, mesic Typic Dystrudepts). Overstory and understory vegetation species lists were also noted at every location. Additional details about the sampling design and laboratory analysis details can be found in Nauman et al. (In Press).

Given the fire history in the study area, efforts were made to search for charcoal within the exposed soil profile and the four satellite O horizon observation points. When charcoal was found, the depth was noted and a representative sample was collected. In the laboratory, the size and shape of the charcoal pieces were recorded before sending them for ^{14}C analysis at the Northern Institute of Applied Climate Science (Michigan Tech. University, Houghton, MI) following the methods of Vogel et al. (1987), with a $\delta^{13}\text{C}$ correction applied to account for isotopic fractionation (Stuiver and Polach, 1977).

Fixed area forest plots employed for data collection were 20 by 20 meters in shape and oriented with the slope aspect. Diameter at breast height (dbh) was measured on all trees greater than 7 cm dbh. From measured dbh values and species tallies, importance (IMP) values (Eq. 2; e.g., Adams et al., 2010) were calculated for red spruce and eastern hemlock.

$$\text{IMP} = 0.5 * ((\text{species basal area} / \text{plot tree basal area}) + (\text{species count} / \text{plot tree count}))$$

[Eq. 2]

At plots, O horizon thicknesses were observed at the soil profile as well as at the center of each plot quadrant (n = 5 per plot). We added the importance of red spruce to that of hemlock to get a ‘conifer importance’ (CNIMP), which Nauman et al., (in Press) showed was likely the primary long term ‘organism’ soil formation driver for podzolization and forest floor thickness in these areas. We also summarized forest types at plots as ‘conifer’ (CNIMP > 0.75), ‘mixed’ (CNIMP 0.25-0.75), and hardwood (CNIMP < 0.25) to help in plotting data.

Studies indicate that current conifer communities are much reduced compared to historic pre-disturbance conditions (e.g., Thomas-Van Gundy et al., 2012), and that current conifer relationship with SI is not as consistent as that with O horizon depth (Nauman et al., In Press). We contend that the Al and Fe accumulations reflected in SI visual cues and smeariness observations are longer lived signs of past vegetation than organic C and O horizons in similar soils (Barrett and Schaetzl, 1998; Hix and Barnes, 1984; Lundström et al., 2000b; Parfitt, 2009). Therefore, we suspected that O horizons have adjusted much more quickly to forest composition changes, and thus would maintain closer correspondence to the current forest state.

6.4.3 *Pedoecological Mapping and Restoration Carbon Sequestration Estimates*

O horizon development in the MNF represents a potentially large pool of C sequestration. Based on our hypotheses that (i) disturbance-based forest conversion to hardwood

and, conversely conifer restoration, will result in changes to O horizon thickness, and (ii) the ability to map those areas using digital mapping methods, we wanted to make estimates of O horizon C stocks that could be restored by returning Spodic areas to the hypothesized historic reference conifer forest state. This could be accomplished by managing hardwood sites with spruce in the understory with overhead spruce release (Rentch et al., 2007; Rentch et al., 2010) or underplanting with spruce and later release if no spruce recruitment is currently present. Figure 2 shows the state and transition model created for the Spodic shale upland conifer forest (SCF) ESD (Teets, 2013). We aimed to estimate the areas in the two logged states (boxes 2 and 3 in Fig.2) and estimate how much O horizon C would be added to the sites when restored to the reference state (box 1 in Fig. 2). To do this we combine analysis of the field point observations, detailed pedon data, plot data, and a forest inventory map (Byers et al., 2013) to map ecological states of the SCF and determine how much O horizon carbon can be restored by managing back to a reference state (Fig. 3).

Areas of SCF were estimated by choosing a spodic probability threshold value of 0.7 from the spatial model in Nauman et al. (In Press), which predicted Spodosol distribution with a reasonable degree of confidence (62-72% user accuracy from transect validation points and forest plot data). Then areas within the SCF currently in a logged hardwood state were estimated in three ways. First, a current forest inventory (Byers et al., 2013) was overlaid with areas in spodic probabilities above the threshold to determine proportion of areas in a hardwood state or mixed conifer-hardwood. We present a map of the overlaid ecological states from the forest inventory and the SCF map as an example of a pedoecological map that provides field scale management prescription units for land managers.

Secondly, field observation sites from the 1/3 withheld validation set in Nauman et al. (In Press) were compared with the validation model spodic probabilities above the threshold to see what proportion of sites had no conifer in the forest overstory species list (i.e., logged state proportion). Thirdly, the fixed-area forest vegetation plots determined to belong to the SCF in Fig. 2 were analyzed to see what proportion fell into alternative logged states and areas that were in transition to the reference state (i.e., mixed composition). The proportions found in these three methods were then multiplied by the area above the spodic probability threshold in the map units sampled by field description locations to determine potential restorable areas in those map units. We also scaled the proportions in the study area out to all areas in WV at elevations above the minimum elevation of our study (880 m) to make extrapolative estimates regarding how much C accumulation might result from restoration of spruce in similar areas in the state. This extrapolative estimate is heavily weighted on assumptions of relationships and proportions being consistent outside of our study area, but we feel it is a conservative estimate based on even thicker O horizons being associated with conifer states on the higher ridgelines in WV (Nauman, , unpublished data, 2013), and also because Byers et al. (2013) show overall conifer composition proportions consistent with our study area across their entire spatial estimate of historic WV red spruce extent. Byers et al. (2013) delineate a more conservative total area of historic red spruce range (532,116 ha) than our 880 meter elevation extrapolation (645,438 ha). However, when compared to the lower elevation observations in witness tree records (Thomas-Van Gundy et al., 2012), we think both these estimates are probably smaller than the true red spruce range.

Once potential restoration areas were identified, we used O horizon laboratory organic C (Method 4H2a; Soil Survey Staff, 2004) and frame bulk density (Method 3B5a; Soil Survey Staff, 2004) estimates of the seven representative pedons sampled for laboratory analysis

for calculating potential new C stocks. Total organic C estimates were averaged for Oi, Oe, and Oa horizons from soil profiles analyzed at the Kellogg Soil Survey Laboratory (Table 3). Average proportions of Oi, Oe, and Oa in O horizons in the three reference state Spodosol profiles sampled were assumed as the proportions in new O horizon formation (Table 2). Total weighted average volumetric C (grams organic C per cubic cm) was used to estimate C additions from a given accrual of O horizon thickness over a certain area. Potential accrual of new O horizon depth was based on the slope of the linear relationship in forest plots between conifer importance and average plot O horizon depth reported by Nauman et al. (In Press).

We set restoration targets to the relative conifer (spruce + hemlock) basal area of 84.4% calculated by averaging the high and low listings in the ESD reference community basal area descriptions (Teets, 2013). Because relative basal area and conifer importance in our data were essentially the same ($R^2 = 0.99$, slope = 0.96) with the best agreement above 80%, we translated the target to a conifer importance (CNIMP) of 84.4% because CNIMP had a better overall relationship with O horizons than relative basal area, although both were significantly correlated to O horizon thickness. We assumed that to reach this target, hardwood sites would need to increase in CNIMP by 76.3% because these sites averaged 8.1% in our corresponding forest plots. Because mixed transition plots averaged 53.1%, we used a 31.3% CNIMP increase for the target. Total areas of logged states and mixed transitions were multiplied by the estimated O horizon depth accruals for those states based on restoration targets and the slope of the O horizon-CNIMP relationship to get a total O horizon accumulation volume. The volume was then multiplied by the total weighted average of C volumetric density of O horizons (0.0572 g/cm³; Table 3) to get a total mass of C predicted to be added to O horizons by meeting those restoration targets. The Oa/A horizons encountered in two laboratory profiles were assumed to be 66% Oa,

and one A/Oa horizon recorded was assumed to be 33% Oa for calculating O horizon depth proportions, and bulk densities were scaled down by those factors as well. To adjust Oa C percentages in these same horizons, which are higher than A horizons in general, the C values were multiplied by 1.33 in Oa/A and 1.66 in A/Oa, which in all cases produced C percentages slightly lower (conservative estimate) than the one uniquely measured Oa horizon C percentage of 47.3%.

6.5 *Results*

6.5.1 *Plot Data*

Plot data supported the choice of the 0.7 probability threshold as a basis for inclusion into the SCF ecological site with the vast majority of Spodosols falling above that value (Fig. 4). The plot data shows a positive trend between observed SI and the predicted spodic property probability. Most plots currently under conifer dominated and mixed hardwood-conifer also fall into the SCF. A few high outliers of SI 0.5, 1.0, and 1.5 fall into the SCF, and one outlier in the SI 2.0 fell out of the SCF. However, the laboratory data suggests that almost all of the field soil descriptions described as an $SI \geq 1.5$ classify as Spodosols, and up to half of the profiles with an SI of 1.0 would barely classify as Spodosols. This seems to suggest that the high outliers are still mostly consistent with the SCF concept. The low outlier plot was examined and has an incorrect slope aspect value attributed to it by the GIS model used in Nauman et al. (In Press) when compared with the field observed aspect. Slope aspect was heavily weighted in the modeling and likely caused an errant probability to be attributed.

6.5.2 *Conifer Importance and O horizons*

Conifer importance at SCF forest plots show positive correlation with the thickness of O horizons (Fig. 5). O horizon response to conifer importance appears to be 0.96 cm of O horizon thickness increase per 10% of conifer importance expansion on average. It is important to note that conifer importance does not include any calculation of site productivity or herbaceous composition; it is solely based on the relative composition of red spruce and hemlock versus other tree species with dbh values greater than seven cm. Therefore, this association is somewhat independent of site productivity and a range of other potential site variability. This relationship was chosen for restoration C sequestration calculations as it was based solely on sites with SI values of 1-2, where we think the reference spruce-hemlock dominated stands described by the SCF ecological site would have been most likely present historically based on the work of Nauman et al. (In Press).

What the O horizon relationship in Figure 5 does not address is the timeframe necessary for O horizon to adjust to forest composition changes. Fortunately, two of the forest plots we sampled were dense, even-aged red spruce stands (CNIMP = 86.3% and 100%) with charcoal evidence of burning after historic harvest. Breast height tree cores of the three biggest spruce at both sites averaged 65 and 60 growth rings with a range of 52 to 70, suggesting stand ages of roughly 60 to 80 years old. Abundant subangular charcoal was found at the interface between the O and E horizons at both sites, indicating that the O horizon had likely burned off before this cohort was established, which is consistent with historic post-logging accounts of long lasting fires in the area (Pauley, 2008). Radiocarbon dates of the charcoal at these sites were 205 ± 25 years and 90 ± 30 years. These dates support modern fires that we postulate followed post-railroad disturbance. Both sites had very similar O horizon thickness averages of 12.1 and

12.5 cm. This contrasts with O horizons at sites in the SCF that are currently under exclusively hardwood cover with similar or older ages (5.4 cm average).

Interestingly, within conifer dominated plots older than 100 years in average tree core counts ($n = 3$ per plot), the average O horizon thickness increases to 18.8 cm, suggesting that over time O horizons may get even thicker. At these older plots, we observed only one site with no charcoal evidence of past fire, and the average O horizon depth there was 26.8 cm with a maximum of 37 cm. This might be suggestive of a climax condition; however, relatively undisturbed sites are hard to find due to the vast extent of historic disturbance and thus it is difficult to establish a representative sample.

6.5.3 *Pedoecological Mapping and O horizon Carbon Accumulation*

From the spodic probability threshold of 0.7 chosen to represent the SCF ecological site, a map was made to determine areas that could be potentially restored from logged alternative states to the reference spruce-hemlock conifer state (Fig. 6). The map created delineated 31% of the study area map units that were originally sampled as SCF. The red spruce cover map (Byers et al., 2013) was intersected with these areas to determine that it was 16.5% conifer, 73.6% hardwood (or small patches of pasture), and 9.9% in mixed conifer-hardwood. Model validation sites from the withheld 1/3 of data points from Nauman et al. (In Press) with greater than 70% spodic probability were analyzed to find that 53.3% of sites had no conifer species in the overstory, which were assumed to be in a hardwood state. Mixed states were not decipherable at these sites due to the qualitative species list observations. Hardwood sites made up 36.4% of SCF forest plots, mixed sites made up 18.25% of plots, with the remaining plots being conifer. The percentages were multiplied by the total area of the study area map units to get estimates of hardwood state and mixed transition areal extents (Table 3).

Potential C sequestration estimates based on restoring to reference state conifer importance levels combined areal estimates with new expected O horizon accumulation (Table 3). We estimated that between 0.29 and 0.52 Tg of C would accumulate in the sampled study area soil map units. This approximation was extrapolated to all areas in WV with an elevation higher than 880 m (the minimum elevation of the study observations) by calculating the ratio of the greater WV area to the area of the study map units and multiplying our study area estimates by that ratio (12.73). Totals for the >880 meter area ranged from 3.74 to 6.62 Tg (Table 3). These estimates are based on the calculation that, on average, areas in hardwood states would add 7.32 cm of O horizon material and mixed transition sites would add 3.0 cm.

6.6 Discussion

Our results demonstrate that understanding how the ecological soil factorial (Eq. 1) drives soil processes can inform our understanding of the ecological history and plausible management responses of a site. We demonstrate how ESD can provide the framework for interpreting these links between site ecology, soil genesis and pedology in eastern U.S. forest systems, just as it has in western U.S. rangelands (Bestelmeyer et al., 2011; Bestelmeyer et al., 2009; Briske et al., 2005; Caudle et al., 2013; Herrick et al., 2006; NRCS, 2014; Teets, 2013). In this case, we build on the link between spodic morphology and historic reference spruce-hemlock communities (Nauman et al., In Press), to show how O horizons have likely changed since railroad era timber harvest related disturbance. The industrial timber harvest and related fire, and resulting forest composition changes, probably caused large losses of soil C stocks in the forest floor, which have somewhat returned in areas where spruce and hemlock have recolonized. However, results also seem to indicate more potential for red spruce restoration which would add potential for accumulation of large amounts of O horizon (and thus C).

We acknowledge that considerable assumptions are made by summarizing all this data into one average across our study area and the state. Further research on using ESD to create these types of restoration potential estimates should address propagating errors through these calculations. This would include errors from the O-horizon thickness regression, O-horizon laboratory data averages, bulk density measurements, and spatial data error (forest inventory and spodic probability surface).

6.6.1 *Timing of O horizon Accumulation*

From the charcoal data at the two even aged red spruce sites, we feel that the most plausible conclusion is that in the 65-80 years of development these stands, a large portion of the ~12 cm of O horizon material has accumulated on top of the charcoal. This seems to match well with the conifer importance relationships with O horizon depth. This would imply that restoration of red spruce at these sites is associated with significant O horizon buildup within a century based on the differing O horizon depths at exclusively hardwood stands (5.4 cm average). This general timeframe is similar to that observed by Schaetzl in O-horizon buildup after fire in northern hardwoods (1994). We acknowledge that there are alternative interpretations of these results, and are uncertain as to how spruce were able to regenerate so dominantly after what appeared to have been an intense fire based on the nearly continuous layer of charcoal found at the O-E horizon interfaces at these plots. Our first impression of these sites was that they were planted, but no records of red spruce plantations exist in the area during that time period to our knowledge. It should also be noted that this is a quicker timeframe than might be expected for successional regeneration of spruce in other many other local areas where they are in the understory beneath hardwood species of mainly red maple, yellow birch, sweet birch, and black cherry. Only very intense, and somewhat unreasonable management actions could set

many areas on a direct course for a similar monoculture and even-aged red spruce stand structure, which would probably not be very desirable for other ecological reasons. More reasonable timeframes of 200+ years are likely for release based restoration as presented in Rentch et al. (2010). However, nearby areas like Cheat Mountain were observed to have thick monoculture spruce thickets recruiting in many areas that will likely produce similar stands, and thus similar forest floor accumulation rates and should be considered in C balance projections for those areas.

