Landsat TM-Based Forest Area Estimation using Iterative Guided Spectral Class Rejection

Jared P. Wayman

Thesis submitted to the Faculty of

Virginia Polytechnic Institute and State University

In partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

in

FORESTRY

Randolph H. Wynne, Chair

Richard G. Oderwald

James B. Campbell

John A. Scrivani

Virginia Polytechnic and State University
Blacksburg, Virginia, 24061

May 15, 2000
Blacksburg, Virginia

Keywords: FIA, Landsat TM, IGSCR, Forest Area Estimation

Copyright 2000, Jared P. Wayman

Landsat TM-Based Forest Area Estimation using Iterative Guided Spectral Class Rejection

Jared P. Wayman

(ABSTRACT)

In cooperation with the USDA Forest Service Southern Research Station, an algorithm has been developed to replace the current aerial-photography-derived FIA Phase 1 estimates of forest/non-forest with a Landsat Thematic Mapper-based forest area estimation. Corrected area estimates were obtained using a new hybrid classifier called Iterative Guided Spectral Class Rejection (IGSCR) for portions of three physiographic regions of Virginia. Corrected area estimates were also derived using the Landsat Thematic Mapper-based Multi-Resolution Land Characteristic Interagency Consortium (MRLC) cover maps. Both satellite-based corrected area estimates were tested against the traditional photo-based estimates. Forest area estimates were not significantly different (at the 95% level) between the traditional FIA, IGSCR, and MRLC methods, although the precision of the satellite-based estimates was lower. The estimated percent forest area and the standard error (respectively) of the estimates for each region and method are as follows; Coastal Plain- Phase 1 66.06% and 1.08%, IGSCR 68.88% and 2.93%, MRLC 69.84% and 3.08%. Piedmont- Phase 1 63.87% and 1.91%, IGSCR 65.52% and 3.50%, MRLC 59.19% and 3.83%. Ridge and Valley- Phase 1 69.74% and 1.22%, IGSCR 70.02%, and 2.43%, MRLC 70.53% and 2.52%.

Map accuracies were not significantly different (at the 95% level) between the IGSCR method and the MRLC method. Overall accuracies ranged from 80% to 89% using FIA definitions of forest and non-forest land use. Given standardization of the image rectification process and training data properties, the IGSCR methodology is objective and repeatable across users, regions, and time and outperforms the MRLC for FIA applications.

ACKNOWLEDGMENTS

I would like to thank Virginia Department of Forestry, The National Council of the Paper Industry for Air and Stream Improvement (Inc.), the USDA Forest Service Southern Research Station, and the Virginia Tech Forestry Department for their generous support towards this research. I would also like to thank Dr. Randolph H. Wynne, Dr. John A. Scrivani, Dr. Richard G. Oderwald, and Dr. James B. Campbell for their enlightening advice, coaxing and guidance; Jan van Aardt and Glenn Fields for early work collecting and editing the helopolys; Rebecca Forest and Jason Braunstein for spending many hours on various aspects of this study; and Sorin Popescu for providing recreational therapy during the busy times. Special thanks to my wife Tressa who had to be without her favorite man while deadlines were being met.

TABLE OF CONTENTS

1. INTRODUCTION . . . . Page 1
1.1 HISTORY OF REMOTE SENSING IN FORESTRY . . . . Page 1
1.2 ANNUAL INVENTORY IN THE SOUTH . . . . Page 2
1.3 AREA ESTIMATES FROM AERIAL PHOTOGRAPHY . . . . Page 3
1.4 SATELLITE-ASSISTED FOREST AREA ESTIMATION . . . . Page 4

2. OBJECTIVES . . . . Page 4

3. HYPOTHESIS . . . . Page 5

4. BACKGROUND . . . . Page 5
4.1 PREPARING FOR A NEW ERA IN FOREST INVENTORY . . . . Page 5
4.2 FIA REMOTE SENSING BAND . . . . Page 6
4.3 IMAGERY . . . . Page 7
4.3.1 Sensor Choice . . . . Page 7
4.3.2 Rectification and Registration . . . . Page 8
4.3.3 Terrain Effects . . . . Page 10
4.3.4 Preprocessing . . . . Page 11
4.3.5 Multi-Date Imagery . . . . Page 12
4.3.6 Derivation of New Bands . . . . Page 12
4.4 REFERENCE DATA . . . . Page 14
4.4.1 Sampling Design . . . . Page 14
4.4.2 Ancillary Data . . . . Page 15
4.5 CLASSIFICATION . . . . Page 16
4.5.1 Classification Methods . . . . Page 16
4.5.2 Unclassified Pixels . . . . Page 21
4.5.3 Post Classification Corrections . . . . Page 21
4.5.4 Spatial Filtering . . . . Page 22
4.6 ACCURACY ASSESSMENT . . . . Page 22
4.7 FROM CLASSIFICATION TO FOREST AREA ESTIMATES . . . . Page 25

5. METHODS . . . . Page 27
5.1 LOCATION . . . . Page 27
5.2 IMAGERY . . . . Page 28
5.3 REFERENCE DATA . . . . Page 30
5.3.1 Helopolys . . . . Page 30
5.3.2 Buffered Points . . . . Page 32
5.3.3 Heads-up Digitizing . . . . Page 33
5.3.4 Harvest Polygons . . . . Page 33
5.3.5 Air Videography and Orthophotography . . . . Page 33
5.3.7 Training Data Summary . . . . Page 34

5.4 IMAGE CLASSIFICATION VIA ITERATIVE GUIDED SPECTRAL CLASS REJECTION . . . . . . Page 34
5.5 ACCURACY ASSESSMENT . . . . . . Page 37
5.6 AREA CORRECTION AND CALCULATION OF PHASE 1 ESTIMATES . . . . . . Page 38
5.7 COMPARISON WITH MRLC CLASSIFICATION . . . . . . Page 39

6. RESULTS AND DISCUSSION . . . . . . Page 41
6.1 ITERATION PROPERTIES . . . . . . Page 41
6.2 MAP ACCURACY . . . . . . Page 41
6.3 FOREST AREA ESTIMATION . . . . . . Page 45

7. CONCLUSIONS . . . . . . Page 47
7.1 TRADITIONAL PHASE 1 vs IGSCR vs MRLC . . . . . . Page 47
7.2 TYPE AND NUMBER OF REFERENCE DATA . . . . . . Page 49
7.3 SCENE SELECTION . . . . . . Page 51
7.4 RECTIFICATION . . . . . . Page 52
7.5 IGSCR PARAMETERS . . . . . . Page 52
7.6 STUDY LIMITATIONS . . . . . . Page 53
7.7 SUMMARY . . . . . . Page 54

8. RECOMMENDATIONS . . . . . . Page 55

9. LITERATURE CITED . . . . . . Page 57

10. APPENDIX . . . . . . Page 62

LIST OF TABLES

Table 1. Landsat 7 Sensor Characteristics . . . . . . Page 9

Table 2. Contingency table for TM map categories and FIA ground plots (truths). Table represents FIA ground plots and classified TM pixels in eight Ridge and Valley counties in Virginia. . . . . . . Page 26

Table 3. TM Images used in study. . . . . . . Page 28

Table 4. Summary of study area and reference data by physiographic region. . . . . . . Page 34

Table 5. SAS output showing which spectral classes (SPEC) met the minimum pixel (SUM) and the homogeneity (PROP) criteria.. . . . . . . Page 35

Table 6. Description of MRLC classification categories. . . . . Page 40

Table 7. Summary of maximum spectral class number (user-specified ISODATA parameter), number of reference pixels and the number of pure spectral classes associated with each iteration. . . . . Page 42

Table 8. Accuracy statistics for Phase 1 estimates and image classifications across all three physiographic regions (3x3 scan-majority filtered and unfiltered and MRLC class 15 as forest and as non-forest). . . . . Page 44

Table 9. Comparison of the three physiographic region accuracy statistics for iterative guided spectral class rejection classification using the 3x3 scan-majority filtered images. . . . . Page 44

Table 10. Area estimation statistics for the Coastal Plain region. Page 45

Table 11. Area estimation statistics for the Piedmont region. . . . . Page 46

Table 12. Area estimation statistics for the Ridge and Valley region. . . . . Page 46

LIST OF FIGURES

Figure 1. Study locations in three physiographic regions of Virginia. . . . . Page 28

Figure 2. Example of the fragmented nature of the Coastal Plain landscape. . . . . Page 29

Figure 3. Simplified flowchart of iterative guided spectral class rejection process. . . . . . Page 36

Figure 4. Example of raw image of three Piedmont counties prior to classification. . . . . Page 37

Figure 5. Example of image recoded into forest (green) and non-forest (tan) following first iteration. . . . . Page 37

Figure 6. Image following removal of pixels labeled as forest or non-forest in iteration 1. This image is then clustered into a new set of spectral classes. . . . . Page 38

Figure 7. Layout of a single FIA plot with a plot center and three subplots. . . . . Page 38

Figure 8. Cumulative frequency of spectral classes by iteration for all physiographic regions. . . . . Page 43

Figure 9. Example of classification results (raw image left, classified image right), dark green pixels and green polygons represent forest, tan pixels and yellow polygons represent non-forest. . . . . Page 49

Figure 10. Example of raw image clearcuts in two stages (top left) with the older cut near the bottom.. . . . . Page 50

Chapter 1
INTRODUCTION

1.1 HISTORY OF REMOTE SENSING IN FORESTRY

Remote sensing allows the discernment of information about some entity or feature without physical contact. A more formal definition is provided by Campbell (1996): "Remote sensing is the practice of deriving information about the earth's land and water surfaces using images acquired from an overhead perspective, using electromagnetic radiation in one or more regions of the electromagnetic spectrum, reflected or emitted from the earth's surface."

Traditional forestry applications of remote sensing have utilized aerial photography, both in the visible and near-infrared portion of the electromagnetic spectrum. The first use of aerial photography in forestry is unknown, but the regular acquisition of aerial photos began during World War I. With the advent of photogrammetry (taking measurements from photographs) in the 1920s, remote sensing evolved into a quantitative and scientific technique used in mapping many types of natural resources. World War II saw the shift from strictly visible range photography to infrared and microwave photography (Campbell, 1996). This war also expedited the development of cameras and of data processing techniques. This jumpstart enabled the application of remote sensing in forestry to boom in the postwar years. One particular advance was the use of "camouflage detection film" in agriculture by Robert Colwell (Campbell, 1996).

The space race provided another sharp advance in remote sensing technology. The National Aeronautics and Space Administration (NASA) established a remote sensing program and, beginning in the 1960s, the National Academy of Science (NAS) also investigated the use of remote sensing in agriculture and science (Campbell, 1996). The launch of Landsat 1 in 1972 provided the first satellite images of the earth's surface on a regular basis. Campbell (1996) also notes that Landsat 1 passed major milestones in remote sensing in two areas: multispectral data and digital analysis. The images from Landsat 1 enabled many people to gain experience in processing and interpreting multispectral data. Landsat 1 also introduced digital analysis into mainstream image processing techniques.

The majority of operational use of satellite imagery in land cover classifications is in regional and larger scale applications. Medium and low spatial resolution sensors such as Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+), the French Satellite Pour l'Observation de la Terre (SPOT), and Advanced Very High Resolution Radiometer (AVHRR) are most common due to their ease of availability and large scene coverages. Increased integration of spaceborne or airborne sensors into natural resource management at smaller scales seems likely given the suite of new sensors being developed and tested. In 1994, the US government granted the first licenses for 1-meter resolution satellites to be built and flown (Glackin, 1998). These higher spatial resolution (1 m panchromatic and 3-4 m multispectral) sensors may enable better cover-type discrimination than current space-borne sensors (Wynne and Carter, 1997). The recent launches of Landsat 7 and SPOT 4 ensure the continuation of medium spatial resolution imagery for at least the next 5 years. Hyperspectral imagery and lidar/radar applications are technologies that both offer the potential for increased use of remote sensing in forest management and in forest inventory.

1.2 ANNUAL INVENTORY IN THE SOUTH

The USDA Forest Service began the Forest Inventory and Analysis (FIA) program in the 1930s. The use of permanent plots measured at approximately ten-year intervals began in the 1950s. The primary focus was on obtaining volume by species but it later shifted to estimates of change. The importance of the South to the nation's forest-based industries and the dynamic nature of change in the region's forest resources led the USDA Forest Service Southern Research Station to replace the traditional FIA surveys with an annual forest inventory system. The annual protocol requires 20% of the permanent plots in a state to be sampled each year (i.e., all the plots are sampled every five years). While Virginia is one of the few southern states in the third year of this program, the US Agricultural Research, Extension, and Education Reform Act of 1998 (Public Law 105-185, 23-Jun-98) now requires the USDA Forest Service to implement this protocol nationwide. In addition, the law requires NASA and the USDA Forest Service to work together to better integrate remote sensing technology into the annual forest inventory. This act recognizes that since the FIA inception there has been a shift in primary forest products from sawtimber in the Northwest to sawtimber and biomass production for pulp in the Southeast. The decadal inventories were deemed too infrequent to adequately estimate the state of the national forest resource because the change in wood products resulted in much shorter rotations. Another issue was that the inventories were disjointed across the South due to the state-by-state nature of the FIA program (Reams and Van Deusen, 1999).

The annual sampling design is described by Reams and Van Deusen, (1999) as "a systematic sample of five interpenetrating grids, also known as annual panels." Each panel is made up of one-fifth of the existing FIA plots and is measured every five years. The increased revisit frequency will allow better detection of growth and removals and eliminate the state-to-state variation in data currency.

1.3 AREA ESTIMATES FROM AERIAL PHOTOGRAPHY

Historically, the USDA Forest Service FIA program has produced area estimates of forest type in two phases using a variation of double sampling (Chojnacky, 1998, Reams and Van Deusen, 1999). The first phase requires a systematic grid of points to be placed on aerial photos (primarily at 1:40,000 scale from the National Aerial Photography Program, NAPP) with a density (in the Southeast) of approximately one point per 230 acres. These points are then classified into forest or non-forest by photo interpretation. The current definition of forest in FIA is land that is at least 10 percent stocked by forest trees of any size, or land formerly having had such tree coverage but not currently developed for non-forest use. The minimum area considered for classification is 1 acre. Forested strips must be at least 120 feet wide.

In the second phase, a subsample of the first phase sample points is visited on the ground to confirm the classification. The sample of ground points refines the estimate and enables the calculation of the standard error of the estimate. Tree and stand level data collected from permanent ground plots represent the third phase of the current system. Reams and Van Deusen (1999) state two shortcomings of the photo-based method. The first is the inability to produce maps, and the second is that it can take up to a year to photo interpret an entire state. A third issue is the difficulty in obtaining current photography.

1.4 SATELLITE-ASSISTED FOREST AREA ESTIMATION

The current air photo-based methodology is proven and reliable. However, the switch to an annual system requires a faster turnaround in the production of land-use and land cover maps. In consideration of these facts, the integration of satellite-based area estimates must, at the minimum, give the same results with the desired precision in a shorter time frame. In the context of FIA, satellite imagery is useful to " a) identify and measure those forest characteristics that cannot be observed easily from the ground, b) identify and measure those characteristics for which we would like more spatially continuous coverage, and c) identify change at intervals interim to the regular inventory cycle" (Riemann and Alerich, 1998). Additional information can also be derived from map output of satellite-based classifications such as patch-size and connectivity. There is also a high potential for use of satellite imagery to improve removal estimates via change detection techniques.

