Spatiotemporal Modeling of Threats to Big Sagebrush Ecological Sites in Northern Utah

Alexander J. Hernandez
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Part of the Environmental Sciences Commons, Forest Sciences Commons, and the Plant Sciences Commons

Recommended Citation
https://digitalcommons.usu.edu/etd/957

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
SPATIOTEMPORAL MODELING OF THREATS TO BIG SAGEBRUSH

ECOLOGICAL SITES IN NORTHERN UTAH

by

Alexander J. Hernandez

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY
in

Ecology

Approved:

R. Douglas Ramsey
Major Professor

Gretchen G. Moisen
Committee Member

Christopher M. U. Neale
Committee Member

Ronald J. Ryel
Committee Member

Eugene W. Schupp
Committee Member

Byron R. Burham
Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2011
Copyright © Alexander Hernandez 2011

All Rights Reserved
ABSTRACT

Spatiotemporal Modeling of Threats to Big Sagebrush
Ecological Sites in Northern Utah

by

Alexander J. Hernandez, Doctor of Philosophy
Utah State University, 2011

Major Professor: Dr. R. Douglas Ramsey
Department: Wildland Resources

This study tested the performance of classification, regression, and ordination techniques to evaluate the spatiotemporal dynamics of threats to big sagebrush ecological sites. The research was focused on invasion by annual exotic grasses and encroachment by woodlands.

We sought to identify those areas that have had a persistent coverage of cheatgrass (Bromus tectorum) in big sagebrush ecological sites. We took advantage of the contrast in greenness between multi-temporal (within one year) remotely sensed vegetation indices captured in the spring and summer to find a distinct phenological signature that allowed mapping cheatgrass. We utilized support vector machines (SVM) to classify three temporal scenarios for which field data sets were available. SVM performed very well with accuracies of 70% (producer's) and 95% (user's) for the class of interest (presence of cheatgrass). This was the focus of chapter 2.

In chapter 3 we report the development of vegetation continuous fields (VCF) for
three years of interest 1996, 2001, and 2007 in order to detect active woodland encroachment. We prepared VCF for shrubs, trees, herbaceous vegetation, and bare ground using a suite of remotely sensed spectral reflectance, vegetation indices, and transformations. We compared the performance of multivariate regression trees (MRT) and random forests (RF) to develop the VCF multi-temporal series. RF outperformed MRT in both accuracy and ability to appropriately map the continuum of percent cover across large landscapes. We estimate that 17,570 hectares of big sagebrush lands showed encroachment by woodlands.

Our goal in chapter 4 was to develop a similarity index for large rangeland landscapes. Trend assessments field sites and a long-term annual series (1984 - 2008) of remotely sensed imagery were used in conjunction with multidimensional scaling (MDS) to measure ecological distance to undesired states such as invasion by exotic annuals and encroachment by woodlands. In this chapter our units of analysis were soil-mapping units, which were predominantly composed of one ecological site (>60%). Our MDS results show that different ecological sites can be identified in the reduced MDS statistical space. The observed transitions and trajectories of mountain, Wyoming, and basin big sagebrush sites correlated well with the ecological expectation in semiarid lands. We anticipate that managers can use our protocols to update ecological site descriptions and state and transition models from a remotely sensed perspective.
DEDICATION

To all my family members and good friends in Honduras, Costa Rica, and the United States. To my beautiful wife Florangeli and our children Alexandra, Josué and Mariangeles, and my mother Maria Teresa. To my grandma Felipa (1908 - 2007) and my brother Sergio Daniel (1964 - 2008), they passed away in Honduras while I was working for my degree. To all of them I dedicate this dissertation.
I have to say that none of this would have been accomplished without the help from many people. I have to thank my major professor, Doug Ramsey. Doug provided guidance and full support for my research here at USU. From the very beginning when we exchanged emails and telephone conversations in 2004 about the chance to come to Utah State for a doctoral degree, Doug furnished unparalleled assistance. He also supported me, and thus my family through assistantships during our stay here in Logan. He gave me the chance to be a research assistant in several projects in the RS/GIS lab, which expanded my professional skills and critical thinking, and he also provided me with the exceptional opportunity to be a teaching assistant and then instructor for his courses in natural resources. It has been quite a fulfilling experience. Gracias Señor! I also want to express my gratitude to the other members of my committee. I thank Gretchen Moisen for her interest to be on my revised committee and also for sharing her constructive opinions and experience in ecological modeling. Ron Ryel taught me about physiological ecology and provided valuable insight about integrating concepts with remote sensing. It was a splendid experience to work with Ron teaching monitoring and assessment. I also have to thank Eugene Schupp for his class on population ecology and all his important comments and criticisms on my papers. Christopher Neale's class on remote sensing of terrestrial surfaces provided me with a conceptual foundation that has helped me through these years of research.

I really need to say thanks from the bottom of my heart to everybody at the RS/GIS lab. Chris McGinty and Chris Garrard have always been there for me since I got here in 2005. They have always provided precious friendship, time to listen, and technical assistance. Thanks guys! Thanks to Ellie McGinty, Lisa Langs, and Ben Crabb too. John Lowry was
quite a good friend and was always there when needed. We suffered together when we were classmates and enjoyed discussing lab projects.

I thank Shane Green and Jamin Johanson at NRCS. Shane was an excellent source of information and discussion of rangeland ecology concepts when I was trying to put my ideas together. Jamin provided me with ESD and STM geospatial data sets and valuable ideas on how to improve my methodology. I also want to thank Elaine York at TNC for sharing field data from the project in the Grouse Creek Mountains.

I must recognize Samuel Rivera for being a good friend all these years. Samuel was my advisor for my B.Sc. He was helpful in bringing my papers to USU when I was trying to contact a professor to work with for my Ph.D. and he also helped me collect data in the desolate areas of Box Elder County. Gracias Samuel.

I definitely need to express my deepest thankfulness to my mother Maria Teresa in Honduras and my American parents. As a widow my mother managed to raise six kids alone with a lot of sacrifice and hard work, and Don and Nancy Glewwe have been there for me since I was 15. I am who I am because of my mother and Don and Nancy.

I need to say a huge thank you to my beautiful wife, Flor, and to my firstborn, Alexandra. These two girls have a lot of patience and love. They have been by my side relentlessly through three degrees in three different countries. Thanks to my son Josué and my baby girl Mariangeles; they have come to help Flor and Alexandra to be an endless fountain of joy to my life. Thanks God for putting them in my life and for all the blessings that I have received every moment and everywhere.

*Gracias a todos!*

Alexander J. Hernandez
## CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT.................................................................................................................................................. iii</td>
</tr>
<tr>
<td>DEDICATION............................................................................................................................................ v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS............................................................................................................................. vi</td>
</tr>
<tr>
<td>LIST OF TABLES........................................................................................................................................ ix</td>
</tr>
<tr>
<td>LIST OF FIGURES...................................................................................................................................... x</td>
</tr>
</tbody>
</table>

### CHAPTER

1. INTRODUCTION......................................................................................................................................... 1

2. USING SUPPORT VECTOR MACHINES AND REMOTELY SENSED DATASETS TO ASSESS DYNAMICS OF CHEATGRASS (BROMUS TECTORUM) EXTENT IN NORTHERN UTAH.......................................................................................................................... 6

3. MONITORING SEMI-ARID RANGELANDS WITH MULTI-TEMPORAL VEGETATION CONTINUOUS FIELDS: MULTIVARIATE REGRESSION TREES VS. RANDOM FORESTS....................................................................................................................................... 50

4. A LANDSCAPE SIMILARITY INDEX: MULTI-TEMPORAL REMOTE SENSING AND MULTI-DIMENSIONAL SCALING TO TRACK CHANGES IN BIG SAGEBRUSH ECOLOGICAL SITES....................................................................................................................... 95

5. SUMMARY AND CONCLUSION................................................................................................................ 144

APPENDIX..................................................................................................................................................... 148

CURRICULUM VITA.......................................................................................................................................... 158
LIST OF TABLES

Table                  Page
2-1. Landsat TM Path039 Row031 dates collected                        33
2-2. Explanatory variables compiled for modeling cheatgrass occurrence 34
2-3. Field sampling data sets used for model building and validation 35
2-4. Parameters γ and C used during modeling                           36
2-5. Areas classified as cheatgrass in the study area                  37
2-6. Validation data and metrics for 2001 and 2007                       38
2-7. Descriptors for the three classes of cheatgrass temporal dynamics 2001-2007 39
2-8. Cheatgrass temporal dynamics and Sagebrush land cover classes      40
3-1. Landsat TM Path039 Row031 dates collected                        77
3-2. Explanatory variables compiled for modeling CFV                   78
3-3. Field sampling data sets used for model building and validation  79
3-4. Summary of Multivariate Regression Trees                         80
3-5. Variables utilized to fit individual CFV using Random Forests     81
3-6. Validation metrics: Multivariate Regression Trees vs. Random Forests 82
3-8. Field points with observed woodland encroachment into big sagebrush areas 84
4-1. Big sagebrush ecological sites included in this study             125
4-2. DWR-RTS Benchmarks                                              126
4-3. Landsat TM Path 39 Row 31 scenes utilized in the study            127
4-4. USGS phenology data used in this research                         128
4-5. Synthesis of major disturbances extracted from DWR-RTS narratives 129
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Study area in Northern Utah and distribution of field observations. The study area is shown in the context of the State of Utah.</td>
</tr>
<tr>
<td>2-2</td>
<td>Sample of SAVI parallel coordinate plots to select the best two dates of imagery to model cheatgrass extent.</td>
</tr>
<tr>
<td>2-3</td>
<td>Box-plots of the SAVI values for the two dates per year chosen to maximize the contrast between the peak and die-off of cheatgrass. CG = Cheatgrass presence, NO-CG = Cheatgrass absence.</td>
</tr>
<tr>
<td>2-4</td>
<td>Random forest variable importance plot (a), and scatter plots for the middle infrared bands (b) for the year 2007.</td>
</tr>
<tr>
<td>2-5</td>
<td>Support Vector Machines classification maps using different options for gamma and cost.</td>
</tr>
<tr>
<td>2-6</td>
<td>Classified Cheatgrass extent for the year 1996.</td>
</tr>
<tr>
<td>2-7</td>
<td>Classified Cheatgrass extent for the year 2001.</td>
</tr>
<tr>
<td>2-8</td>
<td>Classified Cheatgrass extent for the year 2007.</td>
</tr>
<tr>
<td>2-9</td>
<td>Dynamics of Cheatgrass extent for the 2001 – 2007 period.</td>
</tr>
<tr>
<td>3-1</td>
<td>Study area in Northern Utah and distribution of field observations to model multitemporal VCF. The study area is shown in the context of the State of Utah.</td>
</tr>
<tr>
<td>3-2</td>
<td>Cross-Validated Relative Errors for different complexity parameter (cp) values to select the optimum tree size from a MRTS run (2001). A description of each component for the figure can be found in the text.</td>
</tr>
<tr>
<td>3-3</td>
<td>Tree structure for a MRTS run in 2001 – Notice that each final node is a multivariate composite response which can be decomposed into the VCF of interest.</td>
</tr>
<tr>
<td>3-4</td>
<td>Scatter-plots of observed versus predicted percent cover for shrubs and bare ground using Random Forest RF and Multivariate Regression Trees MRTS.</td>
</tr>
<tr>
<td>3-5</td>
<td>Maps of shrub percent cover for 2001 using Random Forests and Multivariate Regression Trees.</td>
</tr>
<tr>
<td>3-6</td>
<td>Maps of bare ground percent cover for 2001 using Random Forests and Multivariate Regression Trees.</td>
</tr>
</tbody>
</table>
3-7. Maps of trees and herbaceous percent cover for 2001 using Random Forests..............91

3-8. Changes in percent cover for (a) Trees, and (b) Shrubs from 1996 to 2007 for the buffers of the sites known to have woodland encroachment.................................................92

3-9. Potential woodland encroachment from 1996 – 2007 for the study area...................93

3-10. Woodland encroachment near Grouse Creek Mountains.........................................94

4-1. Distribution of big sagebrush ecological sites and benchmarks in the study area....130

4-2. Steady states plots for (upper) benchmarks and (lower) ecological site units........131

4-3. Scree plot for two MDS solutions to help determine the number of dimensions.....132

4-4. MDS solution for the period 1984 - 1996: (a) includes all the SMUs, (b) units with a major soil component (>60%)......................................................................................133

4-5. MDS solution for the 1997 - 2008 period. Arrows indicate the direction of change in vector position from 1996 - 2008...........................................................................134

4-6. MDS solution for the 1984-1996 and 1997 - 2008 periods for the DWR-RTS benchmarks. Arrows indicate the direction of change.........................................................135

4-7. Samples of multi-temporal distance plots for (a) distance from certain benchmarks, and (b) randomly chosen ecological site units.................................................................136

4-8. Samples of similarity maps to a benchmark: (a) TS-1-13 (cheatgrass), (b) TS-1-06 (woodland encroachment)......................................................................................137

4-9. Samples of similarity endmember SMUs to a cheatgrass benchmark (TS-13). Arrows indicate the magnitude of similarity (green=Dissimilar, red = Similar)...................138

4-10. Samples of similarity endmember SMUs to a Pinyon-Juniper benchmark (TS-06). Arrows indicate the magnitude of similarity (green=Dissimilar, red = Similar).............139

4-11. Similarity to PJ encroachment (benchmark TS-06). Section of SMU # 5 of R028AY215 with field point 07252007 (UTM East 320640, North 4583520). Photo by Alexander Hernandez.............................................................................................................140

4-12. Similarity to PJ encroachment (benchmark TS-06). Section of SMU # 60 of R028AY215 with field point 07112007 (UTM East 316947, North 4631891). Photo by Alexander Hernandez.................................................................141
4-13.  Similarity to cheatgrass invasion (benchmark TS-13). Section of SMU # 49 of R028AY215 with field point 07242007 (UTM East 322489, North 4633040). Photo by Alexander Hernandez.................................................................142

4-14.  Similarity to cheatgrass invasion (benchmark TS-13). Section of SMU # 2 of R028AY215 with field point 07262007 (UTM East 316256, North 4616998). Photo by Alexander Hernandez.................................................................143
CHAPTER 1

INTRODUCTION

Big Sagebrush (*Artemisia tridentata* ssp) vegetation communities occupy large areas of the western USA and provide a suite of valuable environmental services such as wildlife habitat, recreation, and hydrologic regulation, among others. However, their original spatial extent has been reduced, and their ecological conditions are declining in response to several natural and anthropogenic influences such as climate change, overgrazing, urbanization, invasion by exotics, and encroachment by woodlands (Wisdom et al. 2005a) just to mention a few examples. In the literature there are many documented examples of these negative influences and how they continuously interact to diminish the system’s ecological resistance and resilience. With regards to threats, two were the subject of our interest in this research: the invasion of exotic annual grasses with an emphasis in Cheatgrass (*Bromus tectorum*) that generally occurs on the warmer and drier low elevations, and the encroachment by woodlands (i.e. Pynion-Juniper) in the cooler and wetter highlands (Wisdom et al. 2005b). Due to the effects of these pressures on natural and modified big sagebrush communities, there is a great need to conduct monitoring and assessments of condition so that managers have precise and updated information to direct prevention and restoration activities based on the magnitude of the problem.

Ecological sites (U.S. Department of Agriculture 2008) and their associated state and transition models (Westoby et al. 1989) constitute an appropriate conceptual framework that may be used to conduct periodic monitoring and assessments. Units from the same ecological site are expected to produce the same type and amount of vegetation and respond similarly to disturbances. State and transition models are a relatively new paradigm that “describe the patterns, causes, and indicators of transitions between communities within an ecological site”
Because ecological site units are entities that are spatially correlated to soil survey mapping units (NRCS 2010), which are often large in size (i.e. hundreds of hectares); there exists the need to utilize monitoring methods that are able to cover large areas. These methods should have the ability to discriminate landscape-level attributes such as areas infested by annual grasses, and changes in percent cover and productivity that may be used as surrogates to assess rangeland health. These attributes can be quantitatively collected in a cost-effective manner with the suite of available remotely sensed products and processes that have been widely used to map and assess rangelands (Booth and Tueller 2003; Hunt et al. 2003).

Our work dealt with the generation of spatial attributes that may be used to identify current and past conditions of land affected by invasion of annual exotics and/or woodland encroachment in big sagebrush sites. Retrospective studies of land cover change were done to account for the spatial-temporal variation present in the continuum of sagebrush-cheatgrass and sagebrush-Pynion/Juniper associations found on semiarid landscapes. Here we provide details about work that was implemented to: (a) develop methodologies, and (b) test them in an area of interest in Northern Utah. We present and contrast results against similar published work, and we discuss the management implications of our results.

This dissertation is composed of 3 substantive chapters bounded by this introduction and overall conclusion chapters. In chapter 2 we explore the utilization of multi-temporal (within one year) vegetation indices and a relatively new statistical approach to classification: support vector machines (SVMs). We used the normalized difference soil-adjusted vegetation index NDSAVI and elevation data coupled with SVM to identify areas on the landscape with cheatgrass invasion. This was done for three non-consecutive years: 1996, 2001, and 2007. The focus of chapter 2 was to characterize the dynamics of cheatgrass in big sagebrush
communities in order to detect areas that have had a persistent cover of cheatgrass throughout the three years. Areas with variation (i.e. with cover one year but without on the other) of cheatgrass were also mapped. The discrimination and quantification of these areas is a step towards gaining a better understanding of which shrub communities seem to be more consistently affected by invasion of cheatgrass. Shrub sites with a persistent cheatgrass may be on the pathway to a reduction of plant diversity and eventually become a cheatgrass monoculture.

The focus of chapter 3 was to characterize the temporal variation of major plant communities in semiarid areas. Traditional methods of classifying the land (cover / use) into highly homogenous classes may not appropriately address continuous temporal change. Said classification methods usually rely on the assumption that sharp boundaries discriminate land use/cover classes and that within these boundaries there is no spatial variation, which is seldom the case. If traditional thematic change detection is utilized on these classification products, the generality and homogeneity of these maps may not provide information on subtle changes that are habitually precursors to significant and often irreversible landscape change. An attempt to deal with the mapping of the continuum of vegetation is a relatively new concept. Vegetation continuous fields (VCF) are proportional cover estimates for different vegetation life forms obtained by modeling remotely sensed datasets, and consists of several continuous response surfaces (one for each vegetative or non-vegetated cover type) in which every pixel value corresponds to a percent cover estimate predicted from a regression model. This is a clear advantage over traditional discrete classifications because VCF depict each pixel as a percent of a given vegetation type, and therefore areas of heterogeneity are better represented when compared to traditional land cover/use classifications (Hansen et al. 2002). Here we tested the performance of two somewhat novel
regression methods: random forests and multivariate regression trees to generate VCF for a
semiarid landscape. This was done for shrubs, trees, grasses, and bare ground. As in chapter
2, we also modeled several years to evaluate the variations in percent cover for shrubs and
trees, and with this identify shrublands into which woodlands have potentially expanded.

In chapter 4 we investigated how historic archives of satellite imagery may be used to
assess the spatial-temporal spectral similarity of big sagebrush ecological sites to undesired
conditions. There are three methods that are used to evaluate ecological sites: similarity
index, trend, and indicators of rangeland health (Pellant et al. 2005). Nevertheless, these
methods are designed to assess very specific areas in space and time due to their field data
requirements. Other methods should be explored to gain the capability to evaluate large
landscapes under current and past conditions. We analyzed a long-term imagery data set
(1984 - 2008) using a multidimensional scaling (MDS) ordination technique. Our work with
MDS allowed drawing inferences about how big sagebrush ecological sites migrate in a
reduced ordination space, and how this may be correlated to field observations (benchmarks)
for which there are current evaluations of condition. The comparison of the multispectral
signal of soil map units composed predominantly of one ecological site against that of the
undesired condition benchmarks permitted an assessment of the magnitude of difference
between the two, and with the temporal component, whether landscape change was towards
or away from an undesired condition (i.e. cheatgrass invasion or woodland encroachment).

LITERATURE CITED

Arid Land Research and Management 17:455-467.

HANSEN, M. C., R. S. DEFRIES, J. R. G. TOWNSHEND, R. SOHLBERG, C. DIMICELI,
and M. CARROLL. 2002. Towards an operational MODIS continuous field of


CHAPTER 2

USING SUPPORT VECTOR MACHINES AND REMOTELY SENSED DATASETS TO ASSESS DYNAMICS OF CHEATGRASS (BROMUS TECTORUM) EXTENT IN NORTHERN UTAH

Abstract

The spatiotemporal dynamics of cheatgrass (Bromus tectorum) were analyzed in the northeastern portion of the Great Basin in northern Utah. A novel approach that builds on the concept of variable importance from the Random Forests algorithm, a graphical assessment for correlation problems, and Support Vector Machines (SVM) was used to select the best suite of explanatory variables for modeling. Remotely sensed datasets and vegetation indices in conjunction with topographic layers were used to generate spatially explicit SVM models of cheatgrass occurrence for the years 1996, 2001, and 2007. Multi-temporal (within one year) vegetation indices seemed to capture cheatgrass’ phenological fluctuations adequately. This phenological understanding in conjunction with elevation was found to be the main drivers in cheatgrass classification. Areas classified as cheatgrass accounted for 113,178 hectares in 1996, 240,071 hectares in 2001, and 224,655 hectares in 2007. This spatiotemporal analysis shed light on which areas have exhibited persistence or variation (expansion or reduction) in cheatgrass presence on the landscape. Throughout the entire study area we found that approximately 146,400 hectares had constant coverage of cheatgrass while 78,250 hectares showed expansion to previously unoccupied ground. A relationship of the spatiotemporal results with current land cover conditions showed that Greasewood Flats and Semi-Desert Shrub Steppe were land cover types most affected by cheatgrass invasion, whereas Montane Sagebrush Steppe turned out to be the least impacted. Validation of the
classification models provided a producer’s accuracy of 70% and a user’s accuracy of 95% for the year 2007. We believe that range managers and other parties studying cheatgrass dynamics in semiarid environments can use these results to prioritize land treatments across large landscapes.

**INTRODUCTION**

Even though big sagebrush (*Artemisia tridentata*) communities still occupy large areas of the western USA, their abundance and ecological conditions are declining in response to a set of natural and anthropogenic processes (Wisdom et al. 2005a). Well documented examples of such processes include the spread and increasing invasion of non-native, colonizing herbaceous species like cheatgrass (*Bromus tectorum*) mainly on the warmer and drier lowlands, and the continuous encroachment of woodlands (i.e. Pinyon-Juniper) in the cooler and wetter terrain (Wisdom et al. 2005b). This paper focuses on the spatiotemporal dynamics of cheatgrass as it invades and displaces big sagebrush ecosystems in Northern Utah. Cheatgrass is an exotic aggressive annual grass that has invaded millions of hectares in the Intermountain area, Pacific Northwest, and northern Great Plains (Young and Allen 1997). The invasion success by this grass has been attributed to its superiority over native species in terms of being a generous seed producer, its ability to germinate in late winter or early spring before most natives, and its tolerance to grazing and frequent fires (Pellant 1996). Big sagebrush dominated ecosystems with a history of disturbances such as overgrazing and fire are more likely to be invaded by cheatgrass. The excessive removal of native perennial grasses due to overgrazing and the increase in resource availability in the upper soil profile after a fire may be primary causes of invasion by annual exotics such as cheatgrass (Chambers et al. 2007). Several negative consequences such as shorter return
intervals for wildfires (Pellant 1990), modification of soil temperature and soil water distribution (Norton et al. 2004), increased uptake of nutrients (Pellant 1996) have been reported in the literature. Invariably these consequences negatively impact the overall ecosystem health of big sagebrush communities. In the worst-case scenario the original, native community may ultimately be converted to a monoculture of cheatgrass (Pellant 1996). Without appropriate knowledge of past and current distribution of cheatgrass, it is difficult to characterize the condition and trend of native big sagebrush communities on a given landscape (United States Department of Agriculture 2006). This knowledge of cheatgrass’ spatial distribution may help managers to identify priority areas in which prevention or restoration activities may be carried out. Thus, it becomes important to identify the patterns of spread and extent of cheatgrass across the landscape.

A number of efforts have been carried out to spatially estimate the risks of invasion of cheatgrass. Bradley and Mustard (2005) used field observations of land cover, and a time series of Landsat TM and ETM and Advanced Very High Resolution Radiometer AVHRR to map cheatgrass extent in the Great Basin. Along these lines, Peterson (2005) used Tobit regression with Landsat ETM to estimate cheatgrass cover in Nevada. Suring et al. (2005) modeled the risk of sagebrush and other native vegetation displacement by cheatgrass using topographic variables. These studies have produced regional assessments of cheatgrass extent at one point in time using medium to coarse spatial resolution imagery, and have been used to propose restoration alternatives in the affected areas.

