NASA DEVELOP National Program

Goddard Space Flight Center
Spring 2015

Idaho Disasters II

Using NASA Earth Observations to Identify Savanna and Shrubland Vegetation in Southern Idaho

DEVELOP Technical Report
Final Draft - April 2, 2015

Jeff May (Project Lead)
Andrea Bodenberg
Kiersten Newtoff
Kyle Sowder

Mark Carroll, Goddard Space Flight Center (Science Advisor)
John Schnase, Goddard Space Flight Center (Science Advisor)
Keith Weber, Idaho State University GIS TReC (Science Advisor)

Previous Contributors:
Katy Bradford
Eric Smith

I. Abstract

Wildfires play an important role in ecosystem health, with many native plant species dependent on fire to complete their life cycle. Wildfires also burn dead vegetation, which recycles nutrients back into the soil. However, longer dry periods and the prominence of invasive species (e.g. *Bromus tectorum*) have created favorable conditions in the western United States for larger and more frequent wildfires, which can disrupt ecosystems, human localities, and the critical habitats of endangered wildlife. To prepare for the fire season in Idaho, the Bureau of Land Management (BLM) and the Idaho Department of Lands (IDL) use vegetation moisture measurements from the National Fuel Moisture Database to identify and allocate resources to regions with drier vegetation during the year. In order to supplement their current data products, we created a vegetation map to identify vegetation species with high fire risk and highlight areas of high fuel concentration. The vegetation map was created using a decision tree model on imagery from the Landsat 8 Operational Land Imager throughout the year in southeastern Idaho. The results and data gathered from this study will support IDL and BLM in allocating resources early in the fire season and planning fuel load reduction activities following the fire season.

Keywords

Vegetation map, southern Idaho, cheatgrass, wildfire, savannah

II. Introduction

Wildfires are natural ecological processes that support long-term environmental sustainability and diversity but are also considered major disturbance mechanisms to human society. Key to understanding wildfire regimes is knowing the distribution of vegetation and how fire behaves in the presence of various types of flora. Throughout the rangelands of Idaho, wildfire regimes have grown in frequency due to the introduction of foreign brome grasses, specifically, *Bromus tectorum,* hereby referred to as 'cheatgrass' (Bradley et al., 2009; Mealor et al., 2013). This project is classified in the Disasters application area due to the effect an increase in fire severity and frequency has on the landscape and society (Schneider et al., 2008).

Cheatgrass outcompetes vegetation in the native sagebrush steppe due to its winter and early spring germination cycles and fast-growing shoot and root system. Due to its early phenology, cheatgrass reaches mature and senescent stages much faster than native species and is a fine fuel source for wildfires (Mealor et al., 2013). Cheatgrass is also hazardous because it belongs in the <1-hour fuel moisture class, making it more responsive to day to day climate conditions. The moisture content in vegetation has a direct impact on fire susceptibility and is directly related to environmental conditions and the size of the plant. Larger flora, those greater than 7.6 cm in diameter, are in the 1000-hour fuel class, meaning that it takes 1000-hours or more in order to equilibrate with the moisture content in the air, and takes even longer for living plants (Schoennagel et al., 2004). Larger vegetation species such as trees and some shrubs are not as susceptible to drying out and contributing to fire susceptibility unlike smaller plants in the 1-, 10- and 100-hour fuel classes.

The objective of this study was to create a vegetation classification map that visualized the spatial distribution of cheatgrass and other vegetation classes in southeast Idaho. Although vegetation maps exist for the Great Basin, none are as comprehensive as we proposed and most classify land coverage based on functional land use rather than existing vegetation. Bradley et al. (2008) examined all vegetation species in the Great Basin region of northern Arizona using Normalized Difference Vegetation Index (NDVI); however, the spatial resolution was at 1 km with less than 60% overall accuracy due to pixel mixing. Many researchers have analyzed the distribution of cheatgrass on a 30-250 m scale (Meinke, 2009; Peterson, 2003; PNWRC, 2004; Singh & Glenn, 2009), but the location of all vegetation is needed to understand the fire ecology and restoration processes in the region. Our research seeks to create a methodology that will allow the vegetation map to be updated each year, since fires can change the landscape seasonally.

