Rapid Classification of Landsat TM Imagery for Phase 1 Stratification Using the Automated NDVI Threshold Supervised Classification (ANTSC) Methodology

William H. Cooke and Dennis M. Jacobs

Abstract.—FIA annual inventories require rapid updating of pixel-based Phase 1 estimates. Scientists at the Southern Research Station are developing an automated methodology that uses a Normalized Difference Vegetation Index (NDVI) for identifying and eliminating problem FIA plots from the analysis. Problem plots are those that have questionable land use/land cover information. Four Landsat TM scenes in Georgia have been classified using this methodology. A cross-validation approach was used to assess accuracy. The results are compared with an alternative methodology: the Iterative Guided Spectral Class Rejection (IGSCR) methodology.

Several FIA units have examined methodologies that test the usefulness of pixel-based estimates for Phase 1 stratification. Among these are k-Nearest Neighbor (k-NN) (Franco-Lopez *et al.* 2000), Iterative Guided Spectral Class Rejection (IGSCR) (Wayman *et al.* 2001) and various model-based approaches (Moisen *et al.* 1998). A new methodology developed by scientists at the USDA Forest Service Southern Research Station seeks to combine simple concepts of satellite image data classification with FIA plot data and automate the process. This new methodology compares FIA plot information with spectral information from an NDVI transform, using an automated approach for choosing Euclidean distances used to generate FIA plot-based classification "signatures." An additional component of this methodology was tested that examines crown modeling quantitatively to assess the usefulness of FIA plots for generating signatures over the portion of the NDVI range (150–185) that is most problematic for distinguishing forest from nonforest pixels. The result of these comparisons is the development of efficient Phase 1 classification techniques that meet FIA remote sensing business requirements.

Operational Efficiencies

The Southern Research Station inventories forests in 13 Southern States and requires approximately 131 TM scenes for complete "wall-to-wall" coverage of all States. Phase 1 stratification procedures need to keep pace with changes in forest conditions in the South and with the pace of inventory reporting cycles that require re-measuring all FIA ground plots every 5 years. The rate of change of southern forests is rapid and subject to environmental, social, and economic forces including:

- Clearcutting
- Urbanization
- Landowner assistance programs
- Population shifts

Any classification methodology adopted for FIA should be operationally efficient for FIA purposes and address the following requirements:

- High automation potential
- Straightforward implementation
- High CPU and storage efficiencies
- High repeatability

To date, the various Phase 1 methodologies that have been proposed and tested have failed to meet one or more of these requirements. For example, the IGSCR methodology requires a great deal of subjective interpretation to establish signatures and the iterative nature of the classification requires a great deal of storage space.

Figure 1 indicates the study area for the ANTSC methodology test project. Figure 2 indicates the subset of the study area used for examining crown modeling approaches aimed at refining the NDVI threshold component of the ANTSC methodology. Comparison of the results of the ANTSC methodology with the IGSCR methodology requires examining both methodologies in more detail.