Chapter 2
OBJECTIVES

The primary objective of this research is to develop an algorithm to obtain forest area estimates using Landsat TM imagery that are comparable in area estimation and in precision to the current Phase 1 USDA Forest Service FIA forest area estimates used in the Southern Research Station. The algorithm must be objective, accurate, and repeatable across operators, regions, and time. A secondary objective is the production of forest/non-forest maps (using FIA forest definition) of sufficient resolution and accuracy to enable stratification and spatial analysis.

Chapter 3
HYPOTHESIS

The null hypothesis for this research is that there is no significant difference between the forest area estimates derived via the traditional FIA methods and the satellite-based forest area estimates at the 95% confidence interval.

Chapter 4
BACKGROUND

4.1 PREPARING FOR A NEW ERA IN FOREST INVENTORY

For several decades the prospect of applying remote sensing systems to operational forest/land cover inventory has been researched in hope of increasing efficiency and economy. However, as has been noted elsewhere (Wynne and Carter, 1997), during the twenty-five years that have elapsed since the launch of Landsat-1, satellite remote sensing has not met the demands of forest managers (see also Holmgren and Thuresson, 1998). A decade from now we could easily be repeating the same lament unless we pay careful attention to the needs of resource managers, be they in the public or private sector.

The principal barriers to the effective application of remote sensing systems to forest inventory have originated from the somewhat conflicting needs of resource managers. On the one hand, managers require broad-scale coverage to provide a regional overview; such information is essential in the formation of effective management and strategic plans. On the other hand, managers require spatial resolution and positional accuracy that is adequate to acquire data at useful levels of detail for tactical planning. Managers also require the long term integrity, continuity, and consistency necessary for place-to-place comparisons, and comparisons of data over long intervals of time.

Therefore, our priority should be to design a forest inventory system based upon a close match between the capabilities of remote sensing systems and the needs of resource managers. In today's environment, the range of systems available (or soon to be available) to the community of forest resource managers is much greater than what was available in even the recent past (see Glackin, 1998). These increased capabilities enhance the arsenal of techniques that can be applied to resource management issues.

4.2 FIA REMOTE SENSING BAND

The specific applications of space-borne remote sensors in the USDA Forest Service are outlined in Riemann and Alerich (1998). The FIA Remote Sensing Band was introduced in order to improve cooperation between the separate FIA units by way of sharing data, techniques and expertise (Riemann and Alerich, 1998). Landsat TM, AVHRR, videography, and Synthetic Aperture Radar (SAR) are being investigated for their respective benefits in forest inventory by the USDA Forest Service. Most are research or pilot studies though some are operational. At present there is no clear winner; each has distinct benefits and drawbacks. Regardless of sensor type, methodologies "need to be operational across large areas at regular intervals, within certain budget limitations and with the accuracy, consistency and frequency required of a continual, nationally reporting program" (Riemann and Alerich, 1998). The current projects in which the five FIA units are involved are as follows (from Riemann and Alerich, 1998):

- improving the accuracy of plot locations (all units),
- developing techniques to use FIA's tremendous amounts of ground data in image analysis, with its locational uncertainties, plot design characteristics, and ground-centered attributes (NE, PNW, RM)[1],

[1] NC - North Central, NE - Northeastern, PNW - Pacific Northwest, RM - Rocky Mountains, FC- Fort Collins (RSAC), S - Southern, AK - Alaska (part of PNW)

- developing modeling techniques to predict forest inventory variables from TM-derived information and other available data, through both regression, geostatistical, imputation, and other techniques (NE, PNW, RM),
- using FIA data to help characterize remote-sensing-derived maps and models (NE, RM),
- identifying fragmentation, patch size and context variables (AK, NC, NE, PNW),
- detecting change (from aerial photography and satellite imagery) to identify which sample plots have changed over time (NC, NE, PNW),
- using new sources of remotely sensed data to improve our ability to capture current variables,
- comparing AVHRR (modeled from TM) and TM to current photointerpretation to determine the statistical implications of using these data as a means estimating forest/non-forest area (RM, S, FC),
- assessing classification accuracy in a variety of situations where the sample data are clustered or sampled with unequal probabilities, and the development of software to do this (FC),
- investigating SAR as a potential source of information about stand structure, biomass, and forest composition. (NE, S).

4.3 IMAGERY

4.3.1 Sensor Choice

The three most commonly used sensors for land cover classifications are Landsat TM, SPOT, and AVHRR. Numerous land cover classifications have been carried out with each sensor, but for use in FIA forest/non-forest area estimates, Landsat TM imagery is preferable. Reams and Van Deusen (1999) state that the superior spectral resolution of TM versus both SPOT and AVHRR, and the higher spatial resolution of TM versus AVHRR make it the preferred sensor.

The Landsat series began in 1972 with the Multispectral Scanner (MSS) aboard Landsat 1. Landsat 5 was launched in 1985 and is still operational today, it carries the Thematic Mapper (TM) sensor and the MSS sensor. Landsat 7 was launched in April, 1999 and carries the Enhanced Thematic

Mapper Plus (ETM+) sensor. The following is an excerpt from the Landsat 7 homepage (http://landsat.gsfc.nasa.gov/project/Comparison.html), outlining the attributes of Landsat 7:

> "The earth observing instrument on Landsat 7, the Enhanced Thematic Mapper Plus (ETM+), replicates the capabilities of the highly successful Thematic Mapper instruments on Landsats 4 and 5. The ETM+ also includes new features that make it a more versatile and efficient instrument for global change studies, land cover monitoring and assessment, and large area mapping than its design forebears.
>
> The primary new features on Landsat 7 are: a panchromatic band with 15 m spatial resolution on board, full aperture, 5% absolute radiometric calibration and a thermal IR channel with 60 m spatial resolution. The ETM+ instrument is a fixed "whisk-broom", eight-band, multispectral scanning radiometer capable of providing high-resolution imaging information of the Earth's surface. It detects spectrally-filtered radiation in visible near infra-red (VNIR), short wave infra-red (SWIR), long wave infra-red (LWIR) and panchromatic bands from the sun-lit Earth in a 183 km wide swath when orbiting at an altitude of 705 km.
>
> On board solar calibration and payload correction data allows the ground to radiometrically correct the data to an absolute accuracy of 5% and to geometric register a scene to within 250 meters. Nominal ground sample distances ("pixel" sizes) are 15 meters in the panchromatic band, 30 meters in the VNIR and SWIR bands and 60 meters in the LWIR (Table 1).
>
> Landsat 7 will collect data in accordance with the World Wide Reference System, which has catalogued the world's land mass into 57,784 scenes, each 183 km wide by 170 km long. The ETM+ will produce approximately 3.8 gigabits of data for each scene. An ETM+ scene will have an Instantaneous Field Of View (IFOV) of 30 square meters in bands 1-5 and 7 while band 6 will have an IFOV of 60 square meters on the ground and the new band 8 an IFOV of 15 square meters."

4.3.2 Rectification and Registration

Landsat level 1B imagery is supplied nominally rectified to the UTM coordinate system. In order for the imagery to be better integrated with other spatial data, a more precise geographic rectification must be performed. Cracknell (1998) outlines this resampling operation and presents the sources of error that can arise in the process. Resampling requires the “rubber sheeting” of the image to known ground control points (GCPs). Three interpolation methods are described by Cracknell (1998):

- the nearest neighbor method
- the bilinear interpolation method
- the bicubic interpolation method

The nearest neighbor method is most common because it is fast and does not lead to loss of information by smoothing (Cracknell, 1998). This method simply takes the digital number (DN) from the point nearest the pixel and assigns that value to the pixel. There is the chance of repetition or omission of pixel DNs (Cracknell, 1998). The nearest neighbor method may also result in some scan-line duplication (Itten and Meyer, 1993). The bilinear and bicubic methods smooth the data and therefore some data is lost. The resampling method is seldom mentioned in literature, however both the nearest neighbor (McGwire *et al.*, 1996) and cubic convolution (Muller *et al.,* 1998) are commonly used.

The accuracy to which images are rectified affects the quality of subsequent operations. Recent work has shown that a root mean square error (RMSE) of about 0.2 pixels is necessary to minimize the

Table 1. Landsat 7 Sensor Characteristics (Landsat 7 homepage http://landsat.gsfc.nasa.gov/project/Comparison.html)

ETM+ Bands		
Band	**Bandwidth (micrometers)**	**Resolution (m)**
1	.45 to .515	30
2	525 to .605	30
3	.63 to .690	30
4	.75 to .90	30
5	1.55 to 1.75	30
6	10.40 to 12.5	60
7	2.09 to 2.35	30

error from misregistration (Dai and Khorram, 1998). This RMSE is much lower than traditional values. Many studies are carried out with RMSE values of approximately 1 pixel. For example, Muller *et al*. (1998) had an RMSE of 57.4 m for 50 m pixels (resampled 80 m MSS pixels) and Fuller *et al*. (1994) had an RMSE of 0.8 pixels.

When multiple dates of the same path/row are used in classification, the registration between the images is even more critical. Misregistration can lead to misclassification simply due to the pixel offsets being identified as some kind of phenological or land-cover change. Fuller *et al*. (1994) noted that even minor misregistration results in misclassification, especially with boundary pixels. They also stated that the misregistration results in a greater number of mixed pixels in the composite image, with up to 40% of the pixels representing mixed boundaries. Multi-date imagery is often registered by the image-to-image method, where one image is geographically rectified to known GCPs and the second image is registered to the first. This approach is based on the premise that it is easier to identify corresponding pixels between images than to identify the exact location of a point or line intersection in a medium resolution raster image.

4.3.3 Terrain Effects

Satellite-based classifications of mountainous regions often yield poor results due to the effects of shadowing in the imagery and geometric distortion. Itten and Meyer (1993) performed radiometric and geometric corrections on TM imagery in order to improve forest versus non-forest classification accuracy in Central Switzerland. They derived slope, aspect, illumination, and masks for shadow from Digital Elevation Models (DEMs) with 25 meter horizontal resolution and 10 cm vertical resolution. A shift correction algorithm was applied to the image to correct for geometric distortion and the scene was radiometrically corrected for illumination and atmospheric effects. The slope-aspect corrections gave the greatest improvements in accuracy and, when combined with the atmospheric corrections, forest/non-forest cover could be classified to almost 90%. Their conclusion suggest that a high quality DEM is essential, and that summer scenes are best for mountains due to the higher sun angles.

Traditional image rectification is also more problematic in areas of high relief. Because images are often rectified to vector road coverages and roads mostly occur in valleys in mountainous areas, the problem of locating suitable GCPs increases. This lack of choice in GCP location also results in a poor elevational distribution of GCPs, because most are located at road intersections in valleys, which in turn results in less accurate polynomial transformations. The terrain corrections performed by Itten and Meyer (1993) were effective, however the results can not be repeated unless high quality DEMs (that can be accurately coregistered to images) are available.

4.3.4 Preprocessing

Atmospheric scattering and absorption affect all satellite imagery to some degree. As light passes through the atmosphere it is selectively scattered and absorbed. The scattering of shorter wavelengths such as blue and green is greater than that of longer wavelengths. Absorption of light occurs at longer wavelengths, usually greater than 0.8 µm. Atmospheric scattering adds brightness while atmospheric absorption subtracts brightness from the ground return values, therefore reducing the ability to extract useful information about the terrain from the remotely sensed image (Jensen, 1996). For this reason, preprocessing is often necessary before the data can be transformed into useful information.

Two common radiometric corrections are histogram matching and haze reduction. Histogram matching is the mathematical process of converting the histogram of one image to resemble the histogram of another image. Both images should be of the same area on the Earth's surface so that the range of digital numbers in the image to be matched will be comparable to the digital numbers of the image to which it is matched.

Haze reduction is a preprocessing step that removes the haze and system noise from the image. The identification of the atmospheric effects has been approached in two primary ways: the first assumes a uniform effect over the entire image and the second uses fixed coefficients on a per-pixel approach. The coefficient-based haze reduction method used in ERDAS Imagine Version 8.4 (ERDAS, Inc. Atlanta, Georgia), for example, first derives the haze/noise component via a tasseled cap transformation, then removes that component, and finally carries out the inverse transformation to restore the image. However, Kaufman (1984) notes that haze is often not uniform across an image. From this observation the application of fixed coefficients to every pixel to "remove" the haze component seems counterintuitive. Kaufman (1984) also insists that atmospheric optical characteristics be known and considered before corrections take place. Such an approach was used by Itten and Meyer (1993). The practice of atmospheric correction is seldom carried out (or at least not mentioned) in many studies. McGwire *et al*. (1996) is an example of where correction for the atmospheric path radiance was applied to the imagery. This study did not use atmospheric corrections due to the lack of any robust and repeatable technique.

4.3.5 Multi-Date Imagery

Single-date imagery classifications are at the disadvantage of not containing any information about seasonal vegetative changes. Agricultural areas and urban areas are often confused in single-date imagery, especially if the imagery was captured in the winter. Forest type is also difficult to identify in single-date imagery. Schriever and Congalton (1995) obtained significant improvements in hardwood cover type discrimination when multi-date imagery was used. Fuller *et al*. (1994) combined winter and summer scenes to improve classification accuracy of 25 cover types. The main disadvantage of using more than one image for a given area is the potential error from scene-to-scene misregistration. The change to annual FIA surveys in the Southeast requires improved estimations of removals and mortality (Wynne and Scrivani, 1999). These year to year changes can be identified using anniversary images of the same location via change detection techniques.

4.3.6 Derivation of New Bands

The amount of data provided by the imagery can be increased if new bands are derived from some combination or transformation of the raw bands (San Miguel-Ayanz and Biging, 1997). Popular transformations are band ratios (vegetation indices), principal components analysis (PCA), tasseled cap transformations and texture analysis (Bauer *et al*., 1994, Schriever and Congalton, 1995, and San Miguel-Ayanz and Biging, 1997). It is not uncommon to have as many as 15 bands from which to choose when carrying out a classification (e.g., San Miguel-Ayanz and Biging, 1997); however, 2-6 bands are generally used.

The most common band ratios use some combination of TM bands 3 (red) and 4 (near infrared) and are known as vegetation indices. Images that are produced via vegetation indices are a function of the amount and type of vegetation in the original image. More vegetation results in a lower reflectance in the red (higher absorption by chlorophyll) and a higher reflectance in the infrared (due to the air/cell interface in plant tissue). PCA removes highly correlated (and therefore redundant) image data by combining highly correlated bands into new images, or principal components. The idea is to have a scene with less data dimensionality but with almost the same information as all the original bands combined. Because there is a high correlation between TM bands 1, 2 and 3, and between bands 5 and 7, these bands can be combined with a minimal decrease in image information (e.g., San Miguel-Ayanz and Biging, 1997).

The tasseled cap is similar to a PCA in its application, although the eigenvectors and eigenvalues are replaced by predetermined coefficients. The tasseled cap transformation was originally applied to Landsat MSS data but was later applied to TM data (Jensen, 1996). The tasseled cap transformation is applied to bands 1-5 and 7 of Landsat TM imagery. There are four useful images or indices created by a tasseled cap transformation: the soil brightness index, the green vegetation index, the wetness index and the haze index. The first two indices account for most of the information in the scene in a similar manner as the first two principal components in a PCA.

