Determining Forest Species Composition Using High Spectral Resolution Remote Sensing Data

M. E. Martin, S. D. Newman,* J. D. Aber,* and R. G. Congalton†*

Airborne hyperspectral data were analyzed for the classification of 11 forest cover types, including pure and mixed stands of deciduous and conifer species. Selected bands from first difference reflectance spectra were used to determine cover type at the Harvard Forest using a maximum likelihood algorithm assigning all pixels in the image into one of the 11 categories. This approach combines species specific chemical characteristics and previously derived relationships between hyperspectral data and foliar chemistry. Field data utilized for validation of the classification included both a stand-level survey of stem diameter, and field measurements of plot level foliar biomass. A random selection of validation pixels yielded an overall classification accuracy of 75%. ©Elsevier Science Inc., 1998

INTRODUCTION

The classification of forest cover is an important element in both forest resource management and scientific research issues. The methods used to determine forest species composition vary with both the requirements for a specific application and the cost and availability of data. Existing methods for estimating species distribution include field surveys, photographic interpretation, and, more recently, digital remote sensing. Field measurements of species composition are time-consuming and cannot provide complete coverage of large areas. In contrast, remote sensing instruments provide spectral data for large contiguous areas from several to hundreds of square kilometers per scene. The development of species classification methods utilizing remotely sensed data may be the most effective approach for obtaining assessments of forest cover over large areas at a fine-scale spatial resolution.

Remotely sensed spectral data has been used to identify broad categories of forest cover, for example, conifer versus deciduous stands (Nelson et al., 1985; Shen et al., 1985; Hodgson et al., 1988; Lathrop et al., 1994). Broad band instruments used for land cover identification include the Landsat Thematic Mapper (TM) and Multispectral Scanner (MSS), the airborne Thematic Mapper Simulator (TMS), SPOT HRV, and the Advanced Very High Resolution Radiometer (AVHRR). A number of studies have used these same instruments to classify forest type at a more detailed species resolution with varying degrees of success (Skidmore, 1989; Schriever and Congalton, 1995; Frank, 1988; White et al., 1995; Franklin, 1994). Species classifications have also been made with an airborne multispectral scanner (Rohde and Olson, 1972), and video imagery (Everitt et al., 1987; Thomasson et al., 1994). Within the last decade, a number of instruments with higher spatial and spectral resolution have been developed. These instruments may provide the spectral resolution necessary to improve upon existing classification methods. In addition to broad band data analysis techniques, high spectral resolution instruments produce data suitable for a number of full-spectrum analysis techniques.

The primary goal of this research project was to investigate the application of high spectral resolution remote sensing imagery of the forest canopy to the identification of species composition, using spectral regions correlated with the chemical composition of foliage. In

* Complex Systems Research Center, University of New Hampshire, Durham

† Department of Natural Resources, University of New Hampshire, Durham

Address correspondence to M. E. Martin, Complex Systems Research Ctr., Univ. Of New Hampshire, Durham, NH 03824. E-mail: mem@unh.edu

Received 10 June 1996; revised 10 April 1998.

655 Avenue of the Americas, New York, NY 10010

0034-4257/98/$19.00
PII S0034-4257(98)00035-2

previous work, Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data has been used to identify the nitrogen and lignin concentration in forest canopy foliage (Martin and Aber, 1997b). Foliar nitrogen concentration, mapped with AVIRIS data, was used as a primary driver in a model of ecosystem carbon and water balance (Aber and Federer, 1992; Martin and Aber, 1997a). More accurate model predictions will be realized with the additional development of species classification maps from AVIRIS data. In this paper we use the relationship between AVIRIS spectral data and foliar chemistry, in addition to species specific chemical characteristics, as a basis for classifying species composition.

