Refinement of Automated Forest Area Estimation via Iterative Guided Spectral Class Rejection

Rebecca Forest Musy

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
Stephen P. Prisley
Richard G. Oderwald

Virginia Polytechnic Institute and State University
Blacksburg, Virginia, 24061

April 17, 2003
Blacksburg, Virginia

Keywords: FIA, Landsat ETM+, IGSCR, Forest Area Estimation, Postprocessing

Refinement of Automated Forest Area Estimation via Iterative Guided Spectral Class Rejection

Rebecca Forest Musy

(ABSTRACT)

The goal of this project was to develop an operational Landsat TM image classification protocol for FIA forest area estimation. A hybrid classifier known as Iterative Guided Spectral Class Rejection (IGSCR) was automated using the ERDAS C Toolkit and ERDAS Macro Language. The resulting program was tested on 4 Landsat ETM+ images using training data collected via region-growing at 200 random points within each image. The classified images were spatially post-processed using variations on a 3x3 majority filter and a clump and eliminate technique. The accuracy of the images was assessed using the center land use of all plots, and subsets containing plots with 50, 75 and 100% homogeneity.

The overall classification accuracies ranged from 81.9-95.4%. The forest area estimates derived from all image, filter and accuracy set combinations met the USDA Forest Service precision requirement of less than 3% per million acres timberland. There were no consistently significant filtering effects at the 95% level; however, the 3x3 majority filter significantly improved the accuracy of the most fragmented image and did not decrease the accuracy of the other images. Overall accuracy increased with homogeneity of the plots used in the validation set and decreased with fragmentation (estimated by % edge; $R^2 = 0.932$).

We conclude that the use of random points to initiate training data collection via region-growing may be an acceptable and repeatable addition to the IGSCR protocol, if the training data are representative of the spectral characteristics of the image. We recommend 3x3 majority filtering for all images, and, if it would not bias the sample, the selection of validation data using a plot homogeneity requirement rather than plot center land use only. These protocol refinements, along with the automation of IGSCR, make IGSCR suitable for use by the USDA Forest Service in the operational classification of Landsat imagery for forest area estimation.

ACKNOWLEDGEMENTS

This research was supported by Virginia Tech Department of Forestry, the National Council for Air and Stream Improvement (NCASI), the National Aeronautics and Space Administration (NASA, NA613-01009), the Virginia Department of Forestry, the USDA Forest Service Southern and North Central Research Stations (SRS 01-CA-11330145-470), and the McIntire-Stennis Research Program (VA-136589). I want to thank those organizations as well as my committee members, Dr. Randolph H. Wynne, Dr. Stephen P. Prisley and Dr. Richard G. Oderwald, for their cooperation and patience over all the years it took me to complete this work. My thesis would not have been possible without the intellectual stimulation and camaraderie provided by my advisor, Randy Wynne, and my fellow graduate students; without whom the desire to commute back and forth between Charlottesville and Blacksburg to complete this work would have ebbed from me long ago. I especially want to thank Christine Blinn, whose involvement in this work and contributions to my happiness I cannot leave unrecognized, Zachary Bortolot, who was always there when I needed a study break over a cold one, and Jeff and Rebecca Adams, for being so wonderfully themselves. I also want to thank Brigitte and Ben Parsons, Nina Sengupta and Alok Mallick, and all of the others who took me in while I was visiting Blacksburg. I really appreciate all of the effort you put into making me feel welcome. Finally, I want to thank my family, because they always make me feel exceptional, especially my husband, Pierre, who has shown me nothing but support, constancy, and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
INTRODUCTION

1.1 Forest definition

Before beginning a discussion on forest inventory and the classification of satellite images into forested and nonforested areas, an important question must be asked: what is a forest? The answer is not as obvious as one might think. To some ‘forest’ is merely the collective noun for a group of trees. However, the term ‘forest’ often requires a more precise characterization. A universally accepted definition of a forest does not exist. Lund (1999) reports the results of a survey that found over 130 different definitions for ‘forest’ or ‘forest land’ from more than 30 countries. He notes that the definitions can be divided into three groups – those that refer to administrative or legal units, land cover, or land use.

