Available online at www.sciencedirect.com
SCIENCE DIRECT®
Remote Sensing of Environment 100 (2006) 356 – 362
Remote Sensing of Environment
www.elsevier.com/locate/rse

Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest)

Rick L. Lawrence [a,*], Shana D. Wood [b], Roger L. Sheley [c]

[a] Department of Land Resources and Environmental Sciences, PO Box 173490, Montana State University, Bozeman, Montana 59717, United States
[b] Department of Land Resources and Environmental Sciences, Montana State University, United States
[c] USDA-Agricultural Research Service, Eastern Oregon Agricultural Research Center, United States

Received 10 August 2005; received in revised form 28 October 2005; accepted 29 October 2005

Abstract

Invasive nonindigenous plants are threatening the biological integrity of North American rangelands, as well as the economies that are supported by those ecosystems. Spatial information is critical to fulfilling invasive plant management strategies. Traditional invasive plant mapping has utilized ground-based hand or GPS mapping. The shortfalls of ground-based methods include the limited spatial extent covered and the associated time and cost. Mapping vegetation with remote sensing covers large spatial areas and maps can be updated at an interval determined by management needs. The objective of the study was to map leafy spurge (*Euphorbia esula* L.) and spotted knapweed (*Centaurea maculosa* Lam.) using 128-band hyperspectral (5-m and 3-m resolution) imagery and assess the accuracy of the resulting maps. Beiman Cutler classifications (BCC) were used to classify the imagery using the randomForest package in the R statistical program. BCC builds multiple classification trees by repeatedly taking random subsets of the observational data and using random subsets of the spectral bands to determine each split in the classification trees. The resulting classification trees vote on the correct classification. Overall accuracy was 84% for the spotted knapweed classification, with class accuracies ranging from 60% to 93%; overall accuracy was 86% for the leafy spurge classification, with class accuracies ranging from 66% to 93%. Our results indicate that (1) BCC can achieve substantial improvements in accuracy over single classification trees with these data and (2) it might be unnecessary to have separate accuracy assessment data when using BCC, as the algorithm provides a reliable internal estimate of accuracy.

Keywords: Beiman Cutler classification; Hyperspectral imagery; RandomForest; Leafy spurge; *Euphorbia esula*; Spotted knapweed; *Centaurea maculosa*

1. Introduction

Invasive nonindigenous plants are harmful ecologically and economically to North American rangelands (Bangsund et al., 1999; Sheley et al., 1996; Vitousek et al., 1996). Managers require accurate and timely spatial information to assist with locating and controlling small infestations before they grow too large to eradicate effectively (Johnson, 1999) and to monitor the effectiveness of their management strategies (Cooksey & Sheley, 1997). Traditional survey methods such as hand mapping and Global Positioning System (GPS) receiver mapping (Cooksey & Sheley, 1997) specify sufficiently high accuracies (approximately 80%) for small management areas (Cooksey & Sheley, 1998), but might be financially, technically, and logistically impractical for many managers.

Remote sensing has been used for decades to measure and map the biophysical characteristics of vegetation (Anderson et al., 1976; Lawrence & Ripple, 2000; Treitz et al., 1992; Tucker et al., 1985). Both spatial and spectral resolutions impact the accuracy with which individual species are mapped. Moderate resolution satellite imagery is more suited to mapping at the community level because the spatial resolution is generally too coarse to distinguish individual species unless represented as a monoculture (Dewey et al., 1991; Everitt & Escobar, 1996; Sohn & McCoy, 1997).

In contrast to satellite imagery, aerial photography is capable of producing very high spatial resolution (often less than 1 m). Vegetation reflectance is generally too similar in the visible (VIS) and near-infrared (NIR) wavelengths (Cochrane, 2000; Okin et al., 2001; Woolley, 1971), however,

* Corresponding author. Tel.: +1 406 994 5409; fax: +1 406 994 5122.
E-mail address: rickl@montana.edu (R.L. Lawrence).

0034-4257/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2005.10.014

to delineate spectrally using the limited spectral dynamic range of aerial photography (Lillesand & Kiefer, 2000). Researchers were able to utilize both VIS and NIR spectral characteristics, however, to photointerpret and map landscape-scale infestations of leafy spurge (*Euphorbia esula* L.) with conventional color and color infrared aerial photography in Theodore Roosevelt National Park (Anderson et al., 1996). Airborne digital imagery can achieve spatial resolutions similar to aerial photography, yet methods for processing and spectrally classifying the imagery can be automated, allowing for more efficient landscape-scale coverage. In Idaho, four-band multispectral imagery captured from fixed wing aircraft and with very high spatial resolution has been used to map several invasive plants (Carson et al., 1995; Lass & Callihan, 1997; Lass et al., 1996).

