Decision Tree Classification of a Forest Using AVIRIS and Multi-Seasonal TM Data

Sylvio Mannel, Dong Hua, Maribeth Price, Andreas Knobloch, Xuxia Chen
South Dakota School of Mines and Technology
501 E. St. Joseph St., Rapid City, SD 57701
Sylvio@Mannel.com

Introduction

Inventories of forests are essential for managing and protecting these natural resources. Models to determine carbon cycling, global warming, etc. also depend on accurate and efficient mapping of forested areas. This paper explores different classification methods of multi- and hyperspectral data as a supplement to traditional forest inventories. Results were evaluated towards accuracy, efficiency and costs.

The success of a classification depends on the land cover and the remotely sensed data type, but also on the classification method (Kenk et al. 1988), season (Wolter et al. 1995, Schriever and Congalton 1995), and to a large extend on the reference data (Congalton 1991). In this project we tested the importance of seasonality and the relatively new method of Decision Tree classification on TM and AVIRIS data. As a general guide for preprocessing AVIRIS data we published a white paper on the web at: http://aviris.de.vu In order to increase the number of training sites needed for best decision tree performance, we developed a unique approach to extend the usefulness of the available reference dataset.

Objectives

The main objective was to generate a land cover map of the northern Black Hills, SD. Along with this main goal we pursued the following questions: 1. How well does the new classification technique Decision Tree perform in comparison to other techniques? 2. What role does seasonality play? 3. Can TM data that combines two seasons overcome the drawbacks of multispectral data and provide similar accuracies as single-season hyperspectral data?

Methods

Study Site

The study site lays in the northern Black Hills, SD. The 'black' appearance of the Hills results from dense stands of forest vegetation, particularly ponderosa pine (*Pinus ponderosa*). This conifer covers approximately 84% of the Black Hills and comprises 95% of the forested area (Bennett 1984). Other trees in the study area include, white spruce, aspen and birch. Aspen and birch usually occur as a species association and, like white spruce, form medium to dense stands.

Reference Data Acquisition

Reference (training and test) data of high quality and quantity is the foundation for a successful and robust classification (Congalton 1991). Especially for the classification of AVIRIS imagery it is important to have a sufficient number of training and test data due to the so-called dimensionality problem: A large number of bands can decrease the accuracy without sufficient training data (Swain and Hauska 1977). We sampled species and crown cover of more than 200 plots. 30-m diameter plots were laid out with a range fork and a timber cruise prism following the techniques described in Korhonen (1979), White and Lewis (1982) and Mannel et al. (in progress). With a sighting tube we measured crown cover (as an indicator for density) and species (Ganey and Block 1994, Cook et al. 1995).

Using GIS we overlaid the field data on top of aerial Digital Ortho Quads (DOQ). On the DOQs, we identified the land cover type boundaries depicted by a field-sampled plot and picked additional points in this area. These points were labeled based on the original field plot. This procedure led to a reference data set of more than 3000 points (Figure 1).

