MAPPING WOODY VEGETATION AND EASTERN RED CEDAR IN THE NEBRASKA SAND HILLS USING AVIRIS

Bruce K. Wylie,[1] Dave J. Meyer,[1] M. J. Choate,[1]
L. Vierling,[2] and P. K. Kozak[2].
[1]Raytheon, EROS Data Center (corresponding author:wylie@edcmail.cr.usgs.gov)
[2]South Dakota School of Mines and Tech., Rapid City, SD 57701

1. INTRODUCTION

The Niobrara Valley preserve is owned and managed by The Nature Conservancy and has an annual average precipitation of 558 mm (NCSS 1992). It is located in north central Nebraska, along the scenic Niobrara River. The preserve is unique because of its proximity to the different plant type communities. The Sand Hills Prairie grassland community occurs to the south and the Northern Mixed Prairie to the north. The woodland communities represent both the eastern deciduous forest, which is primarily on the southern canyon walls, and the Rocky Mountain forest community (Ponderosa Pine, *Pinus* ponderosa) occurs primarily on the northern river canyon breaks. Ice age relic Boreal forest species also exist in isolated patches on the moister, cooler southern canyon walls (Kaul et al. 1988). The management goals at the preserve include to: 1) mimic historical ecosystem function through bison grazing and burning and 2) promote ecosystem diversity with a special focus on the areas of unique woodland communities. Eastern Red Cedar (*Juniperus virginiana*) is native to the area, but has greatly expanded its extent since advent of fire control. Ranchers in the region use chemicals and mechanical control to try and maintain and restore rangelands infested with Eastern Red Cedar. The Nature Conservancy has initiated management plans with some of the local ranchers along the Niobrara River to implement management strategies such as burning and harvesting that will help control Eastern Red Cedar.

To facilitate the implementation of these management goals, two major objectives were addressed in this study, 1) the mapping of woody and herbaceous vegetation and 2) the detailed baseline mapping of Eastern Red Cedar. The first objective was driven by two needs. First, the need to estimate livestock carrying capacities and seasonal stocking rates mandates estimation of quantities of forage palatable to livestock. Identification and mapping of woody and subsequently less palatable forage would greatly improve future cattle and bison forage production estimates from remotely sensed data. Secondly, algorithms to estimate biophysical parameters, which are used by the carbon flux and global climate communities, vary with the structure and function of the vegetation (Asner et al. 1998, Gamon et al . 1995, and Gordron et al. 1997). Previous work (Wylie et al 1996 and Wylie et al. in press) on this and other sites have defined algorithms to estimate grassland biophysical parameters from remotely sensed data. However, tree and shrub remote sensing biophysical algorithms are less well defined on this site.

The second objective requires a current Eastern Red Cedar baseline map to assess the subsequent changes in the 10 to 15 year time frame. This will allow assessment of the utility and effectiveness of management control strategies currently being initiated with local ranchers.

2. METHODS

2.1 General

The 3m AVIRIS image was acquired on flight number f990722t02p01_r10 from 15,000ft at approximately 19:53 GMT on July 22 over the Niobrara Preserve. The scene was radiometrically calibrated and geometrically corrected at JPL. The Atmospheric Removal Program (ATREM, version 3.1) was used to correct for the effect of absorbing gases and scattering within the intervening atmosphere. An ozone concentration of 0.30 atm-cm was used in the corrections based on climatology, and an aerosol optical depth at 550 nm of 0.12 was used based on trial and error runs comparing results against a known ground target. The ATREM code returns both the AVIRIS imagery, corrected to surface reflectance, and a water vapor image. The water vapor image clearly showed considerable surface detail, implying that the correction process underestimated the atmospheric transmittance in the water vapor bands, thus overestimating the water vapor content. We continue to explore possible reasons for this result. One hundred and forty nine spectral bands were retained after masking out bands associated with strong atmospheric absorption features. Subsequently, a minimum noise fraction (MNF) transform was performed, and a threshold of 2.0 was used to retain 23 MNF transformed bands.