While it is important to accurately map the current distribution of cheatgrass on the landscape it is also essential to understand the temporal dynamics of such spatial distribution. A sensible assessment of a sagebrush vegetation community should include protocols to understand and identify the spatiotemporal variability of cheatgrass extent. For instance,
areas that have had a continuous and significant component of cheatgrass for a period of time may indicate sections of the landscape that are prone to be converted into a monoculture. On the other hand, those areas that show more variability (i.e. presence and absence through time) may pinpoint zones with different degrees of invasion. In these areas, the manager may suggest different prevention activities to reduce cheatgrass presence and increase native plant diversity.

The thrust of our research was to map the spatiotemporal dynamics of cheatgrass extent in order to understand the continuum of degrees of invasion. Thus our objectives were twofold: a) Model the spatial distribution of cheatgrass in the study area based on multi-temporal (within one year and for different years) vegetation indices and topographic geospatial layers, and b) Assess the multi-temporal (1996, 2001, and 2007) dynamics of cheatgrass extent through classification of infested areas in Northern Utah. We believe that the identification and mapping of these changes on the landscape may provide local managers with updated and objective information to support their decision-making process with regards to implementing prevention or restoration activities.

METHODS

Study Area

We conducted this study in Northern Utah, specifically the northwest portion (114°2’31.2” - 112°43’40.8” West and 41°6’27.36” – 41°59’59.64” North) of Box Elder County, Utah. The area covers an extension of 722,445 hectares excluding barren lands and bodies of water. The vegetation here is predominantly composed of salt desert scrub, big sagebrush steppe and shrublands, as well as pinion-juniper ecosystems (Program 2004). The focus of this work was on the big sagebrush-steppe/shrublands and pinion-juniper
ecosystems. The elevation ranges from 1278 m in the lowlands close to the Great Salt Lake to 3027 m in the Raft River range. The mean elevation is 1520 m. The climate is generally dry, receiving an average of 267 millimeters of precipitation annually typically in the form of winter snows and spring rains. Temperatures are usually cold in the winter (daily average of 26 °F) and moderately hot in the summer (daily average of 69 °F). The yearly average temperature is 46 °F (PRISM Climate Group 2004). There are three distinct physiographic areas: basin floor, piedmont slopes, and mountainous areas. The basin floor consists of playas, salt flats, and beaches that are part of the Great Salt Lake Desert. Here, small dunes of gypsum, oolite or sand can be found. The piedmont slopes consist of alluvial fans, lake terraces, fan terraces, and related fluvial and lacustrine landforms. This physiographic area surrounds the mountains and extends to the playas. The mountainous areas have steep and very steep slopes. The mountain ranges include the Raft River, Grouse Creek, and Goose Creek mountains. The dominant types of rock are limestone, dolomite, quartzite, and igneous rock. The vast majority of streams are intermittent.

The soils range from saline nonproductive in the desert to fertile with a high content of organic matter in the mountains. In much of the area, the soils have a root-inhibiting layer within one meter of the surface (Loerch et al. 1997). The land ownership can be divided into three categories: a) Federal land managed by the Bureau of Land Management (BLM) (41%) and the United States Forest Service (USFS) which manages about 3%, b) Private ownership (43%), and the rest (13%) is State land. The study area has undergone various disturbance regimes ranging from grazing, burning, drought, and flooding events (Sant 2005). Some big sagebrush ecosystems have been converted to exotic annual grasslands or to pinion-juniper environments while an equal area has been maintained as big sagebrush steppe or shrubland (Ramsey 2006).
Field Data

A suite of field data (geo-referenced field points) collected at various times was available to conduct a multi-temporal classification of cheatgrass presence / absence. These points were obtained from four sources (see Fig. 2-1):

(a) Permanent range trend studies from the Utah Division of Wildlife Resources (UDWR) Range Trend Studies (Resources 2010) for the years 1996, 2001, and 2006. This dataset consist of 36 points distributed throughout the study area.

(b) Points collected in the study area by the South West Regional GAP (SWRGAP) project during three (2000, 2001, 2002) fields seasons (Lowry et al. 2007). A total of 175 field points were extracted from this database for 2001.

(c) Points collected by The Nature Conservancy TNC for the Northwest Utah Landscape modeling project in 2007 (Conservancy 2009)

(d) Field points that we collected during the field season of 2007. The combination of the TNC dataset and our data totaled 135 field observations.

The data sets were comparable in the type of information that was collected, namely observations of the presence or absence of cheatgrass. With the exception of the UDWR dataset, which are permanent sample plots, the rest of the field points were visually assessed in terms of cheatgrass percent cover on an area that resembled a 3 x 3 Landsat TM pixel (approximately 90 x 90 meters). We then recoded all the points into presences and absences. In order for a field point to be classified as a presence of cheatgrass it had to have a minimum percent cover of 15% of the exotic. Our sampling on the field did not follow a strict design. We collected information on a purposive way visiting sites located along an elevation range that included Wyoming (Artemisia tridentata ssp. wyomingensis), basin (Artemisia tridentata ssp. tridentata), and mountain big sagebrush (Artemisia tridentata ssp. vaseyana)
communities and that had different coverage of the exotic grass. See Fig. 2-1 that shows the
distribution of the sampled points.

**Geospatial Datasets: Remote Sensing and Topography**

Cheatgrass is known to germinate in late winter or early spring and to become
senescent before most native plants (Pellant 1996). This distinct phenological characteristic
provides an advantage for its spatial recognition using remotely sensed imagery. We
collected multi-temporal Landsat 5 Thematic Mapper imagery (Path39 / Row31) for different
periods of the year but mainly concentrated during spring, summer and fall in order to
capture the phenology of cheatgrass across the growing season. An effort to obtain only
imagery with the best quality (i.e. minimum cloud cover) was made. The imagery
compilation process involved three different years: 1996 (7 dates), 2001 (8 dates), and 2007
(10 dates), see Table 2-1. Selection of the years for analysis was based on the availability of
field observations of cheatgrass presence or absence.

Imagery was rectified and resampled to a common map projection UTM Zone 12
WGS 1984. The imagery was standardized by converting the raw digital numbers to
exoatmospheric reflectance values using an image-based atmospheric correction procedure
(Chavez 1996) with the most up-to-date calibration coefficients for the Landsat TM sensor
(Chander et al. 2009). Once the imagery was standardized, we derived the Soil Adjusted
Vegetation Index (SAVI) for each date. This index takes advantage of the contrast between
the red and near-infrared bands and it also includes canopy background adjustment factor that
minimizes soil brightness variations. SAVI may be calculated as follows:

\[
SAVI = \frac{(\rho_{\text{nir}} - \rho_{\text{red}})(1 + L)}{\rho_{\text{air}} + \rho_{\text{rea}} + L}
\]
In the equation $\rho_{\text{near-infrared}} = \text{near-infrared reflected radiant flux}$, $\rho_{\text{red}} = \text{red reflected radiant flux}$, and $L = \text{adjustment factor}$ (usually a value of 0.5 may be used for different soils).

SAVI has been reported to work well in semiarid ecosystems because it minimizes the soil background effects that are known to affect other indices such as the Normalized Difference Vegetation Index (NDVI) (Huete 1988; Jensen 2007). We used multi-temporal (within a year) SAVI to capture differences in greenness (i.e. onset and die-off) for cheatgrass across the landscape.

In addition to the remotely sensed information (Landsat TM spectral bands, SAVI), we integrated topographic layers, namely a 30-meter digital elevation model (DEM), derived slope, aspect, a heat index (Beers et al. 1966), and a modification to the topographic relative moisture index (TRMI) (Parker 1982). The inclusion of this type of ancillary information has been documented to greatly improve the classification results for cheatgrass in rangelands (Peterson 2005). A list of the explanatory variables may be found in Table 2-2.

**Date Selection**

SAVI values from all available dates of imagery were extracted for cheatgrass field observations and plotted to identify date pairs that best discriminate the peak of cheatgrass greenness in the spring and the senescent period later in the summer (Fig. 2-2). We chose the following Julian dates: 105 and 201 for 1996, 086 and 182 for 2001, and 119 and 183 for 2007. These date pairs were used to model cheatgrass extent for each year utilizing the contrast in SAVI values between peak green and the late summer senesced period.

In this context, a new variable that may be used to indicate cheatgrass intensity and extent was obtained: the normalized difference SAVI or NDSAVI (Ramsey 2009):
The NDSAVI takes advantage of the contrast between the SAVI values of two temporal periods, which, in our case, coincides with the estimated peak green and senescence in cheatgrass. Higher NDSAVI index values correspond with higher SAVI during the spring as opposed to the summer and in these environments, this contrast correlates positively with densities of cheatgrass. The NDSAVI was then included along with all the other explanatory variables in our classification of cheatgrass extent.

Data Preparation

For each year (1996, 2001, 2007), field sample locations were classified as either 1) a site with presence or 2) a site with no occurrence of cheatgrass. Details about the number of presence and absence sites for each year is found in Table 2-3. A Python script was written to extract Landsat TM band reflectance, NDSAVI, and elevation (plus derivatives) for all sites and all years. Each data matrix (one for each year) was used to model cheatgrass extent for that year. The dataset for 2001 was randomly subdivided into two portions: 70% for training and 30% for validation purposes. For the year 2007 we basically had two independent datasets: our field points, and those collected for the TNC study. We decided to use our points to train the model, and the TNC points for validation purposes. This gave us a better sense of independence between training and validation of our models. Because the 1996 dataset was relatively small we decided to use all available points to train the model. Therefore no validation was done for this year.
Classification Using Support Vector Machines

We used support vector machines (SVM) to classify areas infested by cheatgrass. SVM is rooted in statistical learning theory and has acquired a reputation as a robust and accurate classifier even when using small training sets (Gidudu et al. 2007). In remote sensing, SVM has been used with success to: classify land cover types (Pal and Mather 2005), monitor forest disease spread (Liu et al. 2006), discriminate semi-arid vegetation types (Su et al. 2007), estimate soil types (Hahn and Gloaguen 2008), and automate forest cover change analysis (Huang et al. 2008), among others. SVM is a nonparametric statistical technique that non-linearly transforms the training data in the input space to a feature space of a higher dimension through usage of a kernel function. This results in a linearly separable dataset that can be easily split by a linear classifier such as discriminant analysis. With regards to remote sensing this is particularly important, because it allows classifying multispectral data sets, which are typically nonlinear, and thus difficult to separate (Gidudu et al. 2007).

Although SVM have proven to perform well with classification problems, they are negatively affected by redundant explanatory variables (Cutler 2007), thus requiring a process to eliminate highly correlated variables before attempting to train a model. To deal with this issue, we decided to use a novel approach consisting of the variable importance concept as described in the Random Forests (RF) algorithm (Cutler et al. 2007). Variable importance has its foundation on the mean decrease in accuracy concept, and is assessed based on how much poorer the predictions would be if the data for that predictor were permuted randomly. This gives an overall impression of the impact that a specific variable has in decreasing prediction accuracy. A comprehensive explanation of RF and variable importance can be found in the paper by Cutler and colleagues (2007).
We prepared our independent variables (Table 2-2) and dependent variable (cheatgrass presence / absence) for a RF run in order to assess variable importance. We replicated this process 500 times and stored the mean decrease in accuracy in a matrix. We did this because variable importance may change significantly between individual iterations, thus it would be very risky to accept the results of a single run. We plotted the mean decrease in accuracy (see Fig. 2-4 a) and selected the 10 variables that impacted the accuracy the most. This is clearly a subjective approach but so far we have not found literature that proposes a more transparent method.

Whether a steep decline is found in the variable importance plots or the variables with the highest importance are chosen, the fact that these variables may still be highly correlated with each other cannot be overlooked. For any modeling scheme this is particularly problematic because of the instability created by redundant data. It is worse for SVM, which are particularly affected by redundancies. Random Forests does not guarantee that the variables with the highest importance are not correlated because as it randomly permutes the values for the out-of-bag observations it works with one variable at a time (Cutler et al. 2007). To address correlation between variables, we generated scatter plots of all possible combinations of variables and graphically assessed them for correlation. As expected, some of the Landsat TM spectral bands were highly correlated and thus were not included in the model even though they were rated as highly important in the Random Forests process (Fig. 2-4 b). We found that from the original pool of explanatory variables (Table 2-2), a simplified model using only NDSAVI and elevation as primary drivers could provide accurate results. These two variables had the highest values in the variable importance plots, and they were poorly correlated. In addition these two values comprised most of the meaningful information about greenness variation and other physical characteristics such as
precipitation, moisture availability and expected diversity. These topics are explained in more detail in the discussion.

**Gamma and Cost**

When using SVM for classification, there are two important decisions to make: one is which kernel (function to project the data from input space to feature space) to use and the other is which value to use for cost (C), which affects both the complexity of the classifier and the degree to which points are misclassified. We tested two different types of kernels (polynomial and radial) and finally decided to use the radial kernel, which seemed to adjust more appropriately to the available training datasets. Once we had decided to use a radial kernel we then needed to determine appropriate values for gamma (\( \gamma \)) (a parameter needed for radial type kernels) as well as a proper value for C. Gamma controls the flexibility of the SVM classification function. Bigger values of \( \gamma \) will provide function solutions that work better with irregular surfaces, thereby giving more flexibility. On the other hand smaller \( \gamma \) values should give smoother functions. In a modeling exercise, one would want to use bigger \( \gamma \) values in order to have the ability to imitate irregular boundaries. Conversely, smaller values of \( \gamma \) should be used to prevent replicating noise in the samples. Smaller values are often used to avoid overfitting the model during training (Hastie et al. 2009).

These two values are determined prior to conducting the classification. Because the choice of these two parameters is critical, it is necessary to use a tuning or calibration process to estimate them since the choice of \( \gamma \) and C can have a significant impact on the output. In Fig. 2-5 we can see a graphical visualization of SVM classification models in which four different combinations of \( \gamma \) and C have been used. In very simple terms, we can see that \( \gamma \) controls the shape of the predicted surface with higher values resulting in a more complex
outline. Cost on the other hand, affects the expansion of the support vectors. We see that using higher values of C result in the inclusion of areas that previously were not mapped as cheatgrass or a very generalized prediction surface for cheatgrass. In essence we can see that a careless selection of \( \gamma \) and C may result in a classification model that is either too smoothed or generalized (i.e. low values of each parameter) or too irregular thereby producing a very specific prediction surface for cheatgrass (i.e. high values of each parameter). A too specific classification model could potentially cause problems of overfitting. We have seen the need to utilize proper tuning to choose the values of \( \gamma \) and C. Such a tuning process must be able to fit the entire path of SVM solutions for every value of C while minimizing the error rate (Cutler 2007). In this research we used a tuning process that starts with a bivariate grid of values for C and \( \gamma \) and then uses cross validation to find the local minima in the error rate (Hastie et al. 2004). Once the two parameters (C and \( \gamma \)) were identified through the tuning process, a SVM classification was performed for each of the three years. The values for \( \gamma \) and C are reported in Table 2-4.

The SVM tuning process as well as the classification runs were conducted using the R package e1071 (R-Project 2010). Once a model was fit and accepted, we used the package YaImpute (Crookston and Finley 2008) to extract the model for each SVM run, and generate a geospatial response surface (presence / absence of cheatgrass) for each year.

**Accuracy Assessment**

As described earlier, a fraction of the datasets were withheld from the model building and used for model validation purposes. For every year of the analysis and after a successful calibration of the model with the training data, we applied the model using a predict function in R to the validation subset, compared it with the observed presences or absences of
cheatgrass and in this way we were able to obtain a confusion matrix. From the confusion
matrix one is able to identify true positives (observed and predicted presences), false
positives (observed absences but predicted presences), false negatives (observed presences
but predicted absences) and true negatives (observed and predicted absences). With this in
mind we were able to obtain the following metrics:

(a) Percent correctly classified (percentage of all cases correctly classified)
(b) Sensitivity (percentage of true positives correctly predicted)
(c) Specificity (percentage of true negatives correctly predicted)
(d) Kappa (Proportion of specific agreement)

Kappa offers a meaningful numerical value for inter-comparison between models
because it is negligibly affected by prevalence or the frequency of occurrence of the target, in
this case cheatgrass (Manel et al. 2001).

Temporal Dynamics

After we had modeled and validated cheatgrass extent for each year, we intersected
the maps to determine: a) areas that have had a consistent cover of cheatgrass for all three
years, and b) areas that have had variations in the cover of cheatgrass from one temporal
period to another (i.e. from 2001 to 2007). We then subdivided the latter into a) areas that
were not formerly covered by cheatgrass (expansion of cheatgrass to previously unoccupied
land), and b) areas that had cheatgrass cover in one year but not in a subsequent year
(reduction of cheatgrass).

Whether cheatgrass expanded or reduced its areal extent, we consider it essential to
explain that our modeling of cheatgrass was aimed at identifying those pixels that had a
strong NDSAVI signal for cheatgrass. In other words if a pixel was classified as a presence, it
does not necessarily mean that the pixel is entirely occupied by the exotic. The candidate pixel simply has enough coverage to be classified as a presence of cheatgrass. With this said, it is important to look at the results of expansion and reduction with care. This situation will be addressed again during the discussion.

RESULTS

Contrast Between Onset and Die-off NDSAVI Dates

Fig. 2-3 shows that the greenness values for cheatgrass presences are slightly higher than the greenness values for cheatgrass absences for the chosen dates. On the right panel of the same figure (assumed date of die-off) it may be clearer however, that the greenness values for the absences are conspicuously higher than those of the presences. This figure is provided to justify our selection of the dates that we used for the classification in each year (1996, 2001, and 2007).

Cheatgrass Classification: Individual Scenarios

In all RF runs for variable importance, the NDSAVI and DEM variables occupied the highest position in the mean decrease in accuracy matrix. In addition, it was clear from the SVM classification plots that a simple yet effective model could be attained by using these two variables. The final map products for each temporal scenario contained two classes; presence and absence of cheatgrass. Output layers were generated at the same spatial resolution of the explanatory layers (30 meter). These maps are presented in figures 2-6, 2-7 and 2-8 for the years 1996, 2001, and 2007 respectively.

The 1996 classification is clearly the most conservative estimate of cheatgrass distribution when compared to the other two years. The areas classified as cheatgrass were predominantly found between 1400 to 1700 m elevations for 1996. Fig. 2-1 contains the
distribution of training points for the three years of analysis. A close inspection of the 1996 field dataset shows that no points were sampled in the valley flats. This may be the reason why the SVM failed to classify areas of cheatgrass presence in this type of landform. For the years 2001 and 2007, the elevation range of areas classified with cheatgrass presence was 1280 to 2330 m. A close inspection to the landform map of the study area shows that the vast majority (approximately 80%) of the terrain classified as cheatgrass in 1996 corresponds to nearly level plateaus or terraces and gently sloping ridges and hills. For the next two scenarios (2001 and 2007), this was also the case, but a significant portion of the affected area was also found in valley flats. Some of these areas may be confounded with cultivated areas. The areas classified as cheatgrass are summarized per year in Table 2-5.

The SVM classification yielded an overall accuracy of 72.5%, a user’s accuracy of 67%, and producer’s accuracy of 62.5% were obtained for presences of cheatgrass in the year 2001. Better validation results were obtained for the year 2007 where the overall accuracy was 86.9%, a user’s accuracy of 95.8%, and producer’s accuracy of 69.7% were obtained then for the presences of cheatgrass. No validation results were obtained for the year 1996 because all available points were used during the classification. Table 2-6 includes the data used to calculate these accuracies, and also includes measures of sensitivity, specificity and Kappa, which are reported for both years.

**Temporal Dynamics**

A key output from this study was the identification of those areas that show persistence and variation of cheatgrass coverage during the study period. Fig. 2-9 and table 2-7 show the temporal dynamics for the 2001 – 2007 period. We can see that 146,399 hectares have had a persistent coverage of cheatgrass. These lands are mainly concentrated
on the eastern half of the study area. On average, the lands with persistent cheatgrass are located in lower (1550 m) and drier (301 mm/year) terrain compared to those lands that show variation in cheatgrass occurrence between the years (Table 2-7). Those areas that have had consistent cheatgrass coverage for these three periods might be in the process of conversion to a monoculture. Lands that showed a reduction in cheatgrass coverage are widely dispersed throughout the county; those lands with expansion are primarily concentrated on the western and mid sections of the study area. From Table 2-7 we can see that those areas of expansion are predominantly positioned in higher (1660 m) and wetter (375 mm/year) ground. Although it is clear that both persistent and reduction areas have a very similar distribution across landform classes, this is not the case for expansion areas which tend to occupy more gently sloping ridges and hills, and fewer plateaus and valley flats.

Table 2-8 provides a different perspective of cheatgrass temporal dynamics. In this case, the spatiotemporal dynamics have been related to different shrubland land cover classes as determined from the SWREGAP land cover map (Lowry et al. 2007). It seems clear that proportionally, Greasewood Flat is the most negatively affected of the shrub classes with 40.7% of its area under persistent cheatgrass cover from 2001 to 2007. Only 29.8% of this shrub class exhibited no signals of cheatgrass during this period. The Big Sagebrush Shrubland class shows that cheatgrass has expanded to approximately 32,575 hectares (23.7%) of its area. On the other hand, Montane Sagebrush Steppe is least affected by cheatgrass invasion with almost 97% of its area showing no occupation by the exotic. Nevertheless, it is of concern to see that almost 2,491 hectares of this shrub class were invaded during the 2001-2007 period.

Regarding the cheatgrass reduction dynamic, this was more pronounced in the Mixed Salt Desert Scrub land cover class. This class shows an estimated 32,080 hectares (22.3%) of
decline in cheatgrass cover. A noticeable decline was also observed for the Semi-Desert Shrub Steppe (18.5%) and the Greasewood Flat (18%).

**DISCUSSION**

Our research has generated spatially explicit models of cheatgrass extent for three temporal scenarios. To the best of our knowledge, this is the first attempt to map the spatiotemporal dynamics of cheatgrass extent in Northern Utah. We explored the utilization of a new variable, the NDSAVI, to classify areas invaded by cheatgrass. We believe that this variable fulfilled our expectations with regards to discriminating cheatgrass on semiarid landscapes. NDSAVI works as a multi-temporal composite index of greenness that conveys enough information to map the phenological fluctuations of cheatgrass (Ramsey 2006). Peterson (2005) introduced ΔNDVI to model a continuous response of cheatgrass cover in Nevada. This variable is different from NDSAVI in the sense that ΔNDVI is not normalized. In any case, it is clear that two or more dates of a given vegetation index are needed to properly map cheatgrass. Our main assumptions in this study was that sufficient contrast (Fig. 2-3) between two greenness signals may be found within one year, and that such contrast can be used to discriminate cheatgrass from other grass species in the study area.

We have introduced an approach to select the most important variables for modeling. Such an approach was felt necessary in order to use Support Vector Machines, which are significantly affected by redundant data (Cutler 2007). We started our approach with the concept of Random Forests variable importance (Cutler et al. 2007). Nevertheless, we felt the need to go further because there is no guarantee that the variables prioritized by RF will be uncorrelated. This was clearly demonstrated by creating scatter plots (Fig. 2-4) of the most important variables according to RF. In any case, we think that RF variable importance
coupled with a graphical analysis is a good place to start, and then continue with an assessment of SVM classification maps (Fig. 2-5) as it was described before. We were able to show that a simple SVM model that uses NDSAVI and elevation was effective in mapping cheatgrass extent (Table 2-4). We believe that our approach to variable selection complied with the ecological expectation that cheatgrass locations may be identified based on its distinct phenological signature, which may be further refined by introducing elevation during the modeling process. Elevation is a variable that is positively correlated to precipitation, which in turn influence soil moisture availability. At higher elevations the soils in the study area are known to be fertile and with a high content of organic matter (Loerch et al. 1997). In higher elevations the combined effects of greater soil moisture availability with more fertile soils may create conditions for more resistant and resilient big sagebrush communities with native and perennial grasses. Cheatgrass invasibility of big sagebrush ecosystems in the Great Basin varies with elevation gradients with the most susceptible areas found at lower elevations (Chambers et al. 2007). We considered that the variable elevation appropriately accounted for the variability found in precipitation, soil moisture availability and expected plant diversity that may characterize the resistance of big sagebrush communities to invasibility by cheatgrass. Having a composite variable that comprises all this information is an advantage when using a technique such as SVM that as we explained before is negatively affected by redundant information.