6.6.2 Implications of Red Spruce Restoration for Wildlife and Climate Change

Restoring hardwood areas of the SCF ecological site to the reference conifer state will potentially produce significant habitat for rare species in addition to significant C sequestration benefits. The endangered Cheat Mountain salamander (*Plethodon nettingi*) has been associated with red spruce forest communities in parts of our spodic probability model footprint and could benefit from restoration (Dillard et al., 2008a; Dillard et al., 2008b; Pauley, 2008). The formerly endangered Virginia northern flying squirrel (*Glaucomys sabrinus fuscus*) has also been linked to forests with influential red spruce components (Menzel et al., 2004; Menzel et al., 2006a; Menzel et al., 2006b; Odom et al., 2001) and would likely benefit from restoration efforts.

Potential C sequestration calculations associated with ESD restoration scenarios (Table 3) for just the study area map units sampled represent the C equivalent of combusting 4.4 million barrels of oil according to the EPA C equivalents calculator (EPA, 2014). This amounts to about 23% of the 18.89 million barrels of oil used in US in one day (<http://www.eia.gov/tools/faqs/faq.cfm?id=33&t=6>). When this potential is scaled up to all areas above an elevation of 880 meters in WV, this amount increases almost 13-fold to 56.4 million

barrels of oil. We think this represents a conservative estimate of potential C fixation in O horizons because initial data from Cheat Mountain, a higher and wetter area of WV, indicates that conifer composition influence might have twice the impact on O horizon accumulation (i.e., ~2.1 cm O horizon accumulation per 10% increase in conifer importance, $r^2 = 70$, $p=0.0001$, Nauman, unpublished data, 2013). Earlier work on the spruce-hardwood ecotone in Vermont also showed a correspondence between more acidic soils with thicker forest floors and red spruce dominated areas, but didn't report as much specificity between spruce and spodic properties (Siccama, 1974; Young, 1934). However, modern studies must account for the possibility that the vast harvest disturbance of forests associated with European colonization has favored hardwood incursion into formerly conifer influenced areas (Nowacki et al., 2010; Pielke, 1981) that might be reflected in spodic soils currently under hardwood cover.

Our estimates also do not include mineral subsurface C storage, which is significant in Spodosols and can respond quickly to disturbance (Barrett and Schaetzl, 1998; Hix and Barnes, 1984; Hole, 1975; Hole, 1976). Data presented here does not include biomass estimates of carbon stock which may vary with composition and productivity of sites. Further research should be done to fully quantify how all these carbon pools might change in these restoration scenarios within the warmer and wetter climate projections expected (Iverson et al., 2008). Although our C sequestration estimates in WV account for a small portion of global emissions, it is indicative of how temperate forest encroachment into cooler subalpine and boreal conifer systems is a significant potential contributor to atmospheric CO₂ through combinations of human disturbance, as seen in this study, and climate change (e.g., Hamburg and Cogbill, 1988). Lal (2005) showed that the boreal and tundra systems represent significantly larger organic C pools than temperate forests, and that they are potentially the most vulnerable to climate change. Lal (2005) also

points out that as much as two thirds of forest C stocks are in soil organic C, and that in boreal systems this ratio is even higher.

Different studies have indicated that red spruce will mostly disappear from the central Appalachians within the century under even the best climate change scenarios (Butler et al., 2014; Iverson et al., 2008; Prasad et al., 2007; Young et al., 2010) implying that debate regarding the benefits of red spruce restoration is moot. Studies of the red spruce – northern hardwood ecotone in New England have often focused on the elevation of the transition and the associated ecological changes (Siccama, 1974; Beckage et al., 2008). Late twentieth century decreases in the growth of red spruce and upward shifts of the ecotone have largely been attributed to climate warming, but cannot rule out pollution and competition as co-factors (Beckage et al., 2008; McLaughlin et al., 1987). Hamburg and Cogbill (1988) were able to show that climate was probably more influential than air pollution (e.g. acid rain) in red spruce decline since 1800.

However, our results do suggest red spruce restoration could play a role in climate change mitigation and that it might be difficult to discern the effects of climate change on red spruce range because so much of the northeastern U.S. has been intensively disturbed since the industrial revolution. Indeed both the MNF witness tree database (Thomas-Van Gundy et al., 2012) and historical accounts (Hopkins, 1899) indicate that the red spruce range stretched much lower in elevation (500-700 meters) in certain topographies than current distributions would indicate. Several other recent studies show that red spruce populations are actually recovering and expanding (Nowacki et al., 2010; Rollins et al., 2010). Red spruce restoration may also become increasingly important where it co-dominates with eastern hemlock due to the projected loss of hemlock to the hemlock woolly adelgid (*Adelges tsugae*) (Hessl and Pederson, 2013).

Tree cover and species also have an effect on ambient air temperature and soil temperatures. Data from Hobbie et al. (2006) showed that spruce and fir species were associated with lower mean annual soil temperatures and decreased litter decomposition in a common garden experiment with a variety of tree species. Pielke (1981) reported on how the vast destruction of forests circa 1900 increased regional temperature, which then began to lower again around 1940 with the return of the mostly hardwood forest. It is also unclear how large of a role in climate change that land use change surface radiative dynamics play (Pielke, 2001; Pielke et al., 2002). The potential effect of forest mitigation of surface warming should be further investigated with respect to historically native conifer communities, and might uncover further resilience of red spruce communities against warming temperatures, and potential carbon cycle feedbacks.

6.7 *Conclusions*

Understanding how soil properties relate to ecosystem dynamics can help tell the story of a site when current vegetation may not reflect the past due to anthropogenic disturbance. ESD help provide the framework to understand these concepts in a pragmatic manner. In the higher elevation areas of West Virginia, ESD related to spodic soil properties reflect a much different forest composition before the vast ecological disturbance wrought on the land by the railroad timbering era. The key to this understanding is knowing the time scale required for differing soil processes to react to changes in environment. In spodic conifer forests, longer lived subsurface sesquioxide horizons can persist for longer periods than soil organic pools which can shift quickly with disturbance and forest composition change. Carbon pools that respond relatively quickly to forest restoration represent an important potential avenue of C sequestration and habitat renewal in areas where disturbance has caused loss of species that promote soil C

buildup. Although climate change is a daunting challenge and species like red spruce seem to be ill-fated from some perspectives, they also might represent a significant mitigation potential as new data emerges. Alternatively, if red spruce is lost, similar species that promote podsolization and C accumulation including other selected *Tsuga*, *Larix*, *Picea*, *Pinus*, and *Abies* species could serve as alternatives. Restoration of red spruce and similar species represents one of many potential climate mitigation and ecological restoration options that society will need to evaluate in our efforts to balance our global sustainability.

6.8 Acknowledgements

We would like to acknowledge the large group of scientists who contributed to the field soil description efforts for this project. These individuals came from diverse institutions including USDA-NRCS, USDA Forest Service, West Virginia University, West Virginia State University, and Virginia Polytechnic Institute and State University. We also acknowledge the dedicated fieldwork at forest plots by WVU research associate Aaron Burkholder, and the input and assistance from Shane Jones, MNF Biologist. Portions of this research were supported by the United States' Department of Agriculture Natural Resources Conservation Service.

6.9 Literature Cited

- Adams, H.S., S. Stephenson, A.W. Adams, and M.B. Adams. 2010. The isolated red spruce communities of Virginia and West Virginia, pp. 1-12 *The conference on the ecology and management of high-elevation forests in the central and southern Appalachian Mountains*, Vol. WV. Gen. Tech. Rep. NRS-P-64. Department of Agriculture, Forest Service, Northern Research Station, Slatyfork, WV.
- Adams, M.B., and C. Eagar. 1992. IMPACTS OF ACIDIC DEPOSITION ON HIGH-ELEVATION SPRUCE-FIR FORESTS - RESULTS FROM THE SPRUCE-FIR RESEARCH COOPERATIVE. *Forest Ecology and Management* 51:195-205.
- Amundson, R., and H. Jenny. 1991. THE PLACE OF HUMANS IN THE STATE FACTOR THEORY OF ECOSYSTEMS AND THEIR SOILS. *Soil Science* 151:99-109.
- Amundson, R., and H. Jenny. 1997. On a State Factor Model of Ecosystems. *BioScience* 47:536-543.
- Averill, C., B.L. Turner, and A.C. Finzi. 2014. Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage. *Nature* 505:543-545.
- Barrett, L.R., and R.J. Schaetzl. 1998. Regressive Pedogenesis Following a Century of Deforestation: Evidence for Depodzolization. *Soil Science* 163(6):482-497.
- Beane, N.R., J.S. Rentch, and T.M. Schuler. 2013. Using Maximum Entropy Modeling to Identify and Prioritize Red Spruce Forest Habitat in West Virginia, USFS Northern Research Station.
- Beckage, B., B. Osborne, D.G. Gavin, C. Pucko, T. Siccama, and T. Perkins. 2008. A rapid upward shift of a forest ecotone during 40 years of warming in the Green Mountains of Vermont. *Proceedings of the National Academy of Sciences* 105:4197-4202.
- Berg, B., and C. McClaugherty. 2008. *Plant litter* Springer.
- Bestelmeyer, B.T., D.P. Goolsby, and S.R. Archer. 2011. Spatial perspectives in state-and-transition models: a missing link to land management? *Journal of Applied Ecology* 48:746-757.
- Bestelmeyer, B.T., A.J. Tugel, G.L. Peacock Jr, D.G. Robinett, P.L. Shaver, J.R. Brown, J.E. Herrick, H. Sanchez, and K.M. Havstad. 2009. State-and-transition models for heterogeneous landscapes: a strategy for development and application. *Rangeland Ecology & Management* 62:1-15.
- Binkley, D., and R. Fisher. 2012. *Ecology and management of forest soils* John Wiley & Sons.
- Blum, J.D., A. Klaue, C.A. Nezat, C.T. Driscoll, C.E. Johnson, T.G. Siccama, C. Eagar, T.J. Fahey, and G.E. Likens. 2002. Mycorrhizal weathering of apatite as an important calcium source in base-poor forest ecosystems. *Nature* 417:729-731.
- Briske, D.D., S.D. Fuhlendorf, and F.E. Smeins. 2005. State-and-transition models, thresholds, and rangeland health: A synthesis of ecological concepts and perspectives. *Rangeland Ecology & Management* 58:1-10.
- Brundrett, M. 2009. Mycorrhizal associations and other means of nutrition of vascular plants: understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis. *Plant and Soil* 320:37-77.
- Butler, P.R., L. Iverson, F.R.T. III, L. Brandt, S. Handler, M. Janowiak, P.D. Shannon, C. Swanston, J. Bartig, S. Connelly, W. Dijak, K. Karriker, C. Randall, S. Bearer, S. Blatt, A. Brandon, E. Byers, C. Coon, T. Culbreth, J. Daly, W. Dorsey, D. Ede, C. Euler, N. Gillies, L. Lyte, D. McCarthy, D. Minney, D.I. Murphy, C. O'Dea, D. Hix, C. Johnson, R. Orwan, M. Peters, J. Reed, C. Sandeno, T. Schuler, L. Sneddon, B. Stanley, A. Steele, R. Swaty, S. Stout, J. Teets, T. Tomon, J. Vanderhorst, J. Whatley, and N. Zegre. 2014. *Central Appalachians ecosystem vulnerability assessment and synthesis: a report from the Central Appalachians Climate Change Response*

- Framework project. Department of Agriculture, Forest Service, Northern Research Station., Newtown Square, PA.
- Byers, E.A., J.P. Vanderhorst, and B.P. Streets. 2010. Classification and Conservation Assessment of Upland Red Spruce Communities in West Virginia, *In* W. V. D. o. N. R. West Virginia Natural Heritage Program, (ed.). Wildlife Resources Section.
- Byers, E.A., K.C. Love, K.R. Haider, E.J. Burks, and J.E. Rowan. 2013. Red Spruce (*Picea rubens*) Cover in West Virginia, Version 1.0. West Virginia Division of Natural Resources, Central Appalachian Spruce Restoration Initiative, Appalachian Forest Heritage Area Americorps, Monongahela National Forest, and U.S. Fish and Wildlife Service.
- Caudle, D., H. Sanchez, J. DiBenedetto, C. Talbot, and M. Karl. 2013. Interagency Ecological Site Handbook for Rangelands. USDA-NRCS, USDA-FS, & DOI-BLM.
- Clarkson, R.B. 1964. Tumult on the mountains: lumbering in West Virginia, 1770-1920 McClain Printing Company.
- Comas, L.H., and D.M. Eissenstat. 2009. Patterns in root trait variation among 25 co-existing North American forest species. *New Phytologist* 182:919-928.
- Côté, B., and J.W. Fyles. 1994. Nutrient concentration and acid-base status of leaf litter of tree species characteristic of the hardwood forest of southern Quebec. *Canadian Journal of Forest Research* 24:192-196.
- Dillard, L.O., K.R. Russell, and W.M. Ford. 2008a. Macrohabitat models of occurrence for the threatened Cheat Mountain salamander, *Plethodon nettingi*. *Applied Herpetology* 5:201-224.
- Dillard, L.O., K.R. Russell, and W.M. Ford. 2008b. Site-level habitat models for the endemic, threatened Cheat Mountain salamander (*Plethodon nettingi*): the importance of geophysical and biotic attributes for predicting occurrence. *Biodiversity and Conservation* 17:1475-1492.
- Dokuchaev, V.V. 1999. On the concept of natural zones - St. Petersburg, 1899. *Eurasian Soil Science* 32:726-727.
- EPA, U.S. 2014. Greenhouse Gas Equivalencies Calculator [Online] <http://www.epa.gov/cleanenergy/energy-resources/calculator.html>.
- Friedland, A.J., G.J. Hawley, and R.A. Gregory. 1988. Red spruce (*Picea rubens* Sarg.) foliar chemistry in Northern Vermont and New York, USA. *Plant and Soil* 105:189-193.
- Giesler, R., H. Ilvesniemi, L. Nyberg, P. van Hees, M. Starr, K. Bishop, T. Kareinen, and U.S. Lundström. 2000. Distribution and mobilization of Al, Fe and Si in three podzolic soil profiles in relation to the humus layer. *Geoderma* 94:249-263.
- Grazing Lands Technology, I. 2003. National range and pasture handbook [computer file], 2003 Revision ed. [Fort Worth, Tex.] : U.S. Dept. of Agriculture, Natural Resources Conservation Service, Grazing Lands Technology Institute, [1997].
- Hamburg, S.P., and C.V. Cogbill. 1988. Historical decline of red spruce populations and climatic warming. *Nature* 331:428-431.
- Herbauts, J., and E. Buyl. 1981. The relation between spruce monoculture and incipient podzolisation in ochreous brown earths of the Belgian Ardennes. *Plant and Soil* 59:33-49.
- Herrick, J., B. Bestelmeyer, S. Archer, A. Tugel, and J. Brown. 2006. An integrated framework for science-based arid land management. *Journal of Arid Environments* 65:319-335.
- Hessl, A., and N. Pederson. 2013. Hemlock Legacy Project (HeLP) A paleoecological requiem for eastern hemlock. *Progress in Physical Geography* 37:114-129.
- Hix, D.M., and B.V. Barnes. 1984. Effects of clear-cutting on the vegetation and soil of an eastern hemlock dominated ecosystem, western Upper Michigan. *Canadian Journal of Forest Research* 14:914-923.
- Hobbie, S., M. Ogdahl, J. Chorover, O. Chadwick, J. Oleksyn, R. Zytowskiak, and P. Reich. 2007. Tree Species Effects on Soil Organic Matter Dynamics: The Role of Soil Cation Composition. *Ecosystems* 10:999-1018.