ASD spectral data collected on the site for Ponderosa Pine and Eastern Red Cedar indicate lots of spectral overlap with only a few spectral regions that could potentially separate the two species (Figure 1). We posed the question, “Can the identification of useful hyper-spectral properties and appropriate respective spectral thresholds needed for separating woody vegetation from herbaceous vegetation and/or the identification of Eastern Red Cedar be done in an automated fashion?”. Essentially, what is the information that can make the separation and where is it? Because vegetation spectral characteristics are essentially “moving targets” due to plant phenological changes and plant moisture changes over time, an automated technique would allow rapid date specific mapping of target vegetation classes or species from voluminous spectral data sets. The tool that seemed best to fit this need was a data mining technique, known as decision tree analysis (Quinlan 1993).

Decision tree analysis hierarchically subdivides the data set into subsets of greater uniformity, using thresholds of selected spectral data (bands or MNF components) that maximizes subset land cover uniformity for each hierarchical level (Hasnen et al. 1996). An example decision tree is shown in figure 2. A series of hierarchical thresholds are defined in minimum noise fraction components to separate various target woody classes. Arrows to the left are if the threshold is true and to the right if they are false. Decision trees have been successfully used in image classification applications (Hansen et al. 1996) and have been shown superior to maximum likelihood classification (Friedl and Brodley 1997 and Yu et al. 1993). Several of the advantages of decision trees over neural networks are that:1) decision trees are transparent, allowing analyst review of the decision process, 2) influential or inconsequential spectral bands can be identified (Prince and Steininger 1999 and Friedl et al. 1999) and, 3) development of decision trees are rapid and repeatable.

A sample of pixels were selected ad hoc across the AVIRIS scene for training and evaluation purposes. Supplementary data used by the image annalist to assign land cover classes to the sampled pixels included field experience, georegistered multispectral aerial imagery with a 2 meter resolution, and panchromatic digital ortho photos (DOQ quadrangles). Training pixels were labeled with two labels, general land cover classes (shadow, water, bare, grass, woody, and wetland) and specific woody classes (Ponderosa pine, Eastern Red Cedar, shrub, and deciduous forest). Decision trees were developed in a two tiered approach. First to map the general land cover classes and then, subsequentially, decision trees were developed to subdivide the general woody class into the specific woody classes.

2.2 General land cover

For the general land cover, 452 AVIRIS pixels were selected and target general land cover classes identified. These pixels were selected ad hoc across the image probably with minor biases toward pixels that were easily categorized by the image annalist in to land cover classes in either the AVIRIS data, supporting reference data, or knowledge from GPS data and other field visits. 80% of these sampled pixels were randomly selected to train the decision tree and the remaining 20% were for evaluation. The technique of boosting was employed. Boosting uses iterative decision trees, giving training pixels which where misclassified in the previous decision tree, higher weights in the subsequent decision tree. Predictions were then based on voting across all ten of the boosting iterations. Boosting has been showed to increase decision tree accuracies as much as 20-50% with remotely sensed data (Friedl et al. 1999). In this study ten iterations were used in each boosting run. Nine different boosting runs were implemented, each with a unique 80% random training pixel selection. From the nine boosted estimates, five were selected based on overall accuracies on the respective 20% evaluation pixels and applied to the entire spatial extent of the AVIRIS data set. This resulted in 5 different versions of general land cover maps and associated maps of the confidence of land cover estimates. Pixel by pixel voting amongst the five land cover maps was used to produce a general land cover map.

Preliminary accuracy assessment was done by combining the results of the five random sets of 20% evaluation pixels from each of the five selected boosted trees. These points, though independent of the decision tree training pixels, are not from an independent source data set and therefore tended to be a bit over optimistic.

The influential MNF components used by the decision trees was assessed by quantifying for each land cover class, the number of times each NMD component was used to classify each training pixel for all ten boost iterations of each of five selected boosted trees. This resulted in an index of the importance of the NMF components or driving forces for each general land cover class.

To assess the utility of the AVIRIS woody non palatable forage delineation from the AVIRIS general land cover, it was compared to a previous land cover product which was produced for the area using a SPOT HRV scene with a June and an August acquisitions. The AVIRIS 3m land cover was resampled using the majority to 20m and compared to the SPOT HRV classification.

