Modeling Cheatgrass using Quickbird Imagery

Bhushan Gokhale, GIS Training and Research Center, Idaho Sate University, Pocatello ID 83209-8130. (gokhbhus@isu.edu)

Keith T. Weber, GIS Director, GIS Training and Research Center, Idaho Sate University, Pocatello ID 83209-8130 (webekeit@isu.edu)

ABSTRACT

Cheatgrass (*Bromus tectorum*) is of growing concern among farmers, ranchers, and researchers in the intermountain west due to its ability to alter rangeland vegetation communities and their associated fire regimes. Quickbird multi-spectral imagery (spatial resolution 2.4 mpp) was used to detect and model cheatgrass presence in southeastern Idaho. Training sites were created using 2003 field data. The four bands of Quickbird imagery were processed using IDRISI software and various vegetation indices were created using the red and near-infrared bands of the Quickbird imagery. Principal Components Analysis (PCA) was performed using eight vegetation indices and the four original bands of data. Redundant data was minimized using results from PCA. Maximum likelihood classification was then used to classify the imagery to model the presence of cheatgrass. The accuracy of the resulting model was assessed using an error matrix and Kappa statistic. An overall accuracy of 73% was achieved through this process.

Keywords: weeds, vegetation indices, signature extraction, maximum likelihood

INTRODUCTION

Cheatgrass (*Bromus tectorum*) is considered a problematic invasive plant in Idaho's rangelands. This annual plant can germinate in spring or fall and is known to colonize areas quickly following a fire. During most of its life cycle, cheatgrass provides poor forage for livestock. In addition, it is considered a fire promoter as its early senescence provides great quantities of fine fuel to carry late summer wildfires (Vavra M. et al., 1999). Understanding the distribution and density of cheatgrass is essential to management. Due to its widespread, yet patchy distribution (cheatgrass is rarely found in excess of 30% cover within our study area) we chose to use Quickbird multispectral imagery to detect and predict the location of cheatgrass.

The main objectives of this study were as follows-

1. Identify areas with current cheatgrass infestation.
2. Develop spectral signatures for cheatgrass extracted from satellite imagery at those locations of known cheatgrass infestation.
3. Develop a cheatgrass prediction model using maximum likelihood classification.
4. Assess the accuracy of the cheatgrass model.
5. Purify the training sites and repeat classification-assessment.

THE STUDY AREA

The study area, known as the Big Desert, lies in southeastern Idaho, approximately 71 km northwest of Pocatello. The center of the study area is roughly located at 113º 4' 18.68" W and 43º 14' 27.88" N (Figure 1) (Sander and Weber 2005). This area is managed by the Bureau of Land Management (BLM) and exhibits a large variety of native species as well as invasive species. The area is a sagebrush-steppe semi-desert which is bordered by geologically young lava formations to the south and west. Irrigated agricultural lands border the study area to its north, south and east. The area has a history of livestock grazing and wildfire occurrence (Weber and McMahan 2003).

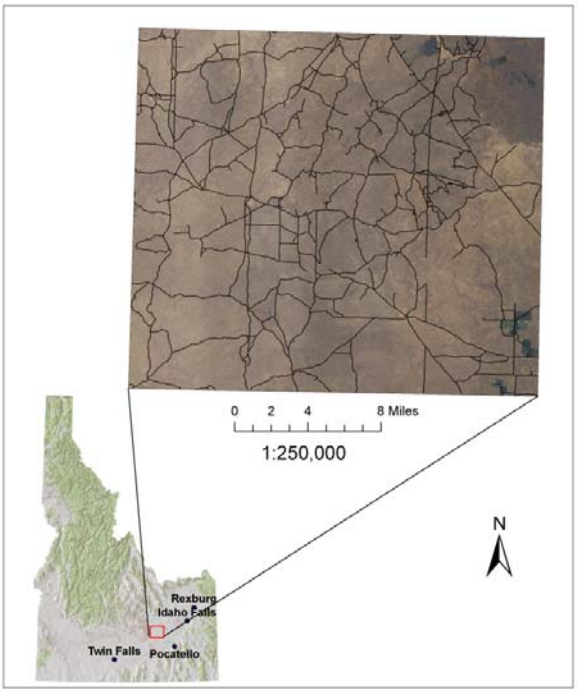

Figure 1. The Big Desert study area in southeastern Idaho.