Band selection is an important step in any classification as it gives the analyst an indication of which band combination provides the highest separability between spectral classes. The most common methods of evaluating spectral class separability are transformed divergence (TD) and Jefferies-Matusita (JM). San Miguel-Ayanz and Biging (1997) found TD and JM were the best of the five band election processes (spectral pattern analysis, TD, JM, weighted TD and weighted JM) they tested on TM and SPOT imagery.

4.4 REFERENCE DATA

4.4.1 Sampling Design

Semi-automated classification of remotely-sensed imagery requires accurate reference data for both training and validation. The methods used to collect reference data vary widely, and include interpretation of higher resolution imagery or hardcopy maps with subsequent ground verification (e.g., Bauer *et al*., 1994; San Miguel-Ayanz and Biging, 1996) as well as the on-site collection of point or polygon samples using global positioning system (GPS) technology (e.g., Rutchey and Vilcheck, 1994). Muller *et al*. (1998) outlines the three main attributes of a sampling strategy: (1) the sampling design, (2) the sampling unit (e.g., points, pixels, or polygons), and (3) the sample size. Each attribute must be considered because the sampling strategy will dictate how the results are analyzed and interpreted (Congalton and Green, 1999).

The statistical attributes of training samples are less important than those of validation samples. The desired properties of training samples is that they be numerous enough and spatially distributed so that they represent the population. Poor training data will result in a poor classification. Statistically, the ideal sample design for validation data is simple random sampling, followed by stratified simple random sampling. Card (1982) notes that "stratified sampling is quite attractive because a simple random or systematic sample drawn from the total map area tends to oversample categories of high frequency and to undersample categories of low frequency." Sample design choice is often hampered by logistical constraints and many studies resort to systematic, stratified-systematic unaligned, cluster sampling, and transect sampling (Muller *et al*., 1998). Reference data collected within a spatially autocorrelated region may introduce bias into the reference sample, because the sample variance may underestimate that of the general population (McGwire *et al*., 1996). This is less important for training data (because no statistical inferences are being made), but must be considered for validation data. Cluster sampling may prove to be more efficient if the cost of a simple random sample is high. Edwards *et al*. (1998) suggest a measure of relative efficiency should be considered prior to finalizing the sampling design. They state that although cluster samples carry less information per unit than simple random sampling, the cost per unit is usually substantially lower for cluster sampling. Stehman (1997) provides a methodology for estimating standard error of map accuracies under cluster sampling. He shows that cluster sampling results in a significant underestimation of standard error if simple random sampling formulas are used, and a slight underestimation if cluster formulas are used. A two stage sample may be appropriate where the training data can be "cheap" data, such as clustered or systematic designs, while the validation data can be based on a stratified random (or similar) design. The two stage method ensures no bias is introduced in the accuracy assessment.

The sample size can also be logistically constrained but a larger sample size will capture more of the image variation and improve classifier performance (Dobbertin and Biging, 1996). Previous work also suggests that training and validation data should be proportional to the actual information class distributions (Richards, 1996; San Miguel-Ayanz and Biging, 1997).

Reference data collection is often a multistaged process (see Bauer *et al*., 1994; Fuller *et al.*, 1994; Schreuder *et al.*, 1995; Kalkhan *et al*., 1998). Bauer *et al.* (1994) utilized primary sampling units (PSUs) in their inventory of Northeastern Minnesota. Each PSU was 88 acres in size (equivalent to 17 by 23 TM pixels) and was derived from 1:9400 scale aerial photography. Approximately 100 cover type classes were interpreted from the photography. Following interpretation, each PSU was visited on the ground or viewed from the air to verify interpretation. The PSUs were then located on USGS 7.5 minute quadrangles, and the cover type boundaries were digitized and then rasterized. The PSUs were then used to train the classifier.

4.4.2 Ancillary Data

Recent classification efforts using Landsat-TM and similar sensors have applied improved and innovative techniques in order to increase stratification accuracies (Schriever and Congalton, 1995; Colby and Keating, 1998; Vogelmann *et al*., 1998(a)). Such efforts have resulted in varying degrees of success, depending on the level of categorical specificity. Forest/non-forest stratification accuracies exceeding 85% are difficult to obtain while utilizing satellite data only. Ancillary data such as digital elevation models (DEMs), tax maps, soils maps, and prior information about the landscape have proved to be useful in some studies (Itten and Meyer, 1993; White *et al*., 1995; McGwire *et al*., 1996;Vogelmann *et al*., 1998(a)). The main restriction to applying ancillary data in applications such as the FIA program is the availability of the data across the area of interest. More general restrictions are related to data quality and the lack of metadata.

4.5 CLASSIFICATION

4.5.1 Classification Methods

Image classification is the process by which image pixels are grouped into classes with similar spectral attributes, and then each spectral class is assigned to an information class. It must be noted from the outset, however, that similar spectral classes do not necessarily represent a single or even similar information classes. A classic example is the grouping of urban and bare soil information classes into the same spectral class. In this study our objectives of forest/non-forest stratification reduce the potential spectral class/information class confusion, but it is still present between information classes such as recent clearcuts and agricultural land.

Traditional methods of image classification require varying degrees of user input and subjectivity. For this reason, two classifications using the same methods on the same image will rarely obtain identical results (e.g., Cardille *et al.,* 1996). Furthermore, the comparison between the many methodologies presented in the literature is difficult due to the effects of study location, the size of the study, the type and amount of reference data and the methods of testing and reporting accuracy (McGwire *et al.*, 1996). The level of generalization is low between classification methods due to the variations in imagery such as illumination (mainly due to topography), atmospheric effects, and land cover. There are, however, fundamental principals that can be applied across many classification efforts such as band selection techniques, objective clustering and classification algorithms.

Fuzzy classifications, where boundaries are not crisp but gradual, seem inherently credible due to the continuous nature of imagery. Fuzzy classifications may be most appropriate where there are a large number of mixed pixels or where some level of probability can be applied to a thematic classification (Foody, 1999). Accuracy improvements have been found (Foody, 1999) when comparing fuzzy and traditional techniques, but as previously stated, there are too many variables and subjective elements to robustly announce the superiority of fuzzy techniques. Cracknell (1998) provides food for thought with respect to hard classifications. He presents a clear synopsis of how pixel digital numbers are obtained and of what error may be present in the pixel digital numbers. This reality check as to what each pixel represents does promote the fuzzy classification technique for those pixels for which a large degree of classification uncertainty remains.

Unsupervised classification requires spectral classes to be defined in some automated fashion, and spectral classes to be labeled by the operator. An often used process of building spectral classes is the iterative self organizing data analysis method, or ISODATA. The ISODATA algorithm is a divisive process, dividing the data into increasingly homogenous spectral classes. Verbyla (1995) gives an example of criteria that may be used to control the clustering. ISODATA performs two "passes" on the data. The first pass (which usually consists of multiple passes) finds the means for each of the predefined number of classes. This is no small task as classes are split and merged depending on their standard deviations, distance between means, class size, and the total number of classes. Once a stopping criterion has been met, pass two classifies each pixel (starting at 1,1) into one of the spectral classes.

The maximum likelihood classifier is a robust supervised classification technique that is commonly used in research and application (McGwire *et al*., 1996, Ediriwickrema and Khorram, 1997). If prior information class membership probabilities are assumed to be equal, the full potential of the maximum likelihood classifier may not be obtained (Ediriwickrema and Khorram, 1997). Prior information can be on an information class basis, such as the percent cover of forest and non-forest, or on a per pixel basis based on predictors such as slope, aspect, elevation and land form. Improvements in classification accuracy due to information class probability inclusion should be expected as information other than that provided by the training samples is being used in the classification. This information can be used to place a pixel into its most likely class if its likelihood of belonging to two classes is the same. However, San Miguel-Ayanz and Biging (1997) found no significant improvements when class probabilities were included in classification efforts.

McGwire *et al*. (1996) provide a concise outline of the assumptions (sample independence, stationarity, and multivariate normality) that the maximum likelihood classifier is based on as well as provide examples of how easily these assumptions may be violated in many image classification studies. They state that an "inappropriate match between classification technique and data characteristics will likely increase the error in derived map products." Fuller *et al*. (1994) is one example of a study that used the maximum likelihood classifier despite uncertainty about the normal distribution.

Prior to classification using the maximum likelihood classifier, or any other supervised classification technique, the spectral classes must be identified. Bolstad and Lillesand (1991) identify spectral training as the most substantial barrier to traditional supervised classifications. Unsupervised classifications remove the operator burden of spectral class training by automatically grouping pixels based on user-defined parameters and spectral statistics. This method also has drawbacks, such as *a posteriori* labeling of spectral classes by the analyst which is time consuming and can introduce bias. For example, San Miguel-Ayanz and Biging (1997) labeled clusters by examining the clusters one at a time and comparing them with the reference map and other information. In response to this overhead, numerous hybrid approaches have been devised that combine both supervised and unsupervised approaches in some manner (e.g., Chilar *et al*., 1998).

If satellite-assisted land cover determination is to become an operational procedure in FIA, the repeatability (and accuracy) of image classification algorithms must be improved and analyst time must be reduced. Multi-stage iterative classifications such as those presented in Jensen *et al.*, 1987, in Rutchey and Vilcheck (1994) and in Miguel-Ayanz and Biging (1996 and 1997) have resulted in improvements in classifier accuracy over single stage methods. Jensen *et al.* (1987) introduced "cluster busting," a technique that separates confused spectral classes by re-clustering pixels in the confused classes. The decision rule for labeling the clusters or for deriving the number of clusters were not specified. In fact, very little of the cluster busting process was described.

A more recent use of cluster busting was by Rutchey and Vilcheck (1994) in mapping the Everglades vegetation. They first generated 30 spectral classes using a minimum spectral distance formula and then identified each cluster and located it in the field using a GPS. Following image rectification, five points were located in each of the 30 spectral classes. The points had to be within the center of a minimum 3 by 3 homogeneous pixel block. In total, 129 reference points were located. The ground truthing consisted of finding the point via GPS and then estimating the percent cover of vegetation within a 20 by 20 meter grid. After further verifying the position of the reference points within the spectral classes, the spectral classes were compressed into19 vegetation classes. Each of the 129 reference points were labeled with a vegetation class and used as seed pixels for generating training signatures. Signatures that were not representative of the class were eliminated following a visual analysis of spectral class ellipses. All remaining signatures were combined into one file for all 19 vegetation classes. Ellipses for all 19 classes were then viewed to assess spectral uniqueness.

A maximum-likelihood (Bayesian) classification was then performed on the original unrectified image using the 19-class signature file. The classified image was then compared with 1:24,000 color infrared photography to assess the performance of the classification. Two vegetation classes (sawgrass/cattail sparse and tree islands) were deemed to be poorly classified. The sawgrass/cattail sparse pixels were broken into five new spectral classes, which were then grouped into two classes following interpretation of large-scale aerial photography. This iteration allowed a new cover type (Periphyton) to be identified, making a total of 20 vegetation classes.

The tree islands were dealt with by digitizing the outlines of the tree islands from the original unrectified image. The tree island areas were then extracted from the unrectified image and cluster busting was performed to produce 10 new classes which were then collapsed into four of the already existing vegetation classes. All vegetation classes were mosaicked into the maximum-likelihood classification output image to create a 20-class vegetation map.

Accuracy assessment was carried out using 241 GPS-located points. The accuracy was assumed satisfactory as long as there was a 5-pixel class majority within the center of a 3 by 3 pixel block, with the site being the center pixel. The 20-class map achieved 70.9% overall accuracy. A 12-class map that combined the 13 density dependent classes into 5 classes achieved 80.9% overall accuracy.

San Miguel-Ayanz and Biging (1997) compared five classification methods for mapping natural resources in Spain, including supervised, unsupervised and a multi-stage approach. They also compared the performance of Landsat TM imagery versus SPOT imagery. The multi-stage methodology required iterations that identified one or two information classes at a time. Each iteration required a band selection that maximized the separability between the remaining cover types. For the Landsat imagery, nine new bands were derived through band ratioing, principal components analysis, and texture analysis. Several spectral separability indices were tested including transformed divergence and Jefferies-Matusita.

The reference data used was a recent forest map of Spain generated from 1:30,000 panchromatic aerial photography. Homogeneous polygons were delineated on the map and then visited on the ground for verification. The 449 cover types identified in the study area were grouped into 21 cover types. A total of 396 training windows of 4 by 4 pixels were chosen, stratified by each cover type.

The iterative method classifies one or two classes at a time with the class that attains the highest average accuracy being masked, and the next iteration being performed on the unclassified part of the image. If two classes obtain high average accuracy, then both are masked. The average accuracy is based on the average of the user's and producer's accuracy for that class. A key component is that optimal band selection is carried out for each iteration to maximize spectral separability. The iterations continue until there is no real improvement in the accuracy of the cover types that remain.

The accuracy assessment of the final stratified map was based on test areas that were the same size as the training samples (4 by 4 pixels). A minimum of 20 samples per cover type was used. A majority rule that assigns labels based on the most frequently occurring class within the test area was used, and the labels were compared with the reference data. The results showed that the TM classification using the multi-stage approach was significantly better than the other approaches tested at the 99% confidence level. Although the overall accuracy of the multi-stage approach was only 65.56%, greater than 72% of the area was classified with 90% accuracy.

The iterative approaches described above make the fullest use of the information content of the image in a repeatable manner. Along similar lines is a new process called Iterative Guided Spectral Class Rejection (IGSCR) (Wayman *et al*., 1999). This technique has been developed and tested with favorable results. User input is minimal, subjectivity is all but eliminated, and accuracy is high. The most substantial benefit of this method is that the cluster labeling associated with unsupervised classification becomes objective and is based on pre-determined criteria such as percent homogeneity (based on user's accuracy) and the number of reference pixels in each spectral class. This results in a potentially large number of spectral classes, each of which has known information class properties. A more detailed description of this process is provided in the methods section.

4.5.2 Unclassified Pixels

Unclassified pixels result from a classification in which a threshold is applied. Thresholding identifies the pixels in a classified image that are most likely to be classified incorrectly (ERDAS, Inc, 1997). The thresholds are most often derived from the distance image histogram and can be selected visually or defined mathematically (ERDAS Inc, 1997). White *et al*. (1995) dealt with spectral classes that could not easily be placed into information classes by finding the normalized difference vegetation index (NDVI) ranges from known information classes and using those ranges to label the unknown spectral classes. Human interpretation can also be used to label pixels based on contextual clues and knowledge of the area under study.

4.5.3 Post Classification Corrections

Some contextual information may be derived from using polygonal reference data or a high density of reference points. However, the use of knowledge-based corrections is the only way to add true context to a classification (Fuller *et al*., 1994, Ton *et al*., 1991). Fuller *et al*. (1994) state that expert opinion does offer a simple solution to the criticism that per-pixel classifiers often make absurd classification errors. One negative is that image analyst intervention adds subjectivity to a classification.

4.5.4 Spatial Filtering

The use of spatial filters to remove noise or "salt and pepper" in a classified image is common practice in image classification (e.g., Bauer *et al*., 1994, Fuller *et al*., 1994, White *et al*., 1995). Most software packages offer filters of different window sizes and filter operations. A commonly used filter is a 3 by 3 scan majority filter. This filter passes a moving window over the image and assigns the class to the center pixel based on the class majority within the neighborhood. This process acts as a low-frequency filter, smoothing the data. However, care must be taken not to remove useful information by using too coarse a filter (Muller *et al*., 1998).