2.3 Specific woody land cover

The specific land cover separation of the generalized woody land cover class was addressed with several different approaches. Two “classical” hyper-spectral algorithms were employed to separate eastern red cedar (a juniper) from Ponderosa pine. The first approach used the standard “spectral endmember” method: 1) find “pure” pixels, 2) identify endmember via “n-dimensional visualization” tools, 3) identify endmembers, and 4) then perform unmixing. Note that the methods used here were based on the algorithms incorporated into the commercial software package ENVI (Environment for the Visualization of Images; ENVI 1997). In the second approach, more supervised approach, sampled pixels of different woody types (juniper, pine, deciduous and sumac) were used in a “matched filtering” scheme on MNF transformed data (Harsanyi and Chang 1994). Used as templates for the matched filter were the average MNF spectra for 25 deciduous, 28 pine, 29 juniper and 23 sumac sites. The training pixel spectra were extracted from the AVIRIS images at locations specified by field GPS surveys and image interpretation. A “matched filter” image was created for each of the four types

A third approach, decision tree analysis, was also implemented to map the specific woody classes. Sampled pixels in the general woody class and matched filter analysis were supplemented with additional woody sampled pixels and their specific land cover classes identified. A total of 217 sampled pixels were used to train and evaluate decision trees which separated the specific woody land cover classes. An approach similar to the general land cover decision trees was employed to select the best 4 boosted decision trees. Subsequent pixel based land cover confidence weighted voting was used to map the general woody class into specific woody land cover classes.

Preliminary accuracy assessment and identification of influential MNF components for the specific woody classes was also done using the same methods as the general land cover.

3. RESULTS

3.1 General land cover

The general land cover map produced (Figure 3) showed good agreement with the AVIRIS image (red = 1650um, green = 1000um, and blue = 650um). Parenthetical specific woody classes should be grouped into the general woody land cover class for this discussion. The greenish tones on the AVIRIS image in the Sand Hill grasslands south of the river near the eastern edge of the image are short grass areas. Visually, agreement appears good even for scattered trees in grasslands.

Preliminary accuracy assessment of the general land cover map derived from the independent 20% test pixels indicated an overall accuracy of 90% (Table 1.). These preliminary estimates are probably artificially high because the test pixels, although independent from the model training pixels, were categorized as land cover using the same reference data sets as the test pixels. Although absolute preliminary accuracies may be upwardly biased, they should serve as relative indicators of strong versus weak classes. The stronger land cover classes based on user accuracies (column totals), barren and wetland, had small sample sizes. The weaker user accuracy classes were woody and shadow.

To understand what MNF components were influential in separating woody from grassland, the relative proportions of MNF components used by the all 50 (10 iterations/boost * 5 selected boosted decision trees) decision trees were compared for MNF components having at least 6% relative proportion used (Table 2). The most influential MNF components for identification of grass and shrub were MNF components 1, 3, 2, and 11.

Where and how much of an improvement in woody estimates was obtained through the use of the AVIRIS data sets over and above that obtained from a two date SPOT HRV classification? The AVIRIS land cover greatly improved classification of shrub clumps in the Sand Hills on the southeastern area of the image (Figure 4). Areas classified in the SPOT land cover as woody but not in the AVIRIS tended to be over estimations of woody components by the SPOT land cover. Examples include scattered areas in the brome hay field north in the northern part of the image and scattered river bottom grass.

3.2 Specific Woody Land Cover

Three approaches (unmixing, matched filters, and decision trees) were investigated to refine the general woody class into the specific woody classes. The linear unmixing analysis used up to 40,000 iterations of the pixel purity index algorithm was run in an attempt to find endmembers associated with the different woody species. While this method achieved separation of conifer from broadleaf trees, it could not discriminate the pines from the junipers and other endmembers. Possibly, the uncertainties in the atmospheric correction parameters impeded this effort.