Figure 1.—*Study area for ANTSC methodology test project.*

Figure 2.—*Subset of the study area used for tests of crown modeling.*

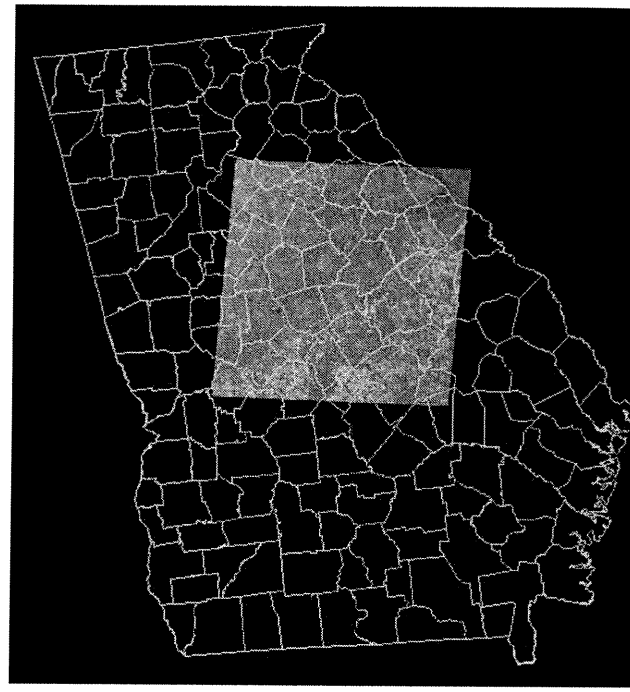
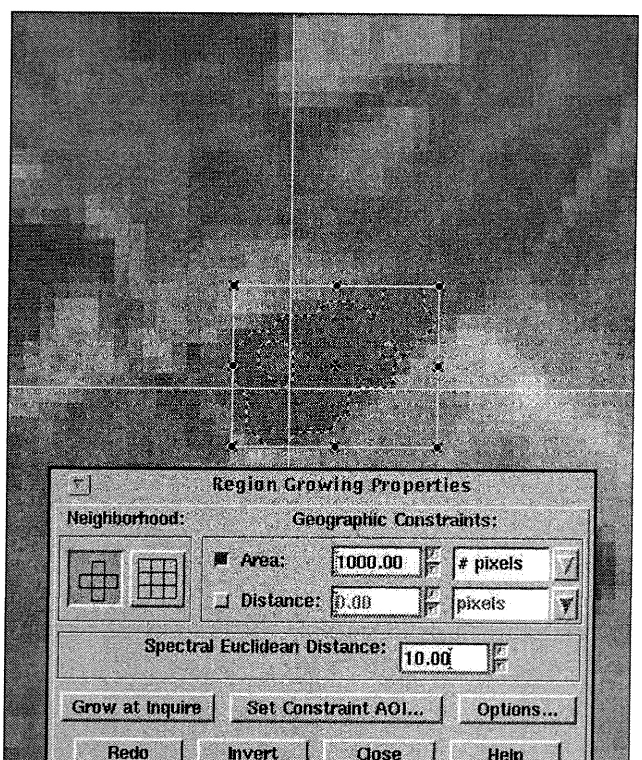


Figure 3.—*Region that was grown using a Euclidean distance of 10.*



IGSCR Methodology

The IGSCR methodology uses FIA plot information for developing statistical signatures. These signatures consist of the mean and variance of the spectral reflectance of the ground conditions in several Landsat TM spectral channels. The analyst views the location of the FIA plot on the image and, at that spot, chooses a pixel (seed) for the signature growing process. Using the pixel collocated at the FIA plot position, the analyst specifies a Euclidean distance in multi-spectral space that captures contiguous pixels to be accepted, if within the Euclidian distance of the same land use condition. Pixels outside the distance are rejected as the same land use condition. The analyst must be able to recognize whether the region included in the signature growing process remains in the land use condition of seed pixel initiation. Figure 3 indicates a region that was grown using a Euclidean distance of 10. The analyst must adjust the Euclidean distance to ensure that the signature does not grow beyond the land use class of initiation, so must frequently zoom in and out of the image to subjectively assess the results of the seed-growing process.

Table 1.—*Georgia IGSCR/ANTSC accuracy assessment comparisons*

TM scene path/row	Forest				Nonforest		
	Overall accuracy	Producers accuracy	Users accuracy	Kappa statistic	Producers accuracy	Users accuracy	Kappa statistic
IGSCR							
17/37	84.79	89.67	88.71	0.6241	73.42	75.32	0.6473
17/38	85.38	92.44	89.08	0.5516	65.28	75.20	0.6649
18/37	84.93	92.17	89.08	0.4855	58.06	66.67	0.5768
18/38	86.71	98.22	83.73	0.5519	66.52	95.51	0.9296
ANTSC							
17/37	90.01	91.03	97.26	0.9884	90.48	73.08	0.7498
18/37	95.28	95.11	99.43	0.9570	96.43	75.00	0.7120
18/38	95.01	95.52	99.58	0.9884	99.32	88.48	0.8182