Land cover has been defined as “that which overlays or currently covers the ground, especially vegetation, permanent snow and ice fields, water bodies, or structures” (USDA-FS, 1989). Image classifications are often based on land cover since they are commonly performed on a pixel-by-pixel basis. Each individual pixel contains only brightness values. The pixels in a satellite image may be interpreted collectively for their spatial patterns; however, this has not been easy to perform in an automated fashion. The classifications performed in this study are based on land use, which goes beyond the ground cover to “the predominant purpose for which an area is employed” (USDA-FS, 1989).

The differentiation between land cover and land use is somewhat difficult to achieve using computers. For example, the land cover on a recently harvested pine plantation is nonforest, since all of the trees have been removed. What will appear on the satellite image is an area with no trees. However, the area might not be converted to a nonforest land use after it is cut. What is more likely is that it will be replanted and return to forest land use, therefore it should be classified as forest land use. There is thus clearly a need to incorporate temporal and spatial data into the classification process.

Lund (1999) states, “A good working definition of forest land would specify a minimum size area to consider and a minimum crown closure threshold…If the definition were land use, then a listing of inclusions and exclusions should be included.” The definition of a forest used by the United States Department of Agriculture (USDA) Forest Service Forest Inventory and Analysis (FIA) program is land that is at least 10 percent stocked by forest trees of any size, or formerly having such tree coverage but not currently developed for nonforest use. The minimum area considered for classification is 1 acre (0.4 ha), and forested strips must be at least 120 feet (36.6 m) wide (USDA-FS, 2001).

Satellite image classifications are usually performed on a pixel-by-pixel basis, however, the FIA definition includes requirements regarding the size and shape of forest areas. This indicates the need for spatial postprocessing to improve the agreement between groups of pixels in classified satellite imagery and the definition of a forest. However, the spatial dimensions of the pixels in many satellite images limit such postprocessing. Fortunately, the size of each pixel in a Landsat Enhanced Thematic Mapper Plus (ETM+) image (the type of imagery used in this study) is 30 m x 30 m. This means that the length of each side of the pixel is approximately 100 feet, and the diagonal measurement of each pixel is approximately 140 feet. Therefore, Landsat pixels are almost ideal for use in enforcing the 120 feet minimum width requirement of the FIA definition. Also, the total area of each pixel is 900 m^2, and since there are 4,047 m^2 in an acre, it takes 4.5 pixels to represent an acre. This information can be used to remove any group of 4 or less "forest" pixels since it does not meet the FIA definition of a forest. Clearly, some form of spatial postprocessing seems necessary to meet the FIA definition of a forest using classified satellite imagery and the closer the dimensions of the pixels are to the definitional limits, the better.

1.2 Forest Inventory and Analysis program

The USDA Forest Service manages the national forest inventory in the United States, the Forest Inventory and Analysis (FIA) program. A description of the FIA program is as follows (from USDA-FS, 1992):

> The Forest Inventory and Analysis mission is to improve the understanding and management of our nation's forests by maintaining a comprehensive inventory of the status and trends of the country's diverse forest ecosystems, their use, and their health…FIA inventories provide the necessary foundation for building a program of land stewardship and serving the people by providing unbiased, accurate, current, and relevant forest resource information that meets their diverse needs…FIA develops basic statistics that are needed as background for many research proposals and problem analyses. Although much of the FIA program focuses on gathering data and reporting statistics, the program includes a dedicated and capable cadre of scientists who evaluate forest resource trends, develop techniques, and adapt the latest technology.

The following is a brief summary of the FIA program, with an emphasis on aspects related to this project. For readers interested in a more complete description of the history of FIA, the role of remote sensing within FIA and statistical sampling procedures it employs, please refer to the December 1999 issue of the Journal of Forestry (Volume 97, Number 12).

1.2.1 History

The FIA program was established in compliance with two federal laws: 1) *The Organic Administrative Act of 1897*, which laid the legal foundation for the National Forests and for the inventory and monitoring of these lands, and 2) the *McSweeny-McNary Forest Research Act of 1928*, which required to Secretary of Agriculture to make "a comprehensive survey of the present and prospective requirements for timber and other forest products of the United States." In 1930, the USDA Forest Service responded to the McSweeny-McNary Act by organizing regional Forest Survey Projects that were conducted on a state-by-state basis beginning in the western United States. By the 1960's, inventories of the lower 48 states were complete, with more heavily forested areas having been inventoried at least twice (USDA-FS, undated).