- Hobbie, S.E., P.B. Reich, J. Oleksyn, M. Ogdahl, R. Zytowski, C. Hale, and P. Karolewski. 2006. Tree species effects on decomposition and forest floor dynamics in a common garden. *Ecology* 87:2288-2297.
- Hoffland, E., T.W. Kuyper, H. Wallander, C. Plassard, A.A. Gorbushina, K. Haselwandter, S. Holmström, R. Landeweert, U.S. Lundström, and A. Rosling. 2004. The role of fungi in weathering. *Frontiers in Ecology and the Environment* 2:258-264.
- Högberg, P., and D.J. Read. 2006. Towards a more plant physiological perspective on soil ecology. *Trends in Ecology & Evolution* 21:548-554.
- Hole, F. 1975. Some relationships between forest vegetation and podzol B horizons in soils of Menominee tribal lands, Wisconsin, USA. *Soviet soil science*.
- Hole, F.D. 1976. *Soils of Wisconsin* University of Wisconsin Press.
- Hopkins, A.D. 1899. Report on Investigations to Determine the Cause of Unhealthy Conditions of the Spruce and Pine from 1880-1893, *In* W. V. A. E. Station, (ed.), Vol. Bulletin 56. Fairmont Index Steam Print, Morgantown, WV.
- Hornbeck, J.W., and R.B. Smith. 1985. Documentation of red spruce growth decline. *Canadian Journal of Forest Research* 15:1199-1201.
- Iverson, L.R., A.M. Prasad, S.N. Matthews, and M. Peters. 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecology and Management* 254:390-406.
- Jenny, H. 1941. *Factors of Soil Formation* McGraw-Hill, New York, New York.
- Jenny, H. 1961. Derivation of state factor equations of soils and ecosystems. *Soil Science Society of America Journal* 25:385-388.
- Jenny, H. 1980. *ECOLOGICAL STUDIES ANALYSIS AND SYNTHESIS VOL. 37. THE SOIL RESOURCE ORIGIN AND BEHAVIOR*, p. XXI+377P-XXI+377P Jenny, H. *Ecological Studies: Analysis and Synthesis, Vol. 37. The Soil Resource: Origin and Behavior. Xxi+377p.* Springer-Verlag: New York, N.Y., USA; Berlin, West Germany. Illus. Maps.
- Johnson, A.H. 1983. Red Spruce Decline in the Northeastern U.S.: Hypotheses Regarding the Role of Acid Rain. *Journal of the Air Pollution Control Association* 33:1049-1054.
- Jongmans, A.G., N. van Breemen, U. Lundstrom, P.A.W. van Hees, R.D. Finlay, M. Srinivasan, T. Unestam, R. Giesler, P.A. Melkerud, and M. Olsson. 1997. Rock-eating fungi. *Nature* 389:682-683.
- Lal, R. 2005. Forest soils and carbon sequestration. *Forest Ecology and Management* 220:242-258.
- Lietzke, D.A., and G.A. McGuire. 1987. Characterization and Classification of Soils with Spodic Morphology in the Southern Appalachians. *Soil Sci. Soc. Am. J.* 51:165-170.
- Lin, H. 2011. Three Principles of Soil Change and Pedogenesis in Time and Space. *Soil Sci. Soc. Am. J.* 75:2049-2070.
- Losche, C.K., and W.W. Beverage. 1967. Soil survey of Tucker County and part of Northern Randolph County, West Virginia. *Soil Survey Reports*. United States Department of Agriculture.
- Lundström, U.S., N. van Breemen, and D. Bain. 2000a. The podzolization process. A review. *Geoderma* 94:91-107.
- Lundström, U.S., N. van Breemen, D.C. Bain, P.A.W. van Hees, R. Giesler, J.P. Gustafsson, H. Ilvesniemi, E. Karlton, P.A. Melkerud, M. Olsson, G. Riise, O. Wahlberg, A. Bergelin, K. Bishop, R. Finlay, A.G. Jongmans, T. Magnusson, H. Mannerkoski, A. Nordgren, L. Nyberg, M. Starr, and L. Tau Strand. 2000b. Advances in understanding the podzolization process resulting from a multidisciplinary study of three coniferous forest soils in the Nordic Countries. *Geoderma* 94:335-353.
- McLaughlin, S., D. Downing, T. Blasing, E. Cook, and H. Adams. 1987. An analysis of climate and competition as contributors to decline of red spruce in high elevation Appalachian forests of the eastern United States. *Oecologia* 72:487-501.

- Melvin, A.M., and C.L. Goodale. 2013. Tree species and earthworm effects on soil nutrient distribution and turnover in a northeastern United States common garden. *Canadian Journal of Forest Research* 43:180-187.
- Menzel, J.M., W.M. Ford, J.W. Edwards, and M.A. Menzel. 2004. Nest tree use by the endangered Virginia northern flying squirrel in the central Appalachian mountains. *American Midland Naturalist* 151:355-368.
- Menzel, J.M., W.M. Ford, J.W. Edwards, and L.J. Ceperley. 2006a. A habitat model for the Virginia northern flying squirrel (*Glaucomys sabrinus fuscus*) in the central Appalachian Mountains US Department of Agriculture, Forest Service, Northeastern Research Station.
- Menzel, J.M., W.M. Ford, J.W. Edwards, and T.M. Terry. 2006b. Home range and habitat use of the Vulnerable Virginia northern flying squirrel *Glaucomys sabrinus fuscus* in the Central Appalachian Mountains, USA. *Oryx* 40:204-210.
- Miles, J. 1985. The pedogenic effects of different species and vegetation types and the implications of succession. *Journal of Soil Science* 36:571-584.
- Nauman, T.W., J.T. Thompson, J. Teets, T. Dilliplane, J. Bell, S.J. Connolly, H.J. Liebermann, and K. Yoast. In Press. Ghosts of the forest: mapping pedomemory to guide forest restoration. *Geoderma*.
- Rentch, J.S. and T. M. Schuler (ed.) 2010. Proceedings from the conference on the ecology and management of high-elevation forests in the central and southern Appalachian Mountains., Slatyfork, WV. USDA-FS Northern Research Station.
- Rentch, J.S. and T. M. Schuler (ed.) 2010. Proceedings from the conference on the ecology and management of high-elevation forests in the central and southern Appalachian Mountains, Slaty Fork, WV. USDA - Forest Service, Northern Research Station.
- USDA-NRCS. 2014. National Ecological Site Handbook.
- Odom, R.H., W.M. Ford, J.W. Edwards, C.W. Stihler, and J.M. Menzel. 2001. Developing a habitat model for the endangered Virginia northern flying squirrel (*Glaucomys sabrinus fuscus*) in the Allegheny Mountains of West Virginia. *Biological Conservation* 99:245-252.
- Parfitt, R.L. 2009. Allophane and imogolite: role in soil biogeochemical processes. *Clay Minerals* 44:135-155.
- Pauley, T.K. 2008. The Appalachian Inferno: Historical Causes for the Disjunct Distribution of *Plethodon nettingi* (Cheat Mountain Salamander). *Northeastern Naturalist* 15:595-606.
- Phillips, J.D., and D.A. Marion. 2004. Pedological memory in forest soil development. *Forest Ecology and Management* 188:363-380.
- Pielke, R.A. 1981. The Distribution of Spruce in West-Central Virginia before Lumbering. *Castanea* 46:201-216.
- Pielke, R.A. 2001. Influence of the spatial distribution of vegetation and soils on the prediction of cumulus convective rainfall. *Reviews of Geophysics* 39:151-177.
- Pielke, R.A., G. Marland, R.A. Betts, T.N. Chase, J.L. Eastman, J.O. Niles, and S.W. Running. 2002. The influence of land-use change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effect of greenhouse gases. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 360:1705-1719.
- Prasad, A., L. Iverson, S. Matthews, and M. Peters. 2007. A climate change atlas for 134 forest tree species of the eastern United States [database]. Northern Research Station, USDA Forest Service, Delaware, Ohio.
- Ranger, J., and C. Nys. 1994. The effect of spruce (*Picea abies* Karst.) on soil development: an analytical and experimental approach. *European Journal of Soil Science* 45:193-204.
- Reich, P.B., J. Oleksyn, J. Modrzyński, P. Mrozinski, S.E. Hobbie, D.M. Eissenstat, J. Chorover, O.A. Chadwick, C.M. Hale, and M.G. Tjoelker. 2005. Linking litter calcium, earthworms and soil properties: a common garden test with 14 tree species. *Ecology Letters* 8:811-818.

- Rentch, J.S., T.M. Schuler, W.M. Ford, and G.J. Nowacki. 2007. Red spruce stand dynamics, simulations, and restoration opportunities in the central Appalachians. *Restoration Ecology* 15:440-452.
- Rentch, J.S., T.M. Schuler, G.J. Nowacki, N.R. Beane, and W.M. Ford. 2010. Canopy gap dynamics of second-growth red spruce-northern hardwood stands in West Virginia. *Forest Ecology and Management* 260:1921-1929.
- Rollins, A.W., H.S. Adams, and S.L. Stephenson. 2010. Changes in Forest Composition and Structure across the Red Spruce-Hardwood Ecotone in the Central Appalachians. *Castanea* 75:303-314.
- Rustad, L.E., and I.J. Fernandez. 1998. Soil Warming: Consequences for Foliar Litter Decay in a Spruce-Fir Forest in Maine, USA. *Soil Sci. Soc. Am. J.* 62:1072-1080.
- Sauer, D., H. Sponagel, M. Sommer, L. Giani, R. Jahn, and K. Stahr. 2007. Podzol: Soil of the Year 2007. A review on its genesis, occurrence, and functions. *Journal of Plant Nutrition and Soil Science* 170:581-597.
- Schaetzl, R.J. 1994. Changes in O horizon mass, thickness and carbon content following fire in northern hardwood forests. *Vegetatio* 115:41-50.
- Schoeneberger, P.J., E. Wysocki, and S.S. Staff. 2002. Field book for describing and sampling soils, Version 3.0 Government Printing Office.
- Schöll, L., T. Kuyper, M. Smits, R. Landeweert, E. Hoffland, and N. Breemen. 2008. Rock-eating mycorrhizas: their role in plant nutrition and biogeochemical cycles. *Plant and Soil* 303:35-47.
- Shigo, A.L. 1972. The beech bark disease today in the northeastern US. *Journal of Forestry* 70:286-289.
- Siccama, T.G. 1974. Vegetation, soil, and climate on the Green Mountains of Vermont. *Ecological Monographs*:325-349.
- Sohet, K., J. Herbauts, and W. Gruber. 1988. CHANGES CAUSED BY NORWAY SPRUCE IN AN OCHREOUS BROWN EARTH, ASSESSED BY THE ISOQUARTZ METHOD. *Journal of Soil Science* 39:549-561.
- Soil Survey Laboratory Staff. 2004. Soil Survey Laboratory Methods Manual. USDA-NRCS, Lincoln, NE.
- Soil Survey Staff. 1999. Soil Taxonomy A Basis System of Soil Classification for Making and Interpreting Soil Surveys, *In* USDA-NRCS, (ed.), 2nd Edition ed. U.S. Government Printing Office, Washington.
- Stanley, S.R., and E.J. Ciolkosz. 1981. Classification and Genesis of Spodosols in the Central Appalachians. *Soil Sci. Soc. Am. J.* 45:912-917.
- Stephenson, S.L., and J.F. Clovis. 1983. Spruce Forests of the Allegheny Mountains in Central West Virginia. *Castanea* 48:1-12.
- Stuiver, M., and H.A. Polach. 1977. Discussion; reporting of C-14 data. *Radiocarbon* 19:355-363.
- Targulian, V.O., and S.V. Goryachkin. 2004. Soil memory: Types of record, carriers, hierarchy and diversity. *Revista Mexicana de Ciencias Geológicas* 21.
- Teets, J. 2013. Ecological Site Description: Spodic Shale Upland Conifer Forest (Draft), *In* USDA-NRCS, (ed.), Morgantown, WV.
- Thomas-Van Gundy, M., M. Strager, and J. Rentch. 2012. Site characteristics of red spruce witness tree locations in the uplands of West Virginia, USA. *The Journal of the Torrey Botanical Society* 139:391-405.
- USDA-SCS, and USDA-FS. 1982. Soil Survey of Randolph County Area, Main Part, West Virginia.
- van Breemen, N., U.S. Lundström, and A.G. Jongmans. 2000. Do plants drive podzolization via rock-eating mycorrhizal fungi? *Geoderma* 94:163-171.
- Vogel, J.S., J.R. Southon, and D.E. Nelson. 1987. Catalyst and binder effects in the use of filamentous graphite for AMS. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 29:50-56.
- West Virginia Geologic and Economic Survey. 1968. Surface Geology - Rock Units, pp. Online Digitized Map: Available at <http://wvgis.wvu.edu/>, *In* T. U. West Virginia DEP, (ed.). William and Heintz Map Corporation.
- Williams, B.H., and H.M. Fridley. 1931. Soil Survey of Randolph County, West Virginia.

- Willis, K.J., M. Braun, P. Sumegi, and A. Toth. 1997. Does soil change cause vegetation change or vice versa? A temporal perspective from Hungary. *Ecology* 78:740-750.
- Young, B., E. Byers, K. Gravuer, K. Hall, G. Hammerson, and A. Redder. 2010. Guidelines for using the NatureServe climate change vulnerability index. NatureServe, Arlington, VA.
- Young, V.A. 1934. Plant distribution as influenced by soil heterogeneity in Cranberry Lake region of the Adirondack Mountains. *Ecology* 15:154-196.

6.10 Tables

Table 1. Description of spodic intensity (SI) classes based on observable field morphology.

SI Class	Description
0.0	No evidence of podzolization.
0.5	Very weak expression of podzolization. There is only slight physical evidence of podzolization. A slightly redder hue and higher value is present at the top of the B horizon, but the hue is less than one Munsell hue redder than an underlying horizon. The soil is non-smeariness*.
1.0	Weak expression of podzolization (spodic intergrade, half of profiles key to Spodosols). Spodic materials are present, but they do not always meet the criteria for a diagnostic spodic horizon. A subtle Bs horizon is present. The Bs horizon is one Munsell hue redder than an underlying horizon. Bhs material is usually absent. An albic E horizon is not present. The spodic materials are sometimes weakly smeariness.
1.5	Moderate expression of podzolization (Spodosol). Spodic materials are present as a diagnostic spodic horizon. A moderately expressed Bs horizon is present, often with pockets of Bhs material. An albic E horizon is not present. The spodic materials are often weakly smeariness.
2.0	Strong expression of podzolization (well-expressed Spodosol). A diagnostic spodic horizon is present usually underlying an albic E horizon. A Bhs or Bh horizon is continuous across at least 85 percent of the pedon. The spodic materials are often moderately smeariness.

* Smeariness (Shoeneberger et al, 2002, page 2-65) is a physical observation about how moistened soil samples fail when they are squeezed and rubbed between the thumb and forefinger. Smeariness can help identify spodic soil materials.

Table 2. Organic C calculations for O horizons from laboratory analysis used in restoration predictions of C sequestration. Standard deviations of averaged values are in parentheses.

Horizon	Average Organic C (wt)	Average Bulk Density (g/cm ³)	Average % of O depth	g C per cm ³	Total weighted average (g C/cm ³) in O horizons
Oi	48% (8.0%)	6.3E-2 (1.8E-2)	21% (6.1%)	3.0 E-2	
Oe	44% (9.8%)	8.8E-2 (4.1E-2)	44% (11%)	3.9E-2	5.7E-2
Oa	43% (5.2%)	2.2E-1 (1.0E-1)	35% (8.7%)	9.6E-2	

Table 3. Organic C accumulation calculations for SCF restoration scenarios. Estimates of total organic C stored in O horizons above 880 m, assuming the same areal proportions of states across those areas and same proportion of spodic areas (31%).