The utilization of SVM for the classification of remotely sensed datasets is relatively new. A potential reason for its somewhat low use may be that SVM rely heavily on the careful selection of the explanatory variables and proper tuning of parameters like $\gamma$ and $C$. However, we believe that SVM performance in identifying features on semiarid landscapes justify spending the time to find the best subset of variables and calibrating the parameters.
Su and others (2007) found that SVM outperformed maximum likelihood classifications in terms of accuracy in semiarid environments. Another reason to use SVM for our research was the fact that our field data sets were relatively small. Other classification algorithms are known to be “data-hungry”, meaning that they require ample training data to obtain an effective model. Our results obtained using SVM seem to agree with the published work by Pal and Mather (2005), who demonstrated that SVM could perform well with small training datasets and high-dimensional data.

A significant result of this study is the identification of areas that show persistence and variation of cheatgrass on the landscape. The variation has been further divided into areas of expansion and reduction. Because cheatgrass can exhibit great temporal and spatial variability that accompany fluctuations in precipitation (Bradley and Mustard 2005), a single snapshot of cheatgrass extent may not provide enough information to assess a site’s condition. We should clarify though that the expansion and reduction of cheatgrass as we presented here might simply be a symptom of variable water years. In other words, there is no definite evidence that expansion is truly expansion per se and may just be temporal variability associated with fluctuations in precipitation during the years that we conducted our research. Furthermore, those areas that show non-persistent cheatgrass may be areas that are not significantly occupied by cheatgrass, but where cheatgrass presents itself intermittently based on changing water availability.

This type of information is particularly valuable if we relate it to a specific shrub community. We were able to identify that Greasewood Flats and Semi-Desert Shrub Steppes had the biggest proportions of their mapped boundaries affected by persistent cheatgrass coverage from 2001 to 2007. It makes sense that these two land cover classes are the most affected given that they are found mainly in the lower and drier sections of the study area and
cheatgrass is known to excel in such environments (Wisdom et al. 2005b). Our modeling and evaluation of posterior spatiotemporal dynamics pinpoint where these conditions are located, thus providing a tool to prioritize treatments by range conservationists and managers.

From a management point of view, perhaps more concerning are those areas that show variable occurrence of cheatgrass. In such areas, cheatgrass may be in the process of establishing and prevention activities may be more feasible than in those areas that show a persistence of the annual grass. Big Sagebrush Shrubland was the land cover class with the highest proportion of cheatgrass expansion (Table 2-8). According to the SWRGAP land cover legend description these shrublands are dominated by Artemisia tridentata ssp. wyomingensis. These shrublands when degraded due to disturbances have been recognized to have low resistance and resilience to invasion by exotic annuals (Wisdom et al. 2005b). The areas of expansion have generally occurred in higher and wetter locations (Table 2-7). This may be another indication of cheatgrass plasticity to adapt to different environments or it may just be a consequence in changes in precipitation patterns from 2001 to 2007.

Our modeling work focused on the detection of those pixels that had a strong NDSAVI index which we equate to a cheatgrass presence. However, it does not mean that the entire extent of the pixel is fully covered with cheatgrass. With this being said, those areas that show a reduction should not be addressed as areas in which cheatgrass has disappeared from the landscape. Rather the NDSAVI index was not as strong as it would be expected for a cheatgrass presence in these areas. In this context we think of the reduction areas as sections of the landscape in which cheatgrass is not as dominant as other areas. On average, the reduction areas are found in transitional locations in terms of elevation and precipitation relative to the persistence and expansion areas. The Mixed Salt Desert Scrub land cover class showed the highest reduction rates (> 22%). We find saline and calcareous
substrates in this land cover class and likely cheatgrass is not a good competitor in these conditions, or again, as in the case of the expansion, there were climatic fluctuations that affected its performance. Out of the seven shrub land cover classes contained in table 2-8 the Montane Sagebrush Steppe turned out to be the least affected by cheatgrass invasion with approximately 97% of its coverage showing no mapped cheatgrass during the period of analysis. This makes sense since this land cover class consists of communities of Artemisia tridentata ssp. vaseyana (mountain sagebrush) which generally occur in higher and wetter conditions which tend to create more diverse communities in which cheatgrass has been so far unable to effectively compete compared to other shrubland areas.

We would like to acknowledge the fact that other plant species may show a similar phenology as cheatgrass. This is a source of potential error in our modeling. So far we have only identified agricultural fields as a potential source of confusion in this study area. We would also like to recognize that the dynamics predicted from these models are sensitive to the way in which field data were collected. Our sampling on the field followed a purposive scheme that aimed at capturing the variability found in different big sagebrush communities across an elevation gradient. We cannot, however, assume that our sampling was comprehensive enough especially for the year 1996 for which our field data set was small. There exists uncertainty about the results for this year because we did not carry out a validation as we did for 2001 and 2007. Nevertheless, the final maps and their correspondent error rates for 2001 and 2007 suggest that using the multi-temporal signal from NDSAVI was successful in modeling cheatgrass. Evidence suggests that SAVI can be very informative in arid and semiarid lands (Tueller 1994). By using SVM and the selection of the best suite of variables and parameters, we consider our results accurate enough for use as a management tool. As a point in fact, the user’s accuracy was higher than the producer’s accuracy in both
years 2001 and 2007 (Table 2-6). For the 2007 map, the user could be certain that 95% of
the time that he/she visits a site that has been classified as cheatgrass a significant component
of cheatgrass will be found at that location. There were differences in the user and producer's
accuracy rates for both years. We think that this may be due to differences in the sample
size. However it may also be a function of water availability. There was a higher
precipitation during the year 2007 for the study area, and this may have influenced that areas
with presence of cheatgrass had a more robust greenness signal and thus more chances to be
classified than they did for the year 2001.

This research provides knowledge about the spatial distribution of cheatgrass in
with a satisfactory level of accuracy were developed and geospatial layers denoting areas of
cheatgrass coverage are available for each year with a 30-meter spatial resolution. Our
results pinpoint areas where cheatgrass has shown persistent cover as well as areas where the
invasive annual has expanded to previously unoccupied ground. Land on which cheatgrass
has reduced its cover was also discriminated. Our spatiotemporal assessment shows that the
dynamics of cheatgrass invasion behave dissimilarly between different plant communities.
The methods and results presented in this paper should be viewed as a new and adaptable
classification technique to model invasive annual grasses.

IMPLICATIONS

Our results have been formatted for use in a GIS environment by range managers or
other parties working in cheatgrass early detection and control. A transparent protocol to map
the invasive species has been presented here. This protocol may be utilized and modified to
map other noxious species in semiarid environments. Our findings relative to the
spatiotemporal dynamics of cheatgrass can provide guidance about areas that show persistence, reduction or expansion of the annual invasive, which can, in turn, prioritize prevention or restoration activities.

We also expect that our results may contribute to further research into state and transition models in the study area. If the spatiotemporal dynamics of cheatgrass extent can be related to ecological site descriptions then it may be possible to determine a site’s state or condition as well as different transitions through time and space. SVM have proven to work well with small training datasets and remotely sensed information, this should help in monitoring efforts for large territories in rangelands of the Intermountain West.

**LITERATURE CITED**


CUTLER, A. Notes from class STAT 6650. 2007


PETE


Table 2-1. Landsat TM Path039 Row031 dates collected

<table>
<thead>
<tr>
<th>Year</th>
<th>Julian Dates</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>105, 153, 169, 201, 217, 233, 281</td>
<td>04/01, 06/01, 06/17, 07/19, 08/04,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>08/02, 09/19, 10/05</td>
</tr>
<tr>
<td>2001</td>
<td>086, 118, 150, 166, 182, 214, 262, 278</td>
<td>03/27, 04/28, 05/30, 06/15, 07/01,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>08/02, 09/19, 10/05</td>
</tr>
<tr>
<td>2007</td>
<td>103, 119, 135, 151, 167, 183, 199, 215, 231, 263</td>
<td>04/13, 04/29, 05/15, 05/31, 06/16,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>07/02, 07/18, 08/03, 08/19, 09/20</td>
</tr>
</tbody>
</table>
Table 2-2. Explanatory variables compiled for modeling cheatgrass occurrence

<table>
<thead>
<tr>
<th>Variable</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landsat TM blue, green, red, near-infrared, middle-infrared 1 and 2 bands</td>
<td>For each year being modeled, we had six bands per date, chose two dates out of table 2-1 that best maximized greenness contrast: equals 12 TM variables of reflectance</td>
</tr>
<tr>
<td>SAVI</td>
<td>Soil-adjusted vegetation index, one per date; two SAVI per year</td>
</tr>
<tr>
<td>NDSAVI</td>
<td>Normalized Difference SAVI, one per year</td>
</tr>
<tr>
<td>DEM</td>
<td>Digital Elevation Model</td>
</tr>
<tr>
<td>DEM Derivatives</td>
<td>Slope, Aspect, Heat Index (Beers et., 1966), Topographic Relative Moisture Index TRMI</td>
</tr>
</tbody>
</table>
Table 2-3. Field sampling data sets used for model building and validation

<table>
<thead>
<tr>
<th>Year</th>
<th>Source</th>
<th>Number of samples</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Presences</td>
<td>Absences</td>
</tr>
<tr>
<td>1996</td>
<td>Big Game Range Trend Studies</td>
<td>10</td>
<td>26</td>
</tr>
<tr>
<td>2001</td>
<td>Big Game Range Trend Studies / SWRGAP</td>
<td>87</td>
<td>114</td>
</tr>
<tr>
<td>2007</td>
<td>Northwest Utah Landscape Modeling Project / RS/GIS field points</td>
<td>49</td>
<td>86</td>
</tr>
</tbody>
</table>
### Table 2-4. Parameters $\gamma$ and $C$ used during modeling

<table>
<thead>
<tr>
<th>Year</th>
<th>Parameters</th>
<th>Explanatory Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gamma $\gamma$</td>
<td>Cost $C$</td>
</tr>
<tr>
<td>1996</td>
<td>0.4</td>
<td>1.0</td>
</tr>
<tr>
<td>2001</td>
<td>0.45</td>
<td>1.5</td>
</tr>
<tr>
<td>2007</td>
<td>0.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

- NDSAVI_1996, Elevation
- NDSAVI_2001, Elevation
- NDSAVI_2007, Elevation
Table 2-5. Areas classified as cheatgrass in the study area

<table>
<thead>
<tr>
<th>Year</th>
<th>Area (Hectares)</th>
<th>Elevation Range (masl)</th>
<th>Major Landforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>113,178</td>
<td>1400 - 1700</td>
<td>Nearly level plateaus or Terraces, Gently Sloping Ridges and Hills</td>
</tr>
<tr>
<td>2001</td>
<td>240,071</td>
<td>1280 – 2180</td>
<td>Nearly level plateaus or Terraces, Gently Sloping Ridges and Hills, Valley Flats</td>
</tr>
<tr>
<td>2007</td>
<td>224,655</td>
<td>1280 - 2330</td>
<td>Nearly level plateaus or Terraces, Gently Sloping Ridges and Hills, Valley Flats</td>
</tr>
</tbody>
</table>
Table 2-6. Validation data and metrics for 2001 and 2007

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Field Sites</th>
<th></th>
<th></th>
<th>User’s accuracy</th>
<th>Overall accuracy:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CG</td>
<td>No – CG</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Cheatgrass</td>
<td>10</td>
<td>5</td>
<td>15</td>
<td>66.7%</td>
<td>72.5%</td>
</tr>
<tr>
<td>No – Cheatgrass</td>
<td>6</td>
<td>19</td>
<td>25</td>
<td>76%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>24</td>
<td>40</td>
<td></td>
<td>42.1%</td>
</tr>
<tr>
<td>Producer’s accuracy</td>
<td>62.5%</td>
<td>79.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Field Sites</th>
<th></th>
<th></th>
<th>User’s accuracy</th>
<th>Overall accuracy:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CG</td>
<td>No – CG</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>CG</td>
<td>23</td>
<td>1</td>
<td>24</td>
<td>95.8%</td>
<td>86.9%</td>
</tr>
<tr>
<td>No – CG</td>
<td>10</td>
<td>50</td>
<td>60</td>
<td>83.3%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>33</td>
<td>51</td>
<td>84</td>
<td></td>
<td>71.1%</td>
</tr>
<tr>
<td>Producer’s accuracy</td>
<td>69.7%</td>
<td>98%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity = Producer’s accuracy for CG

Specificity = Producer’s accuracy for No-CG
Table 2-7. Descriptors for the three classes of cheatgrass temporal dynamics 2001-2007

<table>
<thead>
<tr>
<th>Dynamic</th>
<th>Metric</th>
<th>GSRH</th>
<th>NLPT</th>
<th>VF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Persistent</strong></td>
<td>Mean elevation</td>
<td>1578</td>
<td>1547</td>
<td>1520</td>
</tr>
<tr>
<td></td>
<td>Mean precipitation</td>
<td>313</td>
<td>305</td>
<td>287</td>
</tr>
<tr>
<td></td>
<td>Percent in class</td>
<td>23.4%</td>
<td>57.7%</td>
<td>11.1%</td>
</tr>
<tr>
<td><strong>Expansion</strong></td>
<td>Mean elevation</td>
<td>1760</td>
<td>1615</td>
<td>1602</td>
</tr>
<tr>
<td></td>
<td>Mean precipitation</td>
<td>400</td>
<td>373</td>
<td>353</td>
</tr>
<tr>
<td></td>
<td>Percent in class</td>
<td>30.4%</td>
<td>47.7%</td>
<td>8.2%</td>
</tr>
<tr>
<td><strong>Reduction</strong></td>
<td>Mean elevation</td>
<td>1657</td>
<td>1565</td>
<td>1539</td>
</tr>
<tr>
<td></td>
<td>Mean precipitation</td>
<td>350</td>
<td>342</td>
<td>317</td>
</tr>
<tr>
<td></td>
<td>Percent in class</td>
<td>23.1%</td>
<td>57.2%</td>
<td>10.7%</td>
</tr>
</tbody>
</table>

GSRH: Gently Sloping Ridges and Hills
NLPT: Nearly Level Plateaus or Terraces
VF: Valley Flats
<table>
<thead>
<tr>
<th>Dynamic</th>
<th>XMSS</th>
<th>BSSh</th>
<th>BSSs</th>
<th>GF</th>
<th>MSDS</th>
<th>MSSt</th>
<th>SDSSSt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistent</td>
<td>7612.2</td>
<td>29366.3</td>
<td>108.7</td>
<td>32429.3</td>
<td>42601.5</td>
<td>4.3</td>
<td>3006.1</td>
</tr>
<tr>
<td></td>
<td>10.4%</td>
<td>21.1%</td>
<td>4.1%</td>
<td>40.7%</td>
<td>29.6%</td>
<td>0.0%</td>
<td>35.0%</td>
</tr>
<tr>
<td>Expansion</td>
<td>8538.3</td>
<td>32974.7</td>
<td>366.5</td>
<td>9139.1</td>
<td>13572.5</td>
<td>2490.8</td>
<td>981</td>
</tr>
<tr>
<td></td>
<td>11.7%</td>
<td>23.7%</td>
<td>13.7%</td>
<td>11.5%</td>
<td>9.4%</td>
<td>3.3%</td>
<td>11.4%</td>
</tr>
<tr>
<td>Reduction</td>
<td>6845.8</td>
<td>13470.7</td>
<td>222.4</td>
<td>14335.9</td>
<td>32080.8</td>
<td>143.1</td>
<td>1587.5</td>
</tr>
<tr>
<td></td>
<td>9.4%</td>
<td>9.7%</td>
<td>8.3%</td>
<td>18.0%</td>
<td>22.3%</td>
<td>0.2%</td>
<td>18.5%</td>
</tr>
<tr>
<td>No Cheatgrass</td>
<td>50010.3</td>
<td>63608.9</td>
<td>1986.3</td>
<td>23764.6</td>
<td>55608.0</td>
<td>73476.7</td>
<td>3005.3</td>
</tr>
<tr>
<td></td>
<td>68.5%</td>
<td>45.6%</td>
<td>74.0%</td>
<td>29.8</td>
<td>38.7%</td>
<td>96.5%</td>
<td>35.0%</td>
</tr>
</tbody>
</table>

XMSS: Great Basins Xeric Mixed Sagebrush Shrubland
BSSh: Inter-Mountain Basins Big Sagebrush Shrubland
BSSs: Inter-Mountain Basins Big Sagebrush Steppe
GF: Inter-Mountain Basins Greasewood Flat
MSDS: Inter-Mountain Basins Mixed Salt Desert Scrub
MSSt: Inter-Mountain Basins Montane Sagebrush Steppe
SDSSSt: Inter-Mountain Basins Semi-Desert Shrub Steppe
Figure 2-1. Study area in Northern Utah and distribution of field observations. The study area is shown in the context of the State of Utah.
Figure 2-2. Sample of SAVI parallel coordinate plots to select the best two dates of imagery to model cheatgrass extent
Figure 2-3. Box-plots of the SAVI values for the two dates per year chosen to maximize the contrast between the peak and die-off of cheatgrass. CG = Cheatgrass presence, NO-CG = Cheatgrass absence
Figure 2-4. Random forest variable importance plot (a), and scatter plots for the middle infrared bands (b) for the year 2007
Figure 2-5. Support Vector Machines classification maps using different options for gamma and cost
Figure 2-6. Classified Cheatgrass extent for the year 1996
Figure 2-7. Classified Cheatgrass extent for the year 2001
Figure 2-8. Classified Cheatgrass extent for the year 2007
Figure 2-9. Dynamics of Cheatgrass extent for the 2001 – 2007 period
CHAPTER 3

MONITORING SEMI-ARID RANGELANDS WITH MULTI-TEMPORAL VEGETATION CONTINUOUS FIELDS: MULTIVARIATE REGRESSION TREES VS. RANDOM FORESTS

Abstract

A multi-temporal series of vegetation continuous fields (VCF) was developed in a semi-arid rangeland in order to detect active woodland encroachment into big sagebrush (Artemisia tridentata spp.) communities. VCF consists of proportional estimates of canopy cover for different vegetation types by modeling remotely sensed data. A series of VCF was prepared for shrubs, trees, herbaceous vegetation, and bare ground for three years: 1996, 2001, and 2007. Explanatory variables included spectral reflectance information and various indices derived from Landsat Thematic Mapper imagery, as well as ancillary data sets representing topographic variation. A land cover classification was also used as an independent layer. We tested two relatively new regression methods to model our VCF: Multivariate Regression Trees (MRT) and Random Forests (RF). MRT has the capability to simultaneously model several response variables while RF is gaining a reputation for producing highly accurate results. An assessment of the correlation of observed versus predicted values indicate that RF outperformed the accuracy of MRT using the mean absolute errors and root mean square errors for shrubs, trees, herbaceous and bare ground. It was also observed that RF was able to generate a better spatial depiction of the continuum of percent cover across the entire landscape. We assessed the spatiotemporal dynamics of shrubs, trees and bare ground in order to identify areas of big sagebrush that have experienced change from 1996 to 2007. Approximately 17,570 out of 139,450 hectares of big sagebrush shrub
land were estimated to have intrusion by woodlands. We think that rangeland managers and conservationists may benefit from using our protocols to enhance their spatiotemporal understanding of land cover dynamics across large landscapes. If final users deem the coarse thematic legend that we used appropriate, then our findings may also be utilized to update state and transition models for ecological sites in the study area.

**INTRODUCTION**

The sagebrush ecosystem has been and continues to be under a host of environmental and socioeconomic pressures that negatively influence its condition. An estimated 40% of its pre-European settlement area has been reduced, owing to causes like conversion to agriculture, expansion of infrastructure, energy developments, invasion by exotic plants, and encroachment by woodlands among others (Wisdom et al. 2005). With an accelerating pattern of loss (Hemstrom et al. 2002), it is clear that transparent monitoring tools and decision support systems are needed to properly assess change in time and space, and to evaluate how these changes impact the overall health of the ecosystem. In this paper we report on using remotely sensed information to support monitoring in rangelands.

An Ecological Site Description (ESD) (U.S. Department of Agriculture 2008) and its corresponding State and Transition Model (STM) (Westoby et al. 1989) provides comprehensive information about the biophysical properties of a site. The ESD and STM also supply descriptions of the different plant communities’ structure, composition, and dynamics that may be found in space and time, given different disturbances or management scenarios. For example the ESD of a given big sagebrush site may contain a description of what constitutes a “good” or “average” condition in terms of coverage by shrubs, native grasses and woodlands. A good condition will usually be related to circumstances in which big
sagebrush is dominant with regards to the woody component. On the other hand, the STM for the same site may contain different limits for the occurrence of woodlands above which the sagebrush can no longer be considered dominant, or possibly “over” dominant.

There are limitations in the way ESD and STM work. Traditionally, the descriptions of plant communities correspond to specific field locations that best represent a reference condition or an alternative state according to the range conservationist experience and knowledge. However, it is effectively impossible to have field descriptions (i.e. plant communities’ composition) for the entire extent of a given management unit. Further, it is even more difficult to have a transparent account of all the important spatiotemporal changes that occur on a given landscape. For instance, it is important to identify specific areas on a sagebrush landscape that are in the process of encroachment by woodlands or invasion by exotic annual grasses. By all accounts, it is unfeasible to collect this type of information using field-based methods in a timely manner for large landscape extents.

It is here then, that remote sensing datasets and ecological modeling can be used to generate landscape-level products that could be used as surrogates to discriminate current conditions of the land that approximate a plant community composition. In addition, with the availability of long-term remote sensing datasets such as the Landsat Thematic Mapper (TM), it is possible to study multi-temporal dynamics of change to provide useful information for monitoring sagebrush ecosystems. Nevertheless, traditional methods of classifying the land (land use / land cover) into highly homogenous classes may not be appropriate to address the types of problems mentioned above. Said classification methods typically rely on the assumption that sharp boundaries exist between land use/cover classes. This is typically not the case and there are continuums of change from plant communities to other plant communities. If, for instance, our goal is to discriminate different stages of
dominance or degrees of encroachment between shrublands and woodlands then it is imperative that we be able to map this continuum of change. This is the main thrust of the work we present here. To develop a transparent remote sensing protocol that provides maps of this continuum of sagebrush – woodland relationships that may be associated with developed ESD and STM in rangelands to better understand what has happened and is currently occurring in rangelands of the Intermountain West.

Vegetation Continuous Fields (VCF) is a relatively new concept that attempts to deal with the mapping of this continuum. VCF consists of proportional estimates of canopy cover for different vegetation types by modeling remotely sensed data. Estimates for woody vegetation, herbaceous vegetation, and bare ground are available worldwide (Defries et al. 2000; Hansen et al. 2003b) but this product is only available at very coarse spatial resolutions. A series of VCFs consist of several continuous response surfaces (one for each vegetation cover type) in which every pixel value corresponds to a percent canopy cover estimate predicted through a regression model. The VCF offers an advantage over traditional discrete classifications because each pixel is attributed as a percent of canopy cover. Therefore, areas of heterogeneity are better represented as compared to traditional land cover/use classification (Hansen et al. 2002).

In this context the objectives of our research may be stated as follows: (a) To develop a series of Vegetation Continuous Fields models for shrubs, woodland, herbaceous vegetation and bare ground for a semi-arid shrub-steppe landscape, and (b) To analyze the dynamics of change in the continuum of sagebrush and woodlands by comparing a multi-temporal series of VCF for the years 1996, 2001, and 2007.
METHODS

Study Area

Our research was conducted in the northwest corner of Box Elder County, Utah (114°2’31.2’’ - 112°43’40.8’’ West longitude and 41°6’27.36’’ – 41°59’59.64’’ North latitude). The area covers an extension of 722,445 hectares excluding barren lands and bodies of water. Salt Desert Scrub currently occupies about 143,863 hectares or about one fifth of the area while Big Sagebrush Shrubland covers nearly the same amount (19%). Pinyon-Juniper ecosystems are an important part of the landscape taking up 12% of the area. The remainder consists of Greasewood Flats, Montane Sagebrush Steppe, Xeric Mixed Sagebrush Shrubland and Invasive Annual Grasslands (Program 2004). The elevation ranges from 1,278 m in the lowlands close to the Great Salt Lake to 3,027 m in the Raft River range. The mean elevation is 1,520 m.

Mean annual precipitation is approximately 267 millimeters that typically falls in the form of winter snows and spring rains. The climate is generally dry with temperatures that are usually cold in the winter (daily average of 26 °F) and moderately hot in the summer (daily average of 69 °F). The yearly average temperature is 46 °F (PRISM Climate Group 2004). Three distinct physiographic areas can be identified on the study area: basin floors, piedmont slopes, and mountainous areas. The basin floor consists of playas, salt flats, and beaches that are part of the Great Salt Lake Desert. The mountainous areas have steep and very steep slopes. The northern mountain ranges are the Raft River, Grouse Creek, and Goose Creek mountains. The piedmont slopes consist of alluvial fans, lake terraces, fan terraces and related fluvial and lacustrine landforms. This physiographic area surrounds the mountains and extends to the playas.
The vast majority of the streams are intermittent. The soils range from saline nonproductive in the desert to fertile with a high content of organic matter in the mountains. In much of the area, the soils have a root-inhibiting layer within one meter of the surface (Loerch et al. 1997). The study area has undergone various disturbance regimes ranging from grazing, burning, drought, and flooding events (Sant 2005).