Table 1. Stand Classification Criteria

Stand Type	*Classification Criteria*
0 Open	No trees
1 Red maple	≥80% red maple
2 Red oak	≥80% red oak
3 Mixed deciduous	≥80% mixed deciduous
4 Deciduous/conifer mix	21–79% deciduous
5 White pine	≥80% while pine
6 Hemlock	≥70% hemlock
7 Mixed conifer	≥80% mixed conifer
8 Norway spruce	≥80% Norway spruce
9 Red pine	≥80% red pine
10 Spruce bog	Black spruce wetland
11 Deciduous bog	Mixed deciduous wetland

METHODS

Study Site

The study site for this project is the Prospect Hill tract at the Harvard Forest, in central Massachusetts (Latitude 42°32′N Longitude 72°11′W). This 400 ha research site contains a combination of natural deciduous, mixed deciduous/conifer, hemlock, and white pine stands as well as red pine and Norway spruce plantations.

Foliage Collection

Leaf samples were collected at the Harvard Forest in June 1992 in support of NASA's Accelerated Canopy Chemistry Program (ACCP, documentation and data online at http://www-eosdis.ornl.gov) directed at the study of remotely sensing canopy chemistry using AVIRIS data (Martin and Aber, 1997b). Leaf samples were collected from dominant species on each of 20 50 m×50 m plots. Laboratory analysis of the samples yielded data on nitrogen, lignin, cellulose, and water content. Lignin and cellulose concentrations were determined by a sulfuric acid digest method (TAPPI, 1975; 1976; Effland, 1977; McClaugherty et al., 1985; Newman et al., 1995). Nitrogen concentrations were determined using the CHN combustion method (Page et al., 1982).

AVIRIS Data

Image data were acquired using NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) on 15 June 1992. AVIRIS records data in 224 contiguous spectral bands covering the spectral range of 0.4–2.4 μm with a spectral resolution of 10 nm. The spatial resolution of AVIRIS data is 20 m with a full scene covering 10 km× 10 km (Vane and Goetz, 1988).

Atmospheric corrections of the AVIRIS data were made using the *AT*mosphere *REM*oval Program (ATREM) (Gao et al., 1991; 1992). This program uses information in each AVIRIS radiance spectra to parameterize a radiative transfer model that is then used to convert at-sensor radiance to ground-level reflectance by removing the effects of atmospheric gases, water vapor, and aerosols. Following the ATREM correction, a secondary correction was made, based on differences between paired field measured spectra and the corresponding pixels from an ATREM corrected scene also acquired in 1992 (Clark et al., 1993).

Field Measurement of Species

A geographic information system (GIS) database of species type was obtained from a 1986–1988 stand survey of the study site (http://LTERnet.edu/hfr). In this survey, 252 stands within the study site were identified from aerial photographs. Basal area by species was measured for each stand using a minimum of three variable radius plots located along a transect within each stand. From these data, the relative basal area for each species within each stand was calculated. Based on the relative basal area of each species, each stand was assigned to one of 11 forested categories (Table 1). These categories include stands of pure conifer species (red pine, Norway spruce, and white pine). The majority of deciduous stands are a mix of three or more species. We have included several stands where red maple or red oak basal area is ≥80%; however, there are not enough pixels in these two categories to properly evaluate their classification.

Stand data from the Harvard Forest GIS database were geometrically registered to the AVIRIS image using global positioning system (GPS) data for the study site and the surrounding area, with a root mean square registration error of approximately 1 pixel.

Canopy Chemistry Analysis

In a previous study, multiple linear regression analysis was used to select AVIRIS bands which were closely correlated with field measured canopy nitrogen and lignin concentration for 40 plots located at both Harvard Forest and a second research site at Blackhawk Island, WI (Martin and Aber, 1997b). These 40 plots were located in mature stands with closed canopies and a leaf area index (LAI) of three or greater. For both foliar chemistry and species classification, the AVIRIS reflectance spectra were converted to first difference spectra prior to analysis. The bands selected in this analysis correspond primarily to overtones of mid-infrared absorption by molecules within the chemical constituents and are located in both the visible and near-infrared (NIR) regions of the spectra. AVIRIS channels used in the nitrogen and lignin classifications fell into three spectral regions: 620–820 nm, 1640–1740 nm, and 2140–2280 nm.