Identifying invasive plants in a heterogeneous landscape is difficult with multispectral imagery (Dewey et al., 1991) because healthy vegetation exhibits similar spectral responses in the VIS and NIR portions of the spectrum due to similar cellular chemistry and architecture (Woolley, 1971). An alternative to multispectral imagery is hyperspectral imagery (Ustin et al., 2004). The continuous nature of spectra inherent to hyperspectral imagery might be utilized to differentiate vegetation into taxonomic levels because of greater information content of the data. Researchers have applied hyperspectral imagery, for example, to map leafy spurge in Theodore Roosevelt National Park (O'Neill et al., 2000), undesirable woody vegetation encroaching into grasslands in the Niobrara Valley (Wylie et al., 2000), flowering leafy spurge in northeastern Wyoming (Parker Williams & Hunt, 2002, 2004), flowering leafy spurge in Idaho (Glenn et al., 2005), and hoary cress in Idaho (Mundt et al., 2005). A particular advantage of using hyperspectral imagery for invasive plant mapping is its potential for determining the relative components, or unmix, pixels, which can be especially valuable for determining percent cover of species or detecting sub-pixel size infestations (e.g., Glenn et al., 2005; Mundt et al., 2005; Parker Williams & Hunt, 2004).

We sought to map leafy spurge and spotted knapweed (*Centaurea maculosa* Lam.) at two sites in Madison County, Montana, where infestations occurred at widely varying densities and phenological stages. Three factors provided a substantial challenge in producing precise maps of invasive species in our study. First, previous studies of invasive plants using digital imagery have shown that there is a tendency for the species to be over-classified, that is, more pixels are identified as invasive species than actually exist (e.g., Lass et al., 2002). Second, in our study areas, most infested sites were not uniform, but contained a mixture of invasive species and other vegetation present at the site (referred to herein as "co-occurring vegetation"). Third, there was substantial phenological variability in the invasive species and co-occurring vegetation throughout each of our sites, unlike most previous successful attempts at mapping invasive species with hyperspectal imagery, making it more difficult to identify clear spectral responses. Our attempts to classify hyperspectral imagery for our study area using numerous classification methods (spectral angle mapper, logistic regression, classification trees, boosted classification trees with See5 (Quinlan, 1993), and stochastic gradient boosting (Lawrence et al., 2004)) all met with no success; all classifications were below 70% and included at least one class accuracy in the 40% range, which was deemed well below levels necessary for management purposes (Driscoll, 2002). We decided to test the use of a relatively new statistical method, Breiman Cutler classification (BCC, implemented as the randomForest package in the R statistical program (R Development Core Team, 2005)), based on the assertion that it is particularly powerful when there are many weak explanatory variables, characterized as where no single variable or small group of variables can be expected to distinguish classes (Breiman, 2001). This was the case with our data, where no one of the 128 hyperspectral bands could be expected to distinguish the invasive species. BCC was developed by one of the principal developers of classification and regression trees (CART) and is claimed to be "unexcelled in accuracy among current algorithms" (www.stat.berkeley.edu/users/breiman/RandomForests).

BCC is a bagging (bootstrap aggregation) operation, where multiple classification trees are developed, each one based on a random subset of the training data observations (Breiman, 2001; Lawrence et al., 2004). In addition to this normal bagging function, in BCC, each classification tree split is based on a random subset of the input variables, in our case spectral information. The multiple classification trees then vote by plurality on the correct classification. BCC (under the name randomForest) has been used successfully for studies of clinical drugs (Gunther et al., 2003), structural genomics (Goh et al., 2004), molecular epidemiology (Schwender et al., 2004), vegetation mapping related to climate change (Iverson et al., 2004), and recently in remote sensing studies (Bunn et al., 2005; Pal, 2005).

BCC has several advantages over other classification tree-based approaches (Breiman, 2001; Liaw & Wiener, 2002). Pruning of trees is not necessary and the approach is robust to overfitting, a problem that plagues classification trees. It is easier to use than many other ensemble classification methods, with the only parameters to be set being the number of trees grown and the number of variables used at each tree split; however, it has been shown to be not very sensitive to the setting of either of these parameters. It also is claimed that BCC can provide a reliable estimate of error using the data that is randomly withheld from each iteration of tree development (the "out-of-bag" portion), making it unnecessary to have an independent accuracy assessment data set (Breiman, 2001). We found this claim particularly intriguing, as it would enable all collected data to be used for training and potentially substantially reduce the field effort required for remote sensing studies.