The matched filter approach was then tested. Each matched filter image consisted of matched filter scores at each pixel location, where a score of 1.00 indicates a perfect match. Table 3 indicates what the matched filter (MF) score was for each of the four types (in columns) at the training pixel locations used as training sets for the filtering process (in rows). The table entries are the mean matched filter score for each training pixel, accompanied by the standard deviation of the score in parentheses. The results indicate that it should be possible to construct MF score thresholds that should separate pines from junipers with reasonable success. The results also indicate that juniper, with an average MF score of 0.62 at pine training pixels, would likely be confused somewhat more often for pine than vice versa (pine having a MF score of 0.44 at juniper training pixels). Although this is a preliminary study, it indicates that the matched filter approach would be a much more useful tool for separating different woody types than spectral unmixing, under the conditions described here.

Finally, decision tree analysis was implemented to estimate the specific woody classes. The specific woody land cover estimates were applied to the woody land cover class in the general land cover classification (Figure 3). The shrub clumps in the southern Sand Hills were correctly classified as shrub, primarily smooth sumac (*Rhus glabra)* with some scattered trees. The Ponderosa Pine on the north side of the canyon and selected locations on the south side of the canyon also seem appropriate. The location of deciduous forest on the south canyon side seems right, however, scattered oaks and other deciduous trees occur along the lower reaches of the north canyon forests. These seem to be missclassified as Ponderosa Pine. The estimates of Eastern Red Cedar may be too liberal on the south canyon walls. Given these two potential improvements in the specific land cover classification, the subsequent use of a low altitude AVIRIS data set obtained in November, should greatly reduce this confusion.

Preliminary accuracy estimates derived by pooling independent test results on all selected woody specific boosted decision trees indicated unexpectedly high overall accuracy assessment number of 94% (Table 4). Although the sample sizes here are lower than those used in some of the general land cover classes, Eastern Red Cedar had the lowest user accuracies (column statistics).

The influential MNF components for separating Ponderosa and Eastern Red Cedar as indicated by quantifying the relative proportions of MNF layers used in 40 decision trees (10 iterations per boosted decision tree * 4 selected boosted decision trees) were 1, 11, and 17 (Table 5).

Additional ground observations of actual land cover will be collected in the spring of 2000. These observations will be used to refine specific woody land cover estimates using the November, leaf off, AVIRIS data set. Independent data for accuracy assessment will be collected as well.

4. CONCLUSION

The use of AVIRIS spectral data provided meaningful and useful information that will help Nature Conservancy personnel meet management goals of herd management and ecosystem flora and fauna diversity. Mapping of areas with woody, less palatable species, was improved over existing SPOT HRV land cover maps. This will allow more accurate estimates of grassland biophysical parameters and livestock stocking rate calculations from remotely sensed data. Baseline mapping of Eastern Red Cedar will be useful in a subsequent 10-15 year reassessment that will address the impacts of management interventions to control Eastern Red Cedar both on Nature Conservancy property and lands of cooperating ranchers. Further improvement of base line mapping will be achieved from a November AVIRIS data set.

Decision tree classification was a useful tool in developing land cover information from hyper-spectral data. Not only did decision tree analysis produce good land cover maps, but also produced a preliminary accuracy assessment, produced land cover confidence maps, and identified influential spectral data. The two tiered implementation of decision trees worked well and could also be adapted to map grassland diversity as well. The poor performance of linear unmixing in this study was probable due to perturbations in the data caused by an imperfect atmospheric correction. Matched filter analysis indicated a degree of spectral separability between Ponderosa Pine and Eastern Red Cedar.

ACKNOWLEDGEMENTS

Funded through joint NASA/DOI initiative to utilize hyper spectral data in DOI projects. This is part of a long term collaboration with the Nature Conservancy with special thanks to Mike Behrens, Niobrara Valley Preserve Manager. Work done by Dr. Wylie, Dr. Meyer, and Mr. Choate was performed under U.S. Geological Survey contract 1434-CR-97-CN-40274.

LITERATURE CITED

Asner, G.P., C.A. Bateson, J.L. Privette, N. El Saleous, and C.A. Wessman. 1998. Estimating vegetation structural effects on carbon uptake using satellite data fusion and inverse modeling. J. Geophysical Res. 103:28,839-28,853, D22.

Environment for the Visualization of Images. 1977. Advanced hyperspectral analysis. *In: Tutorial*. ENVI Version 3.0, Dec. better Solutions Consulting, Lafayette, CO, p 205-226.