Field Data

Multi-temporal field estimates of percent canopy cover for shrubs, trees, herbaceous (grasses and forbs) vegetation, and bare-ground were used to develop our VCF models. These data were prepared as geo-referenced field points and were obtained from different sources: (a) 482 points collected by the South West Regional GAP (SWREGAP) project during 2001 (Lowry et al. 2007), (b) Points collected by The Nature Conservancy (TNC) for the Northwest Utah Landscape modeling project in 2007 (Conservancy 2009), (c) Field points that we collected during a field season in 2007. In total, 135 field observations were available for the year 2007. A fourth data set was available from the Utah Division of Wildlife Resources (UDWR) (Resources 2010) for the years 1996, 2001, and 2006. Fig. 3-1 contains the spatial distribution of the different datasets across the study area.

With the exception of the UDWR dataset, which are permanent sample plots, the rest of the field points were visually assessed in terms of percent canopy cover for shrubs, trees, herbaceous vegetation, and bare-ground on an area that resembled a 3 x 3 Landsat TM pixel (approximately 90 x 90 meters). The points were observed either from a position situated in higher terrain or by utilizing a platform that allowed making visual estimates of percent cover for the life forms of interest as well as for bare ground. The percent cover estimates were recorded using 5% increments for each life form and bare ground (i.e. 0%, 5%, 10%, etc.).
The sum of the percent cover for shrubs, herbaceous vegetation, trees, and bare ground totaled 100% at each point. With regards to our sampling schemes we must report that we did not follow a strict design. Rather, we collected information on a purposive way visiting sites located along an elevation range that included Wyoming (*Artemisia tridentata ssp. wyomingensis*), basin (*Artemisia tridentata ssp. tridentata*), and mountain big sagebrush (*Artemisia tridentata ssp. vaseyana*) communities.

Because the study area is primarily composed of open rangeland, a relatively fewer number of samples were available to model trees. In order to overcome this problem, we captured additional samples to enhance the training dataset for trees. To generate these additional samples, an object-oriented classifier algorithm (Laliberte et al. 2004) was used to segment high spatial resolution imagery from the National Agricultural Imagery Program in order to extract woody vegetation percent cover estimates. The intention here was to sample areas on the landscape whose major component was trees, thus guaranteeing a better representation in the model. These data had a response value for tree but not for shrubs, herbaceous or bare ground.

**Explanatory Variables: Remote Sensing and Topography**

Remotely sensed images and topographic datasets were used as explanatory variables for our modeling. With regards to the remotely sensed data, we obtained scenes from the Landsat TM satellite Path 39 Row 31. Due to the underlying differences in phenology that most vegetation types exhibit in semiarid landscapes (Bradley and Mustard 2008), we decided to acquire imagery from multiples dates within the years 1996, 2001, and 2007. These years were chosen to coincide with field data availability. Within year seasonal imagery allowed us to capture major phenological variations that occur during the growing
season. Landsat TM imagery collection was concentrated during late spring, mid summer and early fall. An effort to obtain only imagery with the best quality (i.e. minimum cloud cover) was made. Table 3-1 provides a list of Julian dates that were chosen for each year.

Where necessary, imagery was rectified and resampled to a common map projection UTM Zone 12 WGS 1984. Standardization of the imagery was performed by converting the raw digital numbers to exoatmospheric reflectance values using an image-based atmospheric correction procedure (Chavez 1996) with the most up-to-date calibration coefficients for the Landsat TM sensor (Chander et al. 2009).

The Soil Adjusted Vegetation Index (SAVI) was computed for each seasonal date for every year. SAVI may be calculated as follows:

\[
SAVI = \frac{(\rho_{\text{nir}} - \rho_{\text{red}})(1 + L)}{\rho_{\text{nir}} + \rho_{\text{red}} + L}
\]

In the equation \(\rho_{\text{nir}}\) = near-infrared reflected radiant flux, \(\rho_{\text{red}}\) = red reflected radiant flux, and \(L\) = adjustment factor (usually a value of 0.5 may be used for different soils).

SAVI has been reported to work well in semiarid ecosystems because it minimizes soil background effects that are known to affect other indices such as the Normalized Difference Vegetation Index (NDVI) (Huete 1988; Jensen 2007). It has been widely reported that a vegetation index such as SAVI may be used to follow the phenological trajectory or seasonal and inter-annual change in vegetation growth and activity (Jensen 2007).

We also created a new variable from the multi-temporal SAVI and have named it NDSAVI or the normalized difference SAVI (Ramsey 2009). The NDSAVI takes advantage of the contrast between the spring and the summer SAVI and may be used to enhance our understanding of the phenological dynamics of grasses on the landscape. Higher values in the NDSAVI would correspond with higher greenness during the early spring relative to
summer whereas low values of NDSAVI would relate to areas that become green later in the
growing season. This new variable conveys a multi-temporal signature of greenness
variation that may be used to discriminate among different land cover types and particularly
focus on non-native grasses such as cheatgrass that follows this phonological pattern. Within
this environment, this index allows us to identify areas where cheatgrass is a major
component of the plant community.

We estimated the NDSAVI as follows:

\[
NDSAVI = \frac{SAVI_{\text{spring}} - SAVI_{\text{summer}}}{SAVI_{\text{spring}} + SAVI_{\text{summer}}}
\]

We also generated the Normalized Difference Water Index NDWI (Gao 1996) for
every available date of imagery for every year. NDWI takes advantage of the contrast found
between the near and middle infrared bands to provide information about water content.
Forest disturbances have been successfully detected using the NDWI (Jin and Sader 2005),
and thus we thought it appropriate to test its performance in our regression models. The
brightness, greenness, and wetness (BGW) transformation (Crist and Kauth 1986) was also
derived for each of the available dates of imagery. This transformation has been used
extensively to monitor condition and changes in soil brightness, vegetation, and moisture
content respectively (Jensen 2007) and was successfully used for modeling land cover
conditions for the Southwest Regional SWRGAP project (Lowry et al. 2007).

In addition to the remotely sensed information (Landsat TM spectral bands, SAVI, NDWI, BGW), we also generated derivatives from a 30-meter digital elevation model
(DEM). DEM derivatives included slope, aspect, and landform. Two transformations of
aspect, namely southness and westness indices (Chang et al. 2004) and a modification to the
original topographic relative moisture index TRMI (Parker 1982) were also generated. An
existing land cover map from the SWRGAP project was included as an explanatory variable. The inclusion of this type of ancillary information has been documented to greatly improve classification and regression modeling in rangelands (Peterson 2005). A list of the explanatory variables generated during this study may be found in Table 3-2.

**Modeling with Multivariate Regression Trees**

**Background**

Multivariate Regression Trees (MRT) (De'ath 2002) is an extension to Classification and Regression Trees (CART) (Breiman et al. 1984). CART consists of non-parametric algorithms that require no prior assumption of normally distributed training data. This is an advantage with satellite reflectance information whose distribution is seldom normally distributed. In addition, CART allows including categorical variables and other ancillary datasets such as elevation and its derivatives, which have proven to increase the accuracy in regression as well as classification exercises. For remote sensing datasets, regression trees have been demonstrated to be a robust tool for handling nonlinear relationships (Homer et al. 1997; Lowry et al. 2007). They can also handle complex interactions among covariates. Regression with CART assumes a univariate response while a regression with MRT provides a multivariate response. CARTs use recursive binary splitting of the data to “grow” the classification tree, where each branch (split) is defined by a straightforward binary rule. The splits are generally chosen to minimize the impurity of the resulting two nodes. The terminal nodes are also known as leaves. With a regression tree, the impurity of a node is defined as the total sum of squares (TSS) of the response variable about the node mean. Each split diminishes the TSS within the two nodes formed by the split. Consistently, this maximizes the between-nodes sums of squares (Breiman et al. 1984). Regression with CART can be
extended to MRT by replacing the univariate response with a multivariate response. In order to do this the impurity of a node has to be redefined as the sum of squares about the multivariate mean. Comprehensive details about MRT may be found in De'ath (2002).

While working on a regression problem with CART or MRT, there are two major moments: one is when the tree is grown, and another when the tree is pruned. Pruning is required because tree algorithms typically over fit the final model by partitioning the data into overly small samples that are inadequate to properly differentiate between two response variables. This is also known as “over fitting” or “over learning.” A more generalized tree (with a lower risk of over fitting) can be obtained through several techniques such as cross-validation or V-fold cross-validation. We used cross-validation in our research to deal with pruning, and thus obtained generalized, yet more effective regression trees for each response variable. We decided to test MRT to predict percent cover estimates because it provided us with the opportunity to simultaneously model our four response variables while other regression algorithms (i.e. linear regression, CART, Random Forests) deal with one response variable at a time. Because the driving force of our research was the extraction of percent cover estimates for shrubs, trees, herbaceous vegetation, and bare ground for each individual pixel (in essence a multivariate response for each pixel) it made sense to test the performance of an algorithm that could model all four variables at once.

Tree Growing and Pruning

We used the independent variables (Table 3-2) and our four response variables for training of the model with 70% of our available data. We did this independently for each year: 2001 and 2007. Previous work to identify which set of variables to use was not conducted since CART and MRT may also be used as part of a data-mining process (De'ath
2002), which is often used to discriminate the most important variables. Growing and posterior pruning of the multivariate regression tree was conducted to determine the size of the tree. We used cross-validation which generated a series of trees (500 cross-validations in our case) from which a relative error and a cross-validated error were obtained. The cross-validated error may be used to objectively determine the size of the tree.

Using Fig. 3-2, we can see that with a bigger tree size, the relative error diminishes (green dashed line). On the other hand, the cross-validated relative error (blue dashed line) decreases to a minimum for a tree size of five leaves, and then increases. The vertical bars (blue) indicate one standard error for the cross-validated relative errors, and the solid line (red) indicates one standard error above the minimum cross-validated relative error.

Following a 1-standard error rule: "the smallest tree within one standard error should be selected" (Breiman et al. 1984), we chose to utilize a tree with five leaves. We provide details in table 3-4 about the size of the tree selected; cross-validated relative errors, and most important variables utilized by the model for the years 2001 and 2007. The complexity parameter (cp) is used to optimize the size of the tree. Construction of the tree does not continue when the cost of adding another variable to the current node is above the value of cp (Williams 2010). For instance, if we were to set the value of cp to zero then a tree will be built to its maximum depth, and thus a large tree will be obtained. De'Ath (2002) defines the relative error (RE) as "the total impurity of the leaves divided by the impurity of the root node". RE provides an over-confident estimate of how accurately a tree will predict for new data. The predictive accuracy is better estimated from the cross-validated relative error (CVRE). The range of CVRE is zero for a perfect predictor and approaches one for a poor predictor.
We used the R package MVPART to conduct our data mining and regressions (R-Project 2010). We then used the package YaImpute (Crookston and Finley 2008) to extract the model for each MVPART run, and then applied a predict function to the geospatial explanatory layers to generate a continuous response surface (percent cover estimates for shrubs, trees, herbaceous vegetation and bare-ground) for the entire study area. A new function in R was written to decompose the multivariate response and generate individual VCF maps.

It should be noted that we utilized the MRT model extracted from the 2001 dataset and applied it to the suite of 1996 independent layers to generate VCF maps. Recall that we did not have an appropriate field dataset with which to train a MRT model for 1996. We assumed that a model obtained for the year 2001 could be applied to the 1996 explanatory layers because our imagery was radiometrically corrected and the selected scenes were collected during the same months, thereby reducing differences due to phenology variations.

**Regression with Random Forests**

**Background**

Random Forests (RF) is a relatively new statistical method that emerged from the machine learning literature, and is based on the same philosophy as CARTs. In RF, multiple bootstrapped regression trees without pruning are created. In a typical bootstrap sample, approximately 63% of the original observations occur at least once (Cutler et al. 2007). The data that are not used in the training set are termed “out-of-bag” observations and are customarily used to provide estimates of errors (Prasad et al. 2006). Out-of-bag samples are also used to calculate variable importance (Cutler et al. 2007). In RF, each tree is grown with a randomized subset of predictors, which equal the square root of the number of variables. In
general, 500 to 2000 trees are grown, and averaging aggregates the results. The method is very effective in reducing variance and error in multi-dimensional datasets. One of the strengths of RF is that because it grows a large number of trees, the method tends not to over-fit the data, and because the selection of predictors is random, the bias can be kept low (Prasad et al. 2006). A comprehensive description of the method may be found in Sutton (2005), Lawler et al. (2006), Prasad et al. (2006), and Cutler et al. (2007).

In this research there were two phases for Random Forest. First, we used the algorithm to find the best subset of variables that should be included during training of the model. We did this for each VCF that we wanted to generate: shrubs, trees, herbaceous vegetation, and bare-ground. For this we used the concept of variable importance. Second, once we had determined the group of variables to use, we modeled each VCF individually. This is in contrast to what was done using MRT, where we modeled all VCFs simultaneously. In this context a different set of variables was used for each modeled VCF.

**Variable Importance and Parsimony**

To select the initial set of variables to model each VCF, we followed the underlying principle that the phenological pattern of a given vegetation type should dictate which remotely sensed datasets to use (Bradley and Mustard 2008). For example, it is sensible to use only one scene (mid summer for instance) to model bare ground percent cover due to its relatively constant spectral response throughout the year. On the other hand, it makes sense to utilize two to three images (i.e. mid summer and early fall) to model herbaceous vegetation due to its conspicuous phenological signature which peaks during the summer and then significantly decreases during the fall.
In order to develop a simple yet effective model, we used the concept of variable importance from the RF algorithm (Cutler et al. 2007). Variable importance is based on the mean decrease in accuracy concept, and is assessed based on how much poorer the predictions would be if the data for that predictor were permuted arbitrarily. This gives an overall impression of the impact that a specific variable has in decreasing precision in predictions. For a comprehensive explanation of variable importance, we refer to Cutler et al. (2007).

We prepared our independent variables (Table 3-2) and our dependent variable (percent cover of shrubs, trees, etc.) for a RF run in order to assess variable importance. We ran this process 500 times and stored the percent increment in the mean square error (%IncMSE) in a matrix. We did this because variable importance may change significantly from a specific run to another. We then analyzed plots of the percent increment in MSE. Oftentimes it is easy to know where to set a cutoff point to determine which variables are more important based on an abrupt decline in the %IncMSE from a group of variables to another. Nevertheless, this is not always the case and the user may choose to select 10 or fewer variables with the highest importance. This is clearly a subjective approach to choosing the most important variables but we have not found in the literature a more objective way to do it.

Whether a steep decline is found in the variable importance plots or the “x” variables with the highest importance are chosen, the fact that those variables may still be highly correlated with each other cannot be disregarded. For any modeling scheme, this is particularly bad because of the instability created by redundant data. Random Forests does not guarantee that the variables with highest importance are not correlated because when it randomly permutes the values for the out-of-bag observations it works with one variable at a
time (Cutler et al. 2007). We recognized this potential problem, and we therefore subjected the “most important variables” to additional scrutiny. We generated scatter plots of all the possible combinations of the most important variables and graphically assessed them for association problems. As expected some of the Landsat TM spectral bands were highly correlated and thus were not included in the model even though RF rated them as highly important. Table 3-5 provides a detail of the variables utilized to model each of the VCFs.

**Regression**

We used the R package randomForest (Liaw and Wiener 2002) to conduct our data mining, variable importance, and develop regression trees to calculate the VCF. We ran the regression separately for each of our four response variables (i.e. shrubs, trees, etc.) using the selected subset of variables determined to be most important. Only two user-supplied parameters are needed to run the regression algorithm: the number of trees which in our case we set to 1000, and the number of variables randomly chosen at each split (the \textit{mtry} parameter). It has been shown (Prasad et al. 2006; Walton 2008) that the selection of different values of \textit{mtry} does not affect the performance of the algorithm. Based on this suggestion, we used a value of one (1) for this parameter.

The R package YaImpute (Crookston and Finley 2008) was used to extract the model for each randomForest run, and then applied a predict function to generate a continuous geospatial response surface for the entire study area. To obtain the prediction maps for 1996 we applied the model that had been fitted to the 2001 field data set.

**Validation and Comparison Metrics**

For validation purposes, 30% of the field observations for both the 2001 and 2007 years were withheld during modeling. Correlation between the model’s predictions and
observed (field data) percent cover were calculated. A perfect estimation by the model would render an r-value of 1.0 between observed and predicted values. We calculated Pearson’s correlation coefficients for each of our VCF predictions. Even though each modeling method generates its own validation estimate, we needed an independent metric to effectively compare between the two methods. Using the withheld set of data, we calculated two metrics: mean absolute error (MAE) and root mean square error (RMSE). MAE is the average absolute difference of the predicted value from the field-observed estimate, while RMSE is the square root of the mean squared error (Prasad et al. 2006; Walton 2008). These two metrics are generally correlated, but errors in RMSE tend to be larger due to the square in the term error. We report both metrics to better understand which method performed better in predicting percent cover estimates.

A second piece of evidence to compare the modeling methods was the output maps generated using YaImpute. Recall that one of our main goals is to produce a geospatial representation of the continuum of the sagebrush – woodland percent cover interactions. Therefore, a model that best performs in doing so would be preferred.

**Multi-temporal Dynamics**

Once we determined which method performed better in predicting and best representing the continuum of sagebrush – woodland, we assessed how well the multi-temporal differences corresponded with what had been observed on the ground. Specifically, we wanted to see how well the decrease in shrub cover and the expansion of woodlands into shrubland had been modeled. To do this, we extracted from our 2007 field database only those points that had been identified as showing some degree of woodland encroachment into shrublands. Our assumption was that the model should be able to identify these dynamics in
terms of either the expansion of trees or a reduction in shrubs. We identified 20 samples and for each point, we created a buffer area of 1 hectare. The size of this buffer was approximate to the dimension of a 3x3 grouping of Landsat TM pixels that represented the approximate area sampled on the ground. We assumed that this window would be big enough to summarize the spatiotemporal context of change in shrub and woodland cover. Zonal statistics (mean, standard deviation, median, majority, etc.) were calculated for each output VCF grid for each year (i.e. percent cover for shrubs in 1996, 2001, etc.). We then compared the temporal variations of mean percent cover for shrubs and trees for our period of analysis by plotting the mean response in cover for the VCF of interest against each year.

Because woodland encroachment may be a slow process (Miller et al. 2005), we decided to analyze the dynamics of change from 1996 to 2007. We acknowledge that this is a short time span to detect woodland encroachment, but thought it would be helpful to see and compare the results to known locations of encroachment.

The individual VCF layers (shrubs, trees, herbaceous, and bare ground) for 1996 and 2007 were stacked into a multi-layer GIS grid. In the attribute table of the new grid, the rows represents the different combinations of percent cover, the columns the different VCF for the two years. Out of this array we were able to estimate differences between 1996 and 2007 in terms of percent cover and identify areas that have shown increases or decreases in shrub, trees, herbaceous, or bare ground. We defined encroachment as those pixels that have shown the following simultaneous criteria: decreases in shrub cover, increases in tree cover, and increases in bare ground cover from 1996 to 2007.
RESULTS

Multivariate Regression Trees vs. Random Forests

Multivariate Regression Trees (MRT) provided a model (Fig. 3-3) that used three
variables: NDSAVI, Greenness, and Wetness for Julian date 182. Such a simple model was
not obtained from the Random Forests (RF) run due to the inherent characteristics of the RF
algorithm in which hundreds or thousands of trees are generated to obtain a final model.
Nevertheless, this complexity and inability to completely interpret the RF model was
compensated with a better prediction power over MRT.

Figure 3-4 shows the differences in model fit for two VCF layers: shrubs and bare
ground. Based on these results, RF clearly out-performs the MRT model. Similar
differences were found for herbaceous and trees. This is reinforced with the data shown in
table 3-6 that contain the Pearson’s Correlation Coefficients, mean absolute error (MAE), and
root mean square error (RMSE) for 2001 and 2007. A global average of the correlations was
0.45 for MRT versus 0.65 for RF for 2001 and 0.34 and 0.52 for 2007. It was also interesting
to see that the results for 2001 were quite better than those for 2007. This is probably due to
differences in the size of the training data sets. With the exception of the VCF for trees
(Table 3-6), Random Forests also outperformed Multivariate Regression Trees by showing
lower prediction errors. We include figures 3-5, 3-6, and 3-7 that illustrate samples of our
VCF spatial layers generated using MRT and RF. Notice that the continuum of percent cover
for the different life forms and bare ground is better represented by RF. A careful look at the
maps will reveal that the layers generated by MRT are more generalized and nearly mimic a
classification output as opposed to RF that is able to produce a more specific and continuous
surface. Based on our assessment of the available evidence (Pearson’s correlation, scatter
plots, and spatial predictions) we decided that for our objectives, Random Forests performed better than Multivariate Regression Trees in modeling continuous fields of vegetation.

**Multi-temporal Dynamics: Potential Encroachment**

We were particularly interested in how well the modeled VCF for shrubs and trees represented landscape change. If VCF can effectively represent spatiotemporal dynamics, then areas of reduction in shrub cover as well as areas of increase in tree cover may be identified. This presents an advantage because woodland encroachment into big sagebrush communities may be mapped in relatively large landscapes. Figure 3-8 shows the changes in percent cover for shrubs and trees from 1996 to 2001, and 2007. These changes in percent cover were calculated for those sites for which we had field evidence that woodland encroachment is occurring. Of 20 sites with known encroachment (Fig. 3-8a), approximately half show increases in percent cover of trees from the modeled VCF from 1996 to 2007. With a couple of exceptions in which reductions in tree cover were detected, the rest of the sites show slight increases in percent cover for the 11-year period. Most of the sites show decreases in percent cover of shrubs for the same duration of time (Fig. 3-8b). This does not necessarily mean that active woodland encroachment is to blame however. Big sagebrush sites may also be subject to invasion by annual grasses or other disturbances.

The results of our analysis of woodland encroachment are presented in Fig. 3-9 and Table 3-7. In this figure we also included the three land cover classes from the SWRGAP project (Lowry et al. 2007) that are associated with big sagebrush (*Artemisia tridentata* spp.). Figure 3-10 shows an enlarged section with locations of known woodland encroachment. For plots A and B there is a visible correlation between ground observations and our
spatiotemporal analysis. A list of the 20 sites with known woodland encroachment may be found in Table 3-8.

DISCUSSION

Many examples of the use of regression and classification techniques to depict sub-pixel heterogeneity can be found in the literature. In rangeland environments, work has been conducted to model woody vegetation cover (Danaher et al. 2004), bare ground cover (Weber et al. 2009), and shrub cover and encroachment (Laliberte et al. 2004). With the exception of the MODIS global continuous vegetation maps (Hansen et al. 2003a), the examples above dealt with only one response variable. In our work there were four response variables and we tested two statistical methods: Multivariate Regression Trees and Random Forests. We ultimately decided to use Random Forests due to a better performance in predictions and a more sensible spatial response. We must clarify that Random Forest only dealt with one variable at a time, and therefore we did not model a multivariate output in the end.

MRTS vs. RF: Ease of Understanding and Predictive Power

Since a pixel is in essence an integrated multi-dimensional spectral response of vegetation, bare ground and other features, then it makes sense to attempt to decompose that response to understand the dynamics of a given pixel and surrounding landscape. It is also sensible to utilize a statistical method that is able to model the various landscape components simultaneously. This is the case for Multivariate Regression Trees, but not for Random Forests, which models each component individually. This is one of the big differences between the statistical methods that we tested. With regards to the ease of understanding, we were able to see that every leaf or terminal node in a MRT tree (Fig. 3-3) is a mean multivariate response that can be later decomposed into the continuous fields of interest (i.e.
% shrubs, % trees, etc.). There is a clear advantage with a parsimonious MRT because the model can be interpreted with little effort. Conversely, Random Forests has been described as a “black box” (Prasad et al. 2006) because the individual trees cannot be examined separately due to the sheer number of trees that may be generated. This is another divergence between these methods and it is clear that MRT holds an advantage over RF. Nevertheless, Random Forests provide metrics that may aid in interpretation. One metric is variable importance, which can be used to compare relative importance among predictor variables. Such a feature is not available in MRT and therefore the importance of variables must be determined after a careful data mining process.