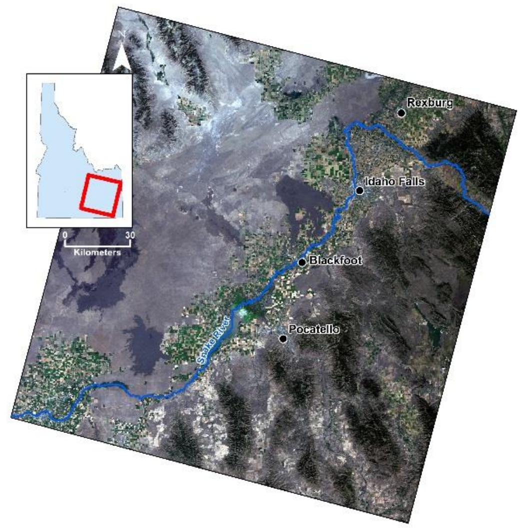


Figure 1 - Study area in SE Idaho located at WRS2 Path 30 row 30

Our study area was comprised of the expansive savannah ecosystems and agricultural operations in southeast Idaho. As a preliminary examination, we used imagery gathered during 2013 and 2014 from Landsat 8 WRS-2 Path 39, Row 30. The spatial extent of the Landsat scene covers 31,450 km^2 centered around Blackfoot, ID. A majority of this region is classified as semi-arid desert scrub, grassland, and agriculture as identified by the 2011 National Land Cover Dataset (NLCD) - most of which is located in The Big Desert (Figure 1). Annual precipitation totals in southern Idaho ranges from 20 to 30 cm, of which 25-50% is snowfall. Vegetation in the savannah ecosystems is a mixture of native species such as sagebrush, rabbit brush, crested winter wheat, and the non-native cheatgrass (Chen et al., 2011). Nearly 100% of the wildfires in this region occur between May and October, peaking in July and August, which are the warmest months in southern Idaho (Westerling et al., 2003).

Results and methodologies obtained from this study will support the Idaho Department of Lands (IDL) and the Bureau of Land Management (BLM) in making resource allocation decisions early in the fire season. The vegetation map will help identify areas with high proportions of fine, easily ignitable fuels that are ideal locations for fuel load reduction activities such as prescribed burns. Not only will vegetation maps aid in fire susceptibility models, decision making, and monitoring invasive species, vegetation distribution also greatly affects faunal distribution and may have important applications to preserving delicate ecosystems (Yensen et al., 2002).

III. Methodology

Data Acquisition

Landsat 8 and WorldView2 satellite imagery were collected through the United States Geological Survey (USGS) using the Earth Explorer web application. Three radiometrically corrected level-1T images were collected from the Operational Land Imager (OLI) instrument on Landsat 8. The images selected for analysis were identified as the earliest dates for 2013 and 2014 that had less than 20% cloud cover and after temperatures were high enough to allow for cheatgrass germination. Early growing season imagery is best for identifying cheatgrass given the earlier phenology of cheatgrass respective to native vegetation (Singh & Glenn, 2009). For this study, imagery was obtained from Landsat 8 OLI for June 16, 2013, April 16th, 2014 and May 2nd, 2014. The 2013 imagery was obtained later in the season due to cloud cover in the region earlier in the year. The 2013 imagery was used to compare to the 2014 imagery and identify areas of agreement. Atmospheric effects were corrected in the Landsat imagery using the IDRISI ATMOSC module prior to data processing. WorldView2 imagery was collected between July 19 and October 14, 2014 for use in validation. This Level 1 data product has 0.5 m resolution; however, the image area is greatly reduced to an average of 3,500 km^2 per image.