State Area Estimate Source	Hardwood State (ha)	Mixed Transition (ha)	Total C seq. (Tg*) in study map units	Total C seq (Tg*) in WV above 880m
Byers et al. (2013)	11758	1581	0.52	6.6
1/3 Validation Set	8513	n/a	0.36	4.5
2013 plots	5814	2907	0.29	3.7
Conifer IMP deficit est.	0.76	0.31		
O horizon increase (cm)	7.3	3.0		

* Teragrams = 10^{12} grams

6.11 Figures

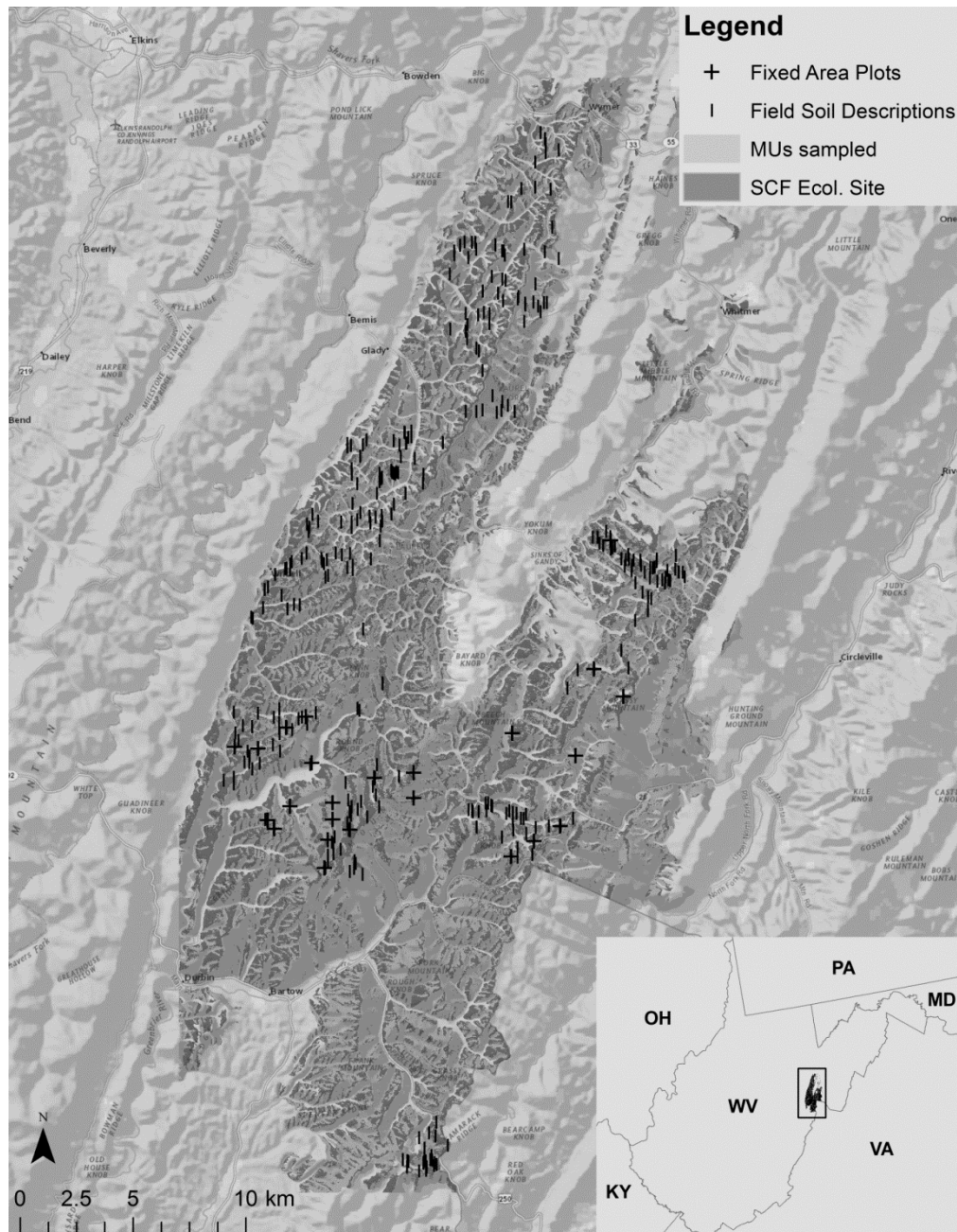


Figure 1. Study area soil map units (MU), areas of the Spodic Shale Uplands Conifer Forest (SCF) ecological site, and data collection locations overlaid on ArcGIS 10 National Geographic mapping baselayer.

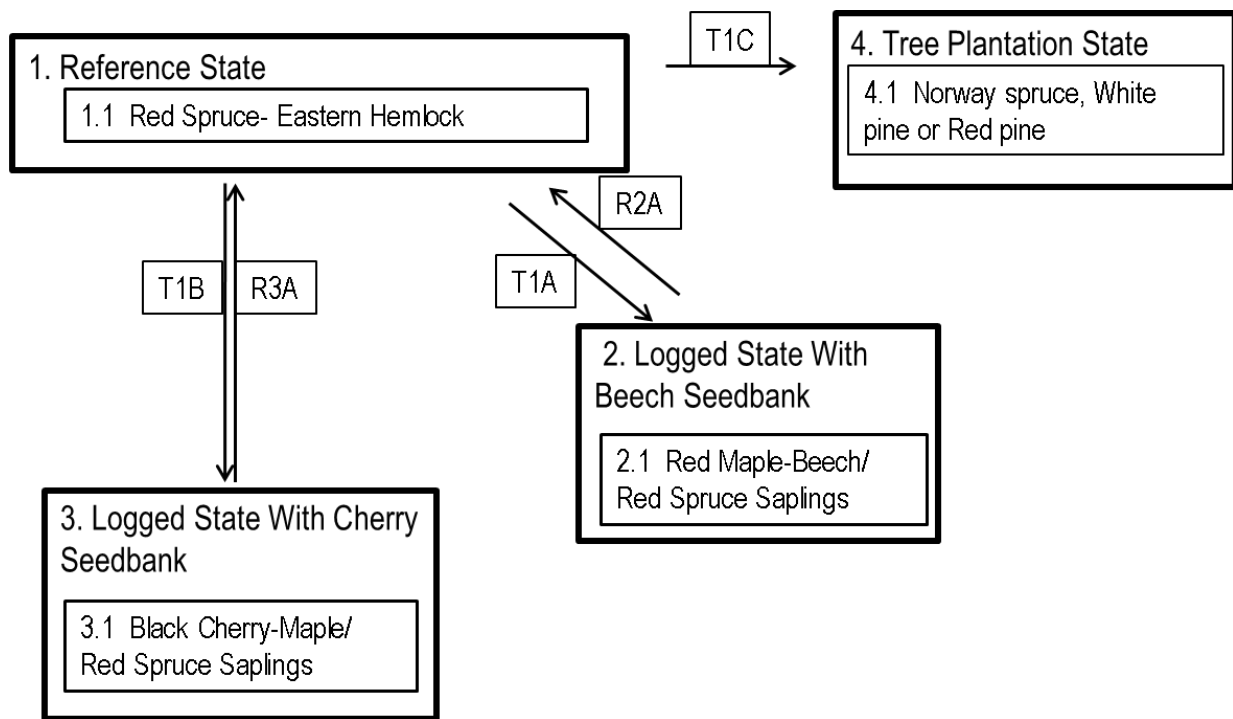


Figure 2. State and transition model from spodic shale upland conifer forest (SCF) (Teets, 2013).

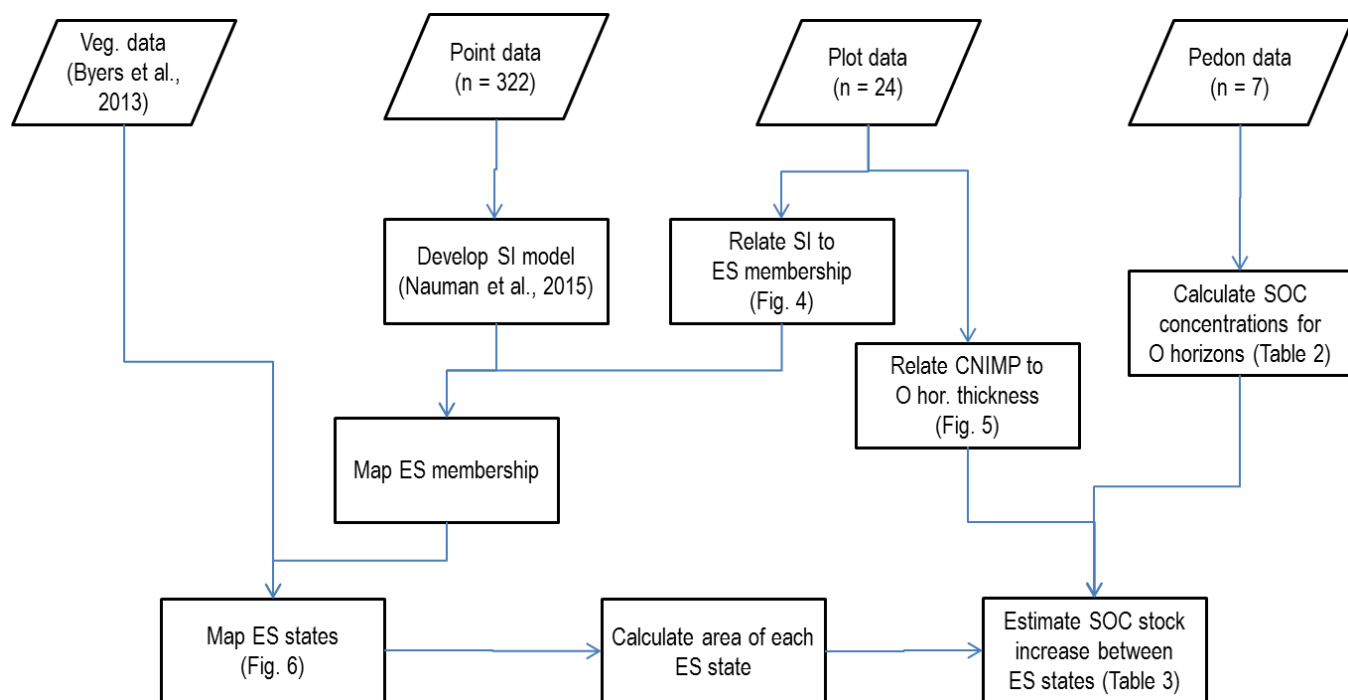


Figure 3. Flow chart of data analysis to create pedoecological areal estimate of SCF extent (Point data) and alternative states (Byers et al., 2013) as well as O-horizon accumulation (Pedon data and Plot Data).

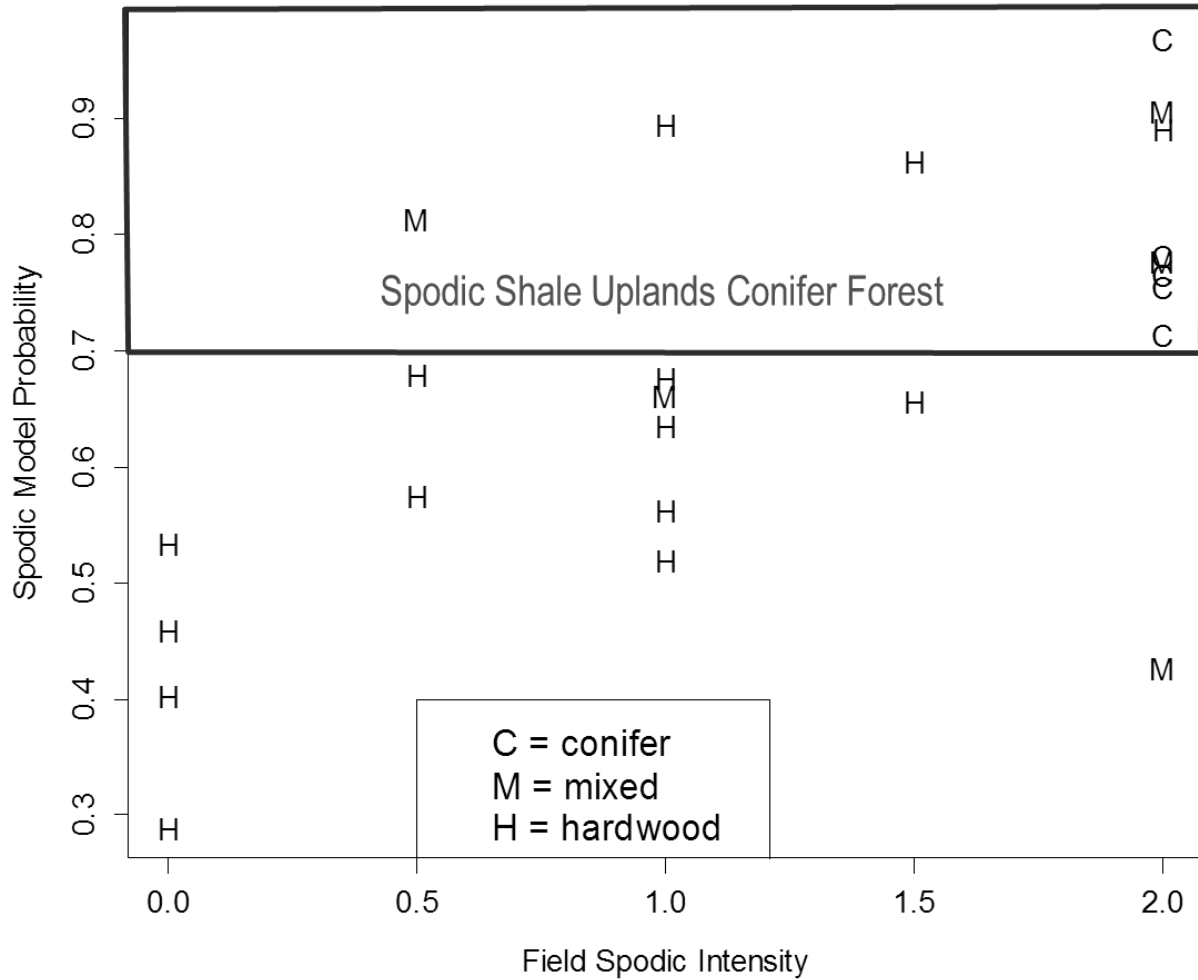


Figure 4. Graph of fixed area plot SI values (x-axis) versus probabilities from the random forest spodic probability model (y-axis). The outlined area at the top of the plot delineates the SCF ecological site, and letters represent dominant tree composition groups (C = conifer, M = mixed, H = hardwood).

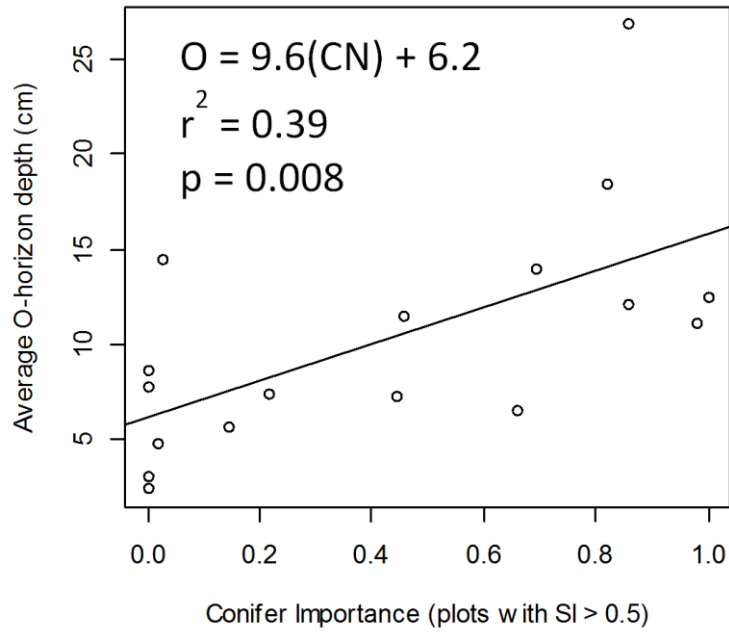


Figure 5. Relationships between conifer importance (CNIMP) with O horizon depth in at plots with SI of 1-2.