METHODS

A total of 253 field samples were collected during 2003 for model development and 171 samples were collected in 2004 for model validation. All sample points were randomly located across the study area. A Trimble GeoXT GPS receiver (+/-1m with a 95% CI) was used to record the location of each point. At each of the sample points data describing percent cover of grasses and shrubs, dominant weed (cheatgrass, leafy spurge etc.) and shrub species, fuel load, sagebrush age, GAP vegetation classification, presence of microbial crust, litter type, and forage availability were recorded. In addition, four photographs were taken at each point. Visual estimation was used to describe percent bareground, litter, shrub and grass into one of nine classes (1. None, 2. 1-5%, 3. 6-15%, 4. 16-25%, 5. 26-35%, 6. 36-50%, 7. 51-75%, 8. 76-95%, and 9. >95% (Sander and Weber 2005)).

Quickbird satellite imagery was used for predictive modeling of cheatgrass. The imagery is comprised of four multispectral bands each with 2.4 x 2.4 meter pixels. The wavelengths sensored by each band were as follows:

- *Blue: 450 to 520 nanometers*
- *Green: 520 to 600 nanometers*
- *Red: 630 to 690 nanometers*
- *Near-Infrared: 760 to 900 nanometers*

ArcGIS and IDRISI software were used to 1) convert field samples into training sites (coded for cheatgrass presence/absence), 2) calculate various vegetation indices, and 3) perform image processing and classification.

Paste in these same things from the RH paper…

1. ***Training sites*** – Training sites were used to allow the image processing software to generate spectral signatures for the target of interest (i.e., cheatgrass). Each training site consists of a geographic location and sample attribute (e.g., 1=cheatgrass presence or 2=cheatgrass absence). To be considered a cheatgrass presence training site, the field sample had to contain $\geq$16% cheatgrass. To be considered a bare ground presence training site, the field sample has to contain $\geq$50% bare ground. The field sampling points (vector shapefiles) were converted into raster training sites (TIFF format) for use in IDRISI. Those pixels containing a training site were given a value (1 (presence) or 2 (absence)) where all other pixels (unknown) retained a value of zero.

2. ***Vegetation Indices-*** Various vegetation indices were calculated to better capture the vegetation characteristics of the study area and potentially improve our modeling results. These indices are relative measures of actively photosynthesizing vegetation present within each pixel (USWCL 2005). Different vegetation indices can be produced using different combinations/ratios of red and infrared bands and by including additional parameters such as slope and intercept of soil line and soil adjustment factors. Normalized Difference Vegetation Index (NDVI) is perhaps the most common, while the soil adjusted vegetation index (SAVI) also considers the reflectance of bare ground and its behavior in the red and near-infrared bands. Thus SAVI may offer a more accurate characterization of communities where bare ground is frequently encountered.

3. Eight vegetation indices were calculated for this study using the VEGINDEX module of IDRISI program. These indices were then evaluated for their relevance using Principal Components Analysis (PCA).

4. ***Principal Components Analysis (PCA) –*** In remote sensing PCA is used as a data reduction and noise removal technique. PCA produces a new set of images from a set of input images. The images of the new set are called components. The components are uncorrelated with each other and are ordered by the amount of variance described in the original input images (Eastman 2003).

   The first two or three components from a PCA typically describe nearly all the variability contained in the input imagery. Subsequent components introduce redundancy, outliers, and noise. Therefore these components can be eliminated from further processing without losing any important data.
   We performed PCA using the four bands of Quickbird imagery along with eight vegetation indices. The PCA analysis produced a set of component images. Based on these results a subset of images (describing about 97% of total variance) was selected to be used for signature extraction.

5. ***Signature Extraction***- Spectral signature is also known as spectral response pattern and the concept is very similar to the human concept of color. A spectral signature provides a description of the magnitude that light energy is reflected by the target in different regions of the electromagnetic spectrum. Thus it plays a key role in the interpretation of remotely sensed data. Figure 3 shows how light is reflected, absorbed, and transmitted from a particular object. The MAKESIG module of IDRISI was used to develop cheatgrass signatures from PCA component images.