The IGSCR process is detailed in Wayman (2001). To begin the IGSCR classification process, an unsupervised classification of 100 classes using a convergence threshold of 0.95 and variance set to one standard deviation was performed for each TM image. Collected signatures were then used to extract the class values that result from the classification process, and output those class pixel values to a text file suitable for statistical analyses. The class information was analyzed for purity (95 percent) and classes deemed pure were removed (masked) from the original TM imagery. The remaining image pixels were then separated into 100 classes for the second iteration of class purity testing. At least three iterations were performed for each image.

Table 1 lists the accuracies obtained for each of the four TM scenes that were classified using the IGSCR methodology. The methodology was relatively accurate for the binary classification of the forest and nonforest conditions, but required significant analyst time and effort for choosing Euclidean distances in the signature collection process. The multiple classifications of the imagery required by IGSCR occupied a lot of storage space. These shortcomings of the methodology prompted the development of a hybrid classification approach combining NDVI-based techniques (Hoppus *et al.* 2000) with the Euclidean distance signature development component of the IGSCR methodology.

ANTSC Methodology

The IGSCR subjective signature generation process relies on visual interpretation of forest and nonforest cover types. Familiarity with the landscape and ecosystem processes is a prerequisite for accurate image classification. At present, the signature collection process is time consuming and tedious, and interpreter fatigue is a real problem.

Euclidean Distance Component

Signature collection in support of the IGSCR methodology resulted in the visual interpretation of over 1,200 signatures for four TM images from 1992 and four TM images from 2000. These results suggested that a Euclidean distance of 13 optimized signature growth for forested conditions but rarely caused the signature to grow out of the condition of seed pixel initiation. A Euclidean distance (D) of 21 gave similar results in nonforest conditions.

Euclidean distance, D:

$$D_{ab} = \left[\sum_{i=1}^{n} (a_i - b_i)^2 \right]^{.5}$$

Where a and b are values of pixels being evaluated and n is the total number of satellite layers.

NDVI Component

A large body of literature exists confirming the usefulness of the Normalized Difference Vegetation Index (NDVI) band transformation for extracting information about forest vegetation (Iverson *et al.* 1989, Anderson *et al.* 1993). Results using an NDVI threshold by the Northeast FIA unit confirmed that NDVI was useful for separating forest from nonforest conditions. Figure 4 illustrates how the NDVI values for FIA plots (subplot 1) compare for a single TM scene.

The search for operationally efficient automated classification methodologies led researchers at the Southern Research Station (SRS) to develop an integrated methodology that used an NDVI threshold with automated signature collection to rapidly classify TM images using a Maximum Likelihood-based "Supervised Classification" approach, dubbed the Automated NDVI Threshold Supervised Classification (ANTSC) method.

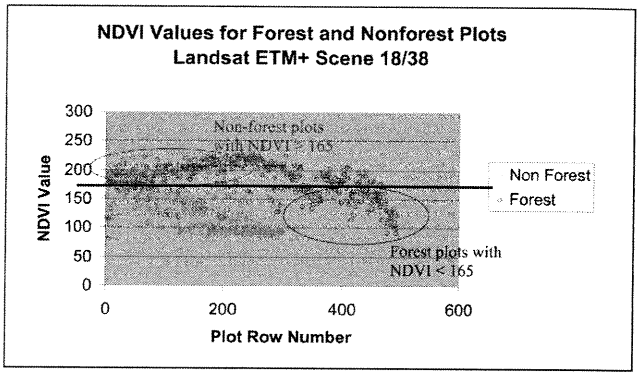
An NDVI threshold of 165 was used to differentiate between forest and nonforest. Each FIA plot's NDVI value was extracted from an NDVI-transformed TM image using a Pixel-to-ASCII extraction program. The NDVI values were compared to the field-derived land use information. Forested plots with NDVI values below 166 and nonforest plots with NDVI values above 165 were considered separate populations of plots that did not represent land cover information contained with the spectral response surface of Landsat TM imagery. Several explanations for the origination of this population of plots may be hypothesized. The following are possible:

- Change based on disturbance
- Land use versus land cover differences (clearcut = forest)
- Pixel/plot mis-registration

It was considered important to the ANTSC process that this NDVI or parity test be conducted to remove these plots from training and accuracy assessment. Certainly, the removal of these plots purifies the training and accuracy assessment pool of plots used in the cross-validation approach. The IGSCR methodology also indirectly purifies the accuracy assessment pool of plots by removing those plots that resulted in poor signatures during the signature generation process. A poor signature was one that did not include a minimum of 9 pixels, or one that grew into a land cover class different from that of the original pixel.

Accuracies for three TM scenes classified using the ANTSC methodology are shown in table 1. A final accuracy assessment test was performed using the accuracy assessment plots from each method (IGSCR, ANSTC) to test the accuracy of the other method. Results showed accuracy differences for the three scenes done by both methods to be less than 5 percent. Differences in operational efficiency between the two methods were obvious. The IGSCR method took 3 to 7 days per scene, while the ANTSC method took less than 1 day. It should be noted that working through the IGSCR methodology enabled the automated specification of Euclidean distances for the ANTSC methodology. It is not known whether the specification of Euclidean distances for forest and nonforest used in this study are stable across a wide variety of ecological conditions or differing image radiometric conditions.

Figure 4.—*NDVI values for FIA plots on a Landsat TM scene at FIA subplot 1.*



Utilizing a hard NDVI threshold of 165 assumes that the NDVI ratio is consistent from image to image and that radiometric differences among images are not reflected in the NDVI transform. To test the concept of using a soft threshold, plots that fell into the range of NDVI values between 145 and 165 were assessed for their correct land use call by using a process of crown modeling. Crown modeling uses the distance and azimuth of each tree tallied on an FIA plot, coupled with regression estimates of crown width derived from Forest Health Monitoring (FHM) data, to calculate the proportion of crown reflectance per FIA subplot. These subplot proportions were compared with the NDVI values at the same location to determine land use/land cover compatibility. A somewhat arbitrary threshold of 16.7 percent crown cover per FIA subplot was chosen as the cutoff between forest and nonforest conditions for the comparisons made in this study.

Crown Modeling

Crown modeling for calculating the average canopy reflectance by subplot follows these steps:

- Develop local regressions that predict crown diameter by species from Forest Health Monitoring data.
- Compute the crown radii for each tree species.
- Use a buffer approach in the GIS software to draw the crowns in their real world locations.
- Intersect the crowns with the subplot circles and calculate proportional reflectance per subplot/plot.

For the subset study area of one TM scene, 28 FIA plots fell within the 145-165 NDVI range. Of these 28 plots, 4 had crown proportion reflectance percentages that were inconsistent with the FIA land use call. The crown models are superimposed on the TM imagery and comparisons shown for 3 of these plots in figures 5, 6, and 7.

Figure 5 shows that for this FIA plot, subplot 4 fell in a forest. The average crown proportion for the four subplots was 19.4 percent. This exceeds the 16.7 percent threshold of canopy reflectance, but the NDVI value (154) for this plot was determined from the pixel that corresponded to subplot 1. Since the calculated average crown reflectance proportion was inconsistent with the NDVI value at subplot 1, the analyst has the option to use the pixel at subplot 1 as a seed for a nonforest signature since the plot was not thrown out on the basis of the NDVI parity test.