Classification Tree Analysis

We implemented Classification Tree Analysis (CTA) to categorize the study area into different vegetation classes. This tool is useful since it does not rely on normally distributed data, uses a variety of different inputs including raw imagery, and generally provides better classification accuracy than other methods (Lawrence et al., 2004). We did not use the raw Landsat 8 imagery as the inputs, instead the modified soil adjusted vegetation index (mSAVI) and Kauth-Thomas tasseled cap transformation (TCT) (Kauth & Thomas, 1976) were calculated. The mSAVI (specifically, mSAVI2) was originally proposed by Qi et al. (1994) to account for the reflectance of soil as a relationship to the percent of vegetation cover, which is beneficial for the semi-arid study region (Singh & Glenn, 2009). The equation uses the near-infrared and red bands in the Landsat 8 imagery:

$$mSAVI2 = 0.5[2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - R)}] \qquad \text{Eqn. 1}$$

The TCT determines the brightness, greenness, and wetness using the Landsat 8 bands in each image. The brightness is the total brightness of each pixel summed through each band. The greenness calculates a vegetation index measuring the photosynthetically active radiation and the wetness index indicates soil moisture. Although TCT is usually displayed as one composite, each of the indices are independent of each other and were used separately as inputs in the CTA.

Classification sites used for training and validation were created using *in situ* cheatgrass data obtained from the GIS Training and Research Center (GIS TReC) in Pocatello, ID, USGS BISON - Biodiversity Information Serving Our Nation (http://bison.usgs.ornl.gov/), and University of Georgia: Center for Invasive Species & Ecosystem Health (http://www.eddmaps.org/). Cheatgrass observations from the GIS TReC were collected during June 2014, BISON observations were from June 2013, and observations from UGA were collected during the summer months in 2004, 2005, and 2007. Other than cheatgrass, the vegetation classification points were created from visual analysis of WorldView 2 and 2013 National Agricultural Imagery Program (NAIP) images and land cover classification products including the 2011 NLCD and the 2001 Northwest Gap Analysis. The other vegetation classes created were bare ground, sagebrush, juniper/montane forest, and riparian (Table 1). An image mask was created from the 2010 Idaho Cropland Data Layer (USDA) and applied to CTA results to remove areas defined as urban, agricultural, basaltic lava formations, and water since these classes are not relevant to the vegetation distribution analysis that was conducted in this study. Five hundred points from the classification sites were randomly subset for each class and used as training sites for CTA. The remaining classification points were used to validate the CTA results. The same set of training and validation sites were used for both the 2013 and 2014 CTA. All imagery for 2013 and 2014 were input into the classification tree analysis in IDRISI. The study area had to be trimmed to best use this tool to an area of 14,222km^2 (see Figure 2 for extent).

Table 1 – Classification Sites Used for Training and Validation

Class	*# Training*	*# Validation*	*Total*
Riparian	485	21830	22315
Bare ground	499	6951	7450
Sagebrush/shrub	412	147030	147442
Cheatgrass	448	352	800
Juniper/Montane forest	76	1126	1202
Total	1920	177289	179209

Validation

The total number of validation points was 177,289 and varied between classes and were dependent on the original number of classification sites for each class (Table 1). The validation points for the riparian, bare ground, juniper/montane forest, cheatgrass, sagebrush/shrub classes were used to compare the vegetation output from the 2013 and 2014 CTA. The 2013 classification output was compared to the results of the 2014 vegetation map to further evaluate the accuracy of this methodology and the outputs. It is assumed that vegetation distribution remained relatively unchanged throughout the study region throughout the 2013 and 2014 growing seasons.