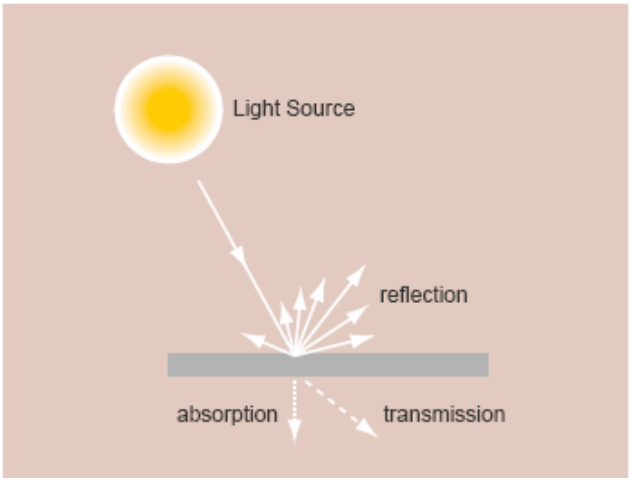

Figure 2. Light behavior of a surface (source: Eastman 2003, IDRISI Kilimanjaro Guide)

6. ***Maximum Likelihood Classification***- Maximum Likelihood Classification was used for image classification. This classification is a type of hard classifier. The hard classifier makes a definitive decision about the land cover class to which each pixel belongs (Eastman 2003).

7. The classification uses information present in the signature file to determine which class each pixel belongs to (i.e., cheatgrass presence or cheatgrass absence). Equal probability was assigned to each of the signature class. Therefore 50 percent probability was assigned to each class.

8. ***Error Assessment-*** This step is also known as validation. While there are different ways to judge the accuracy of a model, we chose to use a standard contingency table (or error matrix) and the Kappa statistic. An error matrix measures the agreement between ground truth data with the developed model.

   The ERRMAT module was used for error assessment of the cheatgrass model. We calculated kappa (how much better or worse the classification performed relative to a chance classification) (Titus et. al. 1984), overall, commission, and omission errors.

9. ***Purification-*** signature purification is sometimes necessary to achieve reliable results. The PURIFY module of IDRISI performs parametric and non-parametric purification. In parametric purification, the training pixel vector is retrieved from the images and the Mahalanobis distance to the mean of the class is calculated. The user is required to specify a typicality threshold which ranges from 0 to 1. A zero typicality threshold indicates that all the pixels will be kept in the data set while value 1 implies that nearly all pixels will be eliminated save for those equal to the mean. Based on the threshold, each pixel's typicality is evaluated (using Mahalanobis distance). A pixel with typicality value less than the threshold is dropped from the new data set (Eastman 2003). Signature purification was performed in the study and a new set of purified signatures was created and used for a second iteration maximum likelihood classification.

RESULTS AND DISCUSSION

Figure 3 shows the cheatgrass model produced using maximum likelihood classification from un-purified training sites. The corresponding error matrix generated using the ERRMAT module of IDRISI is given in table 1.

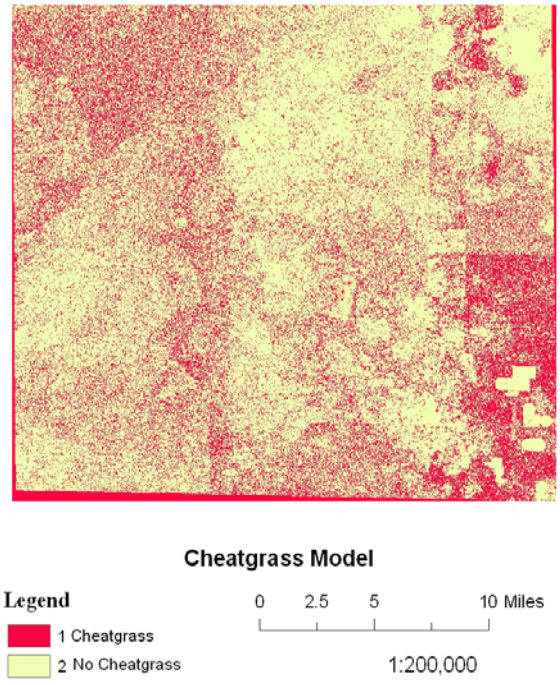

Figure 3. The binary cheatgrass model produced for the Big Desert Study area. Red (dark) pixels represent predicted cheatgrass presense at /or above 16% cover (cover class no. 4).

Table 1 – Error Matrix for a Cheatgrass model developed using un-purified signature files.