Figure 6 shows an FIA plot that is classified as forest in the field, but the calculated average crown reflectance proportion (13.8 percent) is less than the 16.7 percent threshold. The crown models reveal a plot that is in an area that was likely clearcut a few years ago and is reverting to forest. The canopies are small and the crown reflectance proportion calculations are predicated on using FIA tally trees that are 5 inches

Figure 5.—*Crown proportion (19.4 percent) and NDVI value (154) not consistent.*

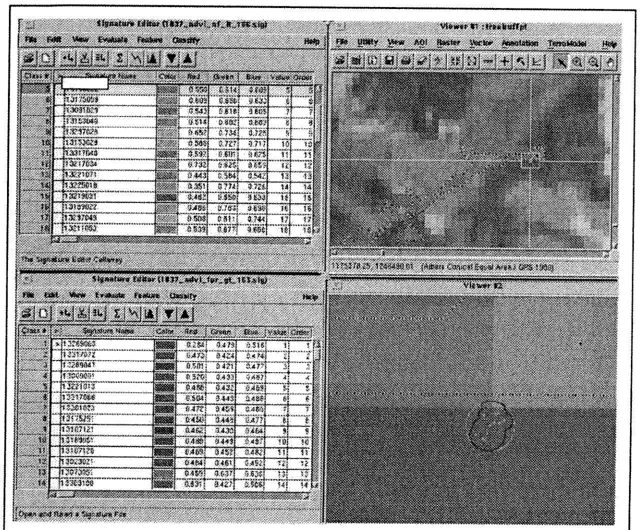


Figure 6.—*Crown proportion (13.8 percent) and NDVI value (174) not consistent.*

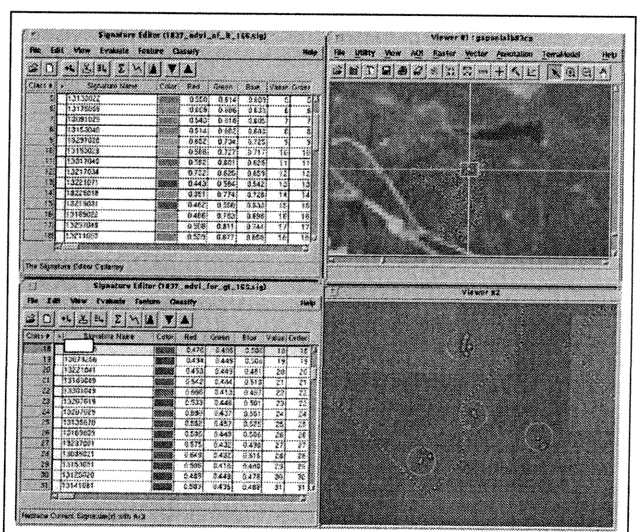
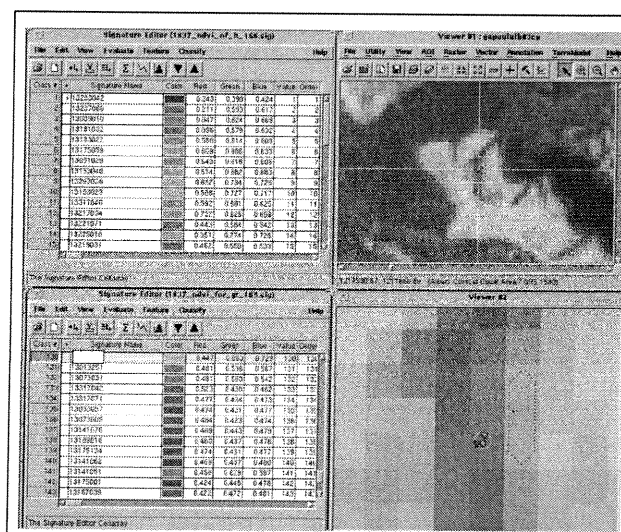


Figure 7.—*Crown proportion (7.9 percent) and NDVI value (170) not consistent.*



d.b.h. or greater. The NDVI value at subplot 1 (174), indicates a forest condition that is consistent with the FIA land use call but inconsistent with the crown modeling-based proportion. The analyst should initiate a seed based on the NDVI value that is consistent with the land use call in the field.