AVIRIS Species Classification

A total of 11 AVIRIS bands were evaluated for the species classification. These bands had been previously used to determine canopy nitrogen and lignin concentration for the study site (Martin and Aber, 1993; 1997b). A supervised classification was done in which polygons from each of the 11 categories were used to extract spectral signatures for each class (5–13 polygons per class, 4–26 pixels per polygon). These supervised classification training pixels were identified using the stand map generated from 1988 field survey data.

Transformed divergence values (ERDAS, 1992) were calculated for all combinations of 4–11 bands to determine which band combinations would provide the best separability of signature classes. A maximum likelihood algorithm was used with a first pass parallelepiped classification to assign all pixels in the image into one of the 11 signature classes using the best band subsets. The parallelepiped classification used maximum and minimum values for each signature to determine class assignments (ERDAS, 1992).

RESULTS AND DISCUSSION

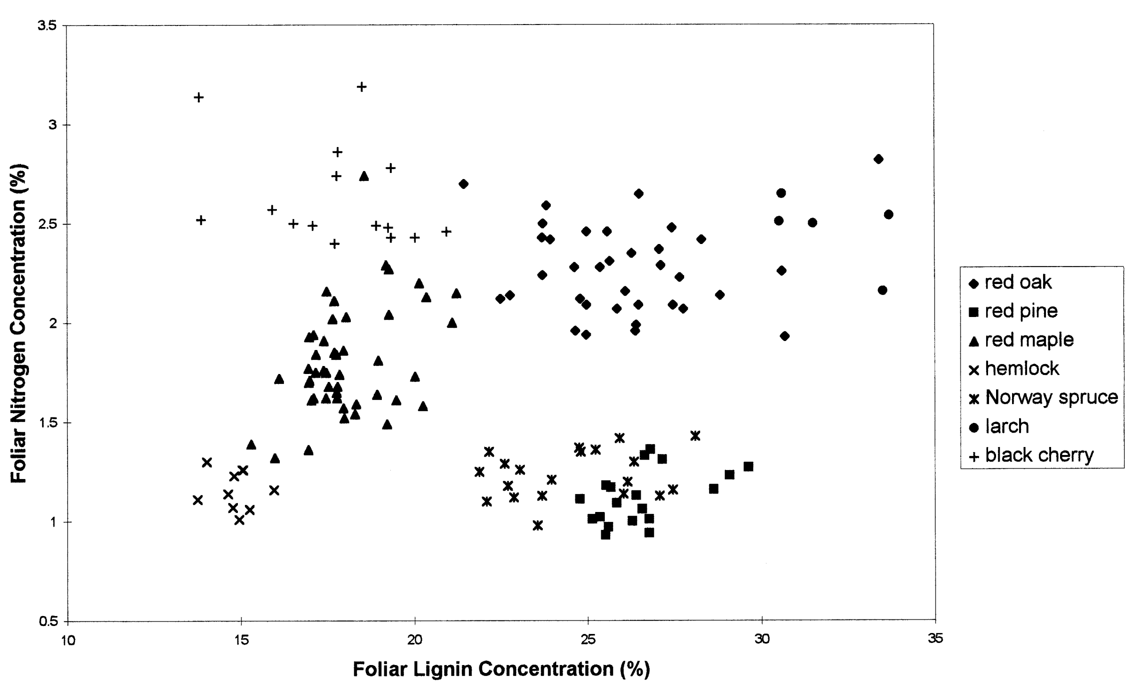
Foliar Chemistry and Species Identification

Nitrogen and lignin concentrations for foliage of different species show that neither lignin nor nitrogen concentration alone is sufficient to identify samples by species. However, these data indicate that species identification can be made on the basis of the combined foliar nitrogen *and* lignin information (Fig. 1). For example, red pine and hemlock have similar nitrogen concentrations but very different lignin concentrations [red pine: 1.11 (±0.13) %N, 26.54 (±1.29) %lignin; hemlock 1.16 (±0.10) %N, 14.82 (±0.62) %lignin], whereas red maple and black cherry have similar lignin concentrations and different nitrogen concentrations [red maple: 1.8 (±0.27) %N, 18.04 (±1.27) %lignin; black cherry: 2.62 (±0.26) %N, 17.79 (±2.08) %lignin].