In this research RF outperformed MRT in terms of prediction power (Table 3-6, Fig. 3-4). The ease of understanding of MRT may be a strong attraction for their utilization if one were just using these models for ecological interpretation of a process. Nevertheless, for our work we were more interested in the predictive power because we wanted to assess additional products (i.e. dynamics of potential encroachment), which depended on the modeled VCF layers. MRT may perform better with larger training data sets; this was not explored during our work however.

**Multi-temporal Dynamics in Big Sagebrush**

An assessment of the performance of VCF for tree encroachment into shrublands (Figs. 3-8, 3-9, 3-10) provided positive feedback on how well these models could characterize this transitional process. We must acknowledge that not all of the decreases in shrub cover may be due to active woodland encroachment since the decline in big sagebrush is not always proportional to the increase in woodlands (Miller et al. 2005). Compounding factors such as errors in the predictions, annual grasses invasion, and wildfires among others
may impact the behavior observed in Fig. 3-8b. Still, we feel that the results appropriately depict the spatiotemporal dynamics in big sagebrush cover.

We present Fig. 3-9 that shows where encroachment has been modeled. We must clarify that this prediction does not directly imply that encroachment actually happened. On the other hand, the clusters of modeled woodland encroachment shown on Fig. 3-10 showed a fair correlation to what had been observed on the ground. We still do not know the age of trees growing in the foreground though. We recognize that 11 years is a relatively short time span to detect encroachment. With only two dates, we cannot provide estimates of the rate of woodland encroachment. As long as 45-50 years may be needed to see evidence of the tree overstory suppressing understory shrubs (Miller et al. 2005). Rather, our work was conducted to use remote sensing and ecological modeling to pin down where on the landscape woodland encroachment may be an issue.

We have developed a collection of continuous vegetation fields for three major life forms and for bare ground. Our comparison of the performance between Random Forests RF and Multivariate Regression Trees MRT to develop the VCF indicated that RF surpassed the accuracy of MRT in predictions for shrubs, trees, grasses and bare ground. In addition RF was more competent than MRT in spatially characterizing the continuum of land cover. If results can be improved with larger training data sets, then the outputs from MRT may provide improved models. The generation of multi-temporal VCF for three different periods provided us with the ability to assess the dynamics of change with a focus on encroachment by trees. Although the time span of our analysis is relatively short (11 years), the understanding of the spatiotemporal changes in life-form occupancy led us to model and quantify areas that have shown potential active encroachment.
IMPLICATIONS

Our development of a multi-temporal collection of VCF may be used to update information about the status or condition of a particular ecological site as well as characterizing the states and transitions for that site. For instance, a specific spatial unit of an ecological site may be characterized in terms of its occupancy by shrubs, grasses, trees, and bare ground using modeled continuous fields. Knowledge about the relative dominance of these life forms in a particular unit may shed light about its current condition relative to a reference condition. Our models of VCF may provide knowledge about usage of the ground by major life forms and bare ground and in this way pinpoint areas that are diverging from a reference condition.

LITERATURE CITED


Table 3-1. Landsat TM Path039 Row031 dates collected

<table>
<thead>
<tr>
<th>Year</th>
<th>Julian Dates</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>105, 201, 281</td>
<td>04/01, 07/19, 10/07</td>
</tr>
<tr>
<td>2001</td>
<td>118, 182, 278</td>
<td>04/28, 07/01, 10/05</td>
</tr>
<tr>
<td>2007</td>
<td>119, 183, 263</td>
<td>04/29, 07/02, 09/20</td>
</tr>
</tbody>
</table>
Table 3-2. Explanatory variables compiled for modeling CFV

<table>
<thead>
<tr>
<th>Variable</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landsat TM blue, green, red, near-infrared,</td>
<td>For each year being modeled, we had six bands per date, and then three dates (table 3-1) per year: equals 18 TM variables of reflectance</td>
</tr>
<tr>
<td>middle-infrared 1 and 2 bands</td>
<td></td>
</tr>
<tr>
<td>SAVI</td>
<td>Soil-adjusted vegetation index, one per date; three SAVI per year</td>
</tr>
<tr>
<td>NDSAVI</td>
<td>Normalized Difference SAVI, one per year</td>
</tr>
<tr>
<td>DEM</td>
<td>Digital Elevation Model</td>
</tr>
<tr>
<td>DEM Derivatives</td>
<td>Slope, Aspect, Heat Index (Beers et., 1966), Topographic Relative Moisture Index TRMI, Southness, Westness</td>
</tr>
<tr>
<td>Distance to roads</td>
<td>A grid with values representing distances to primary, secondary and tertiary roads</td>
</tr>
<tr>
<td>BGW</td>
<td>Brightness – Greenness – Wetness Transformation for each Julian Date: Nine bands per year</td>
</tr>
<tr>
<td>NDWI</td>
<td>Normalized Difference Water Index, one for each Julian Date: Three per year</td>
</tr>
<tr>
<td>SWRGAP</td>
<td>Land cover classes from the Southwest Regional GAP Project</td>
</tr>
</tbody>
</table>
Table 3-3. Field sampling data sets used for model building and validation

<table>
<thead>
<tr>
<th>Year</th>
<th>Source</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>Big Game Range Trend Studies</td>
<td>26</td>
</tr>
<tr>
<td>2001</td>
<td>SWRGAP</td>
<td>482</td>
</tr>
<tr>
<td>2007</td>
<td>Northwest Utah Landscape Modeling Project / Own field points</td>
<td>135</td>
</tr>
</tbody>
</table>
Table 3-4. Summary of Multivariate Regression Trees

<table>
<thead>
<tr>
<th>Year</th>
<th>Size of tree (leaves)</th>
<th>Cross-Validated Error</th>
<th>Variables used during splitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>5</td>
<td>0.756</td>
<td>NDSAVI, Greenness JD182, Wetness JD182</td>
</tr>
<tr>
<td>2007</td>
<td>4</td>
<td>0.857</td>
<td>Greenness JD183, Wetness JD263, Brightness JD183</td>
</tr>
</tbody>
</table>
Table 3-5. Variables utilized to fit individual CFV using Random Forests

<table>
<thead>
<tr>
<th>CVF</th>
<th>Variables Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trees</td>
<td>SWRGAP – Brightness JD 182 – Greenness JD 182 – Wetness JD 182 – NDWI JD 182</td>
</tr>
<tr>
<td>Herbaceous</td>
<td>SWRGAP – Brightness JD 182 – Greenness JD 182 – NDSAVI – Brightness JD 278 – Greenness JD 278 – Wetness JD 278</td>
</tr>
<tr>
<td>Bare ground</td>
<td>Brightness JD 182 – Greenness JD 182</td>
</tr>
</tbody>
</table>

JD: Julian Date from Table 3-1
Table 3-6. Validation metrics between Multivariate Regression Trees (MRTS) and Random Forests (RF).

<table>
<thead>
<tr>
<th>Pearson’s Correlation</th>
<th>MAE *</th>
<th>RMSE **</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>MRTS</td>
<td>RF</td>
</tr>
<tr>
<td>Shrubs</td>
<td>0.49</td>
<td>0.72</td>
</tr>
<tr>
<td>Trees</td>
<td>0.23</td>
<td>0.52</td>
</tr>
<tr>
<td>Herbaceous</td>
<td>0.56</td>
<td>0.77</td>
</tr>
<tr>
<td>Bare ground</td>
<td>0.55</td>
<td>0.62</td>
</tr>
<tr>
<td>Average</td>
<td>0.45</td>
<td>0.65</td>
</tr>
</tbody>
</table>

* Mean Absolute Error

** Root Mean Square Error
<table>
<thead>
<tr>
<th>SWRGAP land cover class</th>
<th>Area with potential encroachment (hectares)</th>
<th>Percentage of the land cover class area affected by encroachment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-mountain basins big sagebrush shrubland</td>
<td>17570.7</td>
<td>12.6</td>
</tr>
<tr>
<td>Inter-mountain basins big sagebrush steppe</td>
<td>349.8</td>
<td>13.0</td>
</tr>
<tr>
<td>Inter-mountain basins montane sagebrush steppe</td>
<td>4001.6</td>
<td>5.3</td>
</tr>
</tbody>
</table>
Table 3-8. Field points with observed woodland encroachment into big sagebrush areas

<table>
<thead>
<tr>
<th>UTM East</th>
<th>UTM North</th>
<th>Source</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>284586.9</td>
<td>4643882.0</td>
<td>TNC</td>
<td></td>
</tr>
<tr>
<td>284856.6</td>
<td>4642757.0</td>
<td>TNC</td>
<td></td>
</tr>
<tr>
<td>291276.0</td>
<td>4642919.0</td>
<td>TNC</td>
<td></td>
</tr>
<tr>
<td>330719.7</td>
<td>4632703.0</td>
<td>TNC</td>
<td>Invasion by annual grasses observed</td>
</tr>
<tr>
<td>276140.7</td>
<td>4634808.0</td>
<td>TNC</td>
<td>Invasion by annual grasses observed</td>
</tr>
<tr>
<td>276274.6</td>
<td>4634564.0</td>
<td>TNC</td>
<td></td>
</tr>
<tr>
<td>268122.8</td>
<td>4615535.0</td>
<td>TNC</td>
<td></td>
</tr>
<tr>
<td>272954.8</td>
<td>4605490.0</td>
<td>TNC</td>
<td></td>
</tr>
<tr>
<td>264862.5</td>
<td>4608384.0</td>
<td>TNC</td>
<td></td>
</tr>
<tr>
<td>253536.9</td>
<td>4615177.6</td>
<td>TNC</td>
<td>Invasion by annual grasses observed</td>
</tr>
<tr>
<td>316947.0</td>
<td>4631891.0</td>
<td>AJHC</td>
<td></td>
</tr>
<tr>
<td>300526.0</td>
<td>4632667.0</td>
<td>AJHC</td>
<td></td>
</tr>
<tr>
<td>262340.0</td>
<td>4629380.0</td>
<td>AJHC</td>
<td></td>
</tr>
<tr>
<td>262130.0</td>
<td>4619710.0</td>
<td>AJHC</td>
<td></td>
</tr>
<tr>
<td>259560.0</td>
<td>4631950.0</td>
<td>AJHC</td>
<td></td>
</tr>
<tr>
<td>255384.0</td>
<td>4638570.0</td>
<td>AJHC</td>
<td></td>
</tr>
<tr>
<td>321250.0</td>
<td>4584520.0</td>
<td>AJHC</td>
<td></td>
</tr>
<tr>
<td>320640.0</td>
<td>4583520.0</td>
<td>AJHC</td>
<td></td>
</tr>
<tr>
<td>303269.0</td>
<td>4614580.0</td>
<td>AJHC</td>
<td></td>
</tr>
<tr>
<td>277695.0</td>
<td>4604143.0</td>
<td>AJHC</td>
<td></td>
</tr>
<tr>
<td>276008.0</td>
<td>4601843.0</td>
<td>AJHC</td>
<td>Invasion by annual grasses observed</td>
</tr>
</tbody>
</table>
Figure 3-1. Study area in Northern Utah and distribution of field observations to model multitemporal VCF. The study area is shown in the context of the State of Utah.
Figure 3-2. Cross-Validated Relative Errors for different complexity parameter (cp) values to select the optimum tree size from a MRTS run (2001). A description of each component for the figure can be found in the text.
Figure 3-3. Tree structure for a MRTS run in 2001 – Notice that each final node is a multivariate composite response which can be decomposed into the VCF of interest
Figure 3-4 Scatter-plots of observed versus predicted percent cover for shrubs and bare ground using Random Forest RF and Multivariate Regression Trees MRTS
Figure 3-5. Maps of shrub percent cover for 2001 using Random Forests and Multivariate Regression Trees
Figure 3-6. Maps of bare ground percent cover for 2001 using Random Forests and Multivariate Regression Trees
Figure 3-7. Maps of trees and herbaceous percent cover for 2001 using Random Forests
Figure 3-8. Changes in predicted percent cover for (a) Trees, and (b) Shrubs from 1996 to 2007 for the buffers of the sites known to have woodland encroachment.
Figure 3-9. Potential woodland encroachment from 1996 – 2007 for the study area
Figure 3-10. Woodland encroachment near Grouse Creek Mountains - Photos: Alexander Hernandez 2007
CHAPTER 4

A LANDSCAPE SIMILARITY INDEX: MULTI-TEMPORAL REMOTE SENSING AND MULTI-DIMENSIONAL SCALING TO TRACK CHANGES IN BIG SAGEBRUSH ECOLOGICAL SITES

Abstract

A similarity index for big sagebrush ecological sites was developed in Northern Utah. In contrast to field measurements used to calculate similarity to reference states, our approach relies on the utilization of historic archives of satellite imagery to measure the ecological distance to benchmarks of undesired conditions such as invasion by exotic annuals and woodland encroachment. Our benchmarks consisted of locations for which there are field data collected for monitoring and evaluation purposes for several periods. We utilized a temporal series of Landsat TM imagery that spanned 1984 to 2008 from which the soil-adjusted vegetation index (SAVI) and other transformations were extracted. Topographic and climatic variables were also included as ancillary data. Multidimensional Scaling (MDS) was used to obtain scores in reduced ordination space for two periods of interest: 1984-1996 and 1997-2008. Inter-annual SAVI mean-variance plots provided evidence that the benchmarks and ecological sites have a distinct temporal response that allows an objective comparison. Our MDS results also show that natural clusters may be identified in the reduced statistical space for ecological sites that are a dominant component of a soil map unit. The two MDS solutions allowed the ordination of ecological sites in two gradients of productivity and bare ground. Interpretations of the transitions and trajectories of mountain, Wyoming, and basin big sagebrush sites correlated well with the ecological expectation. We anticipate that range conservationists and others actively working in rangeland evaluation
may use this application to develop and update ecological site descriptions and state and transition models from a remotely sensed perspective.

**INTRODUCTION**

An Ecological Site Description (ESD) (U.S. Department of Agriculture 2008) and its associated State and Transition Model (STM) (Westoby et al. 1989) provide information about the biophysical properties of a site along with descriptions of the different plant communities that may be found on that landscape. Information about the structure, composition, and dynamics of said plant communities given different disturbances or management scenarios can also be found in the ESD and its corresponding STM (Briske et al. 2005). Generally, an ecological site may be evaluated using three different methods: trend, indicators of rangeland health, and a similarity index (U.S. Department of Agriculture 2008).

Trend is used to determine the direction of change that occurs on a given site. Indicators of rangeland health are qualitative assessments that provide land managers and rangeland specialists with information to evaluate ecological processes, which may be used to identify potential areas of degradation (Pyke et al. 2002). Similarity indices are used to compare the existing conditions with a historic or desired state as defined by the site's STM. The utilization of either trend, rangeland health, or similarity index can provide an indication of disturbances, as well as future management (USDA-NRCS 2006).

These three methods often require comprehensive field surveys to collect the necessary ground data to conduct an ecological site assessment. For instance, in the similarity index, the current method is to collect, classify, and weigh vegetation. This is done because the similarity index currently in use measures how comparable the percentage by weight of the plant community present on the site is to a desired or undesired state. Due to
the high requirement for field data, it is clear that these methods are designed to evaluate specific areas of interest, and that their applicability to assess large landscapes, such as those found in rangelands of the western United States, may be limited due to the costs associated with field surveys.

For any ESD, there may exist a suite of ecological states in its STM. Ecological states are normally distinguished by large differences in plant functional groups, soil properties, ecosystem processes, and consequently in vegetation structure, biodiversity, and management requirements. Ecological states are also distinguished by their reaction to disturbance (Pellant et al. 2005). For rangeland management, it is often necessary to be able to identify where on the landscape particularly undesirable states are present. For instance, for Big Sagebrush ecological sites, it is important to know where those states characterized by invasion of exotic annual grasses or encroachment by woodlands are occurring. It is also important to know which big sagebrush ecological sites are in the process of transitioning to an undesirable state. This type of problem calls for the application of the trend and similarity index methods to shed light about the direction of change, and how similar or dissimilar each site of interest is to undesired states.

Due to the inherent cost of the application of the methods to evaluate ESD, it is unlikely that large regions containing multiple ecological sites may be assessed in a timely manner. It is here then that historic remote sensing data sets can be used to derive quantitative indicators to determine condition in space and time so that trend and similarity of large landscapes may be obtained. An example is the protocol for the ecological monitoring of rangelands using multi-temporal series of SAVI that was prepared for areas of Northern Utah (Washington-Allen 2006). This application demonstrated how historic remote sensing imagery could provide reliable accounts of change in large areas.
In order to detect landscape change, it is necessary to define comparative benchmarks. Benchmarks are standards with which measurements of indicators can be compared (West 1991). Indicators can be composites of a group of measurements that are ideally independent or uncorrelated to each other. If benchmarks can be objectively identified on the landscape and their remotely sensed spectral and temporal signatures are also characterized, then assessments of change can be done for relatively large regions. This is the primary assumption of the research presented in this paper.

To the best of our knowledge, an application that relates historic remote sensing data sets with ecological site descriptions for monitoring and assessment purposes has not been developed. Our objectives may be stated as follows: a) Develop a remote sensing based similarity index for rangelands in the Intermountain West, and b) Assess changes in condition for big sagebrush ecological sites for which preliminary STMs have been prepared.

Even though our similarity index does not have the resolution (spatial and/or thematic) to discriminate individual species, we believe that range conservationists will benefit from our landscape-level assessments that identify which ESD units are likely to be in or approaching an undesired stable state. Work presented in this paper is expected to promote discussion and further methodological refinement on the utilization of remotely sensed datasets for the assessment of ecological sites in rangelands.

**METHODS**

**Study Area**

Our research was conducted in the northwestern corner of the State of Utah, (114°2’31.2” - 112°43’40.8” West and 41°6’27.36” – 41°59’59.64” North) in Box Elder County. We focused our work in the spatial domain of big sagebrush ecological sites that are
contained in the Major Land Resource Area (MLRA) D28A (NRCS 2006), and that have a preliminary or final ESD and STM. Table 4-1 contains a list of the ecological sites that were considered in this study along with a brief description of their main characteristics. Fig. 4-1 depicts the spatial distribution of the ecological sites of interest in the context of the study area. The vegetation in the study area is primarily composed of salt desert scrub, big sagebrush steppe and shrublands, as well as Pynion-Juniper ecosystems (Program 2004). The elevation ranges from 1278 m in the lowlands close to the Great Salt Lake to 3027 m in the Raft River range. The mean elevation is 1520 m. The climate is generally dry, receiving an average of 267 millimeters of precipitation annually typically in the form of winter snows and spring rains. Temperatures are usually cold in the winter (daily average of 26 °F) and moderately hot in the summer (daily average of 69 °F). The yearly average temperature is 46 °F (PRISM Climate Group 2004). The soils range from saline nonproductive in the lower elevations to fertile with a high content of organic matter in the mountains (Loerch et al. 1997). The ownership of the land can be divided into three categories: a) Federal land that is managed by the Bureau of Land Management (BLM) (approximately 41%) and the United States Forest Service (USFS) (about 3%), b) Private ownership accounts for just about 43%, and c) the rest ~13% is owned or managed by the State of Utah. The study area has undergone various disturbances ranging from grazing, burning, drought, and flooding events (Sant 2005).

Ecological Site Units

An ecological site and its description of climate, soils, and vegetation (NRCS 2010a) are related spatially to soil map units (SMUs) delineated by the Natural Resources Conservation Service in the Soil Survey Geographic (SSURGO) database (NRCS 2010b).
Ecological sites are linked to components of one or more soil map unit. Components of map units attempt to capture the variability found within SMUs for soil patches that are below the minimum mapping unit of the SSURGO database. A soil map unit may have up to four components, therefore soil map unit polygons have a one-to-many correspondence with ecological sites (USDA-ARS 2010). Even though this brings about a cumbersome utilization of SMUs to represent ecological sites spatially, the process is possible. The SSURGO database contains an estimate of the percentage of every component that occupies a given SMU. Since the SMUs represents our basic sample unit, and given the potential internal variability of SMUs, we have chosen to develop our models with SMUs that contain a dominant ESD (> 60%). In this way, we can be somewhat assured that soil polygons used to train models are uniform in soil and vegetation. In this paper soil map units SMUs and ecological site units are used interchangeably.

**Benchmarks**

Our rationale is that an assessment of an ecological site’s condition (good, excellent, etc.), and trend (positive, negative, not apparent) may be attempted by comparing the temporally integrated remotely sensed signature of each ecological site unit to the signature of defined benchmarks. We define benchmarks as sites that have been properly identified on the ground and for which there has been credible ecological monitoring. This standard will allow us to assign a particular ecological state to an individual benchmark. Since we were interested in measuring similarity to undesired stable states (i.e. invasion by exotics, encroachment by woodlands), we put more emphasis in obtaining field data sets from which this type of information could be extracted. Field data used to define benchmarks were obtained from the Utah Division of Wildlife Resources, namely the Range Trend Studies
DWR-RTS (DWR 2010). In addition, several ground locations were visited during the growing season of 2007 and 2008. The DWR-RTS sites are surveyed every five years to detect changes in vegetation composition for big game habitats. Established protocols are used to characterize vegetation (species composition, percent cover, density, among others), and its trend (temporal changes in browsing quantity and quality, exotic grasses performance, expansion of woodlands).

For our purposes, data collected during the years 1996, 2001, and 2006 were used to assess the presence of undesired stable states (i.e. Cheatgrass invasion or woodland encroachment) in DWR-RTS plots located within our study area. Table 4-2 and Fig. 4-1 provide information about the sites with a big sagebrush component that were selected as benchmarks in our study. We selected these sites based on an analysis of the narratives and data tables publicly available on the DWR website. These reports offer a comprehensive description of species present on each site and their trends. For instance if a site was described as having an invasion by exotic annual grasses (i.e. Bromus tectorum) and the narrative described a positive trend (i.e. increases in percent cover through time) then that site was selected to be a benchmark for this specific undesired state.

Fieldwork conducted during 2007 and 2008 included visits to several big sagebrush sites (Fig. 4-1) that were clearly affected by either woodland encroachment or annual grasses. The selection of these sites did not follow a strict sampling design; rather these points were collected following an opportunistic sampling scheme. Visual estimates of percent cover for major life forms (grasses, shrubs, trees), and bare ground were obtained at each site and notes were made about the overall condition of the big sagebrush site with regards to undesired stable states. A data set of field points collected in 2007 by The Nature Conservancy TNC (Conservancy 2009) was also available. The field information collected by TNC is quite
similar (i.e. percent cover estimates) to the data we collected in our sites, and it was also
gathered with the intention of identifying critical conditions of big sagebrush ecosystems in
the study area.

Remote Sensing and Ancillary Datasets

We utilized multi-temporal remote sensing datasets as well as climatic and
topographic information to find and describe an integrated spatial response of specific
undesired stable states occurring in each benchmark location. We utilized a time series of
Landsat TM imagery (Path 39 / Row31) from 1984 to 2008. For every year, we obtained one
scene collected during the growing season. An effort to obtain only imagery with the best
quality was made throughout the collection process. Imagery was first rectified and
resampled to a common map projection UTM Zone 12 WGS 1984, and then standardized by
converting the raw digital numbers to exoatmospheric reflectance values using an image-
based atmospheric correction procedure (Chavez 1996) with updated calibration coefficients
for the Landsat TM sensor (Chander et al. 2009). For every year, we derived the Soil
Adjusted Vegetation Index (SAVI) and the Brightness Greenness Wetness (BGW)
components (Crist and Kauth 1986). SAVI has been found to work better than other indices
such as the Normalized Difference Vegetation Index (NDVI) in semiarid environments due
to the minimization of soil background effects (Huete 1988; Jensen 2007) and BGW has been
successfully utilized in the Intermountain West for classification purposes (Lowry et al.
2007). Vegetation indices such as SAVI have been used to follow seasonal and inter-annual
change in vegetation growth and activity (Jensen 2007), thus it made sense to explore its
performance in discriminating features of different stable states in big sagebrush systems.
Table 4-3 contains a list of the dates, elevation angles, and percent cloud cover for the scenes that were used in this study.

In addition to the Landsat TM indices and transformations, we also utilized information generated by the Earth Resources Observation and Science (EROS) data center. EROS has developed a comprehensive suite of remote sensing phenology datasets (USGS-EROS 2010) for the past 20 years (1989 – 2008). This data set includes variables such as (a) beginning and end of measurable photosynthesis in the vegetation canopy, (b) length of photosynthetic activity – the growing season, and (c) canopy photosynthetic activity across the growing season, among others. In total there are nine variables that describe the annual phenological regime throughout the continental United States for the period of record. This type of information was included to enhance our understanding of the spatiotemporal dynamics of vegetation in our study area. To mention an example, those sites that have been invaded by cheatgrass will experience early onsets in photosynthetic activity that may be detected in the multidimensional signal of these phenology products. We provide Table 4-4 containing a list of the phenology variables that were used in this study.