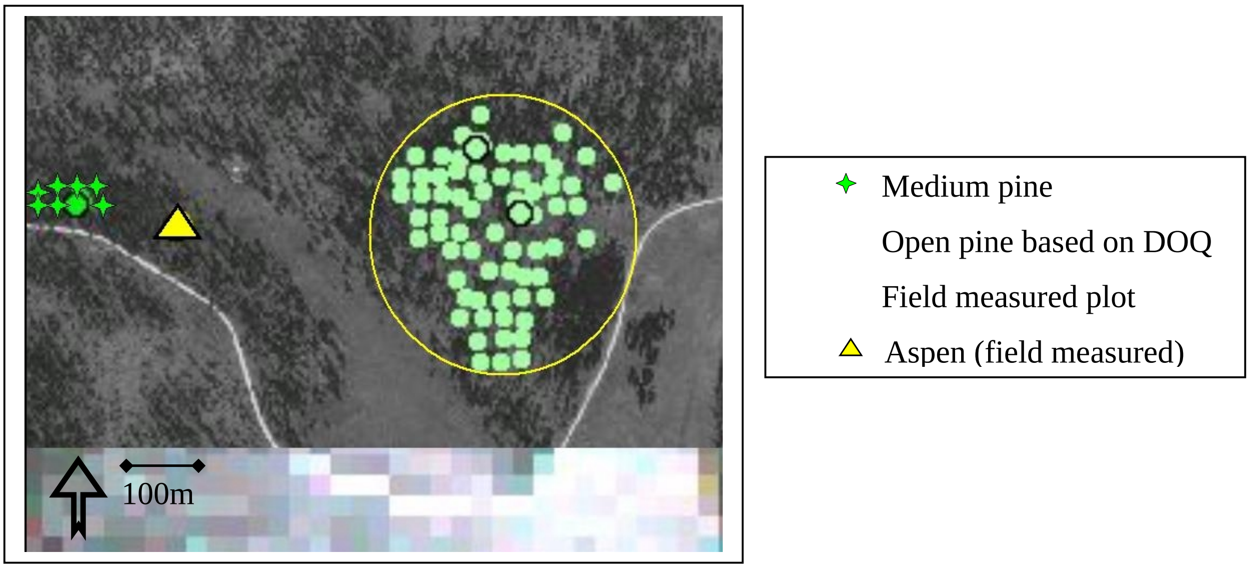

Figure 1. Additional reference points based on ground data collection and DOQ. To avoid autocorrelation, the entire cluster is either training or test data.

Densities of aspen and white spruce were combined because open white spruce and open aspen were rare (only two field plots). Mixed areas were also eliminated because mixed stands were rare and of small size, so that only an insufficient number of reference data could be generated. Furthermore, the database currently employed by the Forest Service does not include mixed classes. The final reference data set contained the following classes: open pine, medium pine, dense pine, meadow, aspen, white spruce, water and non-vegetated areas.

TM /AVIRIS

The high-altitude AVIRIS flight took place in the summer of 2000. The 7 flight lines resulted in 30 scenes. We geo-registered all scenes into the UTM Zone 13 map projection based on DOQs. We employed ACORN to correct for atmospheric absorption. As a general guide on processing AVIRIS data we published a white paper on the web at: http://aviris.de.vu. This guide explains how to prepare AVIRIS data and provides steps for atmospheric correction with ACORN and georeferencing. A good source of information is also the ACORN tutorial (Analytical Imaging and Geophysics, LLC 2001) and ENVI manual (Research Systems, Inc. 2001).

We utilized the solar bands of two Landsat TM5 images from May 5, 1998 (early spring, leaves not yet fully developed) and September 24, 1998 (early fall, “leaves on”). The images were georegistered into the UTM Zone 13 map projection using a 1:24,000 vector layer and 50 ground control points for each images. Each image was dark corrected to reduce atmospheric scatter effects. We built a "hybrid ratio composite" by adding the Normalized Difference Vegetation Index (NDVI) to the solar bands (Lillesand and Kiefer 1994). Both seasons were classified separately and as a multidate scene to test the influence of seasonality.

Decision Tree

Decision Trees for remote sensing applications were already evaluated in the 1970s (Swain and Hauska 1977). Yet, only in recent years did this method gradually emerge from business applications into natural science and provided successful land cover classifications (Hansen et al. 1996, Brodley et al. 1999, Lawrence and Wright 2001, Vogelmann et al. 2001). A decision tree is a representation of branches and nodes. Each node is connected to a set of possible answers that split the cases into subsets corresponding to different test results (Figure 2). Decision trees have similarities to other machine learning approaches. They use recursive partitioning algorithms to derive classification rules from training samples, which is often referred to as data mining (Read 2000). One of the strengths of decision trees is the flexibility in handling large datasets (De’ath and Fabricius 2000), making this approach interesting for hyperspectral data. For this study we employed the decision tree program “See5” distributed by RuleQuest (Quinlan 2002).