AVIRIS Classification of Species

An evaluation of classifications using increasing numbers of bands showed an improvement in classification accuracy with up to nine bands. The species map generated from the AVIRIS classification (Fig. 2), uses nine bands centered at the following wavelengths: 750 nm, 783 nm, 1641 nm, 2140 nm, 2290 nm, 627 nm, 822 nm, 1660 nm, and 2280 nm. These bands are a subset of the bands which were statistically important in the calibration of AVIRIS spectra in determining plot-level foliar nitrogen and lignin concentration at Harvard Forest and Blackhawk Island (Martin and Aber, 1997b). Samples were selected from the 3×3 majority-smoothed classified image

Figure 1. Harvard Forest leaf samples: foliar nitrogen vs. foliar lignin.



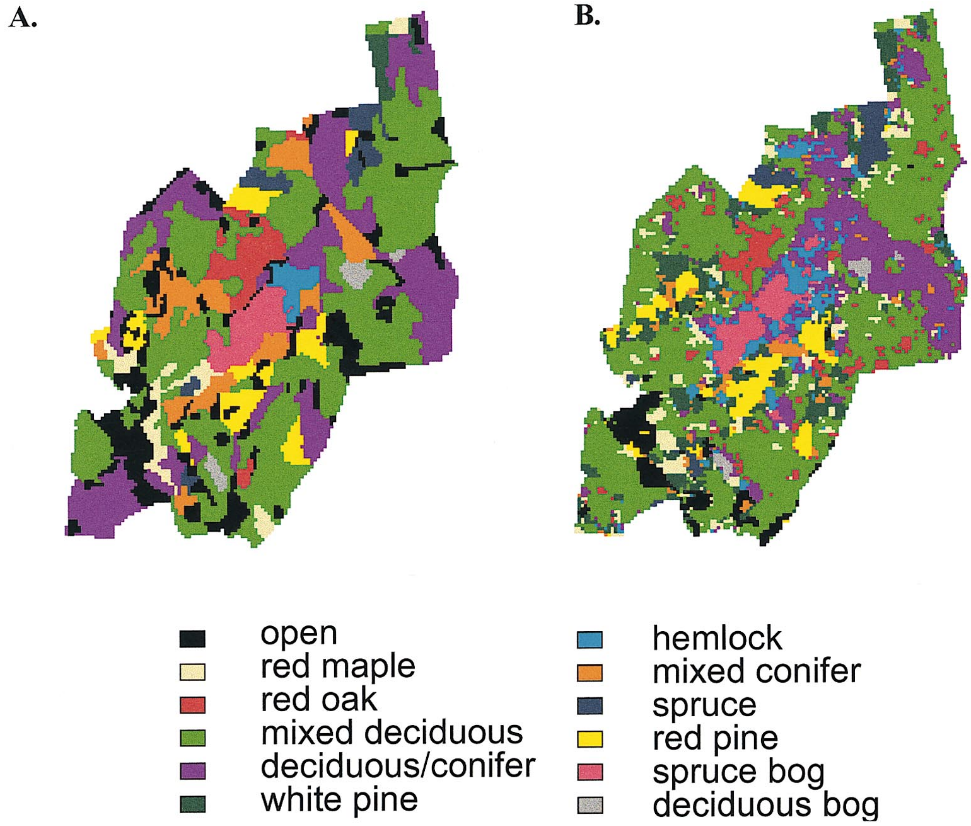


Figure 2. Harvard Forest: a) species stand classification determined from field measurements of basal area; b) species classification derived from AVIRIS data.