A digital elevation model (DEM) and derivatives such as slope, aspect, compound topographic index (CTI), and a modification to the original topographic relative moisture index (TRMI) (Parker 1982) were also included in this analysis. Climatic variables such as annual average precipitation and maximum and minimum temperatures were also prepared using a continental data set (PRISM Climate Group 2004).

**Preparation of the Modeling Dataset**

Soil Map Units dominated (> 60%) by a specific ESD and the defined benchmarks were integrated into a data set that could be used to assess similarity to undesired stable
states. Our benchmarks were made up of specific geographic point locations while polygons (SMUs) defined our ecological site units. To put everything into the same context, we created 1-hectare buffer areas around the coordinates of each benchmark. We attempted to simulate an area equal to three Landsat TM pixels (90 m x 90 m) in square (8100 m²), which is approximately the area that we assessed during our fieldwork. Once we had a combined shapefile of SMU and benchmark polygons, we utilized a zonal statistics technique to extract the local mean and variance for each polygon from our remote sensing, topographic and climatic data sets. We expected that a temporal series of mean and variance of a vegetation index, transformations (i.e. BGW), or phenology variables would provide insight about the spatiotemporal dynamics of vegetation composition in our benchmarks and SMUs. Thus, our modeling data set consisted of a matrix in which the rows consisted of each benchmark and soil mapping unit correlated to a given ecological site while the columns corresponded to means and variances extracted from each spectral, topographic, or climate variable.

**Approach to Integration and Similarity**

We needed to estimate a unique or integrated value for benchmarks and for ecological site units from their multivariate (vegetation indices, topographic, climatic) response. We approached these problems of integration, similarity, and trend from an ordination perspective. Ordination provides a geometric representation of individuals (benchmarks and soil mapping units in our case) in a low dimensional space, so that the distances between the individuals represent their dissimilarity. In addition, this method has been shown to provide insight into whether natural clusters exist or can be generated from a multivariate dataset (Kelly and Basford 2000). Ordination has also been used to reveal
underlying trends in the composition of vegetation communities through the analysis of changes in vector position in reduced space over time (Foran et al. 1986).

There are many techniques that attempt to condense information from a multivariate data set into a reduced dimensional space. Ordination techniques such as principal component analysis (PCA), canonical correspondence analysis (CCA), principal coordinate analysis (PCoA), among others are available for ordination purposes. In our case, we decided to utilize multidimensional scaling (MDS). As with any other ordination technique, MDS concentrates the original information contained in many variables into a suite of ordered scores for a few new attributes that define the dimensions of the new reduced space (Lattin et al. 2003). We selected this technique because MDS has been known to rearrange objects in an efficient manner through the minimization of stress. In multidimensional scaling, stress measures the difference between the original dissimilarity of the individuals and the way in which this is represented as distances on the ordination space (StatSoft 2010). By controlling stress, MDS provides an excellent representation of the data in which most of the relevant information has been preserved with fewer variables (Kelly and Basford 2000).

**Ordination with Multi-dimensional Scaling MDS**

We conducted MDS on our data set for two periods of interest. The first period spanned from 1984 to 1996, and the second from 1997 to 2008. Our criterion to split the data set into these two periods is based on the fact that vegetation composition and trend information contained in the narratives from DWR-RTS are available for 1996, 2001, and 2006. In this way, 1996 seemed like a reasonable year to partition the data set, and then relate the MDS results to the narratives and tables found in the DWR-RTS reports. The end of the second phase (year 2008) is related to our own collection of field data and two years after the
last DWR-RTS assessment of condition and trend. Prior to conducting MDS, we scaled all the variables due the inherent difference in units among remote sensing, topography, and climate variables. While preparing the data set for modeling, we found that several pixels in the EROS phenology variables had very unusual values that were likely errors or missing data. We needed to declare these records as missing data because we did not have a method to interpolate new values from the surrounding pixels. These types of records had to be removed from the data set because the algorithms that are used to run MDS cannot handle missing values in the distance matrix. When running MDS, there are a couple of issues that need to be addressed, one is which type of distance metric will be used to calculate the matrix of dissimilarities, and the other is to determine the number of dimensions in the reduced space. We observed that the type of distance that is used greatly influences the number of dimensions or axes generated. To determine the number of axes, we ran the Kruskal's Metric Multidimensional Scaling (Cox 2001) implementation in R (R-Project 2010).

We generated different $k$-dimensional MDS solutions to measure improvements in fit as we increased the number of dimensions. Each $k$-dimensional configuration is designed to minimize the stress between the input distances and the distances in reduced space, and thus is our measure of fit (Lattin et al. 2003). From each MDS solution, we extracted the value of stress and prepared scree plots to graphically assess the improvements in fit or reductions in stress. We tested this process with several distance methods (i.e. Euclidean, Maximum, Manhattan, Canberra). We decided to utilize the Manhattan method, which seemed to provide better values of stress with fewer dimensions when compared to the other distance methods. Once the optimum number of dimension was selected, we obtained MDS solutions by using the multidimensional scaling implementation in R. We did this for our two periods of analysis (1984 - 1996, 1997 - 2008) independently.
Similarity and Trend

Our primary assumption was that the ordered scores in the reduced MDS space could be considered as the integrated response for a given state in a specific benchmark or SMU at a certain time. In this context, the distance of any SMU to one or more benchmarks on the MDS axes may be used to assess the spectral/temporal similarity in vegetation mean and variance between each SMU and the benchmarks. Because we generated two temporal MDS solutions (1984-96, 1997-2008), we were able to follow the changes in ordination space for a given SMU and/or benchmark, and interpret whether the trajectories are suggesting changes or stability in remotely sensed metrics.

In order to assess trend, we generated biannual (i.e. 1984-1985, 1985-1986, 1986-1987, etc.) matrices of distance between all our SMUs and benchmarks in our data set. For each SMU and benchmark, we assembled a vector of distances. This vector was generated by extracting values (i.e. distance in 84-85, distance in 85-86, etc.) for a particular SMU-benchmark combination from each of the biannual matrices mentioned above. We then developed plots of the temporal distance to a particular benchmark with a specific state observed in 2006. The plotted multi-temporal trend could then be characterized with regards to the distance behavior throughout the years. For instance, those SMUs for which the distance got smaller with time could be characterized as showing directionality towards an undesired state. On the other hand, SMUs in which the distance increased could be typified as displaying directionality away from an undesired state.

Validation

We assessed how our ordination results for SMUs, that is their similarity to undesired states; compared to observed conditions on the landscape. This is not a simple procedure due
to the inherent variability that is found within SMUs, which may cover hundreds and possibly thousands of hectares. We followed a simple approach. For those SMUs whose ordination scores (1997-2008) were very close to benchmarks of either cheatgrass invasion or woodland encroachment, we visually interpreted high-resolution color imagery from the National Agriculture Imagery Program (NAIP) acquired in 2009 (U.S. Department of Agriculture 2010). In our visual inspection we checked for evidence of shrubs being displaced by trees or grasses. We put more emphasis into those SMUs that had been visited during our fieldwork, and for which pictures were available to corroborate what was observed in the NAIP imagery.

RESULTS

Multi-temporal Signatures: Steady States Plots

An exploratory analysis of the modeling data helped us determine how well the benchmarks and ecological site units were partitioned in multi-temporal space. Figure 4-2 shows the natural clusters derived from plotting the mean and variance of the interannual (1984 - 2008) SAVI for (a) some of our benchmarks, and (b) representative polygons for a suite of ecological sites. In this plot, each point represents the mean/variance of greenness captured for a specific year of the analysis. Only a few benchmarks and polygons are included for graphic simplicity. Although overlaps do exist among benchmarks and ecological sites, naturally occurring groups may still be discriminated by using these two parameters.

The distribution of points in spectral space correlates well with what would be anticipated for the selected benchmarks. For instance, site TS-13 (Fig. 4-2 a) where cheatgrass is the dominant vegetation, occupies the bottom center of the plot. A monoculture
of cheatgrass ought to show moderate productivity depending on the time of the year, while
the uniform canopy is expressed by the small variance in greenness. Monocultures such as
cheatgrass should have a narrow range of variability but a wider range of mean greenness
primarily influenced by inter-annual precipitation patterns.

We observe a quite different response for benchmark TS-24 composed primarily of
sagebrush. The wide range in the greenness variance indicates a higher diversity (i.e.
grasses, forbs, and shrubs) compared to the cheatgrass monoculture. Fig. 4-2b provides the
same information but for different soil mapping units that are primarily composed of one
ecological site. This indicates that the spatiotemporal data set offers enough information to
separate the mean response in vegetation composition between ecological sites. With this
piece of graphical evidence, we felt that enough information was contained in our data set to
proceed with the ordination analysis.

**MDS Solutions and Separation of Ecological Sites**

For both MDS efforts (1984-1996, 1997-2008), we determined that two dimensions
could adequately represent the transformed observations. The Scree plots provided in figure
4-3 clearly show that the stress abruptly drops from one to two dimensions for both periods.
Using more than two dimensions evidently provides a better fit, but the gains in stress
minimization do not seem to compensate for the increased complexity. Adding a third
dimension reduces the stress by approximately 2.5 units but at the same time makes the
interpretation of the final solution more difficult.

We also provide Fig. 4-4 showing the 1984-1996 MDS solution for the big sagebrush
SMUs occurring in the study area: R028AY215, R028AY221, R028AY226, and
R028AY306. In this and subsequent MDS plots, each point represents a unique SMU or
ecological site unit that has been previously correlated to a particular ecological site. The upper plot (Fig. 4-4a) shows the MDS scores for all the SMUs regardless of the size of the soil component. We see that even though some clusters can be discriminated in the figure, there exists a significant amount of overlap among ecological sites. The majority of the SMUs shown in the figure does not have a major soil component and therefore cannot be correlated to just one ecological site. If a SMU has multiple components then by definition these components represent different ecological sites on the ground. This may explain the observed overlap and lack of distinct clusters. We see a different situation in figure 4-4b in which a threshold was set for the soil component. In this case, we only show those SMUs that have a major soil component, and therefore ESD, occupying at least 60% of its area. In this case most ecological sites occupy distinct areas in the reduced MDS space. This suggests that the best representations may be obtained from those SMUs that have a dominant ecological site component. The exception is the ESD R028AY226. The SMUs for this ESD occupy three different sections of the plot. None of these SMUs had a dominant component. They were included in the analysis for comparison purposes. This situation emphasizes the need to apply our analysis techniques only to map units that are dominated by one ESD. We presume that if soil map units were generated at a fine enough scale to encompass only one ecological site, the similarity index could be applied to all SMUs.

**Vector Migration and Interpretation of MDS Dimensions**

Because our data sets have been scaled and then ordered, the time trajectories of each ecological site traced through ordination space allows evaluation of spatiotemporal changes. The vector movement in the reduced space from 1996 to 2008 for the different ecological site units is presented in Fig. 4-5. All landscape features are subject to change, but some site's
current scores tend to separate more from the scores obtained for 1984-1996. For instance, all the SMUs correlated to the ecological site R028AY306 seem to have experienced more changes (larger vector migrations) from 1996 to 2008. On the other hand, the majority of SMUs for the ecological site R028AY215 do not show large vector movements. It is interesting though that the SMUs for the selected ecological sites still form distinct clusters in the ordination space for the new period 1997-2008. This may suggest that when faced with a disturbance, the SMUs for this ecological site tend to respond similarly. In Fig. 4-5 we also include the 97-08 MDS scores for the DWR-RTS benchmarks. Sites TS-02, TS-05, and TS-13 are associated with cheatgrass invasion while TS-06 has been documented with increases in Juniper cover. It is evident that SMUs belonging to R028AY215 are more closely related to the benchmarks, but we can also observe SMUs from R028AY221 and R028AY226 in the same vicinity. These types of plots not only provide information about the magnitude of the SMUs vector migrations and thus change in time, but also may be used as a guideline to assess how similar to undesired conditions some of the SMUs have become.

This assessment of the magnitude of change should be accompanied by an interpretation of what the movements in vectors may signify. With so many variables involved in the ordination analysis, it is somewhat difficult to provide a comprehensive interpretation of the MDS dimension's meanings. We decided to utilize the trend narratives found in the DWR-RTS for the field benchmarks as a plausible source to provide some power of explanation for the ordination space. We present Fig. 4-6 that shows all of the Range Trend Studies benchmarks utilized in this study with the ordination scores for both periods of analysis and arrows representing the direction of change. A synthesis of the major observations extracted from the DWR-RTS narratives is provided in Table 4-5. Based on our interpretation of the identified trends in the narratives, it seems that the first MDS dimension
is a measure of productivity and its spatial and temporal variability that increases from right to left while the second dimension seems to be related to the amount of bare ground that increases with the vertical axis. We have identified that those benchmarks located in the lower-right region of the ordination space have generally migrated to monocultures of cheatgrass. Those benchmarks that have shifted towards higher values in the second dimension have generally experienced increases in the proportion of bare ground cover according to the DWR-RTS narratives.

Once MDS dimensions have acquired some interpretability, the migrations observed in Fig. 4-5 should provide the reader with information to assess which ecological sites are showing certain tendencies. Among those tendencies we can mention dynamics such as becoming less productive, increased bare ground, or moving towards a monoculture.

**Trend and Similarity Maps**

So far we have presented results of vector migration in the ordination space, and how this information may be used to assess magnitude and direction of change for ecological site units. We now present results of trend assessment based on the multi-temporal distance among units of interest. Fig. 4-7 shows how some benchmarks and SMUs tend to move towards or away from a pre-defined benchmark of interest. In this case we chose benchmark TS-13, which has evidence of cheatgrass invasion according to the DWR-RTS narratives. The upper part of the figure (a) shows the distance of three benchmarks TS-02, TS-06, and TS-24 to the selected point of reference TS-13. Although there is noise in the data, there seems to be indications that the multi-temporal distance can discriminate the direction of change for the units of interest. For instance, a site with cheatgrass invasion (i.e. TS-02) is becoming more similar to TS-13. On the other hand, a site with no recorded increase in
cheatgrass (TS-24) and a site with reported increases in Juniper cover (TS-06) clearly have a different distance signature. In other words, no trend is apparent for these two sites.

To assess the trend in specific ecological sites, we randomly selected one SMU representing each ecological site to compare against the cheatgrass benchmark (TS-13). Only the SMU UTP40 with the R028AY215 ecological site seems to show evidence of a trend towards the TS-13 condition while the other three SMUs show no distinct tendency towards or away from the benchmark (Fig. 4-7b).

Figure 4-8 shows a geographic depiction of the 1997-2008 similarities for all SMUs dominated by a single big sagebrush ESD to benchmarks TS-13 and TS-06. The DWR-RTS trend assessment indicated a decline in the cover of sagebrush with a significant increase in the presence of cheatgrass (TS-13) and woodlands (TS-06). SMUs with similar conditions to TS-13 are mainly distributed in low elevation regions, whereas those with high similarity to TS-06 are located in mid elevation ranges.

Validation

Our visual inspection of NAIP imagery for those SMUs that had a high-calculated similarity to benchmarks of undesired states showed an adequate correspondence with what would be expected. We provide Figs. 4-9 and 4-10 that present the endmembers (most similar and most dissimilar) SMUs to a cheatgrass invasion and woodland encroachment benchmark, respectively. These SMUs were not visited during our fieldwork. There seems to be a strong correlation between what may be observed on the NAIP imagery and those SMUs identified as most similar to undesired conditions. This is particularly clear for woodland encroachment (Fig. 4-10) in which trees can be readily identified within the SMU. On the other hand, those SMUs identified as most dissimilar correspond to sites that have been completely modified
for agricultural purposes or sites that have no cheatgrass or woodland encroachment. The SMU R028AY310UTP1126 may serve as example of the previous statement. A visual inspection of this SMU suggests that shrubs and grasses are the major components of the landscape. In Figs. 4-11, 4-12, 4-13, and 4-14 we present snapshots of four SMUs correlated to R028AY215. These SMUs had similar ordination scores to benchmarks of woodland encroachment and cheatgrass invasion, and were also visited during our fieldwork. Figures 4-11 and 4-12 show two different degrees of encroachment that occur on SMUs # 5 and # 60 and that correspond to scores in the vicinity of benchmark TS-06 while Figs. 4-13 and 4-14 correspond to similar degrees of observed cheatgrass invasion taking place in SMUs # 49 and # 2. These two SMUs have scores that are similar to benchmark TS-13. The color scale may be used to promptly identify the magnitude of similarity that each SMU has with respect to a specific benchmark.

**DISCUSSION**

The standard concept of a similarity index as defined by the NRCS (USDA-NRCS 2006) requires the collection of field data that are of high thematic (i.e. identification of plant species), and spatial (i.e. a small field plot) resolution. In addition, emphasis is in the measurement of similarity to reference states described in the state and transition model of the ecological site. Our approach to similarity uses coarser resolutions given that we worked with geospatial data sets with a pixel size of 30 meters, and no effort was made to classify the landscape into life forms or land cover types. Moreover, we measured similarity to undesired conditions occurring on the landscape such as invasion by exotic annual grasses. Our rationale was that undesired states can be readily identified on the field, and can be described using an integrated response from a multivariate dataset. From a management perspective, it
makes sense to attempt to identify areas of interest that are on a path to conversion or have already converted to a negative condition so that resources may be allocated for prevention or restoration purposes respectively (Wisdom et al. 2005a; Wisdom et al. 2005b). In spite of the differences and limitations of low thematic precision of our approach, it has been recognized (Brandon et al. 2003; Hunt et al. 2003; Washington-Allen 2006) that remote sensing is a cost-efficient technology used to evaluate the spatiotemporal dynamics of large landscapes.

Our mean-variance plots (Pickup and Foran 1987) were able to separate the inter-annual response of vegetation in each selected benchmark and SMU (Fig. 4-2). Ecological sites are unique due to their response to climatic conditions and disturbance (Briske et al. 2005). This was reflected in our steady states plots since ecological site units occupied distinct areas of the mean-variance greenness space. An example can be the comparison of the SMU correlated to the ecological site R028AY306 (mountain big sagebrush) that is expected to perform differently than SMUs from the ecological site R028AY215 (Wyoming big sagebrush) in terms of plant productivity and diversity. The mountain big sagebrush SMU occupied the upper right section of the mean-variance plot which indicates higher heterogeneity and vegetation canopy cover (Washington-Allen et al. 2008). The Wyoming big sagebrush SMU occupied an area with lower greenness mean and variance, which indicates both lower plant productivity and diversity.

The use of historic archives of satellite imagery and their derivative vegetation indices provides explanatory power to assess ecological sites. This appears to be a recent application in the scientific arena. An effort to identify spectrally anomalous locations in ecological site units was conducted in the Montana plains (Maynard et al. 2007). Landsat ETM+ imagery for three years (2000 to 2002) was classified based on departures from mean values in the Tasseled cap transformation (Crist and Kauth 1986), and then compared to
locations in the field that were inside the norms of productivity and exposed soil according to their ecological site description. Our results indicate that a longer time series spanning 20 or more years seems to be necessary to adequately separate different ecological states based on the inherent year-to-year variance of these ecosystems. At least in our case, the long-term data set seemed to provide natural clusters associated to a specific ecological site.

Furthermore, the distinct spatiotemporal signature that was obtained for the benchmarks of either cheatgrass or woodland encroachment reinforced our objective to develop a similarity index that works at the ecological site unit spatial level.

Ordination techniques to assess and monitor rangelands have been broadly reported in the literature. For example, clustering was used to identify ecological stages in grass prairies (Uresk 1990). Principal component analysis and then clustering was used to classify and monitor Wyoming big sagebrush shrub-steppe habitat (Benkobi et al. 2007), and a two-way indication species classification analysis was used to determine sagebrush–grass states based on species composition data from transects to elaborate state and transition models (Allen-Diaz and Bartolome 1998). Each of these applications was based on field data that only allowed inferences about discrete locations on the ground and for a particular snapshot in time. Our approach, on the other hand, is the first attempt to ordinate ecological sites based on long-term remotely sensed data sets that cover entire landscapes as well as several time periods.

We conducted ordination for two periods (1984-1996 and 1997-2008) to track changes in a reduced statistical space, and also to explain change with plausible supporting data. The transitions (changes in vector position) observed for the benchmarks (Fig. 4-6) had a reasonable direction and magnitude when contrasted against the DWR-RTS narratives. This analysis also provided sensible interpretations for the MDS axes. Our assertion after
this examination and interpretation is that the transitions observed for ecological site units should also be credible since the same data sets were used for benchmarks as well as soil mapping units. Our results seem to be in agreement with the claim that "if data from a series of sites at different times are ordered, then the time trajectories of each site traced through ordination space will allow a successional direction to be indicated" (Austin 1977). In other words, changes in ordination space for a given sphere of observations should produce trajectories that suggest either change or stability.

The SMUs correlated to ecological site R028AY306 migrated to a lower vector position in the second MDS axis but stayed relatively constant in the first dimension. Based on our interpretation of the MDS solution, this means that the proportion of bare ground in these units decreased while keeping their productivity relatively stable from 1996 to 2008. Based on the description for R028AY306 (mountain big sagebrush), this ecological site should not be greatly vulnerable to adverse transitions given its soils, climate regime, and elevation gradient. Our ordination results appear to support this idea because the SMUs correlated to this ecological site are dissimilar to benchmarks of undesired states. Conversely, R028AY215 (Wyoming big sagebrush) is known to be less resistant to disturbances such as fire and overgrazing (Pellant 1996; Wisdom et al. 2005b). Our MDS scores also seem to correspond well with this description. The SMUs for R028AY215 tend to occupy areas with relatively low productivity while at the same time have scores that are similar to some of the benchmarks for cheatgrass. The SMUs for R028AY221 (basin big sagebrush), occupy an area between R028AY306 and R028AY215 in the MDS solution. These units are usually found in deep, well-drained soils whereas Wyoming big sagebrush sites are typically located in shallower soils. In terms of productivity, this may explain the location that SMUs correlated to R028AY221 occupy in the ordination space. One of the
SMUs for this ecological site migrated to a very similar position in the ordination plot occupied by R028AY215, while another drifted down the second axis in what could be interpreted as a reduction of the bare ground portion of that particular SMU. We were not able to draw inferences about what might be happening in the SMUs correlated to R028AY226 because they did not exhibit a congruent cluster in the reduced space. Recall that this ecological site was included for illustration purposes only and we could not locate SMUs correlated to this site that had a dominant soil component. We can only conclude that the non-congruency of this ecological site is due to the diversity of ecological sites occurring within the available SMUs.

An application to assess similarity to undesired states that works for rangeland landscapes has been developed. The procedure is based on the definition of benchmarks that are readily identifiable in the field and that represent undesired conditions against which ecological site units can be contrasted and evaluated. The repeatability of this approach depends on the utilization of high-quality, long-term archives of remotely sensed data coupled with field based monitoring sites for which there have been multi-temporal observations of plant community composition. The underlying premise of our similarity index is that spatiotemporal multivariate signatures clearly discriminate undesired conditions in big sagebrush ecosystems.

Ordination techniques such as multidimensional scaling seem to be appropriate to a) reduce the dimensionality of a large data set, b) estimate an integrated response for both benchmarks and ecological site units, c) provide sensible and interpretable axes that allow describing ecological site units in terms of productivity and proportion of bare ground, and d) track the trajectories of the units of interest in reduced space. This technique permits the
evaluation of similarity to undesired conditions as well as to provide insight about current
states and transition that ecological site units have experienced on the landscape.

The ordination results for the ecological sites evaluated in this work seemed to
comply well with the ecological expectation. The site R028AY306 (mountain big sagebrush)
migrated to regions in the reduced space that are thought to have a greater cover of vegetation
while keeping the productivity relatively constant. This might be explained by conditions in
which there is a relatively high vegetation density but the vigor or quality of said vegetation
is low or decadent. The SMUs correlated to the Wyoming big sagebrush ecological site
R028AY215 tended to occupy a region of relatively low productivity and a wide range of
bare ground conditions. SMUs for this ecological site had the closest proximity to the
benchmarks for cheatgrass and woodland encroachment. No distinct transitions were
observed for ecological site R028AY221 (basin big sagebrush) and R028AY226 (Wyoming
big sagebrush) due to the limited number of units available for analysis and the absence of a
clear pattern of clustering. Since SMUs can consist of up to four ecological sites, and the
spatial location of individual ecological sites within SMUs are not known, we can only applied
this technique to soil mapping units that are dominated by a single ecological site. If
ecological sites within SMUs could be mapped, we assume that this technique could be
applied to the relevant ecological sites in all SMUs.