The specific objectives of our study were to examine the ability of hyperspectral imagery and BCC to classify two invasive species, leafy spurge and spotted knapweed, in southwest Montana rangeland. We examined further whether the internal assessment of accuracy from BCC was as reliable as an independent accuracy assessment data set.

2. Methods

The imagery covered two sites in Madison County, Montana, each site approximately 1024 ha in area (Fig. 1). The leafy spurge site is located 16 km southwest of Twin Bridges at the southern end of the Highland Mountains (UTM extents: 384023E, 5092295N and 387239E, 5039036N; NAD83, Zone 12). Average annual precipitation is 30 cm (Boast & Shelito, 1989). Physiography predominantly consists of upland fans dissected with ephemeral streams and terraces of the Big Hole River. Native vegetation is a mix of grasses, forbs, and shrubs, including *Bouteloua gracilis* (H.B.K.) Lag. (blue grama), *Poa sandbergii* Vasey (Sandberg bluegrass), *Pascopyrum smithii* Rydb. (western wheatgrass), *Pseudoroegneria spicatum* Pursh. (Love) (bluebunch wheatgrass), *Hesperostipa comata* Trin. and Rupr. (needleandthread), *Nassela viridula* Trin. (green needlegrass), *Artemisia tridentata* Nutt. (big sagebrush), and *Chrysthamnus nauseosus* (Pall.) Britt. (rabbitbrush) (Boast & Shelito, 1989). Leafy spurge primarily occupied drainage bottoms and surrounding hillsides and grew in association with native vegetation in low to high density infestations and occasionally grew in dense monocultures. Leafy spurge was growing in various phenological stages on the date the image was captured, ranging from bright yellow-green flowering to early senescence.

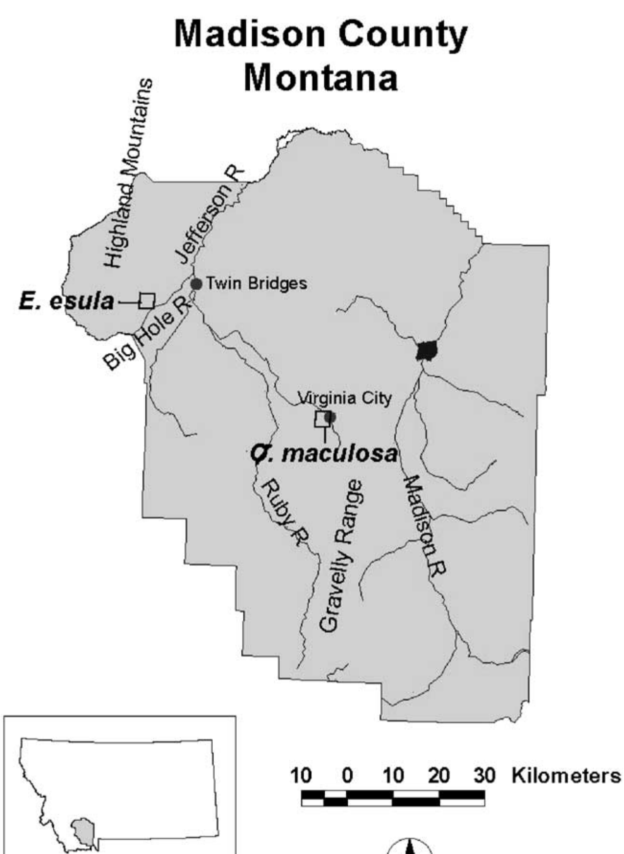

Fig. 1. Leafy spurge and spotted knapweed were mapped at two locations in Madison County, southwest Montana.