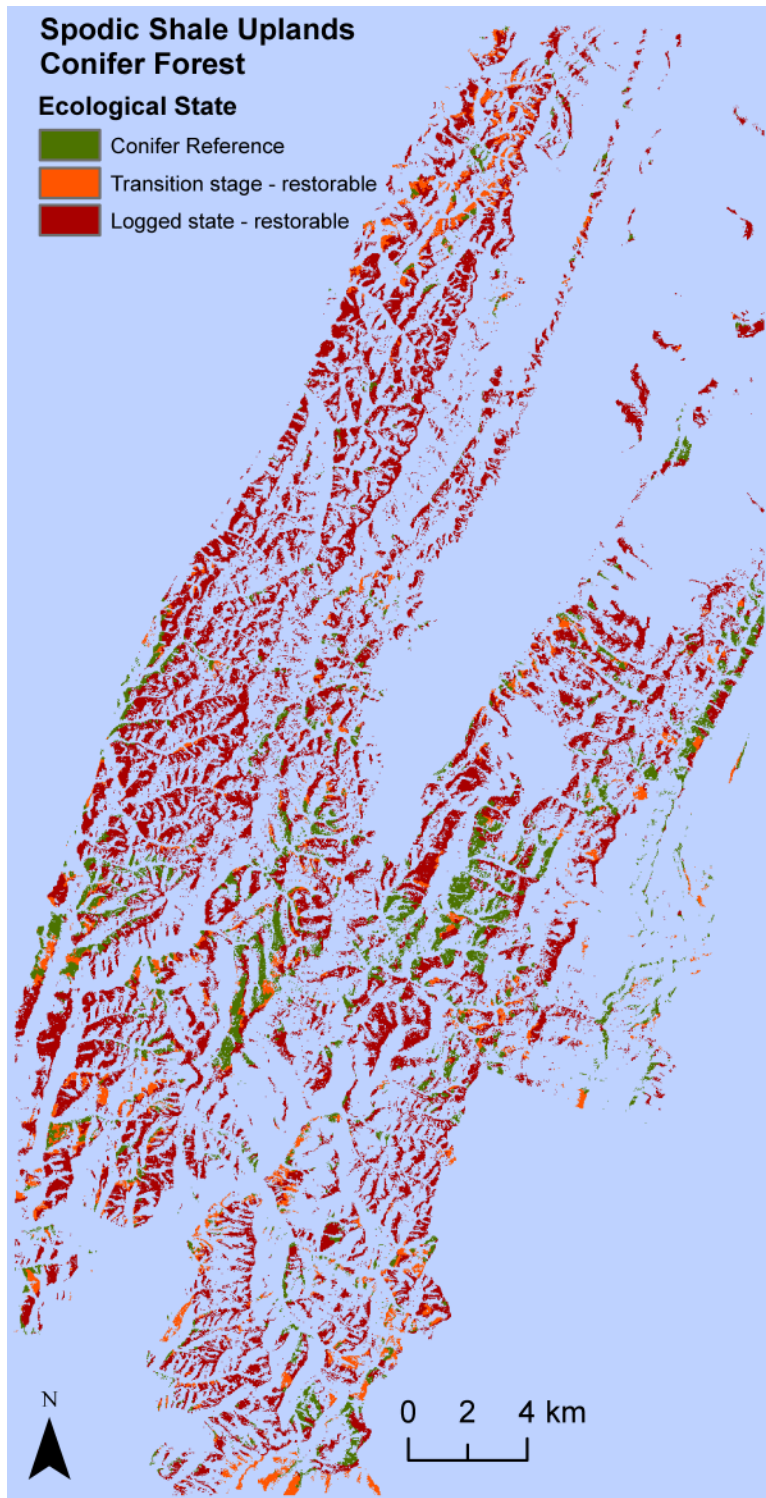


Figure 6. Pedoecological map of the different vegetative states within the spodic shale uplands conifer forest (SCF) ecological site. States are denoted by the different colors.

7 CHAPTER 6: CONCLUSIONS

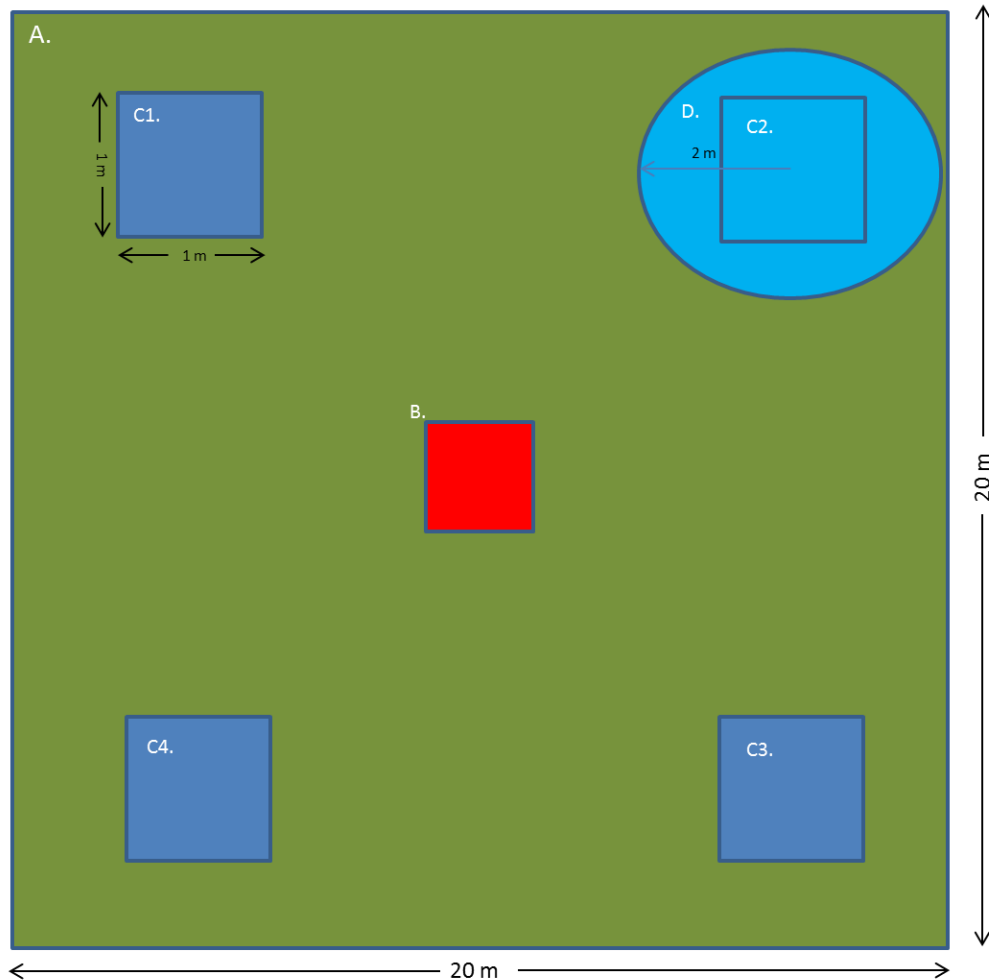
These studies all aim to leverage the predictive aspects of linkages between soils and their surrounding environment. These predictive links enable use of mapping technologies to combine ecological and soils data with available geospatial data sources to create management scale (~10-100 m) raster maps of ecological status and potential using ESD (as one approach) that can be leveraged into other models. In some cases there are very definite connections between soils and factors like vegetation or topography, and these are much easier to map. In other cases, mapping was not able to characterize the variability seen at validation sites. A sizeable contrast in validation accuracies was seen between soil maps (SSURGO and DM) in Arizona (~70%) and West Virginia (25-40%). The updated DM maps showed utility in increasing spatial resolution and creating more transparent keys, but only produced equivalent accuracy to the original soil surveys.

Differences in accuracies between the two DM trials could potentially be from many sources. The mapping intensity and methods could be inconsistent because of the differing eras when the various county soil surveys were completed. Although both areas are relatively remote, the AZ survey is on a National Park Service property where more detail was possibly desired by the client. Or more likely, the soil-landscape relationships were better defined in AZ. The AZ map units generally followed alluvial surface ages of deposits that are relatively easy to identify and generally have discrete breaks along geologic erosional fronts working upslope to the mountains. The WV study area was densely forested with high relief and complicated landforms. There are also aspects of peri-glacial landforms that have shaped some areas creating short order variability. All of these are factors that potentially intensify field scale soil variability in WV.

The variability and uncertainty in soil properties in WV became a barrier to mapping ESD using the DM maps. This was overcome by examining the soil processes dominating the

different ecological sites present. This research uncovered the potential imprint of historic conifer in the form of spodic soils properties as a pedomemory indicator. Pedomemory is a property of the soil that can be linked to and used as a surrogate for a past condition, event, or ecosystem. Spodic soil properties could be mapped due to covariation with slope aspect, surface geometry, and satellite imagery. So a spodic soils map was used in place of the DM and modified to spatialize ESD states and transitions. The states and transitions were then analyzed under assumed restoration scenarios to determine how much historic disturbance related O-horizon loss could potentially be recovered (3.7-6.6+ Tg of C). Further efforts should be undertaken to evaluate the uncertainty in these estimates as well as how they fit into the greater carbon cycle. Although we suspect these values would help shift carbon out of atmospheric pools, carbon pools in biomass and subsurface mineral surface horizons also need to be considered. There are also potential feedbacks between changes in forest composition and soil temperature and moisture that could affect the carbon cycle that merit further investigation in future research.

8 APPENDIX A: PLOT DIAGRAM AND RELATED DATA COLLECTION



- A. 20x20 meter fixed area sample plot
- a. Forest Site Index: height and age of trees by species
 - i. Core 3-4 of each dominant or co-dominant species for site index
 1. Use red spruce growth curves as site 'reference' if there is evidence of podzolization.
 - ii. If necessary core other species for stand age estimates (will be done on minimal basis).
 - b. Basal Area: diameter breast height
 - c. Overall species list (also including plants within visible surroundings)
 - d. Crown Class: dominant, co-dominant, intermediate, suppressed
 - e. Overall canopy cover by spherical densiometer
 - f. Canopy Cover by species by height strata (ocular estimate)
 - i. Stratum breaks (m): 0.5, 1, 5, 10, 20
 - ii. Estimate for larger species: trees, shrubs, dominant herbaceous

- g. Soil Cover Estimates: large debris (e.g. woody debris, large rocks)
- B. Soil pit at plot center
 - a. Full field description (NRCS 232)
 - i. Describe and sample horizons
 - ii. Estimate 'Spodic Intensity' as defined by Tim Dilliplane and Skip Bell (WV NRCS)
 - iii. Pay extra attention to horizon topography and consistency for looking at regressive vs progressive organizational patterns (Johnson and Watson-Stegner, 1987; Minasny et al., 2008).
 - b. Samples analyzed in lab
 - i. Particle size, pH, EC, OC, nutrients, extractions for different groups of sesquioxides (primary Spodic indicator), others?
- C. Vegetation quadrat-primary regeneration quadrat at center of plot quadrants
 - a. Soil cover for smaller types (e.g. bryophytes, bare ground, smaller rock fragments, organic litter, plant basal)
 - b. Canopy cover by species (include all species, but meant to detect smaller organisms)
 - c. Tree regeneration counts.
 - d. O-horizon depth observation (1 just outside of each quadrat)
- D. Secondary regeneration sampling unit. **Only necessary if regeneration is sparse/not present in primary regeneration quadrat.**

9 APPENDIX B CHAPTER 3 DATA

Table B.1 Soil classes at validation sites used to evaluate the disaggregation models in Chapter 3.

NRCS_SITEID	Field_Class	SQRT_Model_Class	LOG_Model_Class	ORIG_Model_Class	X_UTM_E	Y_UTM_N
78WV083003	Berks	Kaymine	Kaymine	Gilpin	571896	4265537
79WV067002	Ernest	Craigsville	Craigsville	Laidig	535578	4232927
79WV067006	Laidig	Gilpin	Gilpin	Gilpin	540966	4239426
79WV067007	Ernest	Laidig	Laidig	Laidig	541282	4239428
79WV067008	Buchanan	Gilpin	Gilpin	Gilpin	532051	4239756
81WV067001	Laidig	Laidig	Kaymine	Laidig	543543	4234507
81WV067002	Laidig	Gilpin	Gilpin	Gilpin	537321	4234538
81WV101001	Cookport	Fenwick	Fenwick	Gilpin	544584	4244007
95WV075001	Gauley	Gauley	Gauley	Gauley	565228	4239801
95WV075002	Gauley	Leatherbark	Leatherbark	Leatherbark	593141	4256352
F06WV071001	Gauley	Mandy	Udorthents	Mandy	627384	4282821
FS06WV081002	Meckesville	Cateache	Snowdog	Cateache	569224	4239814
S03WV-025-001	Cateache	Kaymine	Kaymine	Gilpin	527696	4203325
S03WV-025-002	Shouns	Laidig	Fluvaquents	Laidig	532583	4205852
S03WV071001	Shouns	Mandy	Gauley	Mandy	626191	4282167
S03WV075001	Cateache	Cateache	Belmont	Cateache	599775	4267370
S03WV-083-008	Cateache	Cateache	Cateache	Cateache	594381	4279297
S03WV-101-001	Shouns	Cateache	Cateache	Cateache	561319	4259456
S03WV-101-002	Cateache	Cateache	Cateache	Cateache	561301	4259262
S06WV067002	Fenwick	Fenwick	Fenwick	Gilpin	529717	4228504
S06WV067003	Laidig	Gilpin	Gilpin	Gilpin	528947	4228831
S06WV067004	Fenwick	Fenwick	Fenwick	Fenwick	538592	4234690
S06WV067005	Laidig	Gilpin	Gilpin	Gilpin	536970	4234258
S06WV101-001	Gilpin	Laidig	Laidig	Laidig	534125	4250763
S06WV101-002	Fenwick	Fenwick	Fenwick	Gilpin	544463	4245327
UG06WV075043	Laidig	Mandy	Mandy	Mandy	565228	4247323
UG06WV075084	Cateache	Laidig	Craigsville	Laidig	564104	4246103
UG06WV075101	Shouns	Snowdog	Snowdog	Snowdog	568980	4245673
UG06WV075126	Macove	Cateache	Cateache	Cateache	568146	4243186
UG06WV075130	Shouns	Sensabaugh	Holly	Sensabaugh	569246	4242079
UG06WV075136	Calvin	Shouns	Shouns	Shouns	568943	4238136
UG06WV075137	Hazleton	Cateache	Cateache	Cateache	568153	4237397
UG06WV075138	Laidig	Cateache	Cateache	Cateache	568975	4236913
UG06WV075140	Shouns	Cateache	Udorthents	Cateache	569002	4235624
UG06WV075142	Berks	Cateache	Medihemists	Cateache	570966	4234692