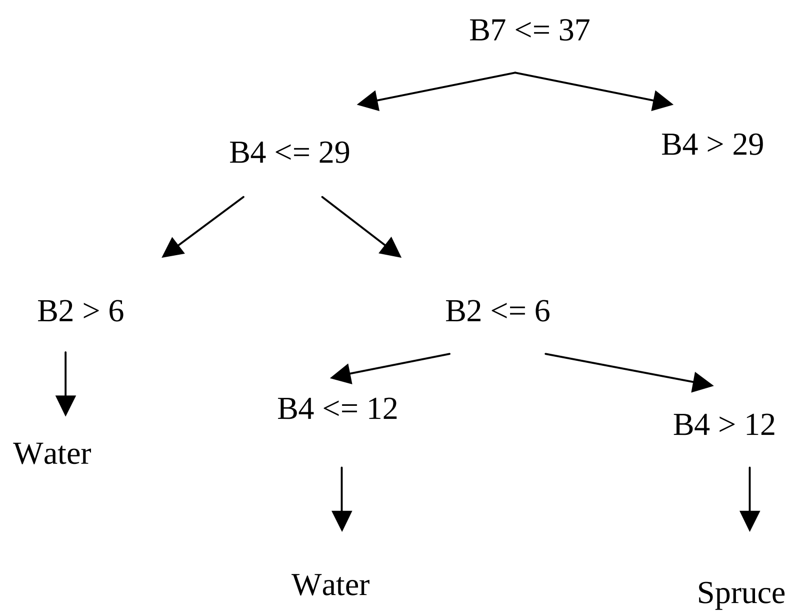

Figure 2. Example of a sub-decision tree that utilizes spring Landsat TM5 data. The tree was trained on reference data based on the summer 2000 survey.

Besides decision trees we tested Maximum Likelihood Classification on TM data and Spectral Angle Mapper (SAM) on AVIRIS data. In the case of SAM, aspen was largely overestimated and this technique was not further investigated. Mixture Tuned Matched Filtering (MTMF) is commonly applied when classifying AVIRIS data. MTMF requires a minimum noise fraction (MNF) transformed input file (Research Systems, Inc. 2001). The MNF is area-specific, i.e. it needs to be applied to the entire study area, because it gives different results for different scenes. However, this approach quickly becomes intractable when applied over a large area. Our available computer power did not permit us to combine the 20 AVIRIS scenes covering the major part of our study area. Thus MTMF was not applied in this study.

Results

Decision tree classification was successfully applied to both TM and AVIRIS data. The overall accuracy of the AVIRIS classification was 82%, and the accuracy of the multi-date TM classification was 86%. AVIRIS was superior to TM in detecting aspen (>90%), but had difficulty separating the different densities of pine (<70% accuracy for medium and dense pine). See Table 1.

Seasonality was decisive. When classifying TM data using decision trees we found that early spring, before leaves were fully developed, gave about 5% better results than using early fall ("leaf on" before senescence set in). However, best results were achieved when combining spring and fall (86% overall accuracy). Decision tree classification of multi-date TM data was superior for non-vegetated bare areas and meadow. Maximum Likelihood Classification of the multi-date TM gave better results for separating densities of pine, but had problems with white spruce and non-vegetated bare areas.

Table 1. Comparison of different classification methods and datasets

Land cover type	Decision tree AVIRIS	Decision tree multi-date TM	Maximum-likelihood multi-date TM
Water	♦♦♦	♦♦♦	♦♦♦
Aspen	♦♦♦	-	♦♦
Pine open	♦♦♦	♦	♦♦♦
Pine medium	- (confusion med/dens)	♦♦	♦♦♦
Pine dense	- (confusion med/dens)	♦♦♦	♦
Spruce	♦♦	♦♦	♦
Non-vegetated	♦♦	♦♦♦	♦
Meadow	♦♦	♦♦♦	♦♦

	Accuracy
♦♦♦	> 90%
♦♦	>80%
♦	>70%
-	<70%

Both Maximum Likelihood and Decision Tree Classification were easily implemented. Classifications based on Decision Trees were more robust than Maximum Likelihood, because the user and producer accuracy were numerically closer, which is considered a sign of robustness (Congalton 1991).

Discussion/Conclusion

The classification technique Decision Tree was easy to implement, efficient and accurate. However, the use of AVIRIS did not offer significant advantages over TM in this particular application. The overall accuracy achieved was similar for both data types. Multiple-season imagery, easily available with TM but less practical for AVIRIS, helped offset the higher spectral resolution of AVIRIS in the classification. It is possible that AVIRIS might have done better if the data were collected in late spring or early fall. According to Schriever and Congalton (1995), summer is not the best season for a forest classification. Other studies confirm that times when leaves are either senescent or not fully developed are best for forest classification. However, the time and computer processing required to classify large areas of AVIRIS proved a significant disadvantage for applications covering many AVIRIS scenes. Atmospheric correction and georeferencing was time and computer intensive—it took about 10 hours of operator time and three CDs (2GB) to prepare one scene for the actual classification; and 20 scenes were used for a total processing time of 200 hours. Moreover, the area classified constituted only a third of the entire forest. For classification work over large areas, AVIRIS does not appear to provide sufficient improvement in results over multi-date TM that would justify the amount of work and cost required. Unfortunately, the usual procedure of reducing the data dimensionality through a MNF transform cannot be applied, because the MNF must be performed on all the scenes simultaneously. This procedure quickly becomes intractable for more than a few scenes. The development of an algorithm to perform an MNF on a collection of individual scenes, rather than on a single scene, would greatly benefit the application of AVIRIS over larger areas.