Figure 7 shows an FIA plot in a recent clearcut that has a low average crown reflectance proportion (7.9 percent) but a relatively high NDVI value at subplot 1. It is obvious that subplot 1 falls in a forest edge while the other 3 subplots fall in the clearcut (nonforest). The crown modeling procedure points out a classic land use/land cover conflict. If the analyst places the seed for this forested plot at subplot 1, the signature will reflect the nonforest condition. If the analyst places the seed for this forested plot at subplot 4, the signature will reflect the forested condition. In this case, the crown proportion calculations raised a red flag that leads the analyst to a closer look at the land use/land cover issue.

Conclusions

Classification accuracies for the ANTSC and the IGSCR methodologies were similar. The ANSTC classification methodology is less subjective and requires no analyst input, making it easy to implement by analysts with minimum remote sensing expertise. Results of the crown modeling experiments indicate that the NDVI threshold of 165 is a good choice but some land use/canopy reflectance inconsistencies exist with the 145-165 NDVI range. The number of inconsistencies was small (<14 percent of the total FIA plots). The additional time spent assessing the problem plots within the 145-165 NDVI range is likely worth the improvement in precision, although a small amount of automation potential may be sacrificed.

It is not known whether the Euclidean distance measures used in the ANTSC methodology will work as well in other States or in different ecological conditions. It is possible that some preliminary work will be required to determine the optimum Euclidean distances for forest and nonforest signatures when ecological conditions are significantly different.

Literature Cited

Anderson, G.L.; Hanson, J.D.; Haas, R.H. 1993. Evaluating Landsat thematic mapper derived vegetation indices for estimating above ground biomass on semiarid rangelands. Remote Sensing of the Environment. 45: 165–175.

Franco-Lopez; H.; Ek, A.R.; Bauer, M.E. 2000. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of the Environment. 77(2): 251–274.

Hoppus, M.; Riemann, R.; Lister, A. 2000. Remote sensing strategies for forest inventory and analysis utilizing the FIA plot database. In: 8th Forest Service biennial remote sensing applications conference; ASPRS, Albuquerque, NM, April 2000.

Iverson, L.R.; Cook, E.A.; Graham, R.L. 1989. A technique for extrapolating and validating forest cover across large regions calibrating AVHRR data with TM data. International Journal of Remote Sensing. 10(11): 1805–1812.

Moisen, G.G.; Edwards, T.C.; Van Hooser, D. 1998. Merging regional forest inventory data with satellite-based information through nonlinear regression methods.

Wayman, J.P.; Wynne, R.H.; Scrivani, J.A.; Reams, G.A. 2001. Landsat TM-based forest area estimation using iterative guided spectral class rejection. Photogrammetric Engineering and Remote Sensing. 67(10): 1155–1166.

USDA

United States Department of Agriculture

Forest Service

GTR-NC-252
February 2005

Proceedings of the Fourth Annual Forest Inventory and Analysis Symposium

Meeting jointly with the Southern Forest Mensurationists

Edited by
Ronald E. McRoberts
Gregory A. Reams
Paul C. Van Deusen
William H. McWilliams
Chris J. Cieszewski

New Orleans, Louisiana
November 19–21, 2002

Sponsored by
Forest Inventory and Analysis Program, USDA Forest Service
National Council for Air and Stream Improvement
Southern Forest Mensurationists
University of Georgia

McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; McWilliams, William H.; Cieszewski, Chris J., eds. 2005. Proceedings of the fourth annual forest inventory and analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 258 p.

Documents contributions to forest inventory in the areas of sampling, remote sensing, modeling, information management and analysis for the Forest Inventory and Analysis program of the USDA Forest Service.

KEY WORDS: annual forest inventory, estimation, sampling, modeling, remote sensing, assessments

Papers published in these proceedings were submitted by authors in electronic media. Editing was done to ensure a consistent format. Authors are responsible for content and accuracy of their individual papers. The content does not necessarily reflect the views or policies of the U.S. Department of Agriculture or the Forest Service.