UG06WV075143	Shouns	Cateache	Cateache	Cateache	569477	4234522
UG06WV075144	Gilpin	Cateache	Cateache	Mandy	569333	4234451
UG06WV075146	Shouns	Cateache	Cateache	Cateache	571549	4234167
UG06WV075147	Cateache	Shouns	Shouns	Cateache	571753	4230584
UG06WV075150	Laidig	Cateache	Cateache	Cateache	568818	4240488
UG06WV075151	Cateache	Cateache	Cateache	Cateache	567640	4237003
UG06WV075153	none	Cateache	Cateache	Cateache	567140	4233748
UG06WV075155	Shouns	Cateache	Shouns	Cateache	567370	4236468
UG06WV075158	Cateache	Shouns	Shouns	Shouns	569991	4233918
UG06WV075159	Shouns	Shouns	Shouns	Shouns	569996	4233834
UG06WV075161	Shouns	Shouns	Shouns	Shouns	569186	4233573
UG06WV075164	Trussel	Snowdog	Snowdog	Snowdog	566146	4239962
UG06WV075166	Macove	Mandy	Cateache	Mandy	568142	4239133
UG06WV075168	Gilpin	Mandy	Mandy	Mandy	567076	4237125
UG06WV075169	Laidig	Cateache	Udorthents	Cateache	568417	4239667
UG06WV075172	Shouns	Mandy	Udorthents	Mandy	567025	4235634
UG06WV075173	Shouns	Cateache	Cateache	Cateache	566803	4234743
UG06WV075174	Shouns	Cateache	Shouns	Cateache	568184	4232055
UG06WV075175	Gilpin	Mandy	Udorthents	Mandy	568728	4231013
UG06WV075181	Shouns	Cateache	Cateache	Cateache	569297	4230319
UG06WV075420	Cateache	Cateache	Cateache	Cateache	570877	4241852
UG06WV075432	Laidig	Snowdog	Snowdog	Snowdog	569972	4246952
UG06WV075436	Laidig	Gilpin	Gilpin	Gilpin	559142	4246005
UG06WV101003	Dekalb	Gilpin	Gilpin	Gilpin	549437	4248690
UG06WV101008	Dekalb	Mandy	Simoda	Mandy	552949	4247457
UG06WV101010	Laidig	Gilpin	Snowdog	Gilpin	551742	4247180
UG06WV101011	Dekalb	Mandy	Gauley	Mandy	554960	4246529
UG06WV101074	Snowdog	Mandy	Simoda	Mandy	548098	4245016
UG06WV101080	Laidig	Laidig	Snowdog	Laidig	552194	4243019
UG06WV101085	Fenwick	Cateache	Meckesville	Cateache	552067	4244610
UG06WV101086	Macove	Gilpin	Cedarcreek	Gilpin	550130	4249619
UG06WV101087	Gilpin	Laidig	Meckesville	Laidig	547174	4249162
UG06WV101089	Laidig	Laidig	Laidig	Laidig	549433	4248467
UG06WV101090	Macove	Gilpin	Gilpin	Gilpin	549720	4247992
UG06WV101092	Laidig	Gilpin	Gilpin	Gilpin	557832	4246085
UG06WV101093	Macove	Gilpin	Gilpin	Gilpin	554702	4245140
UG06WV101094	Laidig	Gilpin	Kaymine	Gilpin	555224	4244739
UG06WV101096	Macove	Mandy	Snowdog	Mandy	562739	4248059
UG06WV101097	Macove	Gilpin	Laidig	Laidig	547861	4247814
UG06WV101105	Gilpin	Gilpin	Gilpin	Gilpin	551280	4244996
UG06WV101107	Atkins	Cotaco	Cotaco	Pope	554131	4243226
UG06WV101110	Gilpin	Gilpin	Gilpin	Gilpin	548340	4247227
UG06WV101113	Gilpin	Gilpin	Gilpin	Gilpin	544678	4246132

UG06WV101114	Dekalb	Gilpin	Gilpin	Gilpin	545256	4244922
UG06WV101121	Macove	Shouns	Cateache	Cateache	562026	4247701
UG06WV101199	Gilpin	Gilpin	Gilpin	Gilpin	550831	4245630
UG06WV101400	Gilpin	Gilpin	Gilpin	Gilpin	547912	4249045
UG06WV101401	Gilpin	Gilpin	Gilpin	Gilpin	548845	4247902
UG06WV101431	Laidig	Gilpin	Gilpin	Gilpin	549333	4246742
UG06WV101434	Dekalb	Gilpin	Dekalb	Gilpin	544863	4246566
UG06WV101435	Laidig	Gilpin	Gilpin	Gilpin	550799	4246075
UG06WV101437	Guyandotte	Laidig	Laidig	Laidig	555531	4246744

10 APPENDIX C CHAPTER 4 DATA

Table C.1 Validation site data from Chapter 4 validation of disaggregated soil survey.

Validation Taxa	SSURGO Map Unit Name	DM Taxa	DM Probability	X UTM E	Y UTM N
Gachado taxdjunct	Gachado extremely cobbly loam, 2 to 8 percent slopes	Gachado	0.63800001	329303	3539929
Gunsight	Gunsight very gravelly loam, 2 to 15 percent slopes	Gunsight	0.99000001	330297	3536882
Cipriano	Cipriano gravelly loam	Cipriano	0.94599998	335249	3538580
Cipriano	Cipriano gravelly loam	Cipriano	0.91799998	336960	3539232
Lomitas taxadjunct	Ajo very gravelly loam, 1 to 5 percent slopes	Ajo	0.71600002	338077	3545915
Cipriano	Ajo very gravelly loam, 1 to 5 percent slopes	Ajo	0.71200001	335669	3541854
Gunsight	Harqua-Gunsight complex	Gunsight	0.75000000	330323	3546071
Cipriano	Cipriano gravelly loam	Cipriano	0.96200001	333198	3551698
Ajo	Ajo very gravelly loam, 1 to 5 percent slopes	Ajo	0.93800002	336826	3549754
Cipriano taxadjunct	Ajo very gravelly loam, 1 to 5 percent slopes	Ajo	0.95800000	336884	3549779
Ajo	Ajo very gravelly loam, 1 to 5 percent slopes	Ajo	0.98199999	333971	3551006
Rillito	Rillito gravelly sandy loam	Rillito	0.74800003	333635	3558643
Rillito	Rillito gravelly sandy loam	Rillito	0.85600001	334209	3563683
Antho	Antho fine sandy loam	Antho	1.00000000	330631	3562527
Antho	Antho fine sandy loam	Antho	0.95599997	328557	3563306
Gilman	Gilman very fine sandy loam	Gilman	0.81599998	326980	3563679
Laveen	Laveen loam	Laveen	0.87400001	321226	3564188
Gilman	Torrifluents	Gilman	0.54600000	320850	3564164
Cherioni	Cherioni gravelly very fine sandy loam, 0 to 8 percent slo	Cherioni	0.88200003	319428	3562915
Cherioni	Cherioni gravelly very fine sandy loam, 0 to 8 percent slo	Cherioni	0.84600002	319098	3562786
Lomitas	Lomitas extremely stony loam, 8 to 40 percent slopes	Lomitas	0.52200001	320809	3545372
Gunsight	Gunsight very gravelly loam, 2 to 15 percent slopes	Gunsight	0.75599998	319441	3545951
Torrifluents	Lomitas extremely stony loam, 8 to 40 percent slopes	Lomitas	0.91600001	317937	3545498
Lomitas	Lomitas extremely stony loam, 8 to 40 percent slopes	Lomitas	0.45400000	314952	3544674
Harqua	Gunsight very gravelly loam, 2 to 15 percent slopes	Gunsight	0.52999997	314853	3544679
Lomitas	Lomitas extremely stony loam, 8 to 40 percent slopes	Gunsight	0.54799998	314952	3544468
Gachado taxadjunct	Harqua-Gunsight complex	Lomitas	0.51800001	315370	3543878
Harqua	Harqua-Gunsight complex	Gunsight	0.64999998	315322	3543858
Harqua	Harqua-Gunsight complex	Gunsight	0.90399998	314839	3543798
Torrifluents	Harqua-Gunsight complex	Gunsight	0.44600001	314478	3543796
Gunsight	Harqua-Gunsight complex	Gunsight	0.47000000	314098	3543772

Gunsight	Torrifluvents	Gunsight	0.47200000	313569	3543193
Torrifluvents	Torrifluvents	Torrifluvents	0.77600002	313482	3543301
Gunsight	Gunsight very gravelly loam, 2 to 15 percent slopes	Gunsight	0.75800002	319460	3545892
Harqua	Harqua very cobbly loam, 0 to 8 percent slopes	Harqua	0.51200002	314960	3533661
Gilmansaline	Harqua very cobbly loam, 0 to 8 percent slopes	Harqua	0.45199999	314925	3533616
Antho	Antho fine sandy loam	Antho	0.75800002	309841	3536616
Antho	Antho fine sandy loam	Antho	0.46799999	309848	3536224
Gunsight	Gunsight very gravelly loam, 2 to 15 percent slopes	Gunsight	0.87599999	326034	3530811
Gunsight	Torrifluvents	Gunsight	0.74400002	328502	3530054
Gilman	Gilman very fine sandy loam	Gilman	0.69000000	324603	3564179
Gilman	Gilman very fine sandy loam	Gilman	0.85399997	326716	3563525
Gilman	Gilman very fine sandy loam	Gilman	0.87000000	326920	3563607
Antho	Antho fine sandy loam	Antho	0.65600002	328026	3563224
Growler, Harqua*	Harqua-Gunsight complex	Harqua	0.76400000	333186	3555192
Cherioni	Cherioni gravelly very fine sandy loam, 0 to 8 percent slo	Cherioni	0.66399997	319036	3562739
Growler, Harqua, Cipriano**	Antho fine sandy loam	Antho	0.53200001	316097	3561232
Antho	Antho fine sandy loam	Antho	0.97200000	310658	3560197
Torrifluvents	Torrifluvents	Torrifluvents	0.64600003	308876	3559551
Antho, Growler*	Growler-Antho complex	Gilman	0.61799997	305501	3557651
Antho	Growler-Antho complex	Harqua	0.23000000	303285	3556843
Growler	Growler-Antho complex	Growler	0.98600000	308683	3559372
Growler	Growler-Antho complex	Growler	0.97200000	307299	3558290
Perryville	Perryville very cobbly fine sandy loam, 0 to 8 percent slo	Perryville	0.85200000	339095	3524608
Harqua	Harqua very cobbly loam, 0 to 8 percent slopes	Gunsight	0.65600002	336484	3525758
Harqua	Gunsight very gravelly loam, 2 to 15 percent slopes	Gunsight	0.92400002	335024	3526551
Lomitas	Gilman very fine sandy loam, saline	Gilmansaline	0.81000000	333719	3527096
Gunsight	Gunsight very gravelly loam, 2 to 15 percent slopes	Gunsight	0.99800003	334702	3529390
Harqua	Gunsight very gravelly loam, 2 to 15 percent slopes	Gunsight	0.86199999	328438	3530418
Gunsight	Gunsight very gravelly loam, 2 to 15 percent slopes	Gunsight	0.94199997	328439	3530510
Gachado	Rock land	Rock land	0.92000002	329176	3540016
Ajo	Ajo very gravelly loam, 1 to 5 percent slopes	Ajo	0.77399999	335019	3550609
Rillito	Rillito gravelly sandy loam	Rillito	0.98400003	333261	3556984
* Site where multiple soils could be identified within the pixel using high resolution imagery to distinguish desert varnish, and corroborating with field notes					
* Site where field notes could only narrow series down to a multiple possibilities.					

11 APPENDIX D CHAPTERS 5 AND 6 DATA

Table D.1 Field transect observations used to train spodic probability model.

UTM East	UTM North	Overstory Species	Understory Species	Spodic Intensity
610423	4287489	ACRU,PRSE2,FAGR	FAGR	0
610879	4287350	PRSE2,FAGR,ACRU,BELE	FAGR	1
610527	4275805	BEALA,TSCA,PIRU,FAGR	RHODO,KALA,TSCA	2
618191	4304293	MAFR,PRSE2,ACRU	TSCA,FAGR	2
618161	4304522	PRSE2,MAFR,ACRU,TSCA	TSCA,FAGR,PIRU,BELE	1.5
617961	4304786	ACRU,PRSE2,BELE,FAGR	TSCA,ACRU,PIRU,BELE	2
609123	4287532	ACRU,PRSE2	FAGR,ACRU	2
618758	4304025	ACSA3,QURU,TIAM,BELE,FAGR,PRSE2	FAGR,BELE,ACSA3,TSCA	0
609832	4287632	PRSE2,ACRU,ACSA3,MAAC	ACRU,ACPE,FAGR	1
610378	4287784	PRSE2,FAGR,ACRU,BELE,MAAC	FAGR	1
609588	4287283	ACRU,PRSE2,BELE,FAGR,MAFR	FAGR,ACRU,ACPE	1
610913	4287819	PRSE2,BELE,ACSA3,FAGR	FAGR	0.5
610946	4287570	PRSE2,ACRU,MAAC,	ACSA3,ACRU,FAGR,PIRU	1
610965	4280342	PIAB (plantation)	--	0
618189	4303902	PRSE2,LITU,FRAM2,ACSA3	ACSA3,FAGR,TSCA	0
609558	4274529	TSCA,PIRU,PRSE2	--	2
611686	4276254	TSCA,PIRU	RHODO,FAGR	2
610100	4282778	TSCA,PIRU,FAGR	--	2
612764	4259104	BELE,BEALA,QUVE	PRSE2,ACPE,BELE	0
612823	4259765	QURU,FAGR	ACPE,PRSE2,BEALA	1
613345	4260916	ACSA3,BEALA,FAGR,PRSE2	ACSA3,HAMAM,TSCA	0
612047	4258925	ACSA3,PRSE2,PRSE2,QURU	ACPE,FAGR	0.5
612792	4259179	ACRU,ACSA3,BELE	FAGR,PIRU	0
613844	4259897	QURU,ACRU,FAGR,PIRU	FAGR	1
612583	4258750	ACSA3,BEALA,FAGR,PRSE2	ACPE,FAGR	1
612867	4259201	PRSE2,BEALA,TIAM,ACSA3	ACPE,ACSA3,BEALA	0
613016	4260446	QUPR2,ACRU,QURU,CARYA	KALA,ILMO	0
613236	4258997	TSCA,PIRU,BEALA	PIRU,ILMO,TSCA	2
613157	4259243	PIRU,BELE,TSCA	PIRU,BELE,ACPE	2
613078	4259258	TSCA,PIRU,BEALA	PIRU,TSCA,ACPE	0
612102	4259181	PIRU,ACRU,PRSE2	FAGR,ILMO,PIRU	1
611856	4259258	PIRU,FAGR,ACSA3	ACPE,PRSE2,PIRU	1
617660	4273403	PIRU,BEALA,TSCA,QURU	FAGR,ACPE,PIRU	1
612579	4260233	ACRU,QURU,BELE,QUPR2	RHODO,ACPE,ILMO	1
613354	4259039	PIRU,BEALA,TSCA	ACPE,ILMO,PIRU	1
612058	4259212	PIRU,BELE,TSCA,ACRU	TSCA,FAGR	1
612917	4258930	PIRU,TSCA,BEALA	TSCA,PIRU,PRSE2,ILMO	1