Friedl, M.A. and C.E. Brodley. 1997. Decision tree classification of land cover from remotely sensed data. Remote Sense. Envr. 61:399-409.

Friedl, M.A., C.E. Brodley, and A.H. Strahler. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Trans. Geoscience and Remote Sens. 37:969-977.

Gamon, J.A., C.B. Field, M.L. Goulden, K.L. Griffin, A.E. Hartley, G. Joel, j. Peneulas, and R. Valentini. 1995. 1998. Relationships between NDVI, canopy structure, and photosynthesis in three California vegetation types. Ecol. App. 5:28-41.

Gordron, N., B. Pinty and M.M. Verstraete. 1997. Theoretical limits to estimation of the leaf area index on the basis of visible and near-infrared red remote sensing data. IEEE Trans. Geoscience and Remote Sens., 35:1,438-1,444.

Hansen, M., R. Dubayah and R. Defries. 1996. Classification trees: an alternative to traditional land cover classifiers. Int. J. Remote Sens. 17:1075-1081.

Harsanyi, J.C. and C.I. Chang. 1994. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projections approach. IEEE Tranc. Geosci. And Remote Sens., 32:779-785.

Kaul, B.R., G.E. Kantak, and S.P. Churchill. 1988. The Niobrara River valley, a postglacial migration corridor and refugium of forest plants and animals in the grasslands of central North America. *The Botanical Review* 54:44-81.

National Cooperative Soil Survey. 1992. Soil survey of Brown County, Nebraska, Soil Conservation Service.

Prince, S.D. and M.K. Steininger. 1999. Biophysical stratification of the Amazon basin. Global Change Biol. 5:1-22.

Quinlan, J.R. 1993. *C4.5 Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA pp302.

Wylie, B.K., D.J. Meyer, L.L. Tieszen, and S. Mannel. In press. Towards satellite mapping of surface biophysical parameters at the biome-scale over the North American grasslands: a case study. Remote Sense Envr.

Wylie, B.K., D.D. DeJong, and L.L. Tieszen. 1996. Grassland canopy parameters and their relationships to remotely sensed vegetation indices in the Nebraska Sand Hills. *Geocarto Int.* 11:39-52.

Yu, X., I.S. Reed, and A.D. Stocker. 1993. Comparative performance analysis of adaptive multispectral detectors. IEEE Tranc. Geosci. And Remote Sens., 41:2639-2655.