The second site, located in the northern foothills of the Gravelly Range and including the town of Virginia City and areas due west and south (UTM extents: 422693E, 5017404N and 425887E, 5014094N; NAD83, Zone 12), was dominated by spotted knapweed. Average annual precipitation is 30 cm (Boast & Shelito, 1989). Physiography predominantly consists of upland hills and ridges of the Gravelly Mountains and floodplains of perennial streams. Native vegetation consisted of mixed rangeland-forest types, including *P. spicatum*, *Festuca idahoensis* Elmer (Idaho fescue), *S. comata*, *Carex filifolia* Nutt. (threadleaf sedge), *Lupinus* L. spp. (lupine), *A. tridentata*, *C. nauseosus*, *Gutierrezia sarothrae* (Pursh) Britt. and Rusby. (broom snakeweed), *Symphoricarpos albus* (L.) Blake (snowberry), *Juniperus scopulorum* Sarg. (juniper), *Populus tremuloides* Michx. (aspen), and *Populus deltoides* Marsh. (cottonwood) (Boast & Shelito, 1989). In addition, the non-native grass *Bromus tectorum* L. (cheatgrass) infested portions of the site. Spotted knapweed was growing in various phenological stages on the date the imagery was captured, ranging from recent flowering to early senescence. Spotted knapweed infestations tended to be mixed with other vegetation and had a higher percentage of bare soil exposed than the leafy spurge site.

The hyperspectral imagery was obtained in August 1999 using the Probe-1 sensor. We converted the imagery to relative reflectance using the internal average relative reflectance algorithm (Kruse, 1988), since we had no ground reference spectra. A minimum noise fraction (MNF) transformation was performed to control for noise in the imagery (Green et al., 1988). Specifications for Probe-1 cite a ground resolution of 5 m when flying at an average height above ground level of 2500 m (Earth Search Sciences, 2004), although resolution at the leafy spurge site was approximately 3 m. Probe-1 records reflected electromagnetic energy in 128 bands in a nearly continuous spectral range from 440 to 2507 nm (from blue through shortwave infrared). Imagery was georegistered (RMSE < 1.0 pixels) to a digital orthophotoquad (DOQ), which in turn had an estimated accuracy of 6 m.

Crews collected ground reference polygons of the target invasive species (leafy spurge or spotted knapweed) and co-occurring vegetation from August 16 to 20, 1999, using GPS receivers that had an accuracy of 2–5 m after differential correction (Marshall, 1996). We randomly located transects within a grid overlaid on the study sites and sampled target invasive species and co-occurring vegetation polygons along each transect by circumnavigating patches where the target species was either present or absent. Field data collected for this study resulted in 120 weed polygons and 64 co-occurring vegetation polygons at the spotted knapweed site and 38 weed polygons and 68 co-occurring vegetation polygons at the leafy spurge site. Sizes of invasive species infestations ranged from 6 to 507 m^2 and percent of plots infested ranged from 5% to over 50%, based on ocular estimations. GPS field data were differentially corrected and were further adjusted to align with the imagery by reference to DOQs.

Table 1
Error matrix and accuracy totals for spotted knapweed based on full data sets

Spotted knapweed accuracy			
Classified data	Reference data		
	Co-occurring vegetation	Spotted knapweed	Row totals
Co-occurring vegetation	540	90	630
Spotted knapweed	42	133	175
Column totals	582	223	805
Producer's accuracy		User's accuracy	
Co-occurring vegetation	93%	Co-occurring vegetation	86%
Spotted knapweed	60%	Spotted knapweed	76%
Overall accuracy	84%		
K_{hat}	0.56		

Accuracy assessments are based on out-of-bag estimates of error. Observations are based on number of pixels within each class.

We used the randomForest package in R (Liaw & Wiener, 2002) to classify the imagery. We developed BCC models with 500 classification trees each; exploratory graphs indicated that error rates became stable well before this number of trees was developed. All 128 bands were included as potential variables for the models. The number of bands used at each tree split was optimized based on out-of-bag estimates of error (Liaw & Wiener, 2002). Models were developed using the entire reference data set and accuracy was evaluated based on the out-of-bag estimation of error. The out-of-bag estimates of error were developed using the one-third portion of the data that was randomly excluded from the construction of each of the 500 classification trees, with a different one-third randomly selected for exclusion from each classification tree. This is an internal estimate of error, as opposed to withholding a portion of the data from all model building for an independent estimate of error after the model is built. The estimation of error was based on correctly classified pixels because the classification, and therefore the out-of-bag estimation of error, was pixel based. Other approaches to error estimation for high spatial resolution data, such as polygon-based estimates, have been advocated (e.g., Glenn et al., 2005), but were not consistent with this classification approach. The larger number of co-occurring vegetation pixels, relative to weed pixels, in each data set resulted in a bias in the estimates of overall accuracy toward the class results for co-occurring vegetation. To evaluate the reliability of the out-of-bag accuracy assessment, we also tested the more traditional approach by randomly dividing the reference data polygons for each site into two equal data sets, developing models using one-half of the data, and predicted classifications for the withheld half of the data. We compared the out-of-bag accuracy assessment of the training data to the accuracy of the predictions for the withheld data.