4.6 ACCURACY ASSESSMENT

Accuracy assessment is an integral component of any classification and the constant stream of literature is a testimony to the importance of a meaningful appraisal of the output map (see Richards, 1996; Muller *et al*., 1998; Stehman and Czaplewski, 1998). Recommended reading on sampling design is Stehman and Czaplewski (1998). They present a thorough review of the fundamental principles of sampling design and analysis for assessing thematic map accuracy.

Detection of thematic map errors depend on both the accuracy of the classification and on the number and type of validation data (San Miguel-Ayanz and Biging, 1997). If there is no restriction on the number of ground points or otherwise known locations, the random sample design is most statistically robust. If some restriction in the number of samples that can be taken applies, then the stratified random method is next best (Muller *et al*.,1998; Congalton and Green, 1999). This ensures that at least some of every class will be represented in the known locations and that there is a good spread of intersample distances.

Sample unit size must also be considered. Most samples are point locations that are assigned to a single pixel on the image. Sample points often have greater spatial detail than the map (Stoms, 1996) but this can be adjusted for by either using a spatial filter or using some majority rule within a given window size (e.g., Rutchey and Vilcheck,1994). There is the option of using multiple pixels of even polygons as the sample unit (e.g., Fuller *et al*., 1994; Stoms, 1996; San Miguel-Ayanz and Biging, 1997). Using polygonal reference data allows boundaries to be captured and makes the assessment more robust. Generally the sample unit should be the same size as the minimum mapping unit of the final output map.

The accuracy of the ground point locations is extremely important in accuracy assessment. Positional inaccuracy of validation data often leads to poor classification results, even though the thematic accuracy may be high. Bauer *et al*. (1994) correctly point out that map accuracy can not be higher than the accuracy of the validation data at the pixel scale.

The rule of thumb value of at least 50 (Congalton, 1991) known ground locations for each information class is generally accepted, though this number increases (to 75-100) if the area is large (>1 million acres) or if there are many classes (>12). The smallest samples are usually 20-30 samples (e.g., Edwards *et al*., 1998) per stratum, however, a higher number narrows the confidence interval of the accuracy assessment (San Miguel-Ayanz and Biging, 1997).

It is important that a distinction be made between classifier performance and map accuracy (Richards, 1996). A classifier may perform well within the training data, but this only extrapolates to the entire image if the sample training data was adequately representative of the population. Robust accuracy assessments are carried out using validation data that is independent of the training data.

The standard for reporting accuracy in most published work includes a contingency matrix (or error matrix) from which values for overall accuracy, Kappa (estimated by KHAT), KHAT variance and Z-scores are derived. Overall accuracy is the sum of the major diagonal (correctly classified points) divided by the total number of points assessed. The KHAT statistic is an estimate of the actual agreement minus the chance agreement (Rutchey and Vilcheck, 1994) and hence gives an indication of how much better the classification is than chance alone. For example, a KHAT of 0.765 means that the classification performed 76.5% better than if the pixels were classified in a random manner. KHAT values range between +1 and -1. A negative KHAT means that the classification was worse that what chance alone would have provided. Three groups of KHAT values are generally recognized: KHAT > 0.80 represents strong agreement (that the classification was better then chance), 0.40-0.80 represents moderate agreement, and < 0.40 represents poor agreement (Congalton and Green, 1999). The KHAT statistic is valuable in assessing the performance of a classifier, because unlike the overall accuracy, it uses the entire contingency matrix and not just the main diagonal values (San Miguel-Ayanz and Biging, 1997).

The KHAT statistics are also used to formulate Z-scores which can be used to test for classification significance or to test if there is a difference between two independent classifications. The Z-scores are the observed values of equations 1 and 2 (from Congalton and Green, 1999). If the Z-score exceeds 1.96 then the single classification is significantly different (at the 95% confidence level) than a random classification (for equation 1) or two independent classifications are significantly different than each other (for equation 2).

$$Z = \frac{\hat{K}_1}{\sqrt{\hat{\mathrm{var}}\left(\hat{K}_1\right)}}$$

Equation 1

$$Z = \frac{\left|\hat{K}_1 - \hat{K}_2\right|}{\sqrt{\hat{\mathrm{var}}\left(\hat{K}_1\right) + \hat{\mathrm{var}}\left(\hat{K}_2\right)}}$$

Equation 2

The user's accuracy, the producer's accuracy, and the overall accuracy are also calculated from the error matrix. The user's accuracy reflects the "probability that a pixel classified into a given category actually represents that category on the ground" (Lillesand and Kiefer, 1994). The producer's accuracy indicates how well reference pixels were classified. The user's accuracy reflects commission errors while the producer's accuracy reflects omission errors.

It is useful to understand and perhaps categorize errors into location, generalization, and genuine misclassification (Stoms, 1996). Muller *et al*. (1998) suggest classing errors as spectral, spatial, or temporal errors. This is important in order to understand where the classifier performed well, and how the performance relates to variables such as type and amount of reference data in each information class.

4.7 FROM CLASSIFICATION TO FOREST AREA ESTIMATES

Estimates of map accuracy calculated from contingency tables can be improved if one has knowledge of the true map category marginal proportions using a variation of double sampling (Card, 1982). Reams and Van Deusen, (1999) describe the current FIA area estimation. A two-phase or double sample design is used where the less accurate data is the satellite-derived map, and the more accurate and costly data are permanent ground plots. A sample of n points/pixels (from the permanent ground plots) is located on the map and the true map categories are determined for each point. The n points are allocated as a simple random sample. This results in a two-way contingency table where n_{ij} is the number of points in the sample i and the map category is j. The derivation of unbiased estimates of classification accuracy is dependent on knowledge of marginal map category proportions that are a by-product of image classifications (Card, 1982). This reduces the variance of estimates of true map category proportions (Wynne *et al*., 2000). The use of simple random sampling or stratified sampling does not affect estimations of true map proportions, only of variance estimates (Reams and Van Deusen, 1999).

An example of forest area correction (following Wynne *et al*., 2000) carried out for the 8 county Ridge and Valley classification is as follows: A contingency table and marginal proportions of 76.87

% forest and 23.13% non-forest are derived from the classified image (Table 2). The true forest map marginal (Card, 1982) is estimated as follows:

Proportion forest = 0.7687 (157/186) + 0.2313 (12/54) = 0.7002

The variance of the percent forest (pf) is:

V(pf) = (0.7687 - 0.6488)(0.6488)/ 184.48 + (0.2313 - 0.0514) (0.0514)/55.52 =0.0005882

This is where/how several numbers in the above formula are derived.

(157/186) (0.7687) = 0.6488

0.7687 (240) = 184.48

(12/54) (0.2313) = 0.0514

0.2313 (240) = 55.52

Table 2. Contingency table for TM map categories and FIA ground plots (truths). Table represents FIA ground plots and classified TM pixels in eight Ridge and Valley counties in Virginia.

	TM Map Categories		
True	Forest	Non-forest	Total
Forest	157	12	169
Non-forest	29	42	71
Total	186	54	240

Map marginal proportions (from classified image): Forest 0.7687 Non-forest 0.2313.

The standard deviation of the percent forest area estimate is 2.43% (sqrt (0.0005882)). This results in an approximate 95% confidence interval estimate of 65.17% to 74.87%.

Area corrections based on AVHRR satellite imagery are compared with the traditional dot-count methods in Lannom *et al*. (1995). They found no significant difference between the two forest area estimates and concluded that AVHRR imagery could be substituted for aerial photography for estimating forest area. Wynne *et al*. (2000) present a perspective on sensor suitability for forest area estimates, comparing the traditional aerial photography dot count method with Landsat TM and AVHRR. They conclude that multispectral earth resource sensors such as TM, ETM+ and SPOT are preferred over AVHRR for forest/non-forest stratification, primarily because there are no acceptable methods of estimating the standard error from AVHRR data, and the resulting maps are more useful in fragmented landscapes characteristic of the eastern U.S. and western Europe.

I compared the error-corrected forest area estimates derived from the TM data with the traditional FIA forest area estimates. This comparison is necessary to assess whether there is any significant difference between the estimates given by the two methods. I also compared the corrected area estimates from the Multi-Resolution Land Characteristics (MRLC) classification with both the Phase 1 estimates and the corrected IGSCR estimates in order to evaluate the utility of an existing national land cover classification program for use in FIA.

Chapter 5
METHODS

5.1 LOCATION

If the objectives of this study (accurate forest area estimates and accurate forest cover maps) are met, the resulting protocol could be implemented over the entire southeastern United States. With this in mind, the study areas needed to be representative of the southeast, while being within the logistics of this study. Three physiographic regions (Coastal Plain, Piedmont, and Ridge and Valley) were selected within Virginia to best represent the landscape of the southeast (Figure 1). Specifically, multi-county areas were chosen within each physiographic region. The counties in the Coastal Plain region were Isle of Wight, Surry, Sussex, Dinwiddie, James City, Charles City, and Prince George (1,499, 089 acres total). The area is fragmented and dominated by crops such as soybeans and corn as well as short rotation timber crops (Figure 2). There is also a large amount of low-density residential land use. The Piedmont area was reduced to three counties due to the presence of clouds along the western side of the area of interest. The Piedmont counties were Louisa, Fluvanna, and Orange (731,778 acres total). The Piedmont has a large proportion of hardwood forest, along with shorter rotation softwood forest. The landscape is also highly fragmented. The Ridge and Valley counties were Scott, Wise, Dickenson, Smyth, Tazewell, Buchanan, Washington and Russell (2,434,529 acres total). The main characteristic of the Ridge and Valley region is that the valleys are dominated by non-forest use such as agriculture, while the ridges are predominantly forested. The image also included an area of the Cumberland Plateau.

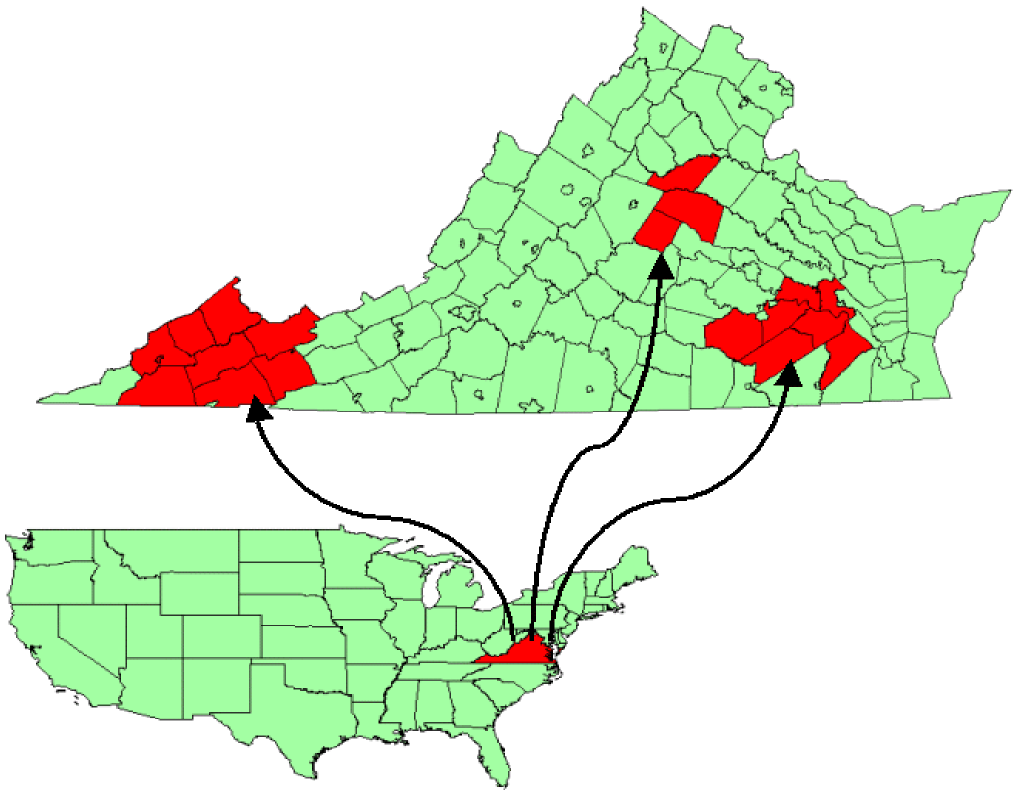

Figure 1. Study locations in three physiographic regions of Virginia.

5.2 IMAGERY

Table 3. TM Images used in study (courtesy USDA Forest Service, NCASI, and Virginia Department of Forestry).

Path	Row	Date
15	33	5/14/98
16	34	9/26/98
18	34	11/11/98

Three path/row combinations (Worldwide Reference System-2) were chosen for this study to coincide with the three physiographic regions that dominate the study area: path 15/row 33 (Coastal Plain), path 16/row34 (Piedmont), and path 18/row 34 (Ridge and Valley) (Table 3). The scenes were rectified to a 1999 Virginia Department of Transportation 1:24,000 digital roads coverage (Lambert projection, NAD 83 datum, units meters) derived from paper based county maps and 10 m SPOT panchromatic imagery. All imagery was

Figure 2. Example of the fragmented nature of the Coastal Plain landscape.

rectified to less than 0.5 pixel RMSE using a first-order polynomial and were resampled to 30 m pixels using the nearest neighbor method. Rather than relying solely on the RMSE values to assess the quality of the rectification, the road coverage was overlayed on the rectified image and a visual analysis was used to judge how well the image was rectified. This was an iterative process with GCPs being moved slightly in order to improve the vector/image correspondence. Although this resulted in a higher error from the checkpoints (13 m for the Coastal Plain, 21 m for the Piedmont and 40 m in the Ridge and Valley) I believe the visual analysis provided a better rectification.

5.3 REFERENCE DATA

5.3.1 Helopolys

In this study, a new data collection technique using precise airborne planimetry was applied. Forest and non-forest polygons were captured in a helicopter equipped with the TrimFlight GPS system. The method is a variation of the primary sampling unit technique (see Bauer *et al*., 1994) that bypasses the use of aerial photography, ground verification and digitizing entirely. The use of helicopters in reference data collection or validation is not new (see Schreuder *et al*., 1995; Stoms, 1996; Muller *et al*., 1998); however, this is (to our knowledge) one of the first uses of helicopters for airborne planimetry. A Bell Jet Ranger 206B3 helicopter equipped with real-time GPS equipment was used to capture thirteen and seventeen forest and non-forest "helopolys" in Louisa (Piedmont) and Isle of Wight (Coastal Plain) counties, respectively. Although a Bell Jet Ranger helicopter was utilized in this study, a piston engine helicopter would provide the same results at approximately half the cost. Collection speeds ranged from near hovering to 50 mph and heights ranged from 500 to 1000 feet above terrain. Speed and flying heights were dependent on polygon properties, cover type, and on proximity to housing and livestock. The TrimFlight DGPS System typically achieves better than one-meter accuracy utilizing real time differential GPS (DGPS). Collection information is provided to the operator through a moving map display so polygons can be viewed as they are collected. This is a useful feature when locating the start/finish location to close large polygons.

The first collection of helopolys was in Louisa County. Pre-planning the helopoly collection in Louisa required the qualitative assessment of spectral class locations within the county to determine their spatial distribution. An ISODATA clustering algorithm (30 classes) was carried out on the image. A location was then visually selected on the 30-class image that contained most of the spectral classes and areas of interest such as water/road interfaces and water/field interfaces. These areas are of specific interest because they allow the sampling of borders between information classes while still meeting the non-forest criterion. Checking the spectral class representation required each spectral class to be highlighted on the image processing software, and the locations noted. Once all spectral class locations were checked, the extent of the collection area was set. One rare spectral class was noted but due to the small area and the nature of the feature (the Lake Anna DOE site) the class was avoided with no real loss in image information.