IV. Results & Discussion

The kappa coefficient for the 2014 vegetation validation was 0.38, representing a fair agreement between the model output and the validation sites (Figure 2A; Appendix 1; Landis & Koch, 1977). Both the 2013 and 2014 CTA results used a majority of the input bands, but the 2014 CTA included more branches to get to the final classification, likely as a

Table 2 – Band Usage in Classification Tree Analysis

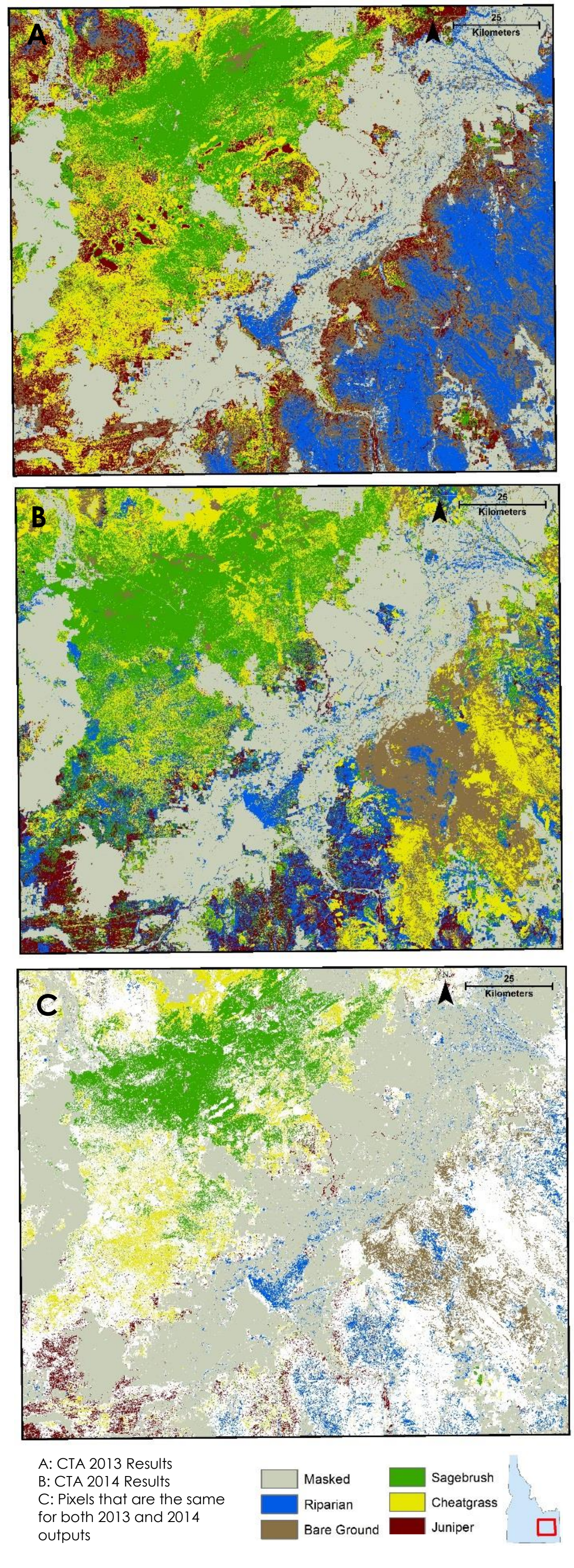
2013			2014		
Date	**Image**	**Used**	**Date**	**Image**	**Used**
Jun 16	mSAVI	10	Apr 16	mSAVI	13
Jun 16	Bright	9	Apr 16	Bright	3
Jun 16	Green	5	Apr 16	Green	7
Jun 16	Wet	1	Apr 16	Wet	3
			May 02	mSAVI	8
			May 02	Bright	3
			May 02	Green	4
			May 02	Wet	0

result of more input bands (Appendix 2). The most useful layers for the 2014 CTA analysis were the mSAVI from both dates and the greenness band from April 16th (Table 1). However, the May 2nd wetness band was not used at all in the analysis. The 2014 vegetation map exhibited 68% overall accuracy between the five vegetation classes (Appendix 1B). In comparison, Bradley *et al*. (2008) used a manually created CTA in this region with only 58% accuracy and the Northwest Gap Analysis (NGA) accuracy with a 70% (Grossmann *et al*., 2008). Although the NGA is slightly more accurate, this data product was created by many thousands of individuals conducting *in situ* data to create accurate training and validation points. Although this product is comprehensive, it is not updated regularly to reflect the changing ecosystems in southern Idaho.