Table 2
Error matrix and accuracy totals for leafy spurge based on full data sets

Leafy spurge accuracy			
Classified data	Reference data		
	Co-occurring vegetation	Leafy spurge	Row totals
Co-occurring vegetation	702	96	798
Leafy spurge	50	189	239
Column totals	752	285	1037
Producer's accuracy		User's accuracy	
Co-occurring vegetation	93%	Co-occurring vegetation	88%
Leafy spurge	66%	Leafy spurge	79%
Overall accuracy	86%		
K_{hat}	0.62		

Accuracy assessments are based on out-of-bag estimates of error. Observations are based on number of pixels within each class.

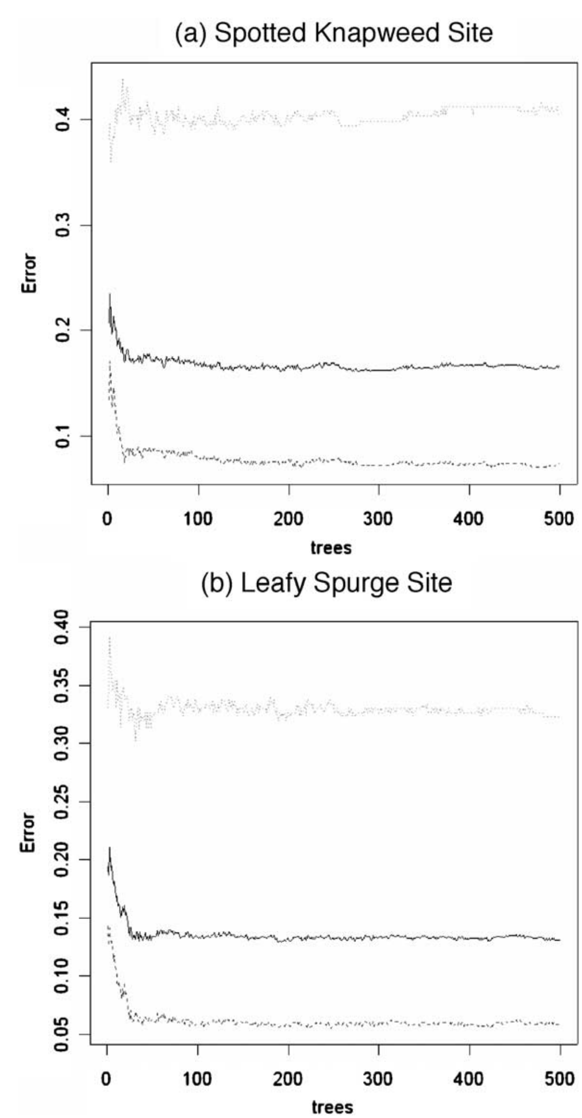

Fig. 2. Graphs showing change in out-of-bag accuracy estimation (in terms of error rates) with increase in number of classification trees. Out-of-bag estimated error rates are computed continuously as additional classification trees are built. Estimates based on the full model (500 classification trees) are hypothesized to be unbiased estimates of error, but estimates based on a small number of trees are likely unreliable. Black lines represent overall accuracy, light gray lines represent weed class accuracy, and medium gray lines represent co-occurring vegetation accuracy.

3. Results and discussion

Estimated overall accuracy from out-of-bag data for the full data sets was 84% (K_{hat}=0.56) for the spotted knapweed site (Table 1) and 86% (K_{hat}=0.62) for the leafy spurge site (Table 2). Class accuracies for the spotted knapweed site ranged from 60% for spotted knapweed producer's accuracy to 93% for co-occurring vegetation producer's accuracy (Table 1). User's accuracy was somewhat more consistent for the spotted knapweed site, 76% for spotted knapweed and 86% for co-occurring vegetation. The largest source of error was spotted knapweed being classified as co-occurring vegetation.

The leafy spurge site had very similar class accuracies to the spotted knapweed site (Table 2). Leafy spurge producer's and user's accuracies were slightly higher, at 66% and 79%, respectively, than spotted knapweed. Co-occurring vegetation producer's accuracy was the same as at the spotted knapweed site, 93%, while user's accuracy was 2% higher at 88%. Again the main source of error was weed locations being incorrectly classified as co-occurring vegetation.