In order to make full use of helicopter time the collection area needed to be as compact as possible. The sample also needed to represent areas of each spectral class. Once we were airborne and over the collection area our choice of a " collect as much as possible as quick as possible" sample design was validated by the almost impossible nature of locating and flying pre-defined polygons with any economy. It was also fortuitous that fragmented landscape was composed of a relatively small number of cover types. For example, in the collection area there were mixed hardwood forests, pine plantations, soybean fields, cornfields, pastureland, water, and developed areas all of which were representative of the surrounding area. One advantage of this two strata collection method is that we could collect large polygons that contained more than one cover type, as long as the cover types were separated into forest or non-forest. The helicopter allowed us access across classes that would usually be inaccessible to traditional field crew GPS collection such as land/water interfaces.

The in-flight protocol required a suitable cover type to be located and points to be taken at each corner of the polygon. No effort was made to further divide the forest or non-forest cover types as emphasis was on maximizing polygon size. This could be amended so that specific forest types could be recorded for higher level classifications. The main assumption of our collection methodology was that if we captured all of the land cover types (and age classes within certain cover types) in the collection area, then we would have a high chance of capturing the spectral variation in the collection area also. Digital photographs of each polygon were taken to aid in polygon labeling.

Thirteen polygons were recorded in Louisa in approximately 90 minutes, however, one non-forest polygon was discarded due to a definitional disparity. The resulting reference data comprised 1343 acres. Seventeen polygons totaling 2192 acres were recorded in Isle of Wight County. Because the vector to raster conversion algorithm included pixels that were outside the reference data boundary, the reference polygons in both counties were buffered inward 50 feet, or half a pixel. This buffering resulted in a better correspondence between the original polygon boundaries and the image.

There are many obvious advantages to this technique including high positional accuracy and access to all cover types on any terrain. Fast turnaround (real time differential correction, output in GIS-ready format), certainty of polygon homogeneity, and use of data for training and (potentially) validation are also benefits. The drawbacks of this method are that better statistical inferences can be drawn from random samples than clusters due to the effects of autocorrelation (Bauer *et al*., 1994; Dobbertin and Biging 1996) and, of course, the cost.

A limiting factor of many image classification studies is that both the training and accuracy assessment data are derived from homogeneous areas or have poor positional accuracy. In extreme cases this results in a classification accuracy that is more related to the reference data than to the classification algorithm (Muller *et al*., 1998). The helopoly reference data has similar attributes of the maplet (Stoms, 1996) in that it is an "exhaustive census of the land cover over a small extent." This affords a high degree of confidence that the helopolys will encompass the variability found in any one cover type, and hence provide excellent training data. The high positional accuracy of helopolys allows some confidence to be placed in the accuracy of the reference data, and also allows the boundary pixels in the image to be included in the spectral class building stage. Another advantage of high positional accuracy is the potential for improvements in registration between the reference data and the image. If there is some misalignment between the data then a correctly classified pixel may be called incorrect, simply due to misregistration (Cracknell, 1998; Dai and Khorram, 1998; Muller *et al*., 1998). This is less of an issue in TM and SPOT imagery as the images themselves are often only geographically corrected to within 0.5 pixels (15 m and 10 m, respectively). However, the forthcoming high resolution satellites will require a higher degree of reference data positional accuracy than is generally accepted today.

5.3.2 Buffered Points

The second data collection method used in this study involved the collection of GPS points in homogeneous areas. The points were then buffered depending on the homogeneity radius around the point. The issues of accuracy, terrain access and context were addressed by varying the buffer radius depending upon the amount of visible homogeneous landscape. Both forest and non-forest reference data were collected in this manner. Due to the difficulty of ensuring homogeneity of forest from the roadside, the forest reference points were acquired following identification of homogeneous forest areas in the field and then locating them on the raw TM imagery. This method was used in the Piedmont and in the Ridge and Valley.

5.3.3 Heads-up Digitizing

The third method for data collection was that of heads-up digitizing on the raw image. The use of expert knowledge in reference data collection is intuitive and is perhaps the most cost-effective method of data collection. The sources of knowledge in this study were USDA Forest Service employees in the Jefferson National Forest at both the Wise and Mt. Rodgers Ranger District Offices in the Ridge and Valley, and VDOF County Foresters in the Piedmont. Areas on the raw image that were known to be of contiguous forest cover were digitized as ArcView (Version 3.0a ESRI Inc, Redlands, California) shapefiles. This method is valid for areas such as National Forests and large private holdings but less valid in the highly fragmented areas such as the Coastal Plain.

5.3.4 Harvest Polygons

The fourth reference data source was that of polygons digitized or colected via GPS by VDOF County Foresters. Most of the reference polygons consisted of areas that were recent (less than four years old) harvests. The data are invaluable because without their inclusion clear-cut areas would not be considered forest by the classifier.

5.3.5 Air Videography and Orthophotography

The fifth and most promising data collection method utilized air-videography and digital orthophotography. This method was tested in the Piedmont physiographic region by VDOF staff. The technique utilizes 1994 digital orthophotography to obtain the location of the attribute that is then updated using current air-videography. Forest and non-forest polygons are heads-up digitized directly over the orthophotography. This reference data collection method shows much promise because it has most of the attributes of helopolys at a discounted cost (if the orthophotography has already been purchased).

5.3.7 Training Data Summary

The training data used in the Coastal Plain consisted of seventeen helopolys, 180 homogeneous forest points and polygons supplied by county foresters totaling 4,695 acres (2,866 acres forest and 1,829 acres non-forest) (Table 4). The Piedmont reference data consisted of helopolys, othophotography-derived polygons and harvest forest area polygons provided by county foresters and totaled 3,746 acres (2,259 acres forest and 1,487 acres non-forest). Training data for the Ridge and Valley region comprised point data and polygons digitized over the imagery and totaled 8,486 acres (4,959 acres forest and 3,527 acres non-forest). For operational purposes, only the helopoly method would be likely unsuitable as reference data given the current costs. This method provides the highest quality data with the fastest turnaround, however similar data can be derived via heads-up digitizing on both digital orthophotography and imagery where expert knowledge (in the form of county foresters or similar) is available. The amount of reference data was equally proportioned across physiographic regions to allow some intercomparision of the results.

Table 4. Summary of study area and reference data by physiographic region.

Physiographic Region	**Total area (acres)**	**Reference Data**			**FIA Points**
		Forest	**Non-forest**	**Total**	
Coastal Plain	1,499,089	2,866	1,829	4,695	121
Piedmont	732,000	2,259	1,487	3,746	103
Ridge and Valley	2,434,529	4,959	3,527	8,486	241

5.4 IMAGE CLASSIFICATION VIA ITERATIVE GUIDED SPECTRAL CLASS REJECTION

Iterative Guided Spectral Class Rejection (IGSCR) is a hybrid classification method that builds and labels spectral classes for use in supervised approaches such as the maximum likelihood classifier (see Wayman *et al*., 1999). The IGSCR algorithm is in essence an objective and guided "cluster busting" (Jensen *et al*., 1987, Rutchey and Vilchek, 1994) approach that uses specific rejection criteria and large numbers of reference pixels. The IGSCR method accepts and labels a spectral class when it meets the desired inclusion threshold and rejects it if it does not. In this case, the inclusion threshold required at least 90% homogeneity for spectral classes with a minimum of at least ten reference pixels. All pixels in spectral classes meeting the 90% homogeneity/minimum pixel test are labeled and removed from the original raw image (Table 5). The unlabeled pixels from the raw image are then clustered into new spectral classes and the next iteration begins (Figure 3). Each of the iterations builds on the number of pixels (and spectral classes) with known identity. Once the iterations are complete (based on user-defined parameters such as the percentage of pixels classified or the classification of all reference pixels), the known spectral classes are combined into single signature file. The pure spectral classes are then used with the maximum likelihood decision rule to classify the image.

The best illustration of the process is provided by a step through of the effects of one iteration on the image of interest (Figures 4-6). ISODATA clustering was performed on the raw images (Figure 4) using 500 classes for the first iteration. The class means were initialized along the principal axis, the convergence threshold was set at 0.975 and the variance was set to one standard deviation. The corresponding spectral classes for each of the reference pixels were then exported to a spreadsheet. A proportion (2%) of the training pixels were separated for validation purposes and the remaining pixels were then sorted by spectral class. The rejection criteria were then applied to each spectral

Table 5. SAS output showing which spectral classes (SPEC) met the minimum pixel (SUM) and the homogeneity (PROP) criteria.

The SAS System 09:22 Monday, May 15, 2000

OBS	SPEC	CLASS	_TYPE_	_FREQ_	HOMO	SUM	PROP
1	18	1	0	2	10	11	0.90909
2	139	1	0	2	10	11	0.90909
3	224	1	0	2	10	11	0.90909
4	280	1	0	2	12	13	0.92308
5	312	1	0	2	9	10	0.90000
6	386	1	0	2	19	20	0.95000
7	397	2	0	14	14	15	0.93333
8	422	2	0	10	10	11	0.90909

class using the SAS System (Release 6.12, SAS Institute Inc. Cary, North Carolina)(Appendix 1). Following the assignment of either forest or non-forest to the spectral classes that met the homogeneity criteria, the ISODATA images were recoded into forest or non-forest. The recoded images (Figure 5) were composed of those pixels whose identity was assigned, along with the remaining pixels that represented either mixed spectral classes or spectral classes that were poorly represented by the reference data. The next step was the extraction of the pixels with assigned identity from the original images (Figure 6). This resulted in an image with zero values in place of the removed pixels. The image with the removed pixels forms the basis for iteration two. This was the final step in the first iteration.

The images used for iteration two are composed of the iteration one residual pixels (that is, the raw image minus the pixels of known identity). To perform the second iteration, the residual image was classified into n ISODATA classes, where n is the number of reference pixels not classified in the previous iteration, divided by ten (the minimum number of reference pixels required for spectral class acceptance). The corresponding reference pixels were then exported and sorted as in iteration one. The spectral

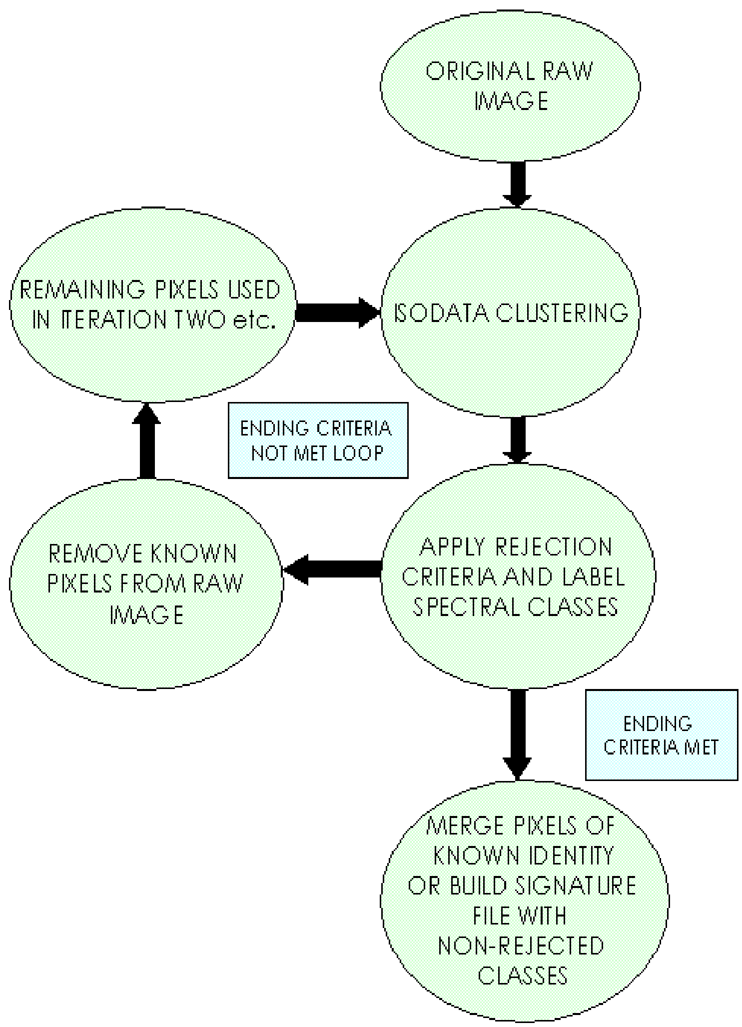

Figure 3. Simplified flowchart of iterative guided spectral class rejection process(Wayman et al., 1999, used with permission)

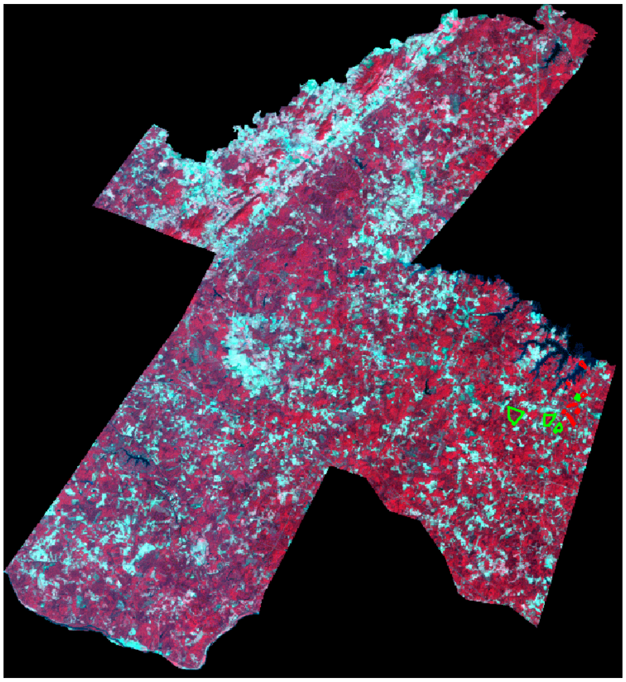

Figure 4. Example of raw image of three Piedmont counties prior to classification.

Figure 5. Example of image recoded into forest (green) and non-forest (tan) following first iteration.

classes that met the homogeneity threshold were then recoded in the second iteration ISODATA image. The labeled pixels were then extracted from the second iteration pre-clustered image in the same manner as in the first iteration. The image with the labeled pixels removed forms the image to be clustered in iteration three. The stopping criterion for this analysis was reached when no spectral class exceeded the rejection criteria in the last iteration.

The signature files for each iteration were edited to remove the spectral classes that were rejected in the spectral class building stage. The resulting spectral classes were merged into one signature file which became the basis of a supervised classification using the maximum likelihood decision rule. Band selection was performed using transformed divergence with all six non-thermal bands resulting in the highest separability. A 3x3 scan majority filter was also applied to the classified image.