Figure 1. Field ASD measurements for Ponderosa Pine and Eastern Red Cedar

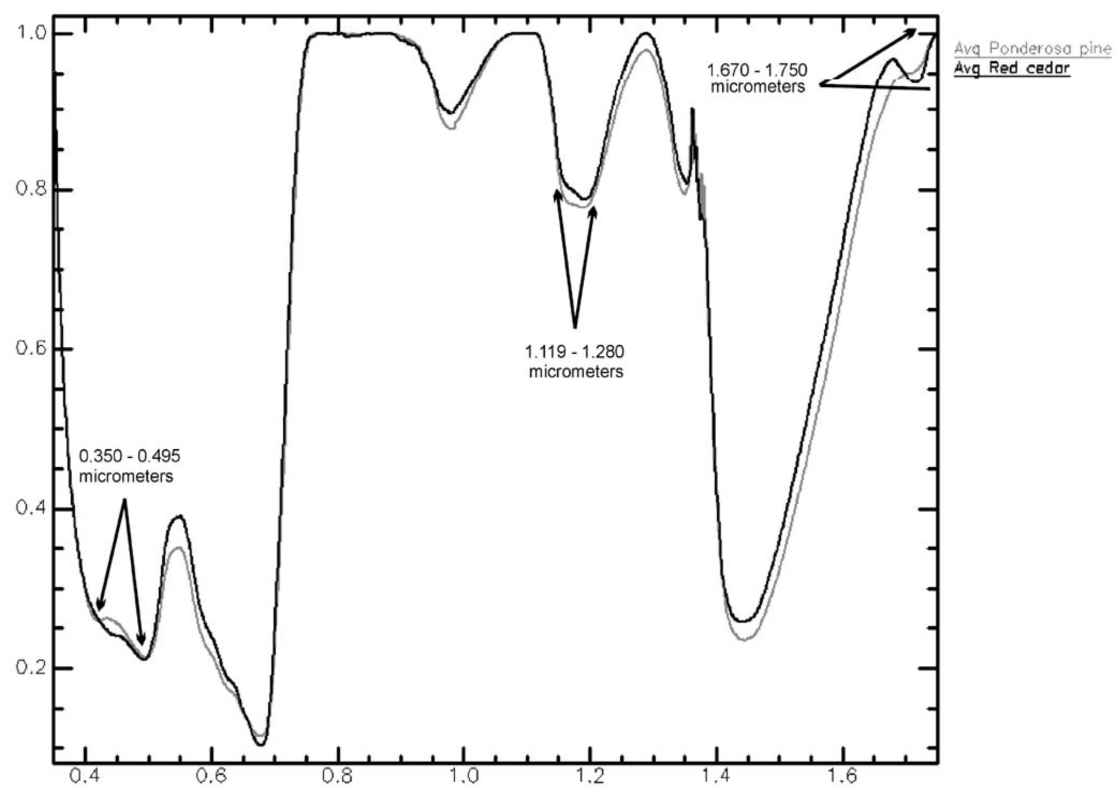

Figure 2. Decision tree example using MNF components to separate woody land cover into deciduous tree (decid), shrub, Ponderosa Pine (P. pine) and E. Red Cedar (cedar)