to assess the accuracy of the classification algorithm. Validation samples were randomly selected from the classified polygons, avoiding edge pixels, with the number of samples per class relative to the total area of the class. An error matrix was calculated by comparing the field measured data with the AVIRIS classification for these samples (Table 2). The boldface diagonal values indicate the number of samples in each class which were classified correctly with AVIRIS data. Producer's and user's accuracy and the kappa statistic are calculated from this table (Congalton, 1991). The overall accuracy of the classification was 75%, with 123 out of 164 samples correctly classified, with a kappa statistic of 0.68.

Due to the species composition of the study site, it was impossible to thoroughly investigate the classification of individual deciduous species—almost all of the deciduous stands in this study area are a mix of three or more species. A red oak and red maple class were included separately in the analysis, but the sample size is too small for proper evaluation. All other stands containing deciduous species were classified as either mixed deciduous or mixed deciduous–conifer. In six of 13 "misclassified" mixed deciduous stands, the AVIRIS classification placed these samples into either red oak or red maple. Thirty percent of the mixed conifer stands were classified correctly, with the remainder classified as either deciduous/conifer mix, white pine, or hemlock. The only "pure" species stands analyzed were conifer stands of hemlock, Norway spruce, white pine, red pine, and black spruce bog. In the random samples chosen for our accuracy assessment, these species were correctly classified by AVIRIS data in 75–100% of the samples.

It is assumed in this analysis that relative foliar biomass is proportional to relative basal area. It is likely that this is not true in the mixed deciduous/conifer stands. We know from field sampling of litterfall that a red pine plot with a basal area of 53 m^2 ha^{-1} has a foliar biomass of 8488 g ha^{-1}, whereas a deciduous site with a basal area of 30 m^2 ha^{-1} has a foliar biomass of 2540 g ha^{-1}. In other words, there is nearly twice the foliar biomass per square meter of basal area in the red pine site as in the deciduous site. Since the remotely sensed reflectance spectra is influenced primarily by the foliar biomass, the use of this stand survey data may not be suitable for mixed deciduous/conifer stands.

The AVIRIS predicted species was also compared to a number of plots for which the canopy biomass of all species had been measured directly by litterfall collection. Ten litter baskets were randomly placed in each plot in early September, and the collection period covered 1 year to capture both fall and spring litterfall of conifer species. Deciduous litter weights represented the foliar bio-

Table 2. Accuracy Assessment of Randomly Selected Pixels.[a] Boldface values indicate the number of correctly classified samples.

Class	*Field Survey Data* 1	2	3	4	5	6	7	8	9	10	11	*Total*	*User's Accuracy*
AVIRIS Classification													
1	**3**	–	5	1	–	–	–	–	–	–	–	9	33.33
2	–	**8**	1	–	–	–	–	–	–	–	–	9	88.88
3	1	2	**54**	7	–	–	–	–	–	–	–	64	84.44
4	–	–	4	**12**	–	–	4	–	–	–	–	20	60.00
5	–	1	3	3	**7**	–	4	–	–	–	–	18	38.88
6	–	–	–	–	–	**1**	2	–	–	3	–	6	16.66
7	–	–	–	–	–	–	**5**	–	–	–	–	5	100.0
8	–	–	–	–	–	–	–	**7**	–	–	–	7	100.0
9	–	–	–	–	–	–	–	–	**13**	–	–	13	100.0
10	–	–	–	–	–	–	–	–	–	**9**	–	9	100.0
11	–	–	–	–	–	–	–	–	–	–	**4**	4	100.0
Total	4	11	67	23	7	1	15	7	13	12	4	**164**	
Producer's accuracy	75.0	72.73	80.6	52.17	100.0	100.0	33.33	100.0	100.0	75.0	100.0	–	75

[a]Categories 1–11 as described in Table 1.