The Forest Ecosystems and Atmospheric Pollution Research Act of 1988 (P.L. 100-521) directs all FIA units to "increase the frequency of forest inventories in matters that relate to atmospheric pollution and conduct such surveys as are necessary to monitor long-term trends in health and productivity of domestic forest ecosystems." This legislation led to the creation of the Forest Health Monitoring (FHM) program.

1.2.2 Recent developments

Due to an increased demand for current data covering a broad scope of forest attributes, Congress enacted legislation which significantly increased the productivity requirements of the forest inventory program. The *Agriculture Research, Extension, and Education Reform Act of 1998 (P.L. 105-185)* mandates the annual measurement of 20% of the sample plots for a state, with a core set of nationally consistent variables measured at each plot. It also requires the compilation of a standardized report of the measurement data for each state every 5 years, including a nationwide analysis of the results and trends over the past two decades.

In response to the *Agriculture Research, Extension, and Education Reform Act (P.L. 105-185)* the USDA Forest Service switched to an annual survey, increasing their capacity to analyze and distribute data, and integrating the FIA and FHM plots into a three-phase sampling system (Van Deusen *et al.*, 1999; Gillespie, 1999). The first phase is an enumeration of the forested and nonforested acreage in the United States, which has been traditionally based on aerial photography. The second phase involves the verification of a subset of the phase 1 photo-interpreted plots. Stand and tree data are also collected at these permanent ground plots. At a subset of the phase 2 plots, broader information regarding forest health factors is gathered. The collection of these data constitutes phase 3 of the new FIA program (USDA-FS, undated).

The photo interpretation in phase 1 will soon be replaced by remotely sensed satellite imagery, also in accordance with the *Agriculture Research, Extension, and Education Reform Act* (P.L. 105-185), which requires that a "process for employing…remote sensing, global positioning systems, and other advanced technologies" in forest inventory be developed and implemented. Recently, a goal of completing the transition from aerial photo interpretation to the use of satellite image analysis by the end of fiscal year 2003 was set by the Director of Science Policy, Planning, Inventory and Information for the United States Forest Service (Guldin, 2000). The classification of satellite images will allow for stratification, which reduces the variance of forest area estimates produced by sample plot data (Riemann and Alerich, 1998).

1.3 Forest area estimation

Precise forest area estimation is the ultimate goal of this research. Forest area estimates and other forest information are of critical significance to a wide variety of interested parties, such as "managers, policymakers, environmental organizations, business interests, consultants, scientists, the media, and citizens who are interested in status, trends, stewardship, and sustainability of the nation's forested ecosystems" (Gillespie, 1999). The FIA program's scope is not local; it does not provide specific information for planning management actions. It is considered to be a strategic, national-level inventory important to broader decision- and policy-making (Gillespie, 1999).

1.3.1 Historical method

The traditional FIA method of producing forest area estimates involves a variation of double sampling (Chojnacky, 1998; Reams and Van Deusen, 1999). The first phase of this process involves photo interpretation of a large systematic grid of sample of points on aerial photographs from the National Aerial Photography Program (Wynne *et al.*, 2000). The points are classified into forest and nonforest based on the FIA definition of a forest. In the second phase, a subset of the phase 1 sample plots is field checked to determine whether or not they were correctly classified. In addition to confirming the classification of the photo-interpreted points, the ground plots are used to estimate tree and stand-level characteristics (Wynne *et al.*, 2000).

In the past, the size of the systematic sample grids varied between FIA research stations (Frayer and Furnival, 1999). In the Southeast, the aerial photo plots each represented approximately 230 acres (Reams and Van Deusen, 1999; Frayer and Furnival, 1999). Each of the ground plots measured in phase 2 represented roughly 5,760 acres (Frayer and Furnival, 1999). These plots were formerly based on the intersections of a 3-by-3 mile grid, but this system was replaced by a set of plots based on a grid of hexagons (McRoberts, 1999). In the Southeast, the sampling scheme for verification of aerial photo-interpreted points is intensified beyond the measured ground plot sample. These supplemental plots at which the photo points are ground checked, and no field measurements are taken, are called intensification plots (Scrivani *et al.*, 2001). These plots represent about 3,840 acres each (Frayer and Furnival, 1999).