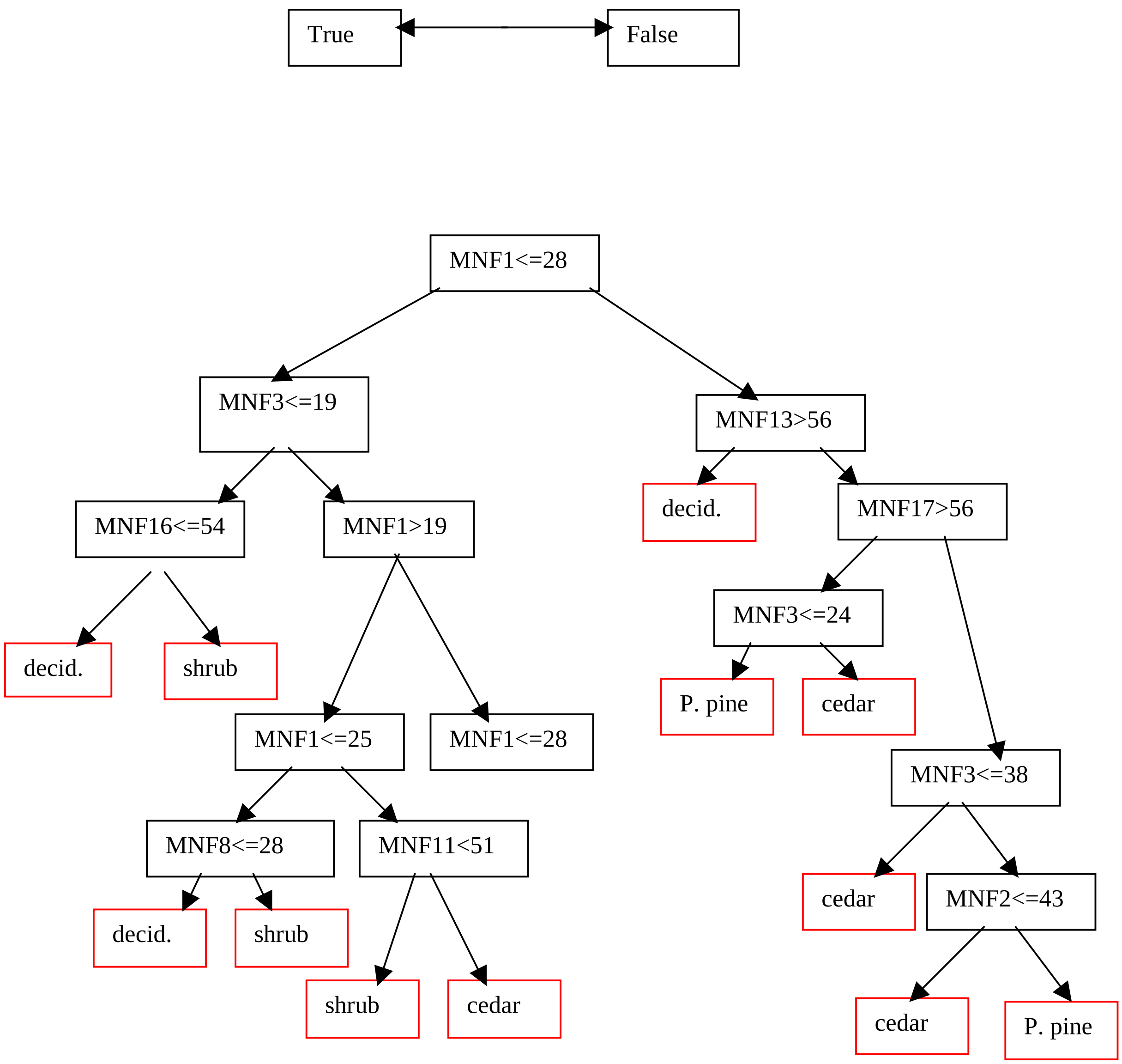

Figure 3. AVIRIS compared to land cover

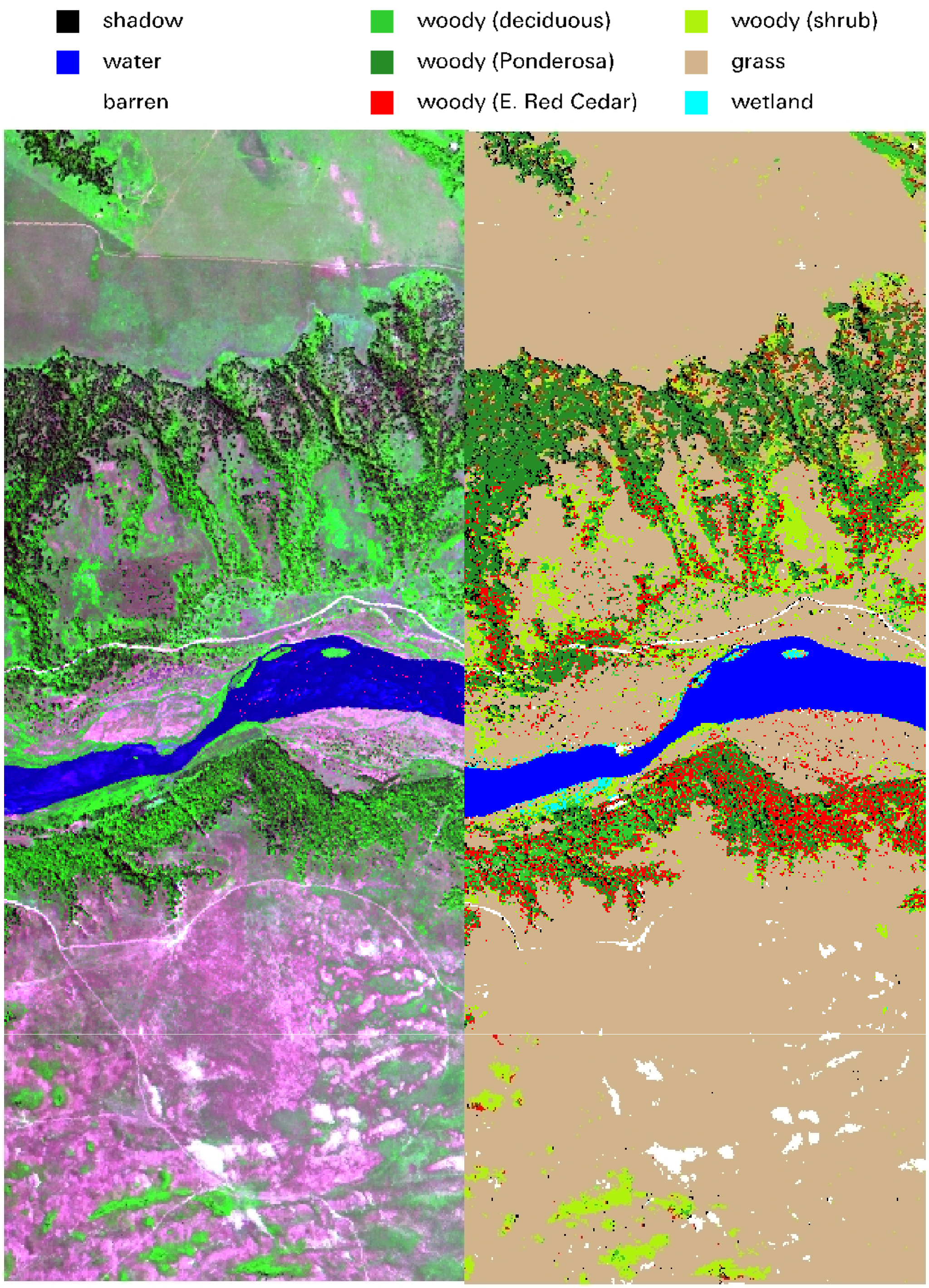

Figure 4. Difference between SPOT HRV and AVIRS woody land cover classes

Table 1. Preliminary accuracy assessment for general classes derived from the test data sets from five boosted decision trees.

		CLASSIFIED							
		shadow	water	bare	woody	grass	wetland	total	percent
	shadow	29	0	0	13	0	0	42	69%
	water	0	21	0	0	0	0	21	100%
	bare	0	0	22	1	2	0	25	88%
REFERENCE	woody	4	0	0	147	8	0	159	92%
	grass	0	0	0	9	174	0	183	95%
	wetland	0	1	0	3	4	12	20	60%
	total	33	22	22	173	188	12	450	
	percent	88%	95%	100%	85%	93%	100%		90%

Table 2. Relative proportions of MNF layers used in 5 boosted trees with 10 iterations each for general land cover classes

MNF	Shadow	Grass	Wetland	Bare	Woody	Water
2	11%	12%	11%	13%	11%	33%
1	14%	13%	10%	12%	12%	22%
7	7%	8%	7%	17%	6%	10%
11	8%	11%	13%	14%	9%	8%
3	7%	11%	10%	11%	12%	6%
20	3%	6%	6%	12%	5%	5%
16	10%	3%	2%	2%	6%	1%
4	6%	6%	8%	1%	9%	1%
5	2%	2%	2%	6%	4%	0%
6	3%	6%	4%	2%	2%	2%
13	1%	3%	5%	0%	1%	0%