Acknowledgements

We would like to acknowledge the efforts of Mark Rumble and the USDA Rocky Mountain Forest Service Research Station for providing vegetation data and support with the reference data collection, Bruce Wylie and Chengquan Huang for assistance with the decision tree analysis, and Doug Baldwin, and Krystal Price for collecting and processing the field data. This research was supported by NASA Food and Fiber Applications of Remote Sensing Program, Grant # NAG13-99021, South Dakota School of Mines and Technology, NASA-EPSCoR, and the SD Space Grant Consortium.

References

Bennett, D.L., 1984, Grazing potential of major soils within the Black Hills of South Dakota. Master of Science Thesis, South Dakota State University, Brookings, SD.

Brodley, C.E., Lane, T., and Stough, T.M., 1999, Knowledge Discovery and Data Mining. American Scientist, 87, 54-61.

Congalton, R.G., 1991, A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, 37, 35-46.

Cook, J.G., Stutzman, T.W., Bowers, C.W., Brenner, K.A., and Irwin, L.L., 1995, Spherical densiometers produce biased estimates of forest canopy cover. Wildlife Society Bulletin, 23, 711-717.

De'Ath, G., and Fabricius, K.E., 2000, Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, 81, 3178-3192.

Ganey, J.L., and Block, W.M., 1994, A comparison of two techniques for measuring canopy cover. Western Journal for Applied Forestry, 9, 21-23.

Hansen, M., Dubayah, R., and Defries, R., 1996, Classification trees: an alternative to traditional land cover classifiers. International Journal of Remote Sensing, 17, 1075-1081.

Kenk, E., Sondheim, M., and Yee, B., 1988, Methods for improving accuracy of Thematic Mapper ground cover classification. Canadian Journal of Remote Sensing, 14, 17-31.

Korhonen, J., 1979, One-person circular plot establishment. The Forestry Chronicle, 55, 51-53.

Lawrence, R.L., and Wright, A., 2001, Rule-Based Classification systems using classification and regression tree (Cart) analysis. Photogrammetric Engineering & Remote Sensing, 67, 1137-1142.

Lillesand, T.M., and Kiefer, R.W., 1994, Remote sensing and image interpretation (New York, Chichester, Brisbane, Toronto, Singapore: John Wiley & Sons, Inc.

Mannel, S., Price, M., Rumble, M., Baldwin, D., and Hua, D., in review, Using a variable-plot technique for establishing fixed-radius circular plots.

Quinlan, R., 2002, See5/C5.0. RuleQuest Research Pty Ltd, St Ives NSW 2075, Australia.

Read, B.J., 2000, Data mining and science?, CLRC Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, UK, http://www.ercim.org/publication/ws-proceedings/12th-EDRG/EDRG12_Re.pdf.

Research Systems, Inc., 2001, ENVI User's Guide, 948 pp.

Schriever, J.R., and Congalton, R.G., 1995, Evaluating seasonal variability as an aid to cover-type mapping from Landsat Thematic Mapper data in the northeast. Photogrammetric Engineering & Remote Sensing, 61, 321-327.

Swain, P.H., and Hauska, H., 1977, The decision tree classifier: design and potential. GE-15, 142-147.

Vogelmann, J.E. et al., 2001, Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper and ancillary data sources. Photogrammetric Engineering & Remote Sensing, 67, 650-662.

White, W.E., and Lewis, C.E., 1982, Establishing circular plot boundaries with a wedge prism and an adjustable target pole. Journal of Range Management, 35, 677-678.

Wolter, P.T., Mladenoff, D.J., Host, G.E., and Crow, T.R., 1995, Improved forest classification in the northern lake states using multi-temporal Landsat imagery. Photogrammetric Engineering & Remote Sensing, 61, 1129-1143.