5.5 ACCURACY ASSESSMENT

An independent validation was carried out by the Virginia Department of Forestry using information from FIA plots (Figure 7). These plots are independent of the training data and are

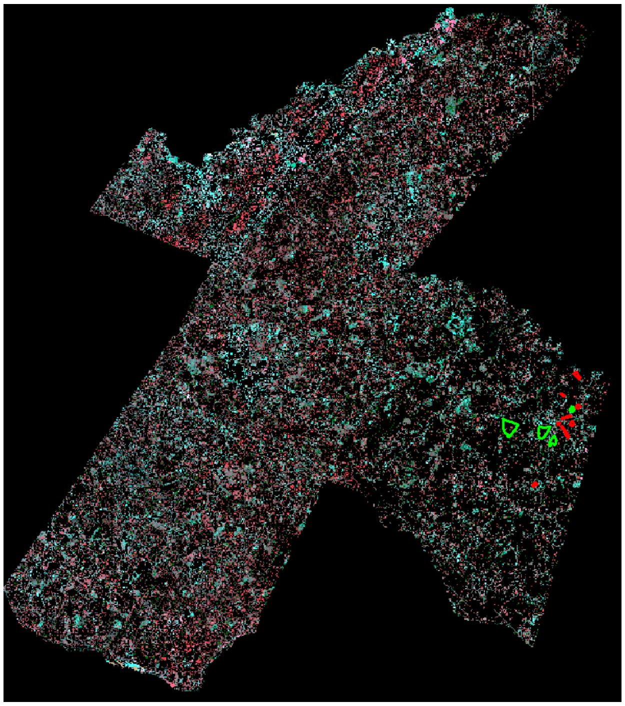

Figure 6. Image following removal of pixels labeled as forest or nonforest in iteration 1. This image is then clustered into a new set of spectral classes.

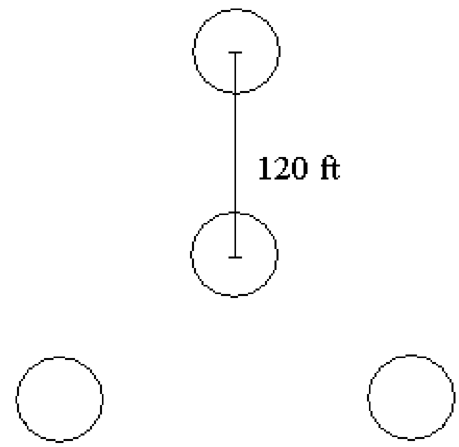

Figure 7. Layout of a single FIA plot with a plot center and three subplots.

randomly distributed over the study area. Land use is defined at plot center, based on FIA definitions. Intensification points derived from 10 m SPOT imagery were also used for accuracy assessment. The intensification points are systematically located verified by field crews. We also checked the aggregate area estimate against the 1991/92 and 1999 FIA estimates of forest cover.

5.6 AREA CORRECTION AND CALCULATION OF PHASE 1 ESTIMATES

Area corrections were carried out by the Virginia Department of Forestry using the methods outlined in Card, 1982. The small sample (or truth data) comprised FIA plots, which were then used to correct the satellite-based estimates. Intensification plots (additional points located on aerial photography and then visited on the ground for verification) were not used to correct the satellite-based estimates due to the positional inaccuracy of the plots

The traditional FIA photo-interpretation method of forest area estimation was applied in Virginia in early 2000. NAPP black-and white photography, mostly taken in 1994, but ranging from 1991-1996, was used. For each permanent ground plot the quadrant of the photo in which the plot fell was selected for photo-interpretation. Within this quadrant a grid of twenty-five points were classed into forest, non-forest, census water, and non-census water categories. The permanent ground plot was also classified. This constituted the large sample, with an intensity of one plot per 192 acres. The small sample was the ground classifications of the visited permanent plots and additional intensification plots. One intensification plot was selected for every two permanent plots and classified by an on-the-ground visit. Since Virginia is implementing an annual inventory system, three-fifths of the permanent plots had been revisited over the period of mid 1997 to early 2000, the time of area estimation. All of the intensification plots had been visited. The resulting small sample intensity is approximately one plot per 4550 acres, or roughly one small sample plot per 24 large sample plots. The small sample is used to adjust the large sample classification proportions, and to obtain standard error estimates, using the formulae provided by Li *et al*. (1992).

5.7 COMPARISON WITH MRLC CLASSIFICATION

The potential utility of existing the most recent (Version 2) land cover classification developed by the Multi-Resolution Land Characteristics Consortium (MRLC) (Vogelmann *et al*., 1998(b)) was also tested. The MRLC classification was based (in Virginia) on 1991-1993 Landsat TM imagery. The MRLC classified land cover into 15 distinct classes (Vogelmann *et al*., 1998(b)) (Table 6). The 15 classes were recoded in to forest and non-forest classes. Two images were created, one with class 15 coded as forest, and one with class 15 coded as non-forest. This was done in order to assess the impact of grouping class 15 as forest or as non-forest. An accuracy assessment was carried out using the same FIA points used in the IGSCR accuracy assessment.

Table 6. Description of MRLC classification categories.

MRLC Class	Description
Class 1	Water
Class 2	Low intensity developed
Class 3	High intensity developed
Class 4	Hay/pasture/grass (areas that "green up" before deciduous tree species leaf out)
Class 5	Row crops
Class 6	Probable row crops
Class 7	Conifer (evergreen) forest
Class 8	Mixed forest
Class 9	Deciduous forest
Class 10	Woody wetlands
Class 11	Emergent wetlands
Class 12	Barren; Quarry areas (excluding spectrally dark coal mines)
Class 13	Barren; Coal mines
Class 14	Barren; Beach areas
Class 15	Barren; Transitional (including clear cut areas)

Chapter 6

RESULTS AND DISCUSSION

6.1 ITERATION PROPERTIES

The Coastal Plain image required 9 iterations and the Piedmont required 17 iterations (Table 7). The Ridge and Valley region required a two step process. After 12 iterations, spectral class number 1 (inthis case shadow and water) was not labeled despite a large number of reference pixels remaining for that spectral class. All pixels that were in spectral class one were extracted from the iteration 13 image (those pixels unclassified after iteration 12) and then a separate IGSCR process was applied to these pixels. This resulted in the one stubborn class being broken into 17 classes that met the acceptance criteria. For operational purposes, some criterion such as "if after iteration *x*, any spectral classes have greater than *y* reference pixels, extract the pixels represented by those classes and perform a separate IGSCR algorithm on them" could be implemented.

The number of pure spectral classes gained in each iteration seems to reach an asymptote at approximately iteration seven for each of the physiographic regions (Figure 8). Computing time per iteration for each physiographic region was approximately 2 hours for the Piedmont, 6 hours for the Coastal Plain and 9 hours for the Ridge and Valley. Analyst time was approximately 1 hour for each iteration once the reference data was in a usable format. Automation of this process will reduce or eliminate analyst time between iterations.

6.2 MAP ACCURACY

The overall accuracies (based upon the independent FIA validation data) for all physiographic regions were very similar (Table 8). The IGSCR with the 3x3 scan-majority filter performed the best in each case. The Kappa statistics and the Z-scores suggest all classifications were significantly better than random classifications. However, within each physiographic region there were no significant differences between the results at the 95% confidence level. There were also no

Table 7. Summary of maximum spectral class number (user-specified ISODATA parameter), number of reference pixels and the number of pure spectral classes associated with each iteration.

	Coastal Plain		Piedmont		Ridge and Valley			
Iteration	Max # of Classes	Pure Classes	Max # of Classes	Pure Classes	Max # of Classes	Pure Classes	Max # of Classes*	Pure Classes*
1	500	246	500	239	500	290	191	10
2	752	55	553	30	1139	127	90	1
3	619	19	486	19	894	29	90	1
4	580	8	457	8	849	18	89	3
5	563	6	447	8	820	9	85	1
6	554	6	438	6	804	9	84	1
7	537	6	431	3	791	10		
8	527	2	428	1	771	4		
9	519	1	418	1	767	3		
10			411	2	763	4		
11			409	2	759	4		
12			406	5	722	2		
13			401	2				
14			397	4				
15			391	4				
16			383	1				
17			381	2				
						509		17
Total		**349**		**337**		**526**		

* Iterations carried out on pixels that were either shadow or water

significant differences between physiographic regions for the best overall accuracies (Table 9). One caveat attached to the accuracy results is that they were calculated using FIA plots. These points are classified as forest or non-forest based on FIA definitions and therefore are perhaps more rigorous than traditional accuracy assessment points. One example is that a recent clearcut is considered forest if the current land use has not changed. Also, the layout of the FIA plots may result in the land use

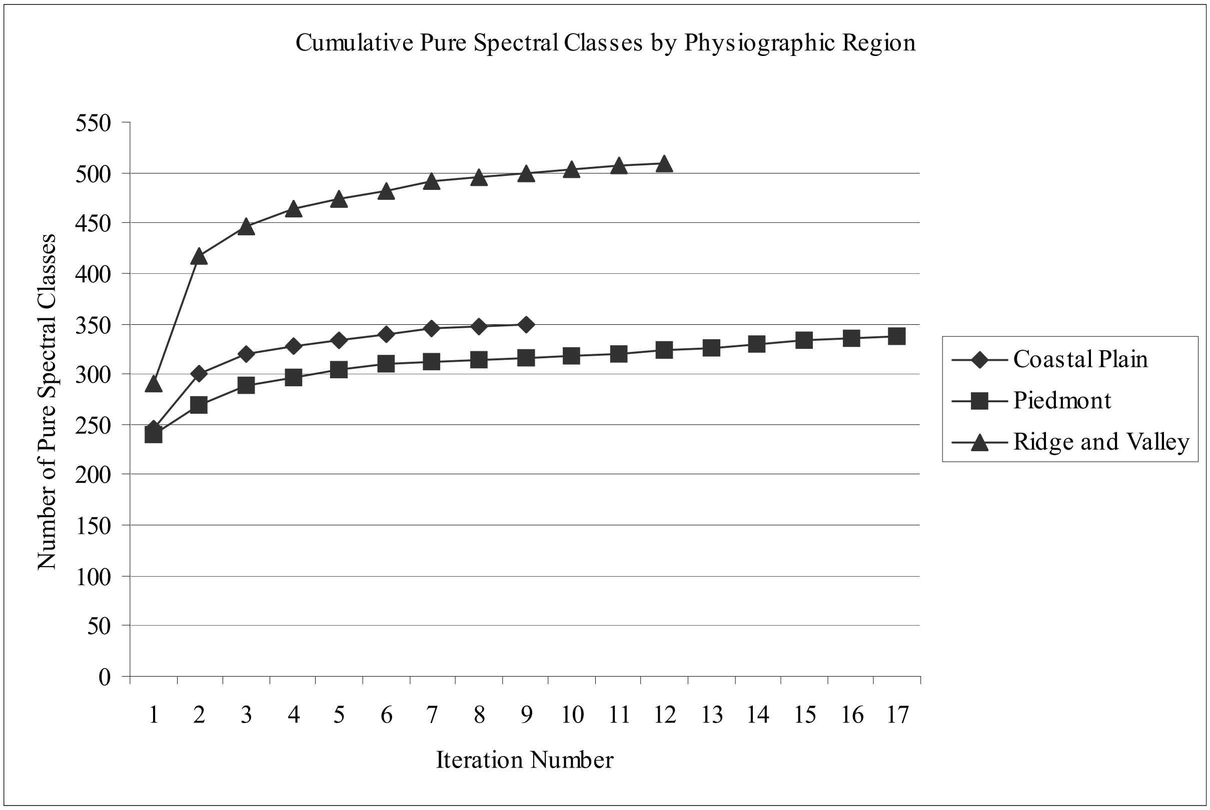

Figure 8. Cumulative frequency of spectral classes by iteration for all physiographic regions.

at plot center being different than the land use at the three surrounding subplots. This could be a source of misclassification error when a majority filter is applied to the classified image. For example, if the plot center is non-forest, the unfiltered image will call it non-forest. But, if the neighboring pixels are classed as forest, the non-forest pixel will be called forest following a 3x3 scan majority filter. This will result in that point being correct in the unfiltered image and incorrect in the filtered image. The case is reversed if the central pixel was forest surrounded by non-forest neighbors. That being said the results show that the 3x3 scan-majority filtered images have consistently better accuracy.

Table 8. Accuracy statistics for Phase 1 estimates and image classifications across all three physiographic regions (3x3 scan-majority filtered and unfiltered and MRLC class 15 as forest and as nonforest).

Method	**Overall Accuracy**				**Pairwise Comparison**			
	Overall	Kappa	Kappa Variance	Z-score	1	2	3	4
Coastal Plain								
1) IGSCR 3x3 Scan Majority Filtered	88.43	0.6969	0.00563	9.29	0.00	1.42	0.28	1.18
2) IGSCR Unfiltered	81.82	0.5328	0.00763	6.10	1.42	0.00	1.15	0.28
3) MRLC Class 15 Forest	86.78	0.6667	0.00587	8.70	0.28	1.15	0.00	0.90
4) MRLC Class 15 Non-forest	81.82	0.5663	0.00666	6.94	1.18	0.28	0.90	0.00
Piedmont								
1) IGSCR 3x3 Scan Majority Filtered	85.29	0.6724	0.00596	8.71	0.00	0.18	1.10	1.21
2) IGSCR Unfiltered	84.31	0.6526	0.00623	8.27	0.18	0.00	0.92	1.03
3) MRLC Class 15 Forest	81.37	0.5470	0.00699	6.54	1.10	0.92	0.00	0.12
4) MRLC Class 15 Nonforest	80.39	0.5323	0.00742	6.18	1.21	1.03	0.12	0.00
Ridge and Valley								
1) IGSCR 3x3 Scan Majority Filtered	82.92	0.5594	0.00362	9.29	0.00	0.21	0.60	0.45
2) IGSCR Unfiltered	81.67	0.5412	0.00367	8.93	0.21	0.00	0.38	0.24
3) MRLC Class 15 Forest	80.83	0.5078	0.00388	8.15	0.60	0.38	0.00	0.15
4) MRLC Class 15 Nonforest	81.25	0.5206	0.00382	8.40	0.45	0.24	0.15	0.00

Table 9. Comparison of the three physiographic region accuracy statistics for iterative guided spectral class rejection classification using the 3x3 scan-majority filtered images.

Method	**Overall Accuracy**				**Pairwise Comparison**		
	Overall	Kappa	Kappa Variance	Z-score	1	2	3
1) Coastal Plain	88.43	0.6969	0.00563	9.29	0.00	0.23	1.43
2)Piedmont	85.29	0.6724	0.00596	8.71	0.23	0.00	1.15
3) Ridge and Valley	82.917	0.5594	0.00362	9.29	1.43	1.15	0.00

6.3 FOREST AREA ESTIMATION

The traditional phase one forest area estimates were captured within the 95% confidence bounds for all classification results (Tables 10-12). This is no surprise given the large standard errors of the estimates. In fact, the standard errors are so large that the phase one estimates were captured within the 99% confidence bounds for all classifications except the Coastal Plain MRLC when Class15 was recoded as non-forest. The large standard error is a function of the reduced number of reference points (as the intensification points could not be used because of imprecise locations) and the reduced proportion correct in the TM-based classifications. A sample size of 500 (Greg Reams, pers comm) is a rule of thumb value for minimizing the standard error of the estimate derived via the inverse estimator (Card, 1982) for area correction. The results also weakly suggest that the IGSCR method might overestimate forest area, while the MRLC may overestimate forest area in all but the Piedmont.

Table 10. Area estimation statistics for the Coastal Plain region.