603918	4276074	BEALA,PRSE2	FAGR	0
605519	4276832	PRSE2,TSCA,BEALA	FAGR,TSCA	0
607618	4278563	FAGR,ACRU,PRSE2	FAGR	0
607999	4279066	ACRU,PRSE2,FAGR	FAGR	0
605047	4276878	PRSE2,ACSA3	FAGR	0
607385	4278874	PIRU,TSCA,PRSE2	ACRU,FAGR	0
604365	4275913	PRSE2	FAGR,ACPE	0
605191	4276581	PRSE2,TSCA	TSCA,FAGR	0
616295	4297332	PRSE2,MAFR,ACRU	FAGR	0
606957	4278328	ACSA3,PRSE2,BEALA	ACPE,ACSA3,PRSE2	0
606374	4278375	FAGR,PRSE2,ACSA3	ACPE,PRSE2,ACSA3	1
610651	4287751	PRSE2,ACSA3	FAGR,ACRU	0
604366	4276206	BEALA,FAGR,TSCA	--	2
604843	4276891	ACSA3,PRSE2	FAGR	0
606096	4277653	ACSA3,BEALA,PRSE2,ACRU	FAGR	1
606421	4277368	ACSA3,BEALA,TSCA,PRSE2	FAGR	0
611767	4290419	PRSE2,ACRU,MAAC,PRSE2	FAGR,AMARA4	0
605434	4279050	FAGR,TSCA,ACRU,PIRU	FAGR,TSCA	0.5
606396	4278141	FAGR,BEALA,ACSA3	PIRU,PRSE2,ACPE	0
606707	4278423	ACSA3,BEALA,PRSE2,MAFR	FAGR,ACSA3,PIRU	0
604435	4278174	PIRU,TSCA,MAFR,ACRU	--	1
623132	4284893	PRSE2,ACSA3,ACRU,PRSE2	FAGR	1
622806	4285042	FRAM2,ACSA3,PRSE2	FAGR,ACPE,ACSA3	0
604572	4277828	PIRU,TSCA,ACRU,BEALA	MAFR,PRSE2,PIRU	2
604774	4278499	PIRU,TSCA,BEALA,FAGR	FAGR,PIRU,BEALA	2
605876	4274297	PRSE2,FRAM2,QURU,BELE	FAGR	0
604365	4279095	TSCA,PIRU,ACRU,PRSE2	RHODO	2
605452	4277497	PIRU,TSCA,FAGR,BEALA	ACPE,PRSE2,BEALA	0
605766	4274345	PRSE2,BELE	--	1
604980	4277358	TSCA,PIRU,BEALA,MAFR	FAGR,TSCA	2
606133	4273954	TSCA	PIRU,FAGR	2
611455	4291133	PRSE2,MAFR,ACRU	MAFR,FAGR,ACRU	1
611710	4290936	ACRU,PRSE2,FAGR	FAGR,ACRU	1
612011	4291022	PRSE2,ACRU	FAGR,ACPE	1
616077	4297676	MAFR,PRSE2,BEALA	FAGR	1
616372	4299356	BELE,PRSE2,FAGR	FAGR,BELE,PRSE2,TIAM	0
612100	4291283	ACSA3,PRSE2,TIAM,FRAM2	ACSA3,FAGR	0
616259	4299545	FRAM2,PRSE2ACRU,PIRU	FAGR,PIRU,ACRU	1
611950	4291479	FAGR,BEALA,ACRU,PRSE2	FAGR	1
611929	4291508	PRSE2,ACRU,TSCA,PIRU,FAGR,BEAL2		1
612379	4290351	FAGR,PRSE2,BEALA,ACRU	FAGR	0
609975	4287941	BEALA,ACSA3,ACRU	FAGR,ACPE,ACRU,TSCA	0
616327	4299550	PRSE2,BELE,ACRU,PIRU	BELE,FAGR,PIRU	1

612750	4289255	PRSE2,ACSA3	ACPE,FAGR	1
612769	4289636	MAFR,PRSE2,BEALA,ACRU	FAGR	2
613631	4291046	PRSE2,TIAM,ACSA3	BEALA,ACSA3,FAGR	0
614638	4292431	PRSE2,ACSA3,MAFR	FAGR	0
615382	4294245	PRSE2,ACRU,MAFR	FAGR	2
615223	4295122	CARYA, FRAM2	ACSA3,FAGR,ACPE	0.5
614702	4295654	FAGR,BEALA,ACSA3	FAGR,ACPE	2
613962	4296079	FAGR,ACSA3,PRSE2	FAGR	0
614719	4296279	ACRU,MAFR,BEALA,BELE	ACRU,BELE,BEALA,PRSE2,TSCA	2
615688	4296330	PRSE2,MAFR,BEALA,ACRU	TSCA,ACRU,FAGR	2
615192	4296709	FAGR,ACSA3,PRSE2	KALA,PIRU	0.5
614639	4296774	FAGR,MAFR,ACSA3	ACPE,PIRU	0
617727	4303482	PRSE2,TSCA,FAGR,PIRU	PIRU	1.5
615444	4296824	PRSE2,FAGR,ACRU	FAGR	0
616520	4301738	PRSE2,MAFR,BELE,PIRU	PIRU,TSCA	0.5
617112	4302307	nhardwoods	TSCA,PIRU	0.5
616406	4298237	ACRU,PRSE2,BEALA,BELE	ACPE,FAGR	1
616266	4298302	PRSE2,TSCA,PIRU,ACRU	ACPE,PRSE2,PIRU,TSCA	0
615801	4298420	PRSE2,MAFR,TSCA,ACRU	TSCA,BELE,PIRU	2
613935	4298515	PRSE2,FAGR,TIAM,TP,MAAC	--	0
615390	4297765	PRSE2,FAGR,BELE,ACRU,MAFR	--	1
611992	4291014	PRSE2,FAGR,ACSA3	FAGR	1
615960	4299817	BELE,PRSE2,FAGR	FAGR,ACPE	1
618408	4302279	PRSE2,TSCA,BEALA	--	0
618476	4300599	MAFR,ACRU,FAGR	BELE,MAFR,ACRU	1
623098	4285775	FAGR	FAGR	1
623624	4284963	nhardwoods (old pasture)	--	1
617725	4302374	nhardwoods	TSCA,FAGR,PIRU	0.5
617234	4299619	nhardwoods, PIRU, TSCA	PIRU,TSCA,FAGR	1
615118	4295232	nhardwoods,PIRU	--	0
614645	4295978	nhardwoods,PIRU,TSCA	PIRU,TSCA	1
622730	4284200	nhardwoods	PIRU	1
623577	4285375	ACRU,ACSA3	FAGR,ACPE	0
610440	4285889	PRSE2,BEALA,ACRU,ACSA3	ACPE,FAGR	0
610832	4286696	ACSA3,MAFR,PRSE2,ACRU	FAGR,ACSP2	0
623347	4284962	FAGR,PRSE2,ACSA3	FAGR	0
611355	4287872	PRSE2,FAGR,ACSA3	FAGR	2
611457	4288320	PRSE2,ACSA3,FRAM2	ACSA3,FAGR	0
609624	4288353	PRSE2,BEALA,ACSA3	FAGR,PIRU	0
610709	4288800	ACSA3,PRSE2,FAGR	FAGR	1
609834	4289226	PRSE2,ACSA3	FAGR	0
622480	4284419	ACSA3,PRSE2	FAGR	0
612232	4291307	PRSE2,BELE,ACSA3,PRSE2	ACSA3,FAGR	0

612234	4291562	PRSE2,PRSE2,BELE,TSCA	FAGR,ACSA3	0
609596	4289643	PRSE2,FAGR,ACSA3	FAGR	0.5
610088	4290948	PRSE2,FAGR,MAFR	FAGR,BELE	2
609585	4289900	PRSE2,FAGR,ACSA3	FAGR,ACPE	0.5
609963	4290445	FAGR,ACSA3,PRSE2	FAGR	0.5
609399	4290857	PRSE2,FRAM2,BELE	ACSA3,ACSP2,FAGR	0.5
609562	4290921	MAFR,PRSE2,BELE	FAGR	1
609398	4291001	PRSE2,BEALA,ACRU	FAGR,ACSA3,PIRU	1
610251	4291160	MAFR,PRSE2,ACSA3	FAGR	2
608250	4285909	FAGR,ACRU,PRSE2	PIRU,ACSP2	1.5
618149	4299502	FAGR,PRSE2,ACRU,BEALA,PIRU	--	2
618766	4299212	PRSE2,FAGR,ACRU	--	2
617234	4296416	ACRU,PRSE2,MAFR,ACSA3	ACPE,ACSA3,FAGR	0
617957	4296850	PRSE2,ACSA3,ACRU,BELE	FAGR,ACRU	0
617847	4297094	clearcut area	FAGR,PRSE2,ACPE	0
617287	4297151	PRSE2,MAFR,ACRU	FAGR	2
617276	4297228	MAFR,FAGR,PRSE2,ACRU	FAGR,MAFR	1
618106	4297246	PRSE2,BELE,ACRU,TSCA	FAGR,ACRU	0
618238	4297253	PRSE2,BELE,ACRU	TSCA,FAGR,PIRU	0
617647	4297277	PRSE2,ACSA3,MAFR	FAGR	0
616945	4297413	PRSE2,ACRU,FAGR	FAGR	0
616982	4297530	FAGR,PRSE2,ACSA3,BEALA	FAGR	0
617734	4298075	PRSE2,ACRU,ACSA3,MAFR	FAGR,ACPE,PIRU	1
614765	4299247	ACRU,BELE,TSCA,PIRU	PIRU,RHODO,FAGR,TSCA	2
614143	4299369	BELE,PRSE2,ACRU,MAFR	FAGR,PIRU	1
614133	4299538	BELE,PRSE2,ACRU,BEALA	RHODO,FAGR,PIRU	2
614391	4299798	PRSE2,BELE,MAFR,TSCA,PIRU,LITU	FAGR,PIRU	0.5
614588	4299943	PRSE2,BELE,ACRU,LITU,MAAC	FAGR,BELE,PIRU	1
614915	4299921	PRSE2,BELE,LITU,FAGR	FAGR,BELE,PRSE2	1
615072	4299921	PRSE2,FAGR,BELE,ACRU,MAFR	FAGR	1
615825	4293131	PRSE2,ACRU,MAFR	FAGR	0
616271	4292872	ACRU,PRSE2,MAFR	FAGR,MAFR	2
616517	4292697	PRSE2,ACSA3,ACRU	FAGR,ACSA3	0
616811	4292433	PRSE2,ACRU,BELE	FAGR,BELE	1.5
616231	4292477	ACRU,ACSA3,PRSE2,BELE	FAGR,ACPE	0
616022	4292381	PRSE2,BELE,ACRU	MAFR,FAGR	2
615386	4292492	ACRU,PRSE2,BELE,MAAC	FAGR	0
615121	4292421	ACSA3,FAGR,PRSE2	FAGR,ACPE	0
624060	4285316	FAGR,ACSA3,ACRU,ACPE	FAGR,ACSA3,ACRU,ACPE	1
612889	4258785	PRSE2,FAGR,BELE,TSCA	ACPE,FAGR	1
613267	4259119	PIRU,BEALA,TSCA,PRSE2,BELE	--	2
609496	4273861	PIRU,TSCA,MAFR,FAGR,BELE	PIRU	2
609515	4286127	PRSE2,MAFR,ACRU,BELE,FAGR	ACRU,FAGR	1

608488	4285853	PIRU,TSCA,PRSE2,MAFR,ACRU,BELE	--	2
616664	4301718	PRSE2,MAFR,ACRU,FAGR	PIRU,ACPE,FAGR,TSCA	1
618330	4274056	ACSA3,PRSE2,MAFR,PIRU	PIRU,KALA,PRSE2	1
623018	4285527	nhardwoods	PIRU,FAGR	1
617326	4273718	PRSE2,PIRU,BELE	PIRU,FAGR	0
623953	4286037	nhardwoods	FAGR	1
623114	4285879	FAGR	FAGR	1
624083	4285303	nhardwoods	FAGR	1
624263	4285303	nhardwoods	FAGR	0
623705	4284930	nhardwoods (old pasture)	--	1
624353	4285099	pasture	--	1
616889	4274620	FAGR,ACSA3,PRSE2	FAGR	0
616662	4272726	PIRU, PRSE2	ACPE,ACSA3,PIRU	1
623928	4285441	PRSE2	FAGR	1
622157	4284838	PRSE2,ACSA3	ACSA3,FAGR,ACPE	1
619395	4274394	PIRU,ACSA3	TSCA	0
610004	4275144	ACSA3,PRSE2,ACRU	FAGR	0
622824	4285502	BEALA,ACSA3,PRSE2,PRSE2	FAGR,ACSP2	1
622868	4284127	old pasture	FAGR	0
608278	4285784	FAGR, ACRU, PRSE2	PIRU,ACSP2	1.5
616747	4273868	PIRU,ACSA3,PRSE2	FAGR,ACSA3	1
623497	4285088	PRSE2,ACSA3	FAGR	0
622712	4283607	PRSE2,ACRU	FAGR	1.5
622149	4283158	BEALA,PRSE2,ACRU	ACPE	0
620335	4281006	--	PIRU	1
621855	4281060	PIRU	--	1
621541	4281833	PIRU,FAGR	--	1
622379	4285836	BEALA,TSCA,ACRU	FAGR,PIRU	2
620866	4287116	BELE,ACRU,TSCA	ACPS2,PIRU,ACRU,TSCA,FAGR	2
620265	4286857	BELE,ACRU,FAGR,PRSE2	BELE,ACRU,ACPS2	2
622405	4285609	BEALA,FAGR	PIRU,TSCA	2
621873	4285869	FAGR,MAFR,PIRU	FAGR,ACRU	2
621050	4286876	FAGR,ACRU,PRSE2,TSCA, BELE	BELE,ACPS2	2
621253	4286472	FAGR,ACRU,PRSE2	FAGR,PIRU,ACPS2	0
621253	4286472	PRSE2,ACSA3,FAGR,ACRU	FAGR,ACRU,ACPS2	0
622087	4285843	FAGR,PRSE2,BEALA	ACPS2,FAGR,PIRU	1
621992	4285371	ACRU,PIRU,FAGR	PIRU,ACRU,FAGR	2
620491	4286848	PRSE2,ACRU,BEAL2,PIRU	FAGR,ACPE	2
622564	4285299	FAGR,ACRU,PRSE2	FAGR	1
621687	4285578	ACSA3,PRSE2	ASCA3,BEALA,FAGR	1.5
621071	4286326	FAGR,PRSE2,ACSA3	FAGR	0
621153	4286393	FAGR,PRSE2,ACRU,PIRU,MAAC	PIRU,FAGR	1
620569	4286589	TSCA,BELE,ACRU	FAGR	2