The cheatgrass vegetation class was slightly more accurate at 73%. The increased accuracy could be due to better classification sites due to many of them being based on *in situ* data and the early mSAVI imagery was used specifically to better delineate cheatgrass. The high accuracy of the cheatgrass locations is the most beneficial to the BLM and IDL since cheatgrass is the most highly monitored. The kappa coefficient for the 2013 vegetation validation was 0.30, similar to the 2014 validation, and represents a fair agreement between the model output and the validation sites (Figure 2B; Appendix 3; Landis & Koch, 1977). In the 2013 CTA, the most used imagery for splits were mSAVI, used 10 times, and brightness, used 9 times (Table 1). The total accuracy for the classes was 55% while cheatgrass had a 64% accuracy. The kappa coefficient, overall accuracy, and cheatgrass accuracy is lower than 2014 which is likely due to using later-season imagery for the input images.

Comparison between 2013 and 2014

The purpose of running the CTA on the 2013 imagery was to see how much agreement existed between the two analyses between years. Although the 2013 vegetation map was less accurate than 2014, the pixels in which both maps agree provides insight and further validation on the



identity of the vegetation class the pixel belongs (Figure 2B). There was 34.7% agreement across all vegetation classes between 2013 and 2014 (Table 3). Specifically, 27.8% of the cheatgrass pixels in 2014 were located in the same location in 2013. These values indicate that there was a lot of change in vegetation between 2013 and 2014, which likely did not occur. Both of the vegetation outputs have errors addressed below, which explains the large disparity between the two maps and the low similarities. Nonetheless, the existing similarities provide strength to the accurate classification of those pixels.

Table 3 – Comparison between 2014 and 2013 Vegetation Output

	2014	2013	Same	% of 2014
Masked	6917	6917	6917	100
Riparian	2715	2972	725	26.7
Bare ground	2306	2654	802	34.8
Sagebrush	3909	2961	1866	47.7
Cheatgrass	3981	3169	1108	27.8
Juniper	1311	2466	437	33.3
Total	14222	14222	4939	
% Agreement			34.7	

*units in km^2

Errors and Uncertainty

One of the major problems encountered was misclassification. Some of the inaccuracy in vegetation classification is due to pixel mixing. Some of the Landsat 8 pixels contain multiple vegetation species, which prevents the decision tree model from accurately categorizing a pixel, a common problem seen in previous research (Atkinson et al., 1997; Bradley et al., 2008). The model usually categorizes a pixel based on the vegetation species most present, but depending on the spectral signature, the pixel may be assigned to a category not representative of any vegetation species in the area; however, this occurrence is rare. Misclassification of pixels is also due to inaccurate training and validation sites, which CTA is particularly sensitive and leads to erroneous results (Friedman, 2001). In the results, the juniper/montane forest in the southeast corner of the study area was mistakenly labeled as either riparian or cheatgrass. This misclassification is likely due to the absence of a conifer forested class and a smaller number of training sites for the juniper/montane forest class.

Future Work

Work in the subsequent term will resolve errors with the vegetation map by incorporating additional *in situ* observations scheduled for collection in May 2015 by GIS TReC staff and sagebrush mapping products currently in development by the USGS (expected delivery July 2015). These additions, combined with previously utilized methods, will result in the creation of more accurate classification sites used for future decision tree models. The Normalized Differenced Bare Soil Index (NDBSI), first proposed by Baraldi *et al*. (2006), will also be assessed for suitability into future decision tree classification models and included if it is observed to assist with the differentiations between soil and other relevant classes. We will also explore using the stochastic gradient boosting technique which works similarly to CTA but places less emphasis on training sites and more emphasis on the relatedness between pixels (Lawrence *et al*., 2004).