total decisions	11939.8	45820.2	6971.4	2277.4	43994.4	1955.2

Table 3. Mean matched filter scores and standard deviations in parentheses for specific woody classes.

		CLASSIFIED			
		decid.	Pond. P.	Cedar	shrub
REFERENCE	decid.	1.03 (0.48)	-0.05 (0.15)	-0.23 (0.13)	0.14 (0.21)
	Pond. P.	-0.14 (0.29)	1.00 (0.30)	0.62 (0.29)	0.03 (0.24)
	Cedar	-0.46 (0.55)	0.44 (0.37)	1.00 (0.47)	0.02 (0.20)
	shrub	0.26 (0.44)	0.02 (0.14)	0.01 (0.18)	1.00 (0.33)

Table 4. Preliminary accuracy assessment for specific woody classes derived from woody class (Table 1) based on from four selected boosted decision trees.

		CLASSIFIED					
		decid.	P. pine	Cedar	shrub	total	percent
REFERENCE	decid.	37	0	0	4	41	0.9024
	P. pine	0	37	3	0	40	0.925
	Cedar	0	0	42	0	42	1
	shrub	0	0	3	46	49	0.9388
	total	37	37	48	50	172	
	percent	1	1	0.875	0.92		0.94

Table 5. Relative proportions of MNF layers used on test data in 4 boosted trees with 10 iterations for Ponderosa Pine and E. Red Cedar

MNF band	P.Pine	Cedar
1	26%	28%
11	14%	17%
17	12%	7%
2	7%	6%
13	5%	6%
3	6%	4%
9	6%	4%