These results are particularly remarkable when compared to other classification attempts with these data (Driscoll, 2002). Spectral angle mapper, for example, had overall accuracies of 40% and 66% for the spotted knapweed and leafy spurge sites, respectively, maximum likelihood had 48% and 50%, and conventional classification trees had 65% and 70%. In all of these attempts, at least one class accuracy was 41% or less. BCC provided graphical output that demonstrated, based on out-of-bag accuracy estimation, how accuracies change as numbers of classification trees are increased (Fig. 2). The starting point is a single classification tree. It can be seen that accuracies increase substantially as trees are added to the model, eventually stabilizing after less than 50 trees. Out-of-bag estimates of accuracy might be less reliable, however, when there are very few trees in the model.

Our evaluation of the reliability of the out-of-bag estimates of accuracy supported the position that these estimates are reliable. All overall and class accuracies from the reduced modeling data sets, based on the out-of-bag estimates, and the withheld data from the reduced data sets, which provided our independent accuracy assessments, were within 3%, and most estimates were less than 1% apart (Tables 3 and 4). These accuracy assessments from the reduced data sets were also nearly identical to the out-of-bag accuracy assessments from the full data sets (compare Table 1 with Table 3 and Table 2 with Table 4). This supports the assertion that, with BCC, it is not necessary to have a separate accuracy assessment data set. We believe that this assertion should receive considerable additional testing with other data before it is accepted as a substitute for reliable independent accuracy assessments. This could provide a tremendous saving of resources for remote sensing studies if sufficient studies show that these out-of-bag estimates of accuracy are reliable. We caution, however, that this conclusion is based on the assumption that there is no bias in the reference data. If there is a bias in the reference data, that bias will be present in the accuracy assessment and the results will not be reliable. We protected against bias in our study by randomly locating all reference plots, but this is not always practical.

Table 3
Error matrix and accuracy totals for spotted knapweed based on reduced data set

Spotted knapweed accuracy			
Classified data	Reference data		
	Co-occurring vegetation	Spotted knapweed	Row totals
Co-occurring vegetation	278/262	43/49	321/311
Spotted knapweed	20/21	62/70	82/91
Column totals	298/283	105/119	403/402
Producer's accuracy		User's accuracy	
Co-occurring vegetation	93%/93%	Co-occurring vegetation	87%/84%
Spotted knapweed	59%/59%	Spotted knapweed	76%/77%
Overall accuracy	84%/83%		
K_{hat}	0.56/0.55		

Numbers represent results for out-of-bag estimates/assessment for withheld data. Observations are based on number of pixels within each class.

Table 4
Error matrix and accuracy totals for leafy spurge based on reduced data set

Leafy spurge accuracy			
Classified data	Reference data		
	Co-occurring vegetation	Leafy spurge	Row totals
Co-occurring vegetation	354/353	47/43	401/396
Leafy spurge	24/28	93/95	117/123
Column totals	378/381	140/138	518/519
Producer's accuracy		User's accuracy	
Co-occurring vegetation	94%/93%	Co-occurring vegetation	88%/89%
Leafy spurge	66%/69%	Leafy spurge	79%/77%
Overall accuracy	86%/86%		
K_{hat}	0.63/0.64		

Numbers represent results for out-of-bag estimates/assessment for withheld data.

Our results might have been adversely affected by, among other things, registration errors, a common issue with high spatial resolution data (e.g., Aspinall et al., 2002; Glenn et al., 2005). Error existed in our reference DOQ (estimated up to 6 m), image registration to the DOQ (estimated at 1 m), and the corrected GPS data (estimated at 2 to 5 m). The compound effect of these sources of error likely resulted in some level of misalignment between the imagery and our ground reference data.

BCC was able to provide reasonable accuracies for what have been difficult data sets to classify. The internally generated out-of-bag accuracy assessments were shown to be reliable, potentially obviating the need to collect separate assessment data. The implementation of BCC in the R statistics package (as the randomForest package) makes it available to analysts free of charge. We believe, based on our relatively high accuracies, ease of use, low cost, and possibly no need for independent accuracy assessment, it is worth considering BCC for remote sensing classification problems in the future.

Acknowledgements

Research was funded by a grant from the Montana Noxious Weed Trust Fund. Earth Search Sciences, Inc. provided imagery for the project, with Montana State University-TechLink assisting with the technology transfer. Authors wish to thank Elizabeth Roberts and her field crew, with logistical support from Madison County weed personnel, for their efforts in collecting the GPS data, and Dr. Andrew Bunn of Woods Hole Research Center and two anonymous reviewers for their thoughtful comments on the manuscript.

References

Anderson, G. L., Everitt, J. H., Escobar, D. E., Spencer, N. R., & Andrascik, R. J. (1996). Mapping leafy spurge (*Euphorbia esula*) infestations using aerial photography and geographic information systems. *Geocarto International*, *11*, 81–89.