V. Conclusions

Using mSAVI and TCT derived from Landsat 8 OLI, this study focused on identifying vegetation distribution throughout southeast Idaho with particular interest in identifying cheatgrass. This work was done in order to create a vegetation map representing the spatial distribution of cheatgrass and other vegetation types and land coverage in the region. Extreme weather events such as El Niño, may encourage cheatgrass expansion across the study area, increasing the susceptibility to wildfire occurrences. Even though phenological stages and plant community structures can influence the accuracy of data derived from remotely sensed imagery, we are confident this product is the most recent and up-to-date tool for use by our end users, the BLM and IDL. Not only can this product be used for identifying fire susceptibility, it can be used to identify changing habitat for endangered and critical species, monitoring responses to climate change and as a guide for restoring ecosystems. Continued use of remote sensing technologies will enable our end users, and others, to promote prevention by characterizing pre-fire conditions and risks. The methods applied in this study are easily applied on an annual basis to update the vegetation map with the most recently available satellite imagery, which is critical in analyzing how the landscape responds to various disturbances and natural biological processes including wildfires.

VI. Acknowledgments

Thank you to Drs. K. Weber, M. Carroll, J. Schnase, and J. Bolten and to our previous contributors E. Smith and K. Bradford. Extended thanks to M. Wooten, R. Gill, and other research associates at the ISU GISTrec for your contributions to this research. This material is based upon work supported by NASA through contract NNL11AA00B and cooperative agreement NNX14AB60A.

VII. References

Atkinson, P. M., Cutler, M. E. J., & Lewis, H. (1997). Mapping sub-pixel proportional land cover with AVHRR imagery. *International Journal of Remote Sensing, 18*(4), 917-935.

Baraldi, A., Puzzolo, V., Blonda, P., Bruzzone, L., & Tarantino, C. (2006). Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images. *IEEE Transactions on Geoscience and Remote Sensing, 44*(9), 2563-2586.

Bradley, B. A., & Mustard, J. F. (2008). Comparison of phenology trends by land cover class: a case study in the Great Basin, USA. *Global Change Biology*, *14*(2), 334-346.

Bradley, B. A., Oppenheimer, M., & Wilcove, D. S. (2009). Climate change and plant invasions: restoration opportunities ahead?. *Global Change Biology*, *15*(6), 1511-1521.

Chen, F., Weber, K. T., Anderson, J., & Gokhal, B. (2011). Assessing the susceptibility of semiarid rangelands to wildfires using Terra MODIS and Landsat Thematic Mapper data. *International Journal of Wildland Fire*, *20*(5), 690-701.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189-1232.

Grossmann, E.B., Kagan, J. S., Ohmann, J. A., May, H., Gregory, M. J., & Tobalske, C. (2008). Final Report on Land Cover Mapping Methods, Map Zones 2 and 7, PNW ReGAP. Institute for Natural Resources, Oregon State University, Corvallis, OR.

Kauth, R. J. & Thomas, G. S. (1976). The Tasselled Cap -- A graphic description of the spectral-temporal development of agricultural crops as seen by LANDSAT. *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*. Purdue University of West Lafayette, Indiana, 4B-41-51.

Landis, J. R., & Koch, G. G. (1977) The measurement of observer agreement for categorical data. *Biometrics 33* (1), 159-174.

Lawrence, R., Bunn, A., Powell, S., & Zambon, M. (2004). Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of the Environment, 90*(3), 331-336.

Mealor, B. A., Mealor, R. D., Kelley, W. K., Bergman, D. L., Burnett, S. A., Decker, … & Fernandez-Gimenez, M. (2013). *Cheatgrass management handbook: Managing an invasive annual grass in the Rocky mountain region*. Laramie, Wyoming: University of Wyoming.