622008	4285598	ACRU,BELE,BEAL2	FAGR,ACPS2,BELE	2
621736	4285868	ACRU,FAGR,PRSE2,BELE,MAAC	FAGR,BELE,BEALA,ACRU	1
620735	4286489	PIRU,ACSA3,MAFR,PRSE2	FAGR,PIRU,ACPS2	1
621533	4285888	ASCA3,FAGR	ACPS2,FAGR,PIRU,PRSE2	2
621847	4286085	TSCA,ASCA3,ACRU,BEALA	PIRU,ACRU	2
621049	4286718	TSCA,PIRU,PRSE2	FAGR	2
608357	4285684	FAGR,ACRU,PIRU	PIRU,FAGR	2
608547	4285157	FAGR,MAAC,ACSA3	FAGR	0
605072	4277143	BEALA,ACSA3	FAGR	2
608430	4285081	FAGR,PRSE2,FRAM2,ACSA3	FAGR	0
607585	4286148	BELE	FAGR,BELE,PIRU	1
607719	4287294	PRSE2,ACSA3,FAGR,BEALA	FAGR,PIRU	2
608102	4287561	PRSE2,ACSA3,FAGR,BEALA	FAGR	0
609636	4285421	FAGR,PRSE2,BEALA,ACSA3	FAGR,MAFR	2
606645	4285456	TSCA,BEALA,PRSE2	PIRU,FAGR,ACSA3	2
606142	4278869	PIRU,TSCA,ACRU,MAFR	MAFR,BEALA,PIRU,ACPE	2
605184	4278453	PIRU,TSCA,FAGR,ACRU	--	2
607383	4285908	PRSE2,BEALA,FRAM2	ACSA3,ACPE,BEALA	0.5
609153	4286071	BEALA,BELE,PRSE2,ACSA3	BEALA,ACSA3	0
607676	4287106	FAGR,PRSE2,ACSA3,BEALA	TSCA,FAGR	1
607853	4287865	PRSE2,ACSA3,FAGR	PIRU,FAGR	1
607289	4283853	PRSE2,FAGR,PRSE2	FAGR	2
609404	4285511	BEALA,MAFR,PRSE2	PIRU	2
609640	4285610	FAGR,BELE,ACSA3	ACPE	0
606483	4284271	PRSE2,ACSA3,BEALA	FAGR,PIRU	2
605879	4284649	BEALA,PIRU,ACSA3	ACSA3,PIRU	1
607605	4287426	PRSE2,FAGR,BEALA,ACSA3	PIRU,FAGR	0
606865	4285405	PRSE2,BEALA,ACSA3	ACPE,TSCA	1
608881	4286155	BEALA,PRSE2	BEALA	0
605140	4283310	TSCA,PIRU,PRSE2,ACSA3	RHODO	2
605665	4283723	TSCA,ACSA3,BEALA,MAFR	PIRU,FAGR,BEALA	0.5
605207	4283256	TSCA	--	0
605731	4284539	PIRU,TSCA,BEALA,PRSE2	PIRU	2
606859	4285668	PIRU,TSCA,BEALA	PIRU	2
608487	4285851	PIRU,TSCA,BEALA,ACRU	PIRU,TSCA	2
606937	4285773	PIRU,TSCA,BEALA	PIRU	2
607036	4283875	PIRU,FAGR,TSCA,PRSE2	PIRU,FAGR	2
606198	4285196	PIRU,TSCA,BEALA	FAGR,PIRU,TSCA	2
606709	4285480	PIRU,TSCA,FAGR,BEALA	FAGR,BEALA	2
617660	4273403	PRSE2,ACSA3	--	2
616247	4273554	ACSA3,PRSE2,TIAM	FAGR,ACSA3	0
606375	4278628	FAGR,ACSA3,PRSE2,MAFR	FAGR,PIRU	2
616729	4274651	FAGR,BEALA,PRSE2,ACSA3	PIRU,FAGR	0

615566	4274729	MAFR,ACSA3,PRSE2	PRSE2,ACPE	0
615643	4275067	ACRU,FAGR,PRSE2,ACSA3	FAGR,ACSA3	0
616940	4273069	ACSA3,PRSE2,BEALA	KALA,ACPE	0
607241	4278800	TSCA,PIRU,FAGR,ACRU	TSCA,PIRU,FAGR	2
617320	4274570	PRSE2,FAGR,ACSA3	ACPE,FAGR	1
614957	4274684	PRSE2,ACSA3,BEALA,MAFR	PRSE2,ACPE	0
615216	4274687	--	--	0
614774	4274777	ACSA3,BEALA,PRSE2	BEALA,FAGR	1
616939	4272662	ACSA3,FAGR,QURU,BEALA	ACPE	0.5
615130	4274002	PRSE2,FAGR,ACRU,BEALA	FAGR,ACSP2	2
618849	4274049	QURU,PRSE2,PIRU,TSCA	PRSE2,TSCA,PIRU	0
617740	4273988	ACSA3,PIRU,PRSE2	FAGR,ACSP2	2
616570	4274725	PIAB,ACSA3	FAGR	0
604406	4277585	PIRU,TSCA,BEALA,MAFR	PRSE2,TSCA,MAFR	2
616242	4273544	PRSE2,TSCA,FAGR,ACSA3	FAGR,ACSA3	0
617090	4274388	ACSA3,PRSE2,BEALA,PIRU	ACPE,PRSE2,BEALA	0
607558	4278943	TSCA,PIRU,BEALA,FAGR	TSCA,ACRU	2
611973	4288794	PRSE2,ACRU,FAGR	ACPE,PRSE2,ACRU,PRSE2	2
617597	4272786	ACSA3,PIRU,PRSE2,BEALA	LIBE3,HAVI4	1
617206	4274380	PIRU,ACSA3,PRSE2	PIRU,FAGR	1
615705	4296774	PRSE2,ACRU,MAFR	FAGR,MAFR	2
615778	4274964	PIRU,ACSA3	PIRU,FAGR	1
615819	4274924	PIRU,ACSA3,TSCA	KALA	1
616441	4274610	PIRU,TSCA,ACSA3	--	0
614927	4274763	PIRU,FAGR,BEALA	PIRU,PRSE2,BEALA	0
615520	4275068	PIRU,FAGR,BEALA,ACSA3	PIRU,FAGR	1
608426	4272178	ACRU,QURU,FAGR,PRSE2	ACPE,PIRU,FAGR	0
608833	4273568	ACRU,NYSY,PRSE2,FRAM2	FAGR,NYSY,ACPE	0
609854	4274781	ACSA3,ACRU	ACSP2,FAGR,ACSA3	0.5
608542	4272522	BEALA,TSCA,FAGR,QURU	ACPE,FAGR,TSCA	1
608785	4272815	FAGR,QUVE,BEALA,ACRU	FAGR,TSCA	0.5
608534	4272756	QURU,FAGR,ACRU	FAGR,ACPE,PIRU	0
610284	4274454	TSCA,FAGR,ACRU,MAAC	TSCA,FAGR,ACPE	1
610659	4275420	PRSE2,TSCA,FAGR	TSCA,BEALA,FAGR	1
609318	4275973	PRSE2,BEALA,TSCA,QURU	FAGR,ACRU,BEALA	2
609505	4272030	QURU(clearcut)	FAGR,BEALA,PIRU	0
608774	4273274	ACRU,NYSY	ACRU,PRSE2,FRAM2	0
609701	4272168	PRSE2,ACSA3,QURU,MAFR	FAGR, ACPE	0
622936	4285487	nhardwoods	PIRU	2
610075	4271902	QURU,QUPR2,ACSA3,PRSE2	FAGR,ACRU	0
608713	4273329	LITU,ACRU,NYSY	ACPE,ACRU	0.5
610616	4276182	FAGR,TSCA,PIRU,BEALA	ACPE,TSCA,FAGR	1
609584	4274904	TSCA,FAGR,BEALA	PIRU,KALA,TSCA	2

609713	4272510	PIRU,TSCA,BEALA,ACRU	BEALA,PIRU,PRSE2,ACPE	1
609455	4273876	PRSE2,TSCA,ACRU,PRSE2,BEAL2	TSCA,FAGR	2
609402	4274193	TSCA,PRSE2,BEALA	FAGR,TSCA,KALA	2
606743	4283670	FAGR,MAAC,ACSA3,PRSE2	FAGR,ACSP2	2
609763	4272393	PIRU,ACRU,BEALA,TSCA	BEALA,PRSE2,PIRU,ACPE	1
609104	4273026	MAAC,TSCA,FAGR,PRSE2,ACRU	FAGR,TSCA,MAAC	2
610709	4276769	PIRU,TSCA,BEALA	FAGR,PIRU	1
609500	4274678	BEALA,FAGR,TSCA	PIRU,TSCA,RHODO	2
610952	4280393	PIAB (plantation)	--	2
609936	4279177	PIRU,ACRU	ACPS2	2
609849	4279269	ACRU,PRSE2, BEALA	FAGR, PIRU, SMILA2	1
607752	4276779	PRSE2,ACSA3,QURU, FRAM2,TIAM	FAGR,ASCA3, POAC4	0
611344	4289817	PRSE2,ACSA3,ACRU	--	1
610829	4289686	PRSE2,ACRU,MAAC	FAGR,ACRU	1
610933	4289512	PRSE2,ACRU,ASCA3	--	1
610900	4289453	PRSE2,MAFR,ACRU,ASCA3	--	1
610829	4289314	PRSE2,MAFR,ACRU,PIRU,TSCA	PIRU,ACRU,TSCA,FAGR, SMILA2	2
611370	4289748	clearcut area	--	1
611391	4289735	clearcut area	--	1
611402	4289711	clearcut area	--	0.5
611454	4289688	clearcut area	--	0.5
611505	4289674	clearcut area	--	1
611525	4289651	clearcut area	--	0
611577	4289741	clearcut area	--	1
611565	4289744	clearcut area	--	1.5
611652	4289707	clearcut area	--	1
616733	4274290	--	--	1

Table D.2 Forest plot data

Spodic Probability	Soil Series	Forest Composition	Conifer Importance	Charcoal presence	Spodic Intensity	Tree rings low	Tree rings high	Tree rings ave	Ave O-horizon thickness (cm)	UTM East	UTM North
0.76	Wildell	CN	0.98	0	2	106	130	120	11.1	609520	4273882
0.78	Wildell	MX	0.46	0	2	80	103	95	11.5	621071	4286720
0.68	Hazleton	HW	0.10	0	1	53	108	79	3.4	608391	4272192
0.81	Kinzua	MX	0.47	0	1	107	184	141	8.3	606174	4273927
0.29	Carrollton	HW	0.00	0	0	101	109	106	2.3	605875	4274308
0.58	Macove	HW	0.00	0	1	90	103	96	6.9	607829	4276852
0.56	Kinzua	HW	0.15	0	1	68	77	74	5.7	618835	4274041
0.68	Carrollton	HW	0.00	1	1	82	99	91	7.7	617669	4273382
0.46	Carrollton	HW	0.26	0	0	82	99	90	2.6	606866	4274929
0.77	Gauley	CN	0.86	1	2	52	72	65	12.1	615788	4274931
0.89	Blandburg	HW	0.22	0	2	82	96	88	7.4	612341	4275296
0.97	Blandburg	CN	0.82	0	2	143	158	151	18.4	604411	4277561
0.66	Mandy	MX	0.44	0	1	94	121	108	7.3	605444	4277486
0.86	Wildell	HW	0.03	1	2	98	111	105	14.5	621634	4279795
0.91	Blandburg	MX	0.66	1	2	103	266	166	6.5	610595	4276185
0.90	Mandy	HW	0.00	0	1	47	65	56	3.0	620331	4281008
0.71	Wildell	CN	0.86	0	2	105	110	110	26.8	607563	4278901
0.40	Hazleton	HW	0.00	0	0	39	41	40	2.5	606699	4278411
0.78	Gauley	CN	1.00	1	2	55	64	60	12.5	616716	4278160
0.66	Wildell	HW	0.00	0	2	82	127	103	8.6	612338	4276414
0.52	Mandy	HW	0.02	0	1	0	0	0	4.8	608534	4273436
0.64	Mandy	HW	0.00	1	1	0	0	0	2.4	608752	4275088
0.43	Gauley	MX	0.69	1	2	109	153	132	14.0	608760	4274340
0.53	Macove	HW	0.00	1	0	0	0	0	2.7	619513	4277167

Table D.3 Laboratory data documenting Al and Fe acid oxalate extractions for different spodic intensity classes.

Pedon ID	Horizon Middle depth (cm)	Al+0.5Fe (% mass)	Spodic Intensity	Analysis Location
S12WV075002	14	0.08	2	NRCS NSSL
S12WV075002	22	1.38	2	NRCS NSSL
S12WV075002	36	1.23	2	NRCS NSSL
S12WV075002	57	0.2	2	NRCS NSSL
S12WV075002	79.5	0.19	2	NRCS NSSL
S12WV075002	100	0.16	2	NRCS NSSL
S12WV075003	5.5	0.27	2	NRCS NSSL
S12WV075003	13.5	0.14	2	NRCS NSSL
S12WV075003	22.5	1	2	NRCS NSSL
S12WV075003	30	1.15	2	NRCS NSSL
S12WV075003	52.5	0.96	2	NRCS NSSL
S12WV075003	88	0.58	2	NRCS NSSL
S12WV075005	15	0.38	2	NRCS NSSL
S12WV075005	26	0.47	2	NRCS NSSL
S12WV075005	33	1.99	2	NRCS NSSL
S12WV075005	43.5	1.79	2	NRCS NSSL
S12WV075005	57	0.73	2	NRCS NSSL
S12WV075005	70.5	0.57	2	NRCS NSSL
S12WV075005	98	0.41	2	NRCS NSSL
S12WV083001	6	0.73	1	NRCS NSSL
S12WV083001	18	0.96	1	NRCS NSSL
S12WV083001	40	0.8	1	NRCS NSSL
S12WV083001	65	0.34	1	NRCS NSSL
S12WV083001	100	0.3	1	NRCS NSSL
S12WV075001	4.5	0.53	1	NRCS NSSL
S12WV075001	10	0.76	1	NRCS NSSL
S12WV075001	16	0.7	1	NRCS NSSL
S12WV075001	30	1.39	1	NRCS NSSL
S12WV075001	73.5	0.4	1	NRCS NSSL
S12WV075001	100	0.24	1	NRCS NSSL
S12WV075004	6.5	0.39	1	NRCS NSSL
S12WV075004	11.5	0.71	1	NRCS NSSL
S12WV075004	30.5	0.94	1	NRCS NSSL

S12WV075004	65	0.37	1	NRCS NSSL
S12WV071001	5	0.5	1	NRCS NSSL
S12WV071001	10	0.99	1	NRCS NSSL
S12WV071001	23.5	1.16	1	NRCS NSSL
S12WV071001	46	1.32	1	NRCS NSSL
S12WV071001	75	0.57	1	NRCS NSSL
130611-0900	18.5	0.299	0	WVU
130611-0900	42.0	0.491	0	WVU
130611-0900	80.0	0.383	0	WVU
130611-0900	115.0	0.301	0	WVU
130710-1000	9	0.750	0	WVU
130710-1000	18.5	0.761	0	WVU
130710-1000	37	0.805	0	WVU
130710-1000	61.5	0.848	0	WVU
130710-1000	87.5	0.673	0	WVU
130710-1000	121.5	0.199	0	WVU
130910-0900	6.5	0.802	0	WVU
130910-0900	23.5	0.689	0	WVU
130910-0900	51.0	0.569	0	WVU
130910-0900	78.0	0.420	0	WVU

Table D.4 O-horizon carbon stock data

Horizon	Pedon ID							average OC	OC Std. Dev.
	s12wv08 3001	s12wv07 5005	s12wv07 5004	s12wv07 5003	s12wv07 5002	s12wv07 5001	s12wv07 1001		
Organic Carbon (OC) (fraction of sample mass)									
Oi	0.402	0.335	0.543	0.540	0.440	0.572	0.494	0.475	0.0800
Oe	0.307	0.545	0.475	0.352	0.540			0.444	0.0977
Oa	0.473	0.404	0.359	0.487				0.431	0.0521
Bulk Density (BD) (g/cm ³)								average BD	BD Std. Dev.
Oi	0.080	0.050	0.040	0.050	0.090	0.070		0.063	0.0180
Oe	0.090	0.100	0.130	0.110	0.010			0.088	0.0412
Oa	0.170	0.106	0.231	0.383				0.222	0.1027
Thicknesses (cm)									
Oi	3	6	1	1	2	1	3		
Oe	0	7	4	3	3	1	0		
Oa	0	7	2	2	5	0	0		
total	3.000	19.600	7.000	5.650	10.000	2.000	3.000		
	Fraction of O total horizon thickness					Averages	g C/ cm ³ soil	weighted g C/ cm ³ soils	wt ave g C /cm ³ soil
Oi	0.306	0.143	0.177	0.200	0.206	0.0301	0.0062	0.0573	
Oe	0.357	0.571	0.531	0.300	0.440	0.0390	0.0172		
Oa	0.337	0.286	0.292	0.500	0.354	0.0958	0.0339		
						Std. Dev.			
						Oi	0.061		
						Oe	0.114		
						Oa	0.087		