		Field Observations			
Model Predictions		Presence	Absence	Total	Commission Error
	Presence	18	55	73	0.75
	Absence	8	152	160	0.05
	Total	26	207	233	
	Omission Error	0.31	0.26		0.27

The model predicted only 18 sites correctly as cheatgrass (Table 1). This resulted in a large error of commission of about 75%. Further examination of the model indicates that eight known cheatgrass training sites were classified as non-cheatgrass sites. This resulted in an error of omission of approximately 30%. The model did well identifying what is not cheatgrass. Considering these results, the overall error of the model was about 27% (making it 73% accurate). The kappa index was 0.2383, indicating little improvement over a simple random classification. Further, such a low Kappa value indicates little reliability in the model. Table 2 shows the error matrix for a second cheatgrass model developed using purified signature files.

Table 2 – Error Matrix for a Cheatgrass model developed using purified signature files.

		Field Observations			
Model Predictions		Presence	Absence	Total	Commission Error
	Presence	20	69	89	0.77
	Absence	5	114	119	0.04
	Total	25	183	208	
	Omission Error	0.20	0.38		0.36

As a result of purification 25 training sites were removed. The overall error of the model was about 36% (or about 64% accurate). The error of omission and commission were very similar to those reported using non-purified signature files (Table 1) and Kappa was 0.2009 indicating the purification process was not beneficial in this case.

A primary source of error in both models is likely poor co-registration between training sites and the Quickbird imagery regardless of our attempts to eliminate the error. While the Quickbird imagery was geo-registered with a ground control shapefile the best RMS achieved was 5.20 m. When considering the pixel size was 2.40 m, this implies a positional error of more than twice the pixel size. In other words, our cheatgrass training sites may have been shifted up to 2 pixels away from the actual pixel containing the target. This error propagated through the model and likely explains the unacceptably low accuracy.

The patchy nature of cheatgrass in the study area is not advantageous to reliable processing. While very common, cheatgrass cover for our training sites ranged between 16 and 36% (no training sites exceeded 36% cover). This means that even for cheatgrass presence training sites, roughly two-thirds of the pixel was covered by non-target objects. It was expected that by using imagery with high spatial resolution (i.e., Quickbird) that the mixed-pixel affect could be minimized and allow for a reliable classification. Indeed, this is not the case. Further, using high-spatial resolution imagery introduced another difficulty; co-registration between imagery and training sites which was discussed earlier.

CONCLUSIONS

While further study is required, cheatgrass currently cannot be modeled reliably (where users' accuracy exceeds 70%) using Quickbird multispectral imagery. The overall accuracy of the

model was 73%; however user's accuracy never exceeded 25%. Purification did not improve accuracy. We speculate that training sites dominated (>50%) by cheatgrass are required to achieve reliable results. This is probably true of all target/species differentiations except in those cases where the target has a distinctive spectral signature or texture. In addition, numerous ground-based control points which are visible within the imagery are required to eliminate co-registration errors and the resulting loss of accuracy propagated by this error.

ACKNOWLEDGEMENTS
This study was made possible by a grant from the National Aeronautics and Space Administration Goddard Space Flight Center. ISU would also like to acknowledge the Idaho Delegation for their assistance in obtaining this grant.

LITERATURE CITED
Eastman, R. J., 2003. IDRISI Kilimanjaro Guide to GIS and Image Processing. Clark University Laboratory.

USWCL 2005. How a vegetation index works. URL=http://www.uswcl.ars.ag.gov/epd/remsen/Vi/VIworks.htm, visited 19-July-2005.

Vavra M., W. A. Laycock, and R. D. Pieper. 1999. Ecological Implications of Livestock Herbivory in the West, Society of Range Management.

Sander L. and K. T. Weber. 2005. Range Vegetation Assessment in the Big Desert, Upper Snake River Plain, Idaho, GIS Training and Research Center. URL = http://giscenter.isu.edu/Research/techpg/nasa_weeds/to_pdf/fieldreport_2003-2004.pdf

Weber K. T. and B. J. McMahan. 2003. Field Collection of Fuel Load and Vegetation Characteristics Wildfire Risk Assessment Modeling: 2002 Field Sampling Report. URL = http://giscenter.isu.edu/Research/techpg/nasa_wildfire/Final_Report/Documents/Chapter2.pdf