Anderson, G. L., Everitt, J. H., Richardson, A. J., & Escobar, D. E. (1976). A land use and land cover classification system for use with remote sensor data. *Professional Paper 964, US Geological Survey, Washington, DC.*

Aspinall, R. J., Marcus, W. A., & Boardman, J. W. (2002). Considerations in collecting, processing, and analyzing high spatial resolution, hyperspectral data for environmental investigations. *Journal of Geographical Systems*, *4*, 15–29.

Bangsund, D. A., Leistritz, F. L., & Leitch, J. A. (1999). Assessing economic impacts of biological control of weeds: The case of leafy spurge in the northern Great Plains of the United States. *Journal of Environmental Management*, *56*, 35–43.

Boast, R. R., & Shelito, R. G. (1989). Soil survey of Madison County area, Montana. *USDA, Soil Conservation Service.*

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Bunn, A. G., Goetz, S. J., & Fiske, G. J. (2005). Observed and predicted responses of plant growth to climate across Canada. *Geophysical Research Letters*, *32*, L16710.

Carson, H. W., Lass, L. W., & Callihan, R. H. (1995). Detection of yellow hawkweed with high resolution multispectral digital imagery. *Weed Technology*, *9*, 477–483.

Cochrane, M. A. (2000). Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing*, *21*, 2075–2087.

Cooksey, D., & Sheley, R. L. (1997). Noxious weed survey and mapping system. *Rangelands*, *19*, 20–23.

Cooksey, D., & Sheley, R. L. (1998). *Mapping noxious weeds in Montana*. EB148, Montana State University Extension Service, Bozeman, Mont. 23 pp.

Dewey, S. A., Price, K. P., & Ramsey, D. (1991). Satellite remote sensing to predict potential distribution of Dyers woad (*Isatis tinctoria*). *Weed Technology*, *5*, 479–484.

Driscoll, S.G. (2002). *Detecting and mapping leafy spurge (Euphorbia esula) and spotted knapweed (Centaurea maculosa) in rangeland ecosystems using airborne digital imagery.* Masters Thesis, Montana State University, Bozeman, MT. 111 pp.

Earth Search Sciences, Inc. (2004). *Earth Search Sciences, Inc. Home Page.* Via the INTERNET. Accessed 2004 June, http://www.earthsearch.com/

Everitt, J. H., & Escobar, D. E. (1996). Use of spatial information technologies for noxious plant detection and distribution on rangelands. *Geocarto International*, *11*, 63–80.

Glenn, N. F., Mundt, J. T., Weber, K. T., Prather, T. S., Lass, L. W., & Pettingill, J. (2005). Hyperspectral data processing from repeat detection of small infestations of leafy spurge. *Remote Sensing of Environment*, *95*, 399–412.

Goh, C. S., Lan, N., & Douglas, S. M. (2004). Mining the structural genomics pipeline: Identification of protein properties that affect high-throughput experimental analysis. *Journal of Molecular Biology*, *336*, 115–130.

Green, A. A., Berman, M., Switzer, P., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transaction on Geoscience and Remote Sensing*, *26*, 65–74.

Gunther, E. C., Stone, D. J., & Gerwien, R. W. (2003). Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. *Proceedings of the National Academy of Sciences*, *100*, 9608–9613.

Iverson, L. R., Prasad, A. M., & Liaw, A. (2004). New machine learning tools for predictive vegetation mapping after climate change: Bagging and Random Forest perform better than Regression Tree Analysis. In R. Smithers (Ed.), *Landscape Ecology of Trees and Forests. Proceedings of the Twelfth Annual IALE (UK) Conference, Cirencester, UK, 21–24 June 2004* (pp. 317–320).

Johnson, D. E. (1999). Surveying, mapping, and monitoring noxious weeds on rangelands. In R. L. Sheley, & J. K. Petroff (Eds.), *Biology and management of noxious rangeland weeds* (pp. 19–35). Corvallis, Ore: Oregon State University Press.

Kruse, F. A. (1988). Use of airborne imaging spectrometer data to map minerals associated with hydrothermally altered rocks in the Northern Grapevine Mountains, Nevada. *Remote Sensing of Environment*, *12*, 31–51.

Lass, L. W., & Callihan, R. H. (1997). The effect of phenological stage on detectibility of yellow hawkweed (*Hieracium pratense*) and oxeye daisy (*Chrysanthemum leucanthemum*) with remote multispectral digital imagery. *Weed Technology*, *11*, 248–256.