mass of the deciduous species; conifer litter weights were multiplied by foliar retention time (1 year of litterfall only represents a portion of the canopy biomass for a conifer species). These measurements were made on 38 plots during 1992 and 1993 within this study site, and represent six of the 11 classes described in Table 1. The number of these plots correctly classified with AVIRIS data is as follows: hemlock—1/1, Norway spruce—4/4, mixed deciduous—24/27, and red pine—2/4, totaling 31 out of 38 correct. The three mixed deciduous plots misclassified were assigned to the red maple and deciduous/conifer mix classes. Two red pine plots were misclassified as mixed deciduous and red oak; however, these plots were each located within 1–2 pixels of red pine classified pixels, within the misregistration of the image and GIS data layers.

The overall appearance of the AVIRIS classified map shows more spatial heterogeneity than the field classified map. The field data are based on as few as three measurements within each stand polygon delineated by photo interpretation (with some stands containing several hundred 20 m×20 m pixels). It is apparent that small scale spatial variation which might be missed with this type of field survey could be measured by remote sensing data in which spectral data are available for every pixel.

CONCLUSIONS

In this work we have used the relationship between foliar chemistry and species identification to demonstrate that high spectral resolution remote sensing data can be used to classify forest species. Remote sensing of canopy chemistry has been possible only in the recent past with the availability of high spectral resolution instruments such as AVIRIS. Additional work must be done to fully explore the potential of high spectral resolution data in determining forest species composition. Selection of signature training sites based on field measured canopy composition should result in a more accurate classification, particularly in deciduous/conifer stands. Field data collection protocols have been developed in support of this ongoing effort, and datasets are currently being developed for further analysis of AVIRIS data. Improvements may also be made in the classification of deciduous/conifer mixed stands by first using leaf on/leaf off data to determine the foliar biomass proportion of each type before attempting species classification.

REFERENCES

Aber, J. D., and Federer, C. A. (1992), A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. *Oecologia* 92:463–474.

Clark, R. N., Swayze, G., Heidebrecht, K., Goetz, A. F. H., and Green, R. O. (1993), Comparison of methods for calibrating AVIRIS data to ground reflectance. In *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop: AVIRIS* (R. O. Green, Ed.), Jet Propulsion Laboratory, Pasadena, CA, Vol. 1, pp. 35–36.

Congalton, R. G. (1991), A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37:35–46.

Effland, M. J. (1977), Modified procedure to determine acid insoluble lignin in wood and pulp. *Tappi* 60:143–144.

ERDAS (1992), *ERDAS Field Guide*, ERDAS, Inc, Atlanta, GA.

Everitt, J. H., Escobar, D. E., Alaniz, M. A., and Davis, M. R. (1987), Using airborne middle-infrared (1.45–2.0 μm) video imagery for distinguishing plant species and soil conditions. *Remote Sens. Environ.* 22:423–428.

Frank, T. D. (1988), Mapping dominant vegetation communities in the Colorado Rocky Mountain Front Range with Landsat Thematic Mapper and digital terrain data. *Photogramm. Eng. Remote Sens.* 54:1727–1734.

Franklin, S. E. (1994), Discrimination of subalpine forest species and canopy density using CASI, SPOT, PLA, and Landsat TM data. *Photogramm. Eng. Remote Sens.* 60:1233–1241.

Gao, B., Goetz, A. F. H., and Zamudio, J. A. (1991), Removing atmospheric effects from AVIRIS data for surface reflectance retrievals. In *Proceedings of the Third Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop* (R. O. Green, Eds.), Jet Propulsion Laboratory, Pasadena, CA, pp. 80–86.

Gao, B., Heidebrecht, K. B., and Goetz, A. F. H. (1992), *Atmosphere Removal Program (ATREM) User's Guide*, Center for the Study of Earth from Space/CIRES, University of Colorado, Boulder.