**IMPLICATIONS**

We developed a methodological approach that may be used by range conservationists
and others actively working in the development of ecological site descriptions and their
 corresponding state and transition models. If the ability to characterize benchmarks exists for
other areas that have correlations of ecological site descriptions to soil mapping units, then
this application may be replicated with relative ease. Changes in ordination space for ecological site units may then be evaluated and interpretations about their trajectories can be drawn. If benchmark data are correlated to MDS scores then it will be possible to assess transitions and stable states. This type of activity may provide multi-temporal quantitative support to existing STMs that have been developed using space-for-time substitutions of field observations.

A new area that is worth exploring by researchers is the prediction of the spatial distribution of ecological site descriptions on a pixel basis. Our results have suggested that a long-term remotely sensed data set provides sufficient information to discriminate sites with a major soil component. This is a relatively new application of remote sensing for rangelands. Hyperspectral imagery has been used to model probabilities of occurrence of a suite of ecological sites in the Patagonia region, South America (Blanco et al. 2010). Their results were used to train a neural network classification model using Landsat ETM data. An application to develop state and transition models has also been reported (Sadler et al. 2010). In this case, principal component scaling of a suite of image metrics derived from a temporal sequence of close range photogrammetry was used to identify phases and transitions in grasslands of northwestern Australia. These two examples of classification and ordination along with the results from our research are evidence that ecological modeling of remotely sense data sets has great potential for the future development of ecological sites and state and transition models. Spatiotemporal analysis of vegetation using remote sensing present many possibilities for the future definition, interpretation, and display of ecological site information in rangelands (Brown 2010).

The similarity index presented here may be used to generate triage maps to categorize the landscape into levels of similarity: a) high similarity to indicate that current conditions are
close to the benchmark of undesired condition, b) moderate similarity to indicate areas that are moving toward undesired conditions, and c) low similarity to designate areas that are significantly different from undesired conditions. The condition in these areas (c) does not necessarily have to be good however.

**LITERATURE CITED**


Table 4-1. Big sagebrush ecological sites included in this study

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Soil characteristics</th>
<th>Physiographic features</th>
<th>Climatic features</th>
</tr>
</thead>
<tbody>
<tr>
<td>R028AY215</td>
<td>Semi desert Gravelly Loam (Wyoming big sagebrush) North</td>
<td>Loam, 60 inches deep, well drained</td>
<td>Elev.(^1): 4300 - 6000 Slope(^2): 2 - 15</td>
<td>MAP(^3): 8 - 12 MAAT(^4): 45 - 50 FFP(^5): 100 - 150</td>
</tr>
<tr>
<td>R028AY221</td>
<td>Semi desert Loam (Basin big sagebrush)</td>
<td>Clay Loam, 60 inches deep, well drained</td>
<td>Elev.: 4400 - 6000 Slope: 1 - 3</td>
<td>MAP: 8 - 12 MAAT: 45 -50 FFP: 100 - 150</td>
</tr>
<tr>
<td>R028AY226</td>
<td>Semi desert Sandy Loam (Wyoming big sagebrush)</td>
<td>Loamy Sand, &gt;60 inches deep, well drained</td>
<td>Elev.: 4500 - 5700 Slope: 1 - 10</td>
<td>MAP: 8 - 12 MAAT: 45 -50 FFP: 100 - 150</td>
</tr>
<tr>
<td>R028AY310</td>
<td>Upland Loam (Mountain Big Sagebrush)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Elevation: feet above sea level  
\(^2\) Slope: Percent rise  
\(^3\) Mean annual precipitation (inches)  
\(^4\) Mean annual air temperature (F)  
\(^5\) Freeze free period (days)
Table 4-2. DWR-RTS Benchmarks

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Dominant Vegetation</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS-1-05</td>
<td>Devil's Playground</td>
<td>Juniper-pinyon woodland / Black sagebrush and Wyoming big sagebrush.</td>
<td>Elevation: 1642 m, slope: 5-10%, aspect: east. Shallow soils, well drained, moderately permeable</td>
</tr>
<tr>
<td>TS-1-06</td>
<td>Bovine Exclosure</td>
<td>Sagebrush-grass / scattered juniper-pinyon woodland.</td>
<td>Elevation: 1950 m, slope: 5-10%, aspect: southeast. Soil is loose and coarse textured but deep.</td>
</tr>
<tr>
<td>TS-1-07</td>
<td>South Side Emigrant Pass</td>
<td>Black sagebrush</td>
<td>Elevation: 1712 m, slope: 5-15%, aspect: southwest. Very deep, well drained, permeable soils</td>
</tr>
</tbody>
</table>
Table 4-3. Landsat TM Path 39 Row 31 scenes utilized in the study

<table>
<thead>
<tr>
<th>Year</th>
<th>Acquisition date</th>
<th>Sun elevation angle</th>
<th>Cloud cover %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>July 18</td>
<td>58.05</td>
<td>0</td>
</tr>
<tr>
<td>1985</td>
<td>July 05</td>
<td>59.81</td>
<td>0</td>
</tr>
<tr>
<td>1986</td>
<td>August 09</td>
<td>53.18</td>
<td>0</td>
</tr>
<tr>
<td>1987</td>
<td>August 12</td>
<td>53.34</td>
<td>0</td>
</tr>
<tr>
<td>1988</td>
<td>August 14</td>
<td>53.36</td>
<td>0</td>
</tr>
<tr>
<td>1989</td>
<td>August 17</td>
<td>52.12</td>
<td>0</td>
</tr>
<tr>
<td>1990</td>
<td>July 03</td>
<td>58.38</td>
<td>10</td>
</tr>
<tr>
<td>1991</td>
<td>June 20</td>
<td>59.60</td>
<td>0</td>
</tr>
<tr>
<td>1992</td>
<td>July 24</td>
<td>56.13</td>
<td>0</td>
</tr>
<tr>
<td>1993</td>
<td>June 25</td>
<td>59.34</td>
<td>0</td>
</tr>
<tr>
<td>1994</td>
<td>July 14</td>
<td>56.85</td>
<td>0</td>
</tr>
<tr>
<td>1995</td>
<td>July 17</td>
<td>54.22</td>
<td>0</td>
</tr>
<tr>
<td>1996</td>
<td>July 19</td>
<td>55.67</td>
<td>0</td>
</tr>
<tr>
<td>1997</td>
<td>June 04</td>
<td>60.57</td>
<td>10</td>
</tr>
<tr>
<td>1998</td>
<td>July 09</td>
<td>60.74</td>
<td>0</td>
</tr>
<tr>
<td>1999</td>
<td>July 12</td>
<td>60.32</td>
<td>0</td>
</tr>
<tr>
<td>2000 *</td>
<td>July 22</td>
<td>60.80</td>
<td>0</td>
</tr>
<tr>
<td>2001 *</td>
<td>July 25</td>
<td>60.01</td>
<td>0</td>
</tr>
<tr>
<td>2002 *</td>
<td>July 28</td>
<td>59.39</td>
<td>0</td>
</tr>
<tr>
<td>2003 *</td>
<td>May 28</td>
<td>62.83</td>
<td>0</td>
</tr>
<tr>
<td>2004</td>
<td>August 10</td>
<td>55.97</td>
<td>0</td>
</tr>
<tr>
<td>2005</td>
<td>August 13</td>
<td>55.95</td>
<td>0</td>
</tr>
<tr>
<td>2006</td>
<td>June 13</td>
<td>64.40</td>
<td>0</td>
</tr>
<tr>
<td>2007</td>
<td>July 02</td>
<td>63.82</td>
<td>0</td>
</tr>
<tr>
<td>2008</td>
<td>June 18</td>
<td>63.45</td>
<td>8</td>
</tr>
</tbody>
</table>

* Scenes from Landsat ETM +
Table 4-4. USGS phenology data used in this research

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Relationship with ESD and STM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOST - Beginning of measurable photosynthesis in the vegetation canopy</td>
<td>Day of year identified as having a consistent upward trend in time series NDVI</td>
<td>A temporal series of this variable may provide insight about significant changes in the beginning of photosynthetic activity. This may be related to early onsets of exotic annual grasses establishment and spread.</td>
</tr>
<tr>
<td>EOST - End of measurable photosynthesis in the vegetation canopy</td>
<td>Day of year identified at the end of a consistent downward trend in time series NDVI</td>
<td>If the ending date tends to become shorter or longer with time this may be an indication of major changes in the vegetation composition of an ecological site.</td>
</tr>
<tr>
<td>MAXT - Time of maximum photosynthesis in the canopy</td>
<td>Day of year corresponding to the maximum NDVI in an annual time series</td>
<td>A small variance in the mean date for the maximum activity in the vegetation canopy may be related to shrubs and trees whereas a high variance could be the result of photosynthetic activity in annual grasses.</td>
</tr>
<tr>
<td>DUR - Length of photosynthetic activity (the growing season)</td>
<td>Number of days from the SOST to the EOST</td>
<td>The spatiotemporal response of different ecological sites should be distinct with regards to the growing season. Longer growing seasons may be correlated to big shrubs and woodlands whereas short durations may be the result of monocultures such as cheatgrass.</td>
</tr>
<tr>
<td>TIN - Time Integrated NDVI Canopy photosynthetic activity across the entire growing season</td>
<td>Daily (interpolated) integration of NDVI above the baseline for the entire duration of the growing season</td>
<td>The integration of daily NDVI throughout the growing season is commonly correlated to primary productivity. This is one of the main attributes that distinguish ecological sites and improve state and transition models.</td>
</tr>
</tbody>
</table>

1 The source for this data was the USGS EROS Center (http://phenology.cr.usgs.gov/)
### Table 4-5. Synthesis of major disturbances extracted from DWR-RTS narratives

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Observed trend 1996</th>
<th>Observed trend 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS-1-02</td>
<td>Improvements in the cover of Wyoming big sagebrush since 1990. Cheatgrass is widely distributed across the site and abundant</td>
<td>Wyoming big sagebrush density declined 22%. Cheatgrass cover was up to 5%, an as nearly half of the total grass cover. Site could be susceptible to fire.</td>
</tr>
<tr>
<td>TS-1-04</td>
<td>Mountain big sagebrush cover slightly down, grass cover up, forbs stable, Cheatgrass present</td>
<td>Cheatgrass cover down, Mountain big sagebrush cover declined, grasses and forbs cover up</td>
</tr>
<tr>
<td>TS-1-05</td>
<td>Cheatgrass abundance low, stable shrub cover, forbs cover up</td>
<td>Shrubs, grasses, and forbs cover declined, Cheatgrass cover up</td>
</tr>
<tr>
<td>TS-1-06</td>
<td>Shrubs and grasses cover stable compared to 1990, forbs cover down</td>
<td>Basin big sagebrush and black sagebrush declined, increase in bare ground and juniper cover, Cheatgrass cover decline</td>
</tr>
<tr>
<td>TS-1-07</td>
<td>Black sagebrush cover stable, grasses and forbs cover up, traces of Cheatgrass cover</td>
<td>Shrub cover declined, grasses and forbs cover down, increase in Cheatgrass and bare ground cover</td>
</tr>
<tr>
<td>TS-1-11</td>
<td>Shrubs and grasses cover down from 1990, Cheatgrass is a significant component</td>
<td>Shrub cover declined, grasses cover stable, Cheatgrass is major component but no significant change from 2001</td>
</tr>
<tr>
<td>TS-1-12</td>
<td>Basin big sagebrush stable, grasses cover declined</td>
<td>Increase in bare ground, Cheatgrass cover declined, shrub cover stable</td>
</tr>
<tr>
<td>TS-1-13</td>
<td>Wyoming big sagebrush recovering, grasses and forbs cover up, cheatgrass is a significant component</td>
<td>Big sagebrush cover declined, significant increases in cheatgrass cover, grasses and forbs cover stable</td>
</tr>
<tr>
<td>TS-1-14</td>
<td>Shrub cover slightly improving, grasses stable and dominant cheatgrass</td>
<td>Mountain big sagebrush declined, grasses down, forbs cover up</td>
</tr>
<tr>
<td>TS-1-15</td>
<td>Mountain big sagebrush improving from 1990, competition with trees, grasses stable and forbs cover up</td>
<td>Big sagebrush recovering after fire, grasses and forbs cover up, Cheatgrass found but only traces</td>
</tr>
<tr>
<td>TS-1-18</td>
<td>Wyoming big sagebrush declined, grasses and forbs cover down, Cheatgrass cover increasing</td>
<td>Big sagebrush declined, Cheatgrass cover increasing, grasses and forbs cover stable</td>
</tr>
<tr>
<td>TS-1-24</td>
<td>Mountain big sagebrush cover stable from 1990, grasses and forbs cover up, cheatgrass declined</td>
<td>Big sagebrush cover increased, cheatgrass declined, grasses and forbs cover up</td>
</tr>
</tbody>
</table>
Figure 4-1. Distribution of big sagebrush ecological sites and benchmarks in the study area
Figure 4-2. Steady states plots for (a) benchmarks and (b) ecological site units
Figure 4-3. Scree plot for two MDS solutions to help determine the number of dimensions
Figure 4-4. MDS solution for the period 1984 - 1996: (a) includes all the SMUs, (b) units with a major soil component (>60%)
Figure 4-5. MDS solution for the 1997 - 2008 period. Arrows indicate the direction of change in vector position from 1996 - 2008.
Figure 4-6. MDS solution for the 1984-1996 and 1997 - 2008 periods for the DWR-RTS benchmarks. Arrows indicate the direction of change.
Figure 4-7. Samples of multi-temporal distance plots for (a) distance from certain benchmarks, and (b) randomly chosen ecological site units.
Figure 4-8. Samples of similarity maps to a benchmark: (a) TS-1-13 (cheatgrass), (b) TS-1-06 (woodland encroachment)
Figure 4-9. Samples of similarity endmember SMUs to a cheatgrass benchmark (TS-13). Arrows indicate the magnitude of similarity (green=Dissimilar, red = Similar).
Figure 4-10. Samples of similarity endmember SMUs to a Pynion-Juniper benchmark (TS-06). Arrows indicate the magnitude of similarity (green=Dissimilar, red = Similar).
Figure 4-11. Similarity to PJ encroachment (benchmark TS-06). Section of SMU # 5 of R028AY215 with field point 07252007 (UTM East 320640, North 4583520). Photo by Alexander Hernandez.
Figure 4-12. Similarity to PJ encroachment (benchmark TS-06). Section of SMU # 60 of R028AY215 with field point 07112007 (UTM East 316947, North 4631891). Photo by Alexander Hernandez.
Figure 4-13. Similarity to cheatgrass invasion (benchmark TS-13). Section of SMU # 49 of R028AY215 with field point 07242007 (UTM East 322489, North 4633040). Photo by Alexander Hernandez.
Figure 4-14. Similarity to cheatgrass invasion (benchmark TS-13). Section of SMU # 2 of R028AY215 with field point 07262007 (UTM East 316256, North 4616998). Photo by Alexander Hernandez.
Ecological site descriptions (ESD) (U.S. Department of Agriculture 2008) together with state and transition models (STM) (Westoby et al. 1989) constitute a conceptual framework for the monitoring and assessment of rangelands in the western United States. Units from the same ecological site are expected to produce the same type and amount of vegetation and respond similarly to management activities and disturbance events. The STM describes "the patterns, causes and the indicators of transitions between communities within an ecological site" (USDA-ARS 2010), and in this way may be thought of as a decision support system that range conservationists and other individuals actively engaged in rangeland management can use to achieve sustainability in these fragile ecosystems. A STM that has been properly correlated to an ecological site provides the opportunity to recognize indicators (i.e. percent cover by invasion of exotics, encroachment, productivity variations, etc.) of certain transitions that lead to undesired states that may impact the ecological services provided by these sites. Therefore, it is of great importance to identify those indicators on the landscape in a timely manner.

The fact that ecological site units are large spatial entities and that STMs have a prominent temporal component call for the utilization of multi-temporal geospatial data sets and ecological models to help in the identification of the indicators mentioned above. For the models to have credibility, and thus applicability, they must be trained and validated with plausible field information. The main thrust of our research was to develop methods that generate those indicators at a landscape level, and that can be used to characterize past and current conditions of ecological sites.
In Chapter 2 we took advantage of the contrast in greenness that may be obtained between remotely sensed vegetation indices captured during the assumed peak and die-off dates for cheatgrass. We found that this contrast in conjunction with elevation were the most important variables to model presence and absence for three temporal scenarios. Generating a multi-temporal classification permitted the identification of those areas that have had a persistent presence of the annual exotic. It is therefore possible to quantify what percentage of an ecological site unit has been under continual pressure, and the spatial location within the units that this has happened.

Additional information to assess ecological site units was generated from research reported in Chapter 3. We utilized relatively novel regression methods to generate a multi-temporal collection of continuous vegetation fields (VCF). VCFs are sub-pixel estimates of percent cover for shrubs, trees, grasses, and bare ground. Better validation results were obtained for the VCF of shrubs and grasses. Having multi-temporal percent cover estimates for shrubs, for example, allows not only knowing current cover conditions for individual ecological site units but also allows us to identify where shrub cover is increasing or decreasing. We also decided to use the combined dynamics of shrubs and trees to identify potential areas of woodland encroachment into big sagebrush areas. The modeled encroachment areas may be associated with modeled cheatgrass persistence from chapter 1 in order to provide even more information about current conditions of sagebrush sites.

In Chapter 4 we explored the utilization of long-term (1984 - 2008) remotely sensed vegetation indices to assess the similarity of ecological site units to undesired states by means of an ordination technique. We found that the spatiotemporal spectral signature of benchmarks (areas of known condition) as well as that of soil map units with a major ecological site component was distinct enough both in the original spectral and the reduced
ordination space to allow us to establish a measure of similarity between the two. Because we split the ordination analysis in two periods (1984-1996 and 1997-2008) we were able to follow trajectories for benchmarks and ecological sites, and also to provide an interpretation of said trajectories in two dimensions. This allowed us to assess not only how similar ecological sites are to current undesired states but also to estimate the potential previous and current condition in terms of productivity and occupancy by bare ground. Our finding in this chapter may be used to update draft STMs.

The work presented in Chapters 2, 3, and 4 have explored novel statistical methods to generate geospatial indicators that may be readily used to enhance understanding of current and past conditions of big sagebrush ecological sites. In those three chapters, the need to have multi-temporal remotely sensed data sets was obvious for three primary reasons: (a) even though field assessments have very fine spatial and thematic resolutions, the inferences that can be drawn are valid only to very specific locations and very explicit times, (b) the pace at which threats advance over shrub communities usually takes years and/or decades thereby rendering one-year studies inadequate to assess the dynamics of threats, and (c) threats to shrub communities may strike from several fronts that may be left out during field monitoring and assessment. Our research does not intend to supersede traditional field monitoring; rather it provides additional insight about spatiotemporal dynamics of threats to sagebrush communities. This insight can potentially enrich current ESD and STM in rangelands of Northern Utah.
LITERATURE CITED


APPENDIX
# Sample R Code for Cheatgrass Classification: Chapter 2

# SVM Classification in R
# Loading required libraries
library(randomForest)
library(e1071)
library(MASS)
library(yaImpute)

################## Year 1996 #######################

# Using all available points to determine variable importance in Random Forests for year 1996
VariablesImp2 = NULL
for (Set in 1:500) {
  # Read the Data
  clauso.train = Belder.cgrass[,5]
  # Fitting Random Forest on a binary response
  rf.out.mul = randomForest(as.factor(clauso.train) ~ ., data = Belder.cgrass, trees = 5000, importance = T)
  # Importance
  importance(rf.out.mul)
  varImpPlot(rf.out.mul)
  VariablesImp1 = data.frame(rf.out.mul$importance[,3])
  VariablesImp2 = append(VariablesImp2, VariablesImp1)
} # Next

Todos = data.frame(VariablesImp2)
TodosSuma = apply(Todos, 1, sum)
TodosSuma = data.frame(TodosSuma)
cbind(colnames(Teton.cover[,3:20]), TodosSuma)
## Most important variables = NDSAVI + Elevation

# Fitting a SVM
# First the regularization to obtain best gamma and Cost
# The SVM is tuned
svm.tune = tune.svm(as.factor(clauso.train) ~ ndsavi + demft, data = Belder.cgrass, sampling = "cross", gamma = 2^(-5:1), cost = 2^(-1:2))
best.gamma = svm.tune$best.parameters$gamma
best.cost = svm.tune$best.parameters$cost
best.gamma
best.cost
## Best Gamma = 0.4 and Cost = 1.0

# The SVM is fitted
# Fit the SVM Model using a binary response with the optimized values of Gamma and Cost
sv.out.mul = svm(as.factor(cgrass) ~ ndsavi + demft, data = Belder.cgrass, gamma = 0.4, cost = 1.0)

# Obtaining a SVM classification plot to assess distribution of Support Vectors
plot(sv.out.mul, Belder.cgrass, ndsavi ~ demft, svSymbol = 1, dataSymbol = 6, color.palette = terrain.colors, grid = 100)

# Opening the SVM model object
sv.out.mul

# Using YaImpute to generate a map of the classification
# First the namelist tells YaImpute about the ASCII grid files: They MUST MATCH the variable names used in the training dataset
namelist <- list("ndsavi.asc", "demft.asc")

# Names of variables used in the SVM modeling
names(namelist) <- c("ndsavi", "demft")

# Name of the output file or GRID to be created with the classification results, it is a text file
# Defining the SVM object to extract the model for classification and the type of response to be generated
AsciiGridPredict(sv.out.mul,xfiles=namelist,outfiles=outfiles, type = "class")

# Process is similar for years 2001 and 2007 with the exception that a training and a validations sub datasets are create prior to Optimization of Gamma and Cost and Fitting of the SVM

######################################### Year 2001 ############################################
#Read the Data

# Dividing the data set into training and testing (validation) sub datasets
n=nrow(Belder.cgrass)
fifth=round(n/5)
reorder = sample(1:n,replace=FALSE)
test.cover  = Belder.cgrass[reorder[1:fifth],] # 20% of the data
train.cover = Teton.cover[reorder[(fifth+1):n],] # 80% of the data

#The SVM is fitted only on the Training data in this case 80%
#Fit the SVM Model using a binary response
sv.out.mul2=svm(as.factor(cgrass)~ndsavi+demft,data=train.cover,
gamma = 0.45, cost = 1.5)
sv.out.mul2

# Predicting on the training data: 80% data
# Confusion Matrix
pred.sv.train2=predict(sv.out.mul2, train.cover)
table(pred.sv.train2,train.cover$cgrass)

# Calculating the error rate
errate.sv.train2=mean(pred.sv.train2!=train.cover$cgrass)
errate.sv.train2

# Predicting on the test data: 20%
# Confusion Matrix
pred.sv.test2=predict(sv.out.mul2, test.cover)
table(pred.sv.test2,test.cover$cgrass)
errate.sv.test2=mean(pred.sv.test2!=test.cover$cgrass)
errate.sv.test2

######################################### Year 2007 ############################################

#Read the Data
Belder.cgrass2<-subset(Belder.cgrass, source == "TNCFSEASON")
Belder.cgrass3<-subset(Belder.cgrass, source == "ALEXFSEASON")

# The SVM is tuned
svm.tune = tune.svm(as.factor(cgrass)~ndsavi+demft,data=Belder.cgrass3,
sampling ="cross",gamma = 2^(-5:1), cost=2^(-2:3))
best.gamma = svm.tune$best.parameters$gamma
best.cost = svm.tune$best.parameters$cost
best.gamma
best.cost

# Best values of Gamma = 0.5 and Cost 1.5

#The SVM is fitted only on the Training data in this case is the Alex's Points
#Fit the SVM Model using a multiple response
sv.out.mul1=svm(as.factor(cgrass)~ndsavi+demft,data=Belder.cgrass3,
gamma = 0.5, cost = 1.5,probability = TRUE)
# Predicting on the training data: Alex
# Confusion Matrix
pred.sv.train=predict(sv.out.mull, Belder.cgrass3, probability = TRUE)
table(pred.sv.train,Belder.cgrass3$cgrass)

# Calculating the Error rate
errate.sv.train=mean(pred.sv.train!=Belder.cgrass3$cgrass)
errate.sv.train

# Predicting on the test data: TNC
# Confusion Matrix
pred.sv.test=predict(sv.out.mull, Belder.cgrass2, probability = TRUE)
table(pred.sv.test,Belder.cgrass2$cgrass)

# Estimating the Error rate
errate.sv.test=mean(pred.sv.test!=Belder.cgrass2$cgrass)
errate.sv.test

# The process of generating the map is similar to the 1996 Year usage of YaImpute
# Must use the SVM classification object that was obtained for each year, and tell R
# where the ASCII Grid files are located in disk, Must Match the names of the variables
# used for model fitting
### Sample R code to run Multivariate Regression Trees - Chapter 3