Lass, L. W., Carson, H. W., & Callihan, R. H. (1996). Detection of yellow starthistle (*Centaurea solstitialis*) and common St. Johnswort (*Hypericum perforatum*) with multispectral digital imagery. *Weed Technology*, *10*, 466–474.

Lass, L. W., Thill, D. C., Shafh, B., & Prather, T. S. (2002). Detecting spotted knapweed (*Centaurea maculosa*) with hyperspectal remote sensing technology. *Weed Technology*, *16*, 426–432.

Lawrence, R. L., 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 Environment*, *90*, 331–336.

Lawrence, R. L., & Ripple, W. J. (2000). Fifteen years of revegetation of Mount St. Helens: A landscape-scale analysis. *Ecology*, *81*, 2742–2752.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2/3*, 18–22.

Lillesand, T. M., & Kiefer, R. W. (2000). *Remote sensing and image interpretation*, fourth edition. New York, NY: John Wiley and Sons, Inc. 724 pp.

Marshall, G. (1996). *Characterizing accuracy of Trimble Pathfinder mapping receivers*. Trimble Surveying and Mapping User Conference and Exposition, San Jose, Cal.

Mundt, J. T., Glenn, N. F., Wever, K. T., Prather, T. S., Lass, L. W., & Pettingill, J. (2005). Discrimination of hoary cress and determination of its detection limits via hyperspectral image processing and accuracy assessment techniques. *Remote Sensing of Environment*, *96*, 509–517.

Okin, G. S., Roberts, D. A., Murray, B., & Okin, W. J. (2001). Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. *Remote Sensing of Environment*, *77*, 212–225.

O'Neill, M., Ustin, S. L., Hager, S., & Root, R. (2000). Mapping the distribution of leafy spurge at Theodore Roosevelt National Park using AVIRIS. In R. O. Green (Ed.), *2000 AVIRIS Airborne Geoscience Workshop, Pasadena, Cal.*

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*, 217–222.

Parker Williams, A., & Hunt Jr., E. R. (2002). Estimation of leafy spurge from hyperspectral imagery using mixture tuned matched filtering. *Remote Sensing of Environment*, *82*, 446–456.

Parker Williams, A., & Hunt Jr., E. R. (2004). Accuracy assessment of detection of leafy spurge with hyperspectral imagery. *Journal of Range Management*, *57*, 106–112.

Quinlan, J. R. (1993). *C4.5 programs for machine learning*. Morgan Kaufmann Publishers, San Mateo, California.

R Development Core Team. (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, ISBN: 3-900051-07-0 URL http://www.R-project.org

Schwender, H., Zucknick, M., Ickstadt, K., & Bolt, H. M. (2004). A pilot study on the application of statistical classification procedures to molecular epidemiological data. *Toxicology Letters*, *151*, 291–299.

Sheley, R. L., Svejcar, T. J., & Maxwell, B. D. (1996). A theoretical framework for developing successional weed management strategies on rangeland. *Weed Technology*, *10*, 766–773.

Sohn, Y., & McCoy, R. M. (1997). Mapping desert shrub rangeland using spectral unmixing and modeling spectral mixtures with TM data. *Photogrammetric Engineering and Remote Sensing*, *63*, 707–716.

Treitz, P. M., Howarth, P. J., & Suffling, R. C. (1992). Application of detailed ground information to vegetation mapping with high spatial resolution imagery. *Remote Sensing of Environment*, *42*, 65–82.

Tucker, C. J., Townshend, J. R. G., & Goff, T. E. (1985). African land-cover classification using satellite data. *Science*, *227*, 369–375.

Ustin, S. L., Roberts, D. A., Gamon, J. A., Asner, G. P., & Green, R. O. (2004). Using imaging spectroscopy to study ecosystem processes and properties. *BioScience*, *54*, 523–534.

Vitousek, P. M., D’Antonio, C. M., Loope, L. L., & Westbrooks, R. (1996). Biological invasions as global environmental change. *American Scientist*, *84*, 468–478.

Woolley, J. T. (1971). Reflectance and transmittance of light by leaves. *Plant Physiology*, *47*, 656–662.

Wylie, B. K., Meyer, D. J., Choate, M. J., Vierling, L., & Kozak, P. K. (2000). Mapping woody vegetation and eastern red cedar in the Nebraska Sand Hills using AVIRIS. In R. O. Green (Ed.), *AVIRIS Airborne Geoscience Workshop, Pasadena, Cal.*