Hodgson, M. E., Jensen, J. R., Mackey, J. E., Jr, and Coulter, M. C. (1988), Monitoring wood stork foraging habitat using remote sensing and geographic information systems. *Photogramm. Eng. Remote Sens.* 54:1601–1607.

Lathrop, R. G., Aber, J. D., Bognar, J. A., Ollinger, S. V., Casset, S., and Ellis, J. M. (1994), GIS development to support regional simulation modeling of northeastern (USA) forest ecosystems. In *Environmental Information Management and Analysis* (W. Michener, J. W. Brunt, and S. Stafford, Eds.), Taylor and Francis, London, pp. 431–451.

Martin, M. E., and Aber, J. D. (1993), Measurements of canopy chemistry with 1992 AVIRIS data at Blackhawk Island and Harvard Forest. In *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop: AVIRIS* (R. O. Green, Ed.), Jet Propulsion Laboratory, Pasadena, CA, Vol. 1, pp. 113–116.

Martin, M. E., and Aber, J. D. (1997a), Estimating canopy characteristics as inputs for models of forest carbon exchange by high spectral resolution remote sensing. In *The Use of Remote Sensing in the Modeling of Forest Productivity* (H. L. Gholz, K. Nakane, and H. Shimoda, Eds.), Kluwer Academic, Dordrecht.

Martin, M. E., and Aber, J. D. (1997b), Estimation of forest canopy lignin and nitrogen concentration and ecosystem processes by high spectral resolution remote sensing. *Eco. Appl.* 7:431–443.

McClaugherty, C. A., Pastor, J., Aber, J. D., and Melillo, J. M. (1985), Forest litter decomposition in relation to soil nitrogen dynamics and litter quality. *Ecology* 66:266–275.

Nelson, R. F., Latty, R. S., and Mott, G. (1985), Classifying northern forests using Thematic Mapper Simulator data. *Photogramm. Eng. Remote Sens.* 50:607–617.

Newman, S. D., Soulia, M. E., Aber, J. D., Dewey, B., and Ricca, A. (1995), Analysis of forest foliage I: Proximate carbon fraction and nitrogen analyses for the Accelerated Canopy Chemistry Program: Methods and quality control. *J. Near Infrared Spectrosc.* 2:5–14.

Page, A. L., Miller, R. H., and Keeney, D., Eds. (1982), *Methods of Soil Analysis*, American Society of Agronomy, Madison, WI, Vol. 2.

Rohde, W. G., and Olson, C. E., Jr. (1972), Multispectral sensing of forest tree species. *Photogramm. Eng. Remote Sens.* 38:1209–1215.

Schiever, 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. *Photogramm. Eng. Remote Sens.* 61:321–327.

Shen, S. S., Badhwar, G. D., and Carnes, J. G. (1985), Separability of boreal forest species in the Lake Jennette area, Minnesota. *Photogramm. Eng. Remote Sens.* 51:1775–1783.

Skidmore, A. K. (1989), An expert system classifies eucalypt forest types using Thematic Mapper data and a digital terrain model. *Photogramm. Eng. Remote Sens.* 55:1449–1464.

TAPPI (1975), *Water Solubles in Wood and Pulp*, Technical Association of the Pulp and Paper Industry, Atlanta, GA.

TAPPI (1976), *Alcohol–Benzene and Dichloromethane Soluble in Wood and Pulp*, Technical Association of the Pulp and Paper Industry, Atlanta, GA.

Thomasson, J. A., Bennett, C. W., Jackson, B. D., and Mailander, M. P. (1994), Differentiating bottomland tree species with multispectral videography. *Photogramm. Eng. Remote Sens.* 60:55–59.

Vane, G., and Goetz, A. F. H. (1988), Terrestrial imaging spectroscopy. *Remote Sens. Environ* 24:1–29.

White, J. D., Kroh, G. C., and Pinder, J. E. (1995), Forest mapping at Lassen Volcanic National Park, California, using Landsat TM data and a geographical information system. *Photogramm. Eng. Remote Sens.* 61:299–305.