#### Loading required libraries
```r
library(mvpart)
library(foreign)
library(MASS)
library(yaImpute)
```

#### Read the data
```r
```

#### Testing Multivariate Regression Trees

```
# Dividing the database in training % and test %

n=nrow(Belder.cfvo1)
fifth=round(n/5)
reorder = sample(1:n,replace=FALSE)
Belder.cfvo1_4 = Belder.cfvo1[reorder[1:fifth],] # Test subset
Belder.cfvo1_5 = Belder.cfvo1[reorder[(fifth+1):n],] # Training subset

# Fitting a multivariate regression tree

##### First on the training

```r
belder.svmpart<-mvpart(data.matrix(Belder.cfvo1_5[,4:7]) ~
bgw01b_182+bgw01g_182+bgw01w_182+bgw01b_278+bgw01g_278+bgw01w_278+ndaavi+savi01_182+tm01rd_182+tm01mir1_278+tm01mir2_182,data=Belder.cfvo1_5,xv="pick",xvmult=500,use.n=TRUE,all=TRUE,text.add=TRUE)
```

##### Predicting on the 20%
```r
ttc100<-predict(belder.svmpart,Belder.cfvo1_4)
```

#### Validation (TREES, SHRUBS, GRASSES, BAREGROUND)
```r
ttc101<-cbind(Belder.cfvo1_4[,4],ttc100[,1],Belder.cfvo1_4[,5],ttc100[,2],Belder.cfvo1_4[,6],ttc100[,3],Belder.cfvo1_4[,7],ttc100[,4])
ttc102<-c("trees_field","trees_pred","shrubs_field","shrubs_pred","grasses_field","grasses_pred","bground_field","bground_pred")
```
```r
colnames(ttc101)<-ttc102
ttc101
```

#### Scatter plot to assess predicted vs. observed values
```r
plot(ttc101[,1],ttc101[,2],main="MRTS-Trees 2001",xlab="Observed",ylab="Predicted") # TREES
plot(ttc101[,3],ttc101[,4],main="MRTS-Shrubs 2001",xlab="Observed",ylab="Predicted") # SHRUBS
plot(ttc101[,5],ttc101[,6],main="MRTS-Grasses 2001",xlab="Observed",ylab="Predicted") # GRASSES
```

#### Correlation coefficients
```r
cor(ttc101[,1],ttc101[,2],method="pearson") # Shrubs
cor(ttc101[,3],ttc101[,4],method="pearson") # Shrubs
cor(ttc101[,5],ttc101[,6],method="pearson") # Grasses
cor(ttc101[,7],ttc101[,8],method="pearson") # Bground
```

#### Now code to extract the model and produce a map (continuous surface)

#### Using yaImpute
```r
# First we defined "myPred" function that will extract predictions from the
multivariate mean response
```
```r
myPred = function (obj,newdata) {
x=predict(obj,newdata)
x[,1] # Here "1" is the predicted response for trees, then we change to "2" to get
shrubs, "3" to get grasses, and "4" to get bare ground
}
```
myPred(belder.svmpart, Belder.cfv01_5)

## Setting yaImpute
## First the names of the ASCII files that represent our independent geospatial variables
to be used for prediction (MUST match the names used during model fitting)
namelist <- list("bgw01b_182.asc", "bgw01g_182.asc", "bgw01w_182.asc", "bgw01b_278.asc", "bgw01g_278.asc", "bgw01w_278.asc", "ndsavi.asc", "savi01_182.asc", "tm01red_182.asc", "tm01mir1_278.asc", "tm01mir2_182.asc")

names(namelist) <- c("bgw01b_182", "bgw01g_182", "bgw01w_182", "bgw01b_278", "bgw01g_278", "bgw01w_278", "ndsavi", "savi01_182", "tm01red_182", "tm01mir1_278", "tm01mir2_182")

# We defined the name of the output map (one per CVF)
outfiles=list(predict="mvpart_trees_y2001_.txt")

# Finally run the AsciiGridPredict function to get the map
## Need to define the svmpart object, the name list, the name of the output file, and the prediction function
AsciiGridPredict(belder.svmpart, xfiles=namelist, outfiles=outfiles, myPredFunc=myPred)

#########################################################################
#########################################################################
####### Sample R code to run Random Forests
###### Loading required libraries
library(randomForest)

## Here we have to fit models for each VCF
## First for Shrubs
#########################################################################
### SHRUBS ##############################################################
RF.out.shrub<-randomForest(sh07~swrgap+savi07_183+tm07red_183+ndsavi+tm07mir1_263+bgw07g_183+bgw07b_263
tm07mir2_183, data=Belder.cfv07_5, trees =5000)

### Now we can predict on the validation subset
ttc17<-predict(RF.out.shrub, Belder.cfv07_4)
nombres.shrubs<-c("shrubs_field","shrubs_pred")
field.shrubs.pred<-cbind(Belder.cfv07_4$sh07, ttc17)

colnames(field.shrubs.pred)<-nombres.shrubs

#field.shrubs.pred
plot(field.shrubs.pred[,1],field.shrubs.pred[,2])# A scatter plot to check how good the fit is between observed and predicted values

### Now the yaImpute to produce a map (continuous response for shrubs)
## We do not need to redefine the independent geospatial variables again because it had been done above
outfiles=list(predict="CVFShrubs_RFreg_2007.txt")

AsciiGridPredict(RF.out.shrub, xfiles=namelist, outfiles=outfiles, type="response")

#########################################################################
#########################################################################
#### HERBACEOUS #########################################################
RF.out.grasses<-randomForest(gr07~swrgap+bgw07b_183+bgw07g_183+bgw07w_183+ndsavi+bgw07b_263+bgw07g_263+bgw07w_263, data=Belder.cfv07_5, trees =5000)

### Now we can predict on the validation subset
ttc19<-predict(RF.out.grasses, Belder.cfv07_4)
nombres.grasses<-c("grasses_field","grasses_pred")
field.grasses.pred<-cbind(Belder.cfv07_4$gr07, ttc19)

colnames(field.grasses.pred)<-nombres.grasses

#field.grasses.pred
plot(field.grasses.pred[,1],field.grasses.pred[,2],xlim=c(0,60),ylim=c(10,45)) A scatter plot to check how good the fit is between observed and predicted values

### Now the yaImpute to produce a map (continuous response for herbaceous)
We do not need to redefine the independent geospatial variables again because it had been done above.

```r
outfiles=list(predict="CVFGrasses_RFreg_2007.txt")
AsciiGridPredict(RF.out.grasses,xfiles=namelist,outfiles =outfiles, type="response")
```

```
### BARE GROUND

RF.out.bground<-randomForest(bg07~bgw07b_183+bgw07g_183,data=Belder.cfv07_5,trees =5000)
#### Now we can predict on the validation subset

```r
ttc21<-predict(RF.out.bground, Belder.cfv07_4)
nombres.bground<-c("bground_field","bground_pred")
field.bground.pred<-cbind(Belder.cfv07_4$bg07, ttc21)
colnames(field.bground.pred)<-nombres.bground

plot(field.bground.pred[,1],field.bground.pred[,2]) #xlim=c(0,60),ylim=c(10,45)) #A scatter plot to check how good the fit is between observed and predicted values
```
```
## Now the yaImpute to produce a map (continuous response for bare ground)
## We do not need to redefine the independent geospatial variables again because it had been done above

```r
outfiles=list(predict="CVFBground_RFreg_2007.txt")
AsciiGridPredict(RF.out.bground,xfiles=namelist,outfiles =outfiles, type="response")
```
```
### TREES

```
## Read the data
Belder.cfvtrees<-
Belder.cfvtrees<-subset(Belder.cfvtrees, trees > 0)

###### Dividing the database in training and test

n=nrow(Belder.cfvtrees)
fifth=round(n/5)
reorder = sample(1:n,replace=FALSE)
Belder.cfv01_4 = Belder.cfvtrees[reorder[1:fifth],] # Validation subset
Belder.cfv01_5 = Belder.cfvtrees[reorder[(fifth+1):n],] # Training subset

#### Now we can predict on the validation subset

```r
ttc15<-predict(RF.out.trees, Belder.cfv01_4)
nombres.trees<-c("trees_field","trees_pred")
field.trees.pred<-cbind(Belder.cfv01_4$trees, ttc15)
colnames(field.trees.pred)<-nombres.Trees

plot(field.trees.pred[,1],field.trees.pred[,2],main="RF Trees 2001",xlab="Observed",ylab="Predicted") #xlim=c(0,60),ylim=c(10,45)) #A scatter plot to check how good the fit is between observed and predicted values
```
```
## Now the yaImpute to produce a map (continuous response for trees)
## We do not need to redefine the independent geospatial variables again because it had been done above

```r
outfiles=list(predict="CVFTrees_RFreg_2007.txt")
AsciiGridPredict(RF.out.trees,xfiles=namelist,outfiles =outfiles, type="response")
```
## Sample R Code for Similarity Index - Chapter 4

### Code to generate Mean-Variance Plots

**Load Dataset containing the benchmarks points with SAVI for 25 years (1984 - 2008)**

```r
# Load dataset containing the benchmarks points with SAVI for 25 years (1984 - 2008)
data.field1 <- read.csv('/users/alex_hernandez/desktop/modelingeco/ecositesmod/d28d25_y20071ha_nodupli3.csv', header=T)

# Taking care of Unknown values
data.field2 <- unknownToNA(x=data.field1, unknown=-9999)
data.field2 <- unknownToNA(x=data.field2, unknown=-1000)
data.field2 <- unknownToNA(x=data.field2, unknown=1000)
data.field2 <- unknownToNA(x=data.field2, unknown=255)
data.field2 <- data.field2[complete.cases(data.field2),]  # Removing NA cases

# Rearranging the data set: Keeping CODEF and PLOT identifiers and the reshaping Years sequentially in a column
# By Variable
savi <- data.field2[,c(1,seq(8,60,by=2))]
savi1 <- reshape(savi, v.names="SAVI", idvar="codef", varying=list(2:28), direction="long", timevar="Year", times=seq(1984,2010,by=1))
savi.sd <- data.field2[,c(1,seq(9,61,by=2))]
data.field3<-cbind(savi1,savi1.sd[,3])
colnames(data.field3)<-c("codef","Year","Savi","SaviSD")

# Order the data frame by plot and year
data.field3=data.field3[order(data.field3$codef,data.field3$Year),]
# make sure Year is numeric
data.field3$Year = as.numeric(as.character(data.field3$Year))

# Extracting some plots of interest
D5087 = subset(data.field3,codef==5087)  # A benchmark with "good health" characteristics
plot(D5087$Savi, D5087$SaviSD,pch=15,xlab="Interannual Mean Greenness",ylab="Interannual Variance Greenness",xlim=c(0.13,0.40),ylim=c(0.01,0.10))

D5081 = subset(data.field3,codef==5081)  # A benchmark with Cheatgrass characteristics
points(D5081$Savi, D5081$SaviSD,pch=21)

D5046 = subset(data.field3,codef==5046)  # A benchmark with Encroachment characteristics
points(D5046$Savi, D5046$SaviSD,pch=17)
```

### Code for the MDS solutions

**Loading required libraries**

```r
library(gdata)
library(rgl)
library(MASS)
library(foreign)

data.field3=data.field3[order(data.field3$codef,data.field3$Year),]
make sure Year is numeric
data.field3$Year = as.numeric(as.character(data.field3$Year))

# Extracting some plots of interest
D5087 = subset(data.field3,codef==5087)  # A benchmark with "good health" characteristics
plot(D5087$Savi, D5087$SaviSD,pch=15,xlab="Interannual Mean Greenness",ylab="Interannual Variance Greenness",xlim=c(0.13,0.40),ylim=c(0.01,0.10))

D5081 = subset(data.field3,codef==5081)  # A benchmark with Cheatgrass characteristics
points(D5081$Savi, D5081$SaviSD,pch=21)

D5046 = subset(data.field3,codef==5046)  # A benchmark with Encroachment characteristics
points(D5046$Savi, D5046$SaviSD,pch=17)
```

**Load Data: This data does not contain descriptive attributes per SMU**

```r
ecosites1<-read.csv('/users/alex_hernandez/desktop/modelingeco/ecositesmod/d28d25_y20071ha_nodupli3.csv',header=T)
head(ecosites1)
dim(ecosites1)

# Taking care of Unknown values = -9999, -1000, 1000, and 255 are categorized as Unknown
# Must be deleted because the algorithm for MDS cannot handle missing values
ecosites2<-unknownToNA(x=ecosites1, unknown=-9999)
ecosites2<-unknownToNA(x=ecosites2, unknown=-1000)
ecosites2<-unknownToNA(x=ecosites2, unknown=1000)
ecosites2<-unknownToNA(x=ecosites2, unknown=255)
```
dim(ecosites2) # Dimensions should be identical to ecosites1
names(ecosites2)
ecosites2X<ecosites2[complete.cases(ecosites2),] # A vector that tells us which records
are not NA / missing Removing NA cases
dim(ecosites2)
names(ecosites2)

# Loading the descriptive attributes per SMU
SMUatt<-
read.csv('/users/alex_hernandez/desktop/modelingeco/ecositesmod/D28withstm_add.csv',header =T)
dim(SMUatt)

# Loading the descriptive attributes of Y2007 Benchmarks with known problems of
Cheatgrass and Woodland Encroachment
Y2007att<-
read.csv('/users/alex_hernandez/desktop/modelingeco/ecositesmod/Y2007cgpj_add.csv',header =T)
dim(Y2007att)

# A dataset with only the SMU WITH descriptive attributes is obtained
ecositesSMU<merge(SMUatt,ecosites2,by.x="codef",by.y="codef")
dim(ecositesSMU)

# A dataset with only the Y2007 field points WITH descriptive attributes is obtained
ecositesY2007<merge(Y2007att,ecosites2,by.x="codef",by.y="codef")
dim(ecositesY2007)

# a vector of the ECOSITE's names + PLOTS names
econamespolys<as.vector(ecositesSMU$shapeid)
plotnames<as.vector(ecositesY2007$PLOT)
nombres.filas<-c(econamespolys,plotnames)

#### MDS Years 1984 - 1996 ####
# First must generate a data set to calculate ecological distance
# Independent variables go up to 1996
# First extract only those polygons from ecosites with STM and field points
ecosites8496<-
ecosites2[(c(1:312),c(800:862)),(c(8:33),seq(70,85,by=2),seq(110,125,by=2),seq(150,165,
by=2),seq(190,205,by=2))]

# Then assign the name of the polygon and name of the field plot to each record
row.names(ecosites8496)<nombres.filas[(c(1:312),c(800:862))]

##### Starting a MDS #####
# First must generate a data set to calculate ecological distance
# Distances between SMU and benchmarks is obtained
ecosites8496.dist<dist(scale(ecosites8496),method="manhattan")

## How many dimensions to use?
# Using Kruskal Procedure
ecosites8496.mds1<isoMDS(ecosites8496.dist,k=1)
ecosites8496.mds2<isoMDS(ecosites8496.dist,k=2)
ecosites8496.mds3<isoMDS(ecosites8496.dist,k=3)
ecosites8496.mds4<isoMDS(ecosites8496.dist,k=4)
ecosites8496.mds5<isoMDS(ecosites8496.dist,k=5)
ecosites8496.mds6<isoMDS(ecosites8496.dist,k=6)
ecosites8496.mds7<isoMDS(ecosites8496.dist,k=7)

stress.behavior<-
cbind(c(1:7),c(ecosites8496.mds1$stress,ecosites8496.mds2$stress,ecosites8496.mds3$stress ,ecosites8496.mds4$stress,ecosites8496.mds5$stress,ecosites8496.mds6$stress,ecosites8496.mds7$stress))
stress.behavior

### A scree plot to determine how many dimensions to use
plot(stress.behavior,type="b",ylab="Stress Value",xlab="Number of dimensions",main="MDS
Stress Minimization",col=1,pch=19,ylim=c(5,30))

### Two dimensions seems to work fine
### Now, a MDS with 2 dimensiones
# Plotting SMU from Ecosites 215, 221, 226 y 306 // For illustrative purposes and check whether two ecosites are different in reduced statistical space
ecosites8496.mds.bsb<-ecosites8496.mds[c(c(38:110),c(211:220)),]
nom.ecosites<-as.factor(substr(rownames(ecosites8496.mds.bsb),1,9)) # To have different colors for ecosites
nom.ecosites<-as.numeric(nom.ecosites)
nom.ecosites<-nom.ecosites + 19
nom.poligonos<-substr(rownames(ecosites8496.mds.bsb),10,14) # Assigning the SMU unique identifier
temp215.221.226.306<-as.matrix(cbind(ecosites8496.mds.bsb,nom.ecosites,nom.poligonos)) # A dataframe is created that includes MDS solutions, colors + SMU identifiers

# Un plot in B/W for the FOUR ecosites described above: PCoA 1 versus PCoA 2
plot(temp215.221.226.306[,1],temp215.221.226.306[,2],pch=unclass(as.numeric(temp215.221.226.306[,3])),cex=1.3,main="Big Sagebrush Sites - MDS 1984-1996",xlab="PCoA 1",ylab="PCoA 2")
legend("topleft", c("R028AY215", "R028AY221", "R028AY226", "R028AY306"),pch=c(20,21,22,23),cex=1.2)

# Extracting SMU with 60% components for cosines
temp215.221.226.306.dos<-temp215.221.226.306[c(2,35,36,56,57,69,70,71,72,73,74,75,76),]

# Un plot in B/W for the FOUR ecosites described above: PCoA 1 versus PCoA 2: only SMU with 60% soil components
legend("topleft", c("R028AY215", "R028AY221", "R028AY226", "R028AY306"),pch=c(20,21,22,23),cex=1.2)

####### ANNIOS 1997 - 2008 ###################################
### First must generate a data set to calculate ecological distance
### Independent variables go up to 2008
# First extract only those polygons from ecosites with STM and field points
ecosites9710<-ecosites2[c(c(1:312),c(800:862)),c(c(34:61),seq(86,108,by=2),seq(126,148,by=2),seq(166,188,by=2),seq(206,228,by=2))]
# Then assign the name of the polygon and name of the field plot to each record
row.names(ecosites9710)<-nombres.filas[c(c(1:312),c(800:862))]

### Starting a MDS ###
### First must generate a data set to calculate ecological distance
# Distances between SMU and benchmarks is obtained
ecosites9710.dist<-dist(scale(ecosites9710),method="manhattan")

### HOW MANY DIMENSIONS TO USE? ###
## Using Kruskal Procedure
ecosites9710.mds1<-isoMDS(ecosites9710.dist,k=1)
ecosites9710.mds2<-isoMDS(ecosites9710.dist,k=2)
ecosites9710.mds3<-isoMDS(ecosites9710.dist,k=3)
ecosites9710.mds4<-isoMDS(ecosites9710.dist,k=4)
eosites9710.mds5<-isoMDS(ecosites9710.dist,k=5)
eosites9710.mds6<-isoMDS(ecosites9710.dist,k=6)
eosites9710.mds7<-isoMDS(ecosites9710.dist,k=7)

stress.behavior9710<-cbind(c(1:7),c(ecosites9710.mds1$stress,ecosites9710.mds2$stress,ecosites9710.mds3$stress,ecosites9710.mds4$stress,ecosites9710.mds5$stress,ecosites9710.mds6$stress,ecosites9710.mds7$stress))
stress.behavior9710

### A scree plot to determine how many dimensions to use
plot(stress.behavior9710,type="b",ylab="Stress Value",xlab="Number of dimensions",main="MDS Stress Minimization 1997-2010",col=3,pch=19)

### Two dimensions seems to work fne, just like with 1984-1996
RESEARCH AND TEACHING INTERESTS
Remote sensing-based land-use/cover mapping and temporal dynamics; Ecological Site Descriptions and State and Transition Models; Invasive species mapping; Climate change and ecosystem services; Geographic information systems and spatiotemporal modeling; Integrated watershed management; Landscape ecology modeling.

EDUCATION
• Ph.D. in Ecology, Utah State University, Logan, Utah. 2011 Dissertation: Spatiotemporal Modeling of Threats to Big Sagebrush Ecological Sites in Northern Utah. GPA 3.97.


PROFESSIONAL EXPERIENCE
o RESEARCH ASSISTANT (Fall 2005 – currently) at the RS/GIS laboratory. Department of Wildland Resources, Utah State University, Logan, Utah. Mainstream of research: Developing Remote Sensing protocols to assess rangeland condition and trend at the landscape level to provide managers better inputs for decision making.

Project involvement at the RS/GIS Lab:
• States and transition models preparation for Rich County, Utah (NRCS)
• Remote sensing based rangeland monitoring protocol for Box Elder County, Utah (BLM)
• Land cover and ecosystem mapping in Honduras (ESNACIFOR)
• Assessment of ecological condition at the Grand Teton National Park (National Park Service)
• Vegetation change tracker implementation in Utah, New Mexico and forest region one (US Forest Service)

o TEACHING ASSISTANT (Fall 2006 – currently) at the Department of Wildland Resources, Utah State University, Logan, Utah. Classes taught: Applied Remote Sensing WILD6750 (Graduate level), Assessment and Synthesis in Natural Resources WILD4910 (Undergraduate level). Principal Instructor for WILD6750 during the fall of 2008.


FORESTRY/GIS TECHNICIAN (June 1996 – Dec 2000) Honduran Forest Service AFE/COHDEFOR, Tegucigalpa, Honduras. Field data capture using GPS for forest management plans; forest roads update, micro watershed characterization, and land use planning using GIS.

CONSULTING EXPERIENCE

Hydrologic modeling. ESA Consultores – MARENA-IDB. Modeling of water yield spatial and temporal dynamics in the sub watersheds of Lake Yojoa and Reitoca River using physically-based models i.e. BASINS/SWAT. 2005.

Cadastral applications. CINSA-DEC. Design of a cadastral GIS application for 3 counties of Northern Honduras. Integration of field and CAD data into an Arcview Avenue application and training of cadastral technicians in the usage of it. November 2004 – May 2005.


Database Assessment. MARENA-SERNA-IDB. Implementation of a hydro-meteorological database systematization process in the MARENA (natural resources management in prioritized sub watersheds program) region. Final products used in subsequent water resources management plans for allocation purposes. July thru September 2004.

Hydrological modeling. ROCHE ITEE-IDB. Study on water yield estimation in the Tegucigalpa (capital of Honduras) Municipality Sub watersheds by means of hydrological modeling. Water balance and streamflow data were simulated and validated with the Soil and Water Assessment Tool (SWAT) model in order to be employed in a water resources administration plan. May - June 2004.

GIS Specialist. PBPR-WB. Co-Execution of land use planning studies in a municipalities association (South of Honduras) counseled by the Rural Productivity and Forest Project PBPR. The spatial database and developed tools were used for the establishment of a watershed authority institution. March - June 2004.


GIS Consultant. CIAT-IDB. Land use multitemporal dynamics in the Humuya River Watershed. Generation of socioeconomic and biophysical indicators and decision making tools (part of an overall consultancy adapted to appraise PROCUENCULA’s (Program in charge of “El Cajón” watershed management) activities impact on local economy and program economic profitability. Tegucigalpa and Siguatepeque, Honduras. July – November 2000.


PAPERS AND POSTERS AT PROFESSIONAL MEETINGS


FORMAL TRAINING

- Image classification techniques for the development of accurate, detailed, quantitative land-cover data. ASPRS, April 28th, 2008.
- Image classification techniques for the development of accurate, detailed, quantitative land-cover data. ASPRS, May 7th, 2007.

ACADEMIC HONORS / EDUCATIONAL HIGHLIGHTS

- Utah State University - Ecology Center Research Support Funds - April 2011
- 2nd Place Oral Presentation College of Natural Resources – Intermountain Graduate Research Symposium. Graduate Student Senate, Utah State University. 2009.
- Graduate Student Senate Enhancement Award for outstanding research and teaching experience at Utah State University. 2008.
- Organization of American States OAS Scholarship to fund graduate studies. 2008. Was not accepted.
- Research assistantship to conduct doctoral studies at the Wildland Department, College of Natural Resources, Utah State University. 2005-2010.
- ASDI/FOCUENCAS. Tegucigalpa, Honduras. SCHOLARSHIP to attend M.Sc. studies in Turrialba, Costa Rica. For outstanding academic background.
JOURNAL REVIEWER

Rangeland Ecology and Management REM

PROFESSIONAL MEMBERSHIPS

American Society for Photogrammetry and Remote Sensing (ASPRS) 2005-present
Society of Range Management (SRM) 2011 - present
Honduran Foresters Association COLPROFORH 0664.

LANGUAGES

Spanish (native), English (Fluent TOEFL 633 03/2004)