Meinke, C.W., Knick, S. T., & Pyke, D. A. (2009). A spatial model to prioritize sagebrush landscapes in the intermountain west (U.S.A.) for restoration. *Restoration Ecology, 17*(5), 652-659.

Pacific Northwest Regional Collaboratory (PNWRC). (2004). Final Report for Idaho State University (ISU): Landscape Data Fusion and Assessment: Improved Feature Extraction using Multivariate Stacking. Retrieved from GIS Training and Research Center, *url:* http://giscenter.isu.edu/research/Techpg/pnwrc/index.htm

Peterson, E. B. (2003). Mapping percent-cover of the invasive species *Bromus tectorum* (cheatgrass) over a large portion of Nevada from satellite imagery. *Report for the US Fish and Wildlife Service, Nevada State Office, Reno, by the Nevada Natural Heritage Program, Carson City*.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*(2), 119-126.

Schneider, P., Roberts, D. A., & Kyriakidis, P. C. (2008). A VARI-based relative greenness from MODIS data for computing the Fire Potential Index. *Remote Sensing of Environment*, *112*(3), 1151-1167.

Schoennagel, T., Veblen, T. T., & Romme, W. H. (2004). The interaction of fire, fuels, and climate across Rocky Mountain forests. *Bioscience, 54*(7), 661-676.

Singh, N., & Glenn, N. F. (2009). Multitemporal spectral analysis for cheatgrass (*Bromus tectorum*) classification. *International Journal of Remote Sensing*, *30*(13), 3441-3462.

Westerling, A. L., Gershunov, A., Brown, T. J., Cayan, D. R., & Dettinger, M. D. (2003). Climate and wildfire in the western United States. *Bulletin of the American Meteorological Society*, *84*(5), 595-604.

Yensen, E., Quinney, D. L., Johnson, K., Timmerman, K., & Steenhof, K. (1992). Fire, vegetation changes, and population fluctuations of Townsend's ground squirrels. *American Midland Naturalist, 128*(2), 299-312.

Appendix 1 – Error Matrix for 2014 Classification Tree Analysis

VIII. Appendices

	Riparian	Bare ground	Sagebrush	Cheatgrass	Juniper	Total	ErrorC
Riparian	17907	356	17976	27	198	36464	0.5089
Bare ground	153	6183	3558	14	24	9932	0.3775
Sagebrush	888	194	93890	45	143	95160	0.0133
Cheatgrass	1717	210	26086	256	78	28347	0.991
Juniper	1164	7	5520	9	682	7382	0.9076
Total	21829	6950	147030	351	1125	177285	
ErrorO	0.1797	0.1104	0.3614	0.2707	0.3938		0.3292
90% Confidence Interval = +/- 0.0018 (0.3274 - 0.3311)							
95% Confidence Interval = +/- 0.0022 (0.3270 - 0.3314)							
99% Confidence Interval = +/- 0.0029 (0.3263 - 0.3321)							
Overall Kappa = 0.375							

Appendix 2 – Classification Tree Analysis in 2014 (A) and 2013 (B)

Appendix 3 – Error Matrix for 2014 Classification Tree Analysis

	Riparian	Bare ground	Sagebrush	Cheatgrass	Juniper	Total	ErrorC
Riparian	18970	1046	72	1	103	20192	0.0605
Bare ground	1990	4672	4023	11	163	10859	0.5698
Sagebrush	11	173	73552	69	19	73824	0.0037
Cheatgrass	269	280	42480	226	209	43464	0.9948
Juniper	589	779	26903	44	631	28946	0.9782
Total	21829	6950	147030	351	1125	177285	
ErrorO	0.131	0.3278	0.4997	0.3561	0.4391		0.4469

90% Confidence Interval = +/- 0.0019 (0.4450 - 0.4489)

95% Confidence Interval = +/- 0.0023 (0.4446 - 0.4492)

99% Confidence Interval = +/- 0.0030 (0.4439 - 0.4500)

Overall Kappa = 0.2981