Method	**% Forest**	**Std Error**	**Approx 95% CI**		**Proportion Correct**	**n**	**Unadjusted % Forest**
			Lower	**Upper**			
Traditional FIA Phase I	**66.06%**	**1.08%**	**63.90%**	**68.22%**	**97.31%**	**260**	**67.27%**
IGSCR (3x3 filter)	**68.88%**	**2.93%**	**63.01%**	**74.75%**	**88.43%**	**121**	**70.76%**
IGSCR Non-filtered	69.45%	3.50%	62.46%	76.45%	81.82%	121	68.36
MRLC Class 15 Forest	69.84%	3.08%	63.69%	76.00%	86.78%	121	68.40%
MRLC Class 15 Non-forest	72.14%	3.34%	65.46%	78.83%	81.82%	121	66.70%

Table 11. Area estimation statistics for the Piedmont region.

Method	% Forest	Std Error	Approx 95% CI		Proportion Correct	n	Un-adjusted % Forest
			Lower	Upper			
Traditional FIA Phase I	**63.87%**	**1.91%**	**60.04%**	**67.70%**	**93.48%**	**184**	**63.45%**
IGSCR (3x3 Filter)	**65.52%**	**3.50%**	**58.51%**	**72.53%**	**85.29%**	**102**	**85.29%**
IGSCR Non-filtered	65.44%	3.59%	58.27%	72.62%	84.31%	102	70.20%
MRLC Class 15 Forest	58.12%	3.59%	50.94%	65.30%	81.37%	102	75.10%
MRLC Class 15 Nonforest	59.19%	3.83%	51.54%	66.84%	80.39%	102	72.77%

Table 12. Area estimation statistics for the Ridge and Valley region.

Method	% Forest	Std Error	Approx 95% CI		Proportion Correct	n	Unadjusted % Forest
			Lower	Upper			
Traditional FIA Phase I	**69.74%**	**1.22%**	**67.29%**	**72.19%**	**92.49%**	**493**	**65.75%**
IGSCR (3x3 filtered)	**70.02%**	**2.43%**	**65.17%**	**74.87%**	**82.92%**	**240**	**76.87%**
IGSCR Non-filtered	70.05%	2.48%	65.10%	75.01%	81.67%	240	73.95%
MRLC Class 15 Forest	70.53%	2.52%	65.48%	75.57%	80.85%	240	77.28%
MRLC Class 15 Nonforest	70.70%	2.50%	65.70%	75.70%	81.25%	240	77.17%

Chapter 7
CONCLUSIONS

7.1 TRADITIONAL PHASE 1 vs IGSCR vs MRLC

The primary objective of this research was to develop an algorithm to obtain forest area estimates using Landsat TM imagery that are comparable (in accuracy and in precision) to the current Phase 1 estimates. The results show that both IGSCR and MRLC can provide corrected area estimates to satisfy the accuracy component of the objective. I therefore conclude that there is insufficient evidence (at the 95% confidence level) to reject the null hypothesis. There is no difference in forest area estimates derived via the traditional FIA method those derived via the satellite-based methods. The precision (represented by the standard error) of the estimate, however, is less than that provided by the photo-based method (partially due to fewer points being used in the area correction). The IGSCR method provides the best estimate, though it is not significantly different than the MRLC results for a 95% confidence interval. The precision of the area estimates can be improved with a higher number of ground truth points (possibly in the form of intensification points taken in the field), and also by increasing the map accuracy.

The main advantage of the IGSCR over the MRLC for area estimation is that it can be repeated every 3-5 years, while the MRLC is, at present, only available once every 7-10 years. Another benefit of the IGSCR method is that it uses the FIA definition of forest. Both satellite-based methods have a distinct advantage over the Phase 1 method, they have a map output. The accuracy of the IGSCR and MRLC maps were not significantly different from each other, however the IGSCR method had consistently higher accuracy than the MRLC. The highest overall map accuracies were obtained with the IGSCR algorithm and the 3x3 majority filtered images.

The consistent best performance of the 3x3 majority filtered image may be explained in two ways. The first is that the FIA definition requires a forest to be at least 1 acre in size, and at least 120 feet wide. A single pixel is approximately 1/4 acre in size, and is therefore less than the FIA minimum mapping unit. A 3x3 majority filter increases the minimum mapping unit of the image to approximately 2 1/4 acres and therefore it is able to capture the FIA minimum mapping unit within its bounds. However, a non-forest area that met the FIA minimum mapping unit with 4 pixels in an unfiltered image would be filtered out if those four pixels occurred within an area that was forest in majority. The second possible explanation is that there is a higher likelihood in a 3x3 majority filtered image that if a point on the ground is forest, then the neighboring pixels will be forest. This means that the point is more likely to be classified correctly.

High map accuracy contributes to the accuracy of the area estimate but when area corrections are applied the map accuracy is less of a concern. The pairwise comparison (Table 7) shows that neither the two classifications (IGSCR and MRLC) nor the spatial filtering affected the overall accuracy individually at the 95% confidence level.

Overall map accuracies below 90% may not be adequate for forest managers and therefore further work is needed in order to improve map accuracy. We have identified several possible reasons for the low overall accuracies. As mentioned earlier, the current definition of forest in FIA is land at least 10 percent stocked by forest trees of any size, or formerly having had such tree coverage and not currently developed for non-forest use. The minimum area considered for classification is 1 acre and forested strips must be at least 120 feet wide. This definition is problematic for classifications using Landsat TM and similar sensors primarily due to the difference between what the sensor sees (land cover) and what is recorded on the ground (land use).

A 10 percent stocked forest is less forested than some non-forest urban areas. Unless there is a clear spectral difference between the poorly stocked forest area and the low density urban area, some misclassification is likely. The direction of misclassification (into forest or non-forest) will depend on the next most similar spectral class. A visual example of definitional errors is shown in figure 9. The tree-covered non-forest areas were correctly labeled as non-forest, which in turn resulted in low-density forest (with the same spectral signature) also being labeled as non-forest. This event is

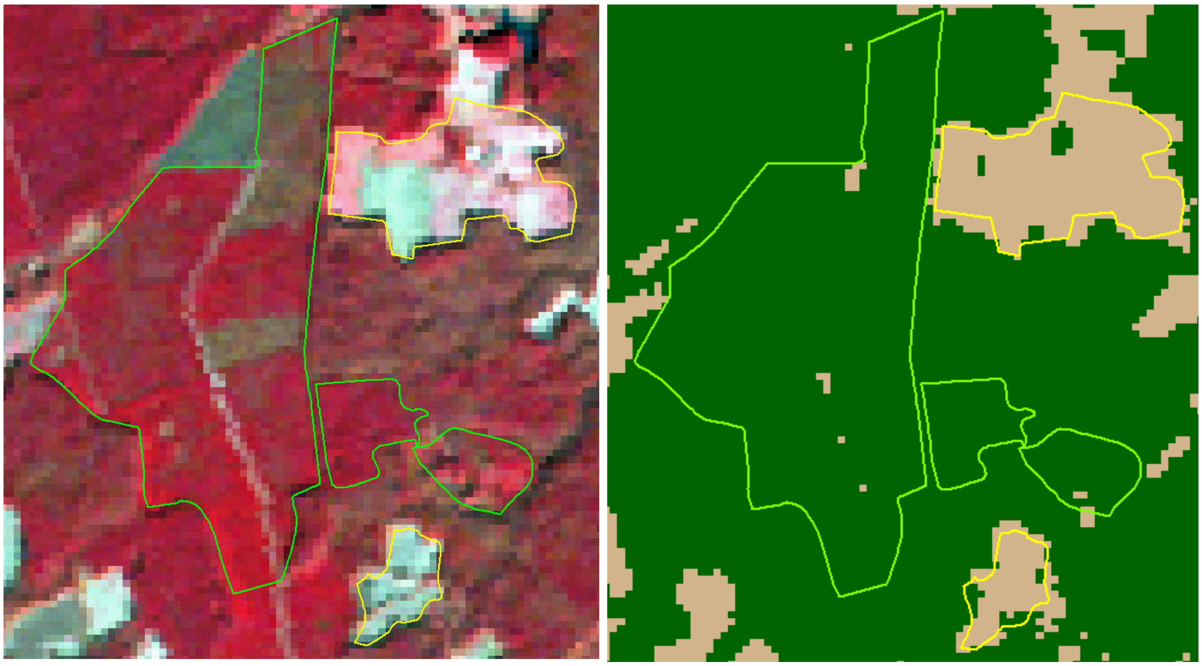

Figure 9. Example of classification results (raw image left, classified image right), dark green pixels and green polygons represent forest, tan pixels and yellow polygons represent non-forest.

likely a function of the training data and not the classifier. This situation may arise when there is a higher proportion of non-forest training data when compared to the non-forest that is actually present on the ground. Because the classifier relies on homogeneity of spectral classes, an uneven sample may result in a mis-labeling of spectral classes, even though they are homogeneous. The best safeguard against this occurrence is to have training data proportions that are approximately equal to the spectral class proportions, while concurrently capturing the full range of spectral variability in the area of study. There are certainly cases where two or more spectral classes are inseparable, in that case, no amount of training data can rectify the situation.

7.2 TYPE AND NUMBER OF REFERENCE DATA

All image classifications rely heavily on training data amount and quality. This study shows that land cover classes that cause the most problems must be adequately represented in the training data. An example of this is the difficulty in correctly labeling clearcut areas as forest. Recent clearcuts were

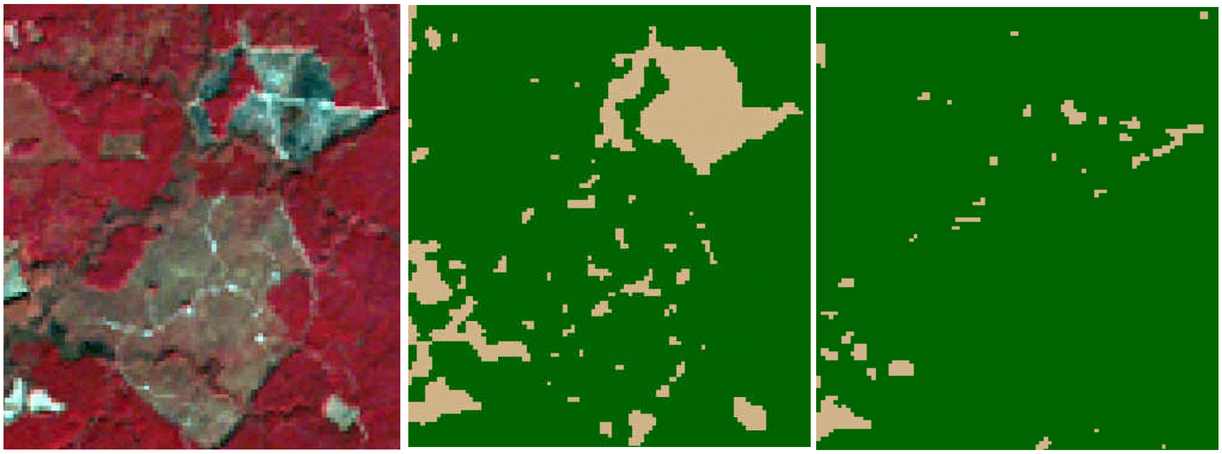

Figure 10. Example of raw image clearcuts in two stages (top left) with the older cut near the bottom. The classified image above (green forest, tan non-forest) shows the older cut is classified well but the newer cut is not. The classified image to the left is a result of increasing the minimum number of pixels required for homogeneity status from 4 to 10 and also adding clearcut polygons into the training data. The clearcuts are now more correctly classified.

not detected by the IGSCR algorithm when clearcut data were not in the training data set (Figure 10). However, when harvest polygons were added, the recent clearcuts were more accurately classified.

Although high quality training data are required for this hybrid classifier to perform well, the algorithm worked equally well with different training data sources. At present, the air-videography and orthophotography method is most promising due to the high positional accuracy, high spatial resolution, and currency of the videography update. The amount of training data needed has not been tested, but I believe that proportions approximating those used in this study are appropriate and would be easy to obtain operationally. The training data should capture as much spectral variability in the landscape as possible.

The use of a subset of FIA plots for training data may be feasible if (1) a number of intensification points are also collected and/or (2) the plot subset (e.g., the Phase III Forest Health Monitoring (FHM) plots) is buffered and mapped. The Ridge and Valley non-forest training data comprised approximately 250 buffered points compared with the 240 FIA (forest and non-forest)points used as validation. In order to use FIA point data as training and still retain separate points for validation, the number of points must be increased. The collection of non-forest points is an easier task than collecting homogeneous forest points and therefore the point-based method alone may not be ideal for forest training data. Forest training data digitized directly off the imagery proved to be cost-effective in the Ridge and Valley. The use of expert knowledge does introduce subjectivity into the classification process. However, the tradeoff between subjectivity and incomplete training data must be weighed.

Another issue concerning the use of FIA plots is the information on the plot homogeneity. The mapped plot data could potentially be used in some form of weighting of land use. This may improve the map accuracy and perhaps improve the area estimates.

7.3 SCENE SELECTION

The importance of a cloud-free scene cannot be overrated. The Coastal Plain and Piedmont classifications used leaf-on imagery, while the Ridge and Valley classification used leaf-off imagery. The Piedmont classification was hampered by clouds along the western edge of the region which resulted in a smaller area of study being classified. From the Ridge and Valley results, I conclude that a high quality image (i.e., cloud and snow free) may be more important than the season in which the imagery was acquired.

Multiple images have been shown to improve accuracies (e.g., Fuller *et al*., 1994; Schriever and Congalton, 1995), especially when a higher level of categorical specificity is desired. The main pitfall with multiple dates is that images must be coregistered to a high degree of accuracy. The rectification experience gained in this study leads to the conclusion that multiple dates may not provide additional benefits in mountainous regions (due to misregistration errors), but perhaps may be useful on the Coastal Plain and on similar topography.

7.4 RECTIFICATION

Image rectification has the potential to change an objective and repeatable process into a subjective and unrepeatable process, especially in regions of high relief. The rectification methodology followed in this study was not objective, however I feel that the method resulted in the best possible rectification short of terrain correction. Standardization of image rectification is a daunting task, yet the process is integral to the classification process. It is clear that standard protocols such as those used by MRLC and the Gap Analysis Program may provide a starting point to development of a protocol for FIA. Rectification is a key area in need of additional research, especially if multiple images are to be used in the classification process.

7.5 IGSCR PARAMETERS

There are numerous unexplored paths which may lead to improved map accuracy and more precise area estimates. This study used the most basic approach possible and therefore it provides a solid baseline from which to measure improvements in results. An obvious source for accuracy gains is in refinement of the IGSCR parameters, specifically the number of spectral classes in each iteration, the homogeneity threshold, and the minimum number of pixels.

Setting the initial number of spectral classes at 500 was somewhat arbitrary and “out of the box”. Traditional unsupervised approaches use fewer ISODATA classes primarily due to the difficulty in labeling the classes in an objective manner. The IGSCR process labels the classes objectively and therefore the traditional ceiling no longer applies. This attribute may be what distinguishes the IGSCR technique the most. There is a possibility that the high number of spectral classes is not needed, this will be investigated in further work.

There is some potential of maximizing the internal accuracy of the classifier by using a random subsample of the training data to set the parameters. This would only be possible if the algorithm were automated and may be a time consuming process. The main benefit to this approach may be in quantifying which parameters have the most effect on the output, and in quantifying what potential range in values should be considered for any given parameter.

Another alterative to using the combined signatures in a maximum likelihood classifier is to use the maximum likelihood classifier only on the pixels that remain confused after the final iteration. This would result in the pixels that were classified in each iteration staying in the spectral classes to which they were assigned to in that iteration. With the current methodology there is a possibility that, by combining all the spectral classes into one signature file, some pixels may be assigned to a different spectral class than to which they were originally assigned. This would pose no problems if a pixel did not change information classes, but it is an area in need of further investigation.

The training data and the inherent spectral variability in the image affect how many iterations are needed before the stopping criterion is met. The optimal number of iterations for the algorithm is unclear, further research needs to be conducted on the gains in accuracy and in the precision of area estimates. At present, a classification for a full scene would take less than a month to complete. This time could potentially be cut in half if the number of iterations was reduced, with possibly no net loss in the accuracy of the results.

7.6 STUDY LIMITATIONS

Along with the goal of obtaining accurate and precise area estimates, I wanted to develop an objective methodology, one that was repeatable across operators, time, and regions. The biggest limitation to this study is in the rectification process. As previously mentioned the methodology used was subjective and non-repeatable. The rectification error in the Ridge and Valley region was on the order of a pixel (30 m). This alone can cause spurious results in the reported classification accuracy. The use of a 3x3 scan majority filter may decrease the effects of misregistration. However, if repeatability is the goal then this error will result in two different answers if the process were to be repeated.

Another limitation is the type of reference data used. I took the approach of using whatever was available. This is less appropriate for an operational process, however it did serve to provide some indication of what reference data source works and what does not

The IGSCR parameters, such as the initial number for spectral classes, the spectral class numbers for successive iterations, the homogeneity criterion, and the minimum number of pixels, will all affect the classification outcome. I cannot conclude which setting is optimal for each parameter, I can only conclude that the parameters I used worked well.

The iterative guided spectral class rejection method developed and tested in this study is objective (save image rectification and reference data collection), repeatable, and subject to being nearly completely automated. Unlike many other classification techniques it does not require highly trained image analysts except, possibly, for geometric corrections and/or reference data collection. In-house code is being developed to enable the automation of the process. This should further enhance the repeatability by eliminating steps, such as recoding, which are often a source of operator error.

7.7 SUMMARY

The satellite-based methods can provide forest area estimates comparable to the traditional photo-based method. The IGSCR and MRLC classifications are comparable to each other in the corrected area estimates and in the errors of the estimates. Accuracy and precision of the satellite-based area estimates can be improved by increasing the number of points used in the area correction process.

Standardization of the rectification process and the reference data properties will increase the repeatability of the IGSCR method. Scene quality (i.e., the absence of clouds and snow) may be more important than phenological stage. However, scene selection may be more critical in areas of rapid change. In these areas, imagery should be chosen that maximizes the separability between forest and non-forest land use classes that are traditionally confused in satellite-based classifications.

The overall map accuracies for both satellite-based methods were between 80% and 89%, with the IGSCR method achieving consistently higher accuracies. Accuracy of the IGSCR map product may potentially be increased by altering the algorithm parameters such as the number of spectral classes per iteration, the minimum number of training pixels required per spectral class, and the percent homogeneity.

The MRLC is an attractive option to the IGSCR method since it is an existing program that obtains similar results. However, there are two main drawbacks to the MRLC. First, it may be too infrequent to be useful for FIA purposes due to the rapid changes in the landscape. Second, it will not produce map products that meet FIA definitions

Chapter 8
RECOMMENDATIONS

Following the completion of this study, several recommendations can be made for further research on the IGSCR algorithm/protocol:

- Investigate the IGSCR parameters such as homogeneity threshold and minimum number of reference pixels per spectral class
- Test effects of iteration number on area estimation and on classification accuracy
- Explore effects of validation sample number on current uncorrected area estimates
- Explore maximizing the accuracy of the classification using a random subsample of training pixels
- Make rectification more objective and repeatable
- Determine whether FIA plots are sufficient for training and validation (and, in general how to obtain suitable training data through an objective protocol)
- Determine if current use of the maximum likelihood classifier is a better approach than using maximum likelihood on the unclassified pixels alone
- Try the IGSCR approach in different physiographic regions of United States and in areas of rapid change

- Explore the possibility of higher categorical specificity, particularly within forested areas.
- Identify/characterize classification errors into components such as registration, mixed pixels, definitional, etc.

Chapter 9
LITERATURE CITED

Bauer, M.E., T.E. Burk, A.R. Elk, P.R. Coppin, S.D. Lime, T.A. Walsh, D.K. Walters, W. Befort, and D.F. Heinzen, 1994. Satellite Inventory of Minnesota Forest Resources, *Photogrammetric Engineering & Remote Sensing*, 60(3):287-298.

Bolstad, P.V., and T.M. Lillesand, 1991. Semi-automated Training Approaches for Spectral Class Definition, *International Journal of Remote Sensing*, 13(16):3157-3166.

Campbell, J.B., 1996. *Introduction to Remote Sensing* Second Edition, The Guilford Press, New York, 622 p.

Card, D.H., 1982. Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy, *Photogrammetric Engineering & Remote Sensing*, 48(3):431-439.

Cardille, J.A., D.W. Bolgrien, and R.H. Wynne, 1996. Variation in Landscape Metrics Derived from Multiple Independent Classifications, In: *Proceedings, Eco-Informa '96*, Lake Buena Vista, Florida, Vol. 11:749-754.

Chilar, J., Q. Xaio, J. Chen, J. Beaubien, K. Fung, and R. Latifovic, 1998. Classification by Progressive Generalization: A New Automated Methodology for Remote Sensing Multichannel Data, *International Journal of Remote Sensing*, 19(14):2685-2704.

Chojnacky, D.C., 1998. Double Sampling for Stratification: A Forest Inventory Application in the Interior West, Research Paper RMRS-RP-7, USDA Forest Service Rocky Mountain Research Station.

Colby, J.D. and P.L. Keating, 1998. Land Cover Classification Using Landsat TM Imagery in the Tropical Highlands: The Influence of Anisotropic Reflectance, *International Journal of Remote Sensing*, 19(8):1479-1500.

Congalton, R.G., 1991. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data, *Remote Sensing of the Environment*, 37:35-41.

Congalton, R.G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, CRC Press-Lewis Publishers, Boca Raton, Florida, 180 p.

Cracknell, A.P., 1998. Synergy in Remote Sensing - What's in a Pixel?, *International Journal of Remote Sensing*, 19(11):2025-2047.

Dai, X., and S. Khorram, 1998. The Effects of Image Misregistration on the Accuracy of Remotely Sensed Change Detection, *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1566-1577.

Dobbertin, M., and G.S. Biging, 1996. A Simulation Study of the Effects of Scene Autocorrelation, Training Sample Size and Sampling Method on Classification Accuracy, *Canadian Journal of Remote Sensing*, 27(4):360-367.

Ediriwickrema, J., and S.K. Khorram, 1997. Hierarchical Maximum-Likelihood Classification for Improved Accuracies, *IEEE Transactions on Geoscience and Remote Sensing*, 35(4):810-816.

Edwards, T.C., G.C. Moisen, and D.R. Cutler, 1998. Assessing Map Accuracy in a Remotely Sensed, Ecoregion-Scale Cover Map, *Remote Sensing of the Environment*, 63:73-83.

ERDAS Inc., 1997. *ERDAS Field Guide* Fourth Edition, ERDAS Inc., Atlanta, 655p.

Foody, G.M., 1999. The Continuum of Classification Fuzziness in Thematic Mapping, *Photogrammetric Engineering & Remote Sensing*, 65(4):443-451.

Fuller, R.M., G.B. Groom, and A.R. Jones, 1994. The Land Cover Map of Great Britain: An Automated Classification of Landsat Thematic Mapper Data, *Photogrammetric Engineering & Remote Sensing*, 60(5):553-562.

Glackin, D.L., 1998. International Space-Based Remote Sensing Overview: 1980-2007, *Canadian Journal of Remote Sensing*, 24(3):307-314.

Holmgren, P., and T. Thuresson, 1998. Satellite Remote Sensing for Forestry Planning - A Review, *Scandinavian Journal of Forest Resources*, 13:90-110.

Itten, K.L., and P. Meyer, 1993. Geometric and Radiometric Correction of TM Data of Mountainous Forest Areas, *IEEE Transactions on Geoscience and Remote Sensing*, 31(4):764-770.

Jensen, J.R., 1996. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice Hall, New Jersey, 316 p.

Jensen, J.R., E.W. Ramsey, H.E. Mackey, Jr., E.J. Christensen, and R.R. Shartz, 1987. Inland Wetland Change Detection using Aircraft MSS data, *Photogrammetric Engineering & Remote Sensing*, 53(5):521-529.

Kalkhan, M.A., R.M. Reich, and T.J. Stohlgren, 1998. Assessing the Accuracy of Landsat Thematic Mapper Classification Using Double Sampling, *International Journal of Remote Sensing*, 19(11):2049-2060.

Kaufman, Y.J., 1984. Atmospheric Effects on Remote Sensing of Surface Reflectance, *Proceedings of SPIE- The International Society for Optical Engineering*, 475:20-32.

Lannom, K.B., D.L. Evans, and Z. Zhu, 1995. Comparison of AVHRR Classification and Aerial Photography Interpretation for Estimation of Forest Areas, Research Paper SO-292, USDA Forest Service Southern Forest Experiment Station.

Li, H.G., H. T. Schrueder, D. D. Van Hooser, and G. E. Brink, 1992. Estimating strata means in double sampling with corrections based on second-phase sampling. *Biometrics* 48:189-199.

Lillesand, T.M., and R.W. Kiefer, 1994. *Remote Sensing and Image Interpretation*, Wiley and Sons, New York, 750 p.

McGwire, K.C., J.E. Estes, and J.L.Star, 1996. A Comparison of Maximum Likelihood-Based Supervised Classification Strategies, *Geocarto International*, 11(2):3-13.

Muller, S.V., D.A. Walker, F.E. Nelson, N.A. Auerbach, J.G. Blockheim, S. Guyer, and D. Sherba, 1998. Accuracy Assessment of a Land-Cover Map of the Kuparuk River Basin, Alaska: Considerations for Remote Regions, *Photogrammetric Engineering & Remote Sensing*, 64(6):619-628.

Reams, G.A., and P.C. Van Deusen, 1999. The Southern Annual Inventory System, *Journal of Agricultural, Biological, and Environmental Statistics*, 4(3):108-122.

Richards, J.A., 1996. Classifier Performance and Map Accuracy, *Remote Sensing of the Environment*, 57:161-166.

Riemann, R., and D.J. Alerich, 1998. The Remote Sensing Band-Bringing Together Remote Sensing Application and Research Within the Forest Inventory and Analysis Program, In: *Natural Resources Management Using Remote Sensing and GIS, Proceedings of the Seventh Forest Service Remote Sensing Applications Conference*, Nassau Bay, Texas, pp. 201-209.

Rutchey, K., and L. Vilchek, (1994). Development of an Everglades Vegetation Map using a SPOT Image and the Global Positioning System, *Photogrammetric Engineering & Remote Sensing*, 60(6):767-775.

San Miguel-Ayanz J., and G.S. Biging, (1996). An Iterative Approach for Mapping Natural Resources from Satellite Imagery, *International Journal of Remote Sensing*, 17(5):957-981.

San Miguel-Ayanz J., and G.S. Biging, 1997. Comparison of Single-stage and Multi-stage Classification Approaches for Cover Type Mapping with TM and SPOT data, *Remote Sensing of the Environment*, 59:92-104.

Schreuder, H.T., V.J. Labau, and J.W. Hazard, 1995. The Alaska Four-Phase Forest Inventory Sampling Design Using Remote Sensing and Ground Sampling, *Photogrammetric Engineering & Remote Sensing*, 61(3):291-297.

Schriever, J.R., and R.G. Congalton, 1995. Evaluating Seasonal Variability as an Aid to Cover-Type Mapping from Landsat Thematic Mapper Data in the Northeast, *Photogrammetric Engineering & Remote Sensing*, 61(3):321-327.

Stehman, S.V., 1997. Estimating Standard Errors of Accuracy Assessment Statistics under Cluster Sampling, *Remote Sensing of the Environment*. 60:258-269.

Stehman, S.V., and R.L. Czaplewski, 1998. Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles, *Remote Sensing of the Environment*, 64:331-344.

Stoms, D.M., 1996. Validating Large-Area Land Cover Databases with Maplets, *Geocarto International*, 11(2):87-95.

Ton, J., J. Sticklen, and A.K. Jain, 1991. Knowledge-Based Segmentation of Landsat Images, *IEEE Transactions on Geoscience and Remote Sensing*, 29(2):222-231.

Verbyla, D.L. 1995, *Satellite Remote Sensing of Natural Resources*, Lewis Publishers, New York, 224 p.

Vogelmann, J.E., T. Sohl, and S.M. Howard, 1998(a), Regional Characterization of Land Cover Using Multiple Sources of Data, *Photogrammetric Engineering & Remote Sensing*, 64(1):45-57.

Vogelmann J.E., T. Sohl, P.V. Campbell, and D.M. Shaw, 1998(b), Regional Land Cover Characterization Using Landsat Thematic Mapper data and Ancillary Data Sources, Environmental Monitoring and Assessment 51(1-2):415-428.

Wayman, J.P., R.H. Wynne, and J.A. Scrivani, 1999. Satellite-Assisted Forest Cover Mapping in the Southeastern United States using Iterative Guided Spectral Class Rejection, In: *Proceedings of Second International Conference-Geospatial Information in Agriculture and Forestry*, Lake Buena Vista, Florida, 2:355-362.

White, J.D., G.C. Kroh, and J.E. Pinder III, 1995. Forest Mapping at Lassen Volcanic National Park, California, Using Landsat TM Data and a Geographical Information System, *Photogrammetric Engineering & Remote Sensing*, 61(3):299-305.

Wynne, R.H and D.B. Carter, 1997. Will Remote Sensing Live Up To Its Promise for Forest Management?, *Journal of Forestry*, 95(10):23-26.

Wynne, R,H., R.G. Oderwald, G.A. Reams, and J.A. Scrivani, 2000. Optical Remote Sensing for Forest Area Estimation, *Journal of Forestry,* 98(5):31-36.

Wynne, R.H., and J.A. Scrivani, 1999. Improving SAFIS Forest Removal Estimates Through Spectral Change Detection, In: *Proceedings of Second International Conference-Geospatial Information in Agriculture and Forestry*, Lake Buena Vista, Florida, 1:200-205.

APPENDIX
Appendix A

Appendix A. SAS code used to extract pure spectral classes

```
data test;
infile 'your_file_name.txt' dlm= '09'x missover;
input spec class;    */reads in input columns, first is spectral class, second is information class/*
proc summary data = test;
        by spec class;
        var class;
output out = test2 n = homo;
run;

proc summary data = test2;
        by spec;
        var homo;
output out = test3 sum = sum;
run;

data test4;
merge test2 test3;
by spec;
prop = homo/sum;
run;

data igscr;
set test4;
if prop<0.9 or sum<10 then delete; */homogeneity criteria/*
proc print;
run;

data fileme;
set igscr;

file 'your_output_file.txt'; */prints output spectral classes/*
put spec 1-4 class 6-10 sum 12-16 prop 18-25 .5;
run;

proc summary data=igscr;
var sum;
output out=sumsum sum=sumofsum;
proc print;
run;
```