The classical double sampling method used for stratification by the FIA involves two samples (the phase 1 photo plots and the phase 2 ground plots), where the second sample is a stratified subsample of the first (Chojnacky, 1998). The photo-interpreted data are used to calculate the proportions of the first phase sample units in forest and nonforest. The proportions calculated are the stratum weights for the sample, which are estimates of the population stratum weights.

In phase 2, observations of random variables in the sample units are made. The two types of variables involved in the second phase sample are 1) resource attributes and 2) other variables used to define subpopulations of interest (Chojnacky, 1998). Resource attributes include forest area, number of trees, volume, growth and canopy cover. These variables are the focus of the estimation process. Variables that are collected for defining subpopulations comprise forest type, ownership, site class, habitat type, species, crown descriptions, and disease ratings, among others (Chojnacky, 1998). Subpopulation variables are just meaningful categories into which the attribute variables are divided. The observations made in phase 2 are used to calculate stratum (sample) means for the random variables. The sample means are used in conjunction with the estimated stratum weights to produce an unbiased estimate of the population mean.

Two shortcomings of the photo-based method are that it cannot produce maps of forest and nonforest area, and it may take up to a year to photo interpret an entire state (Reams and Van Deusen, 1999). In addition, the future of the National Aerial Photography Program is uncertain

(Wynne *et al.*, 2000). Despite these limitations, deriving forest area estimates using aerial photographs is a proven, reliable method (Wynne *et al.*, 2000).

1.3.2 Satellite image-based method

Wynne *et al*. (2000) enumerate the benefits of using satellite imagery over aerial photography. They state that satellite imagery permits more frequent data collection and creates a map product, which generates good area estimates and can be used to estimate stratum sizes and calculate ground-plot expansion factors. They also remark that a map product facilitates additional analyses pertinent to forest inventory such change detection studies. Riemann and Alerich (1998) state that the USDA Forest Service is interested in using satellite imagery in traditional stratification procedures partly because it would make the method of stratification more consistent between administrative units.

Those who are involved with remote sensing research within the FIA program have made two recommendations for plot data collection that would assist in the use of satellite imagery. First, they recommend that accurate GPS (Global Positioning System) data be collected as a core variable at all phase 2 plots (Riemann and Alerich, 1998). Second, they have suggested that a characterization of the land cover at each plot be made in addition to the land use description that is currently made. Implementation of ideas such as these will improve the likelihood of successful integration of satellite image-based forest area estimation into the FIA program. However, in order to satisfactorily replace the traditional methodology, a system of satellite image classification must be developed that produces forest area estimates which meet the precision standard for forest area estimates of 3 percent per million acres of timberland set by the Forest Inventory and Analysis (FIA) program (Scrivani *et al.*, 2001). The method should also be objective, repeatable and low in cost.

Reams and Van Deusen (1999) suggest the use of a two-phase method of forest area estimation based on known map proportions from a thematic map. However, the use of maps to calculate known stratum weights is not new to FIA. Chojnacky (1998) indicates that stratified sampling was first applied by FIA in the Interior West using forest type maps to calculate stratum weights. He states that this practice was abandoned around 1965 due to the limited availability of usable resource maps. Employing a new double sampling theory for forestry, a shift to the use of aerial photo sample points for stratification was made. This application of double sampling in FIA is that which remains in use today.

Using a classified satellite image, a rough estimate of the total forested area can be obtained by simply multiplying the number of census acres in the area by the marginal map proportion for forest (Wynne *et al*., 2000). However, the use of this calculation to estimate forest area would require the inappropriate assumption that the image classification was without error. The most suitable means to determine the error associated with a thematic image is to use a contingency table to assess the accuracy of the classification (Card, 1982). In a contingency table, a cross-tabulation of map category versus true category for a sample of points is created (Table 1). It is important to note that unbiased estimates of error probabilities derived using contingency tables are relative to the sampling scheme used (Card, 1982). The equations given below are based on simple random sampling of reference data.

Table 1. Example of a contingency table used to assess the accuracy of, and calculate the true proportions and variance of, a thematic image classification. The abbreviations used are as follows: t = true, m = map, f = forest, nf = nonforest, p = proportion, Σ = sum (or total).

True	Classified image (map) categories		
	Forest	Nonforest (nf)	Total
Forest	True forest-Map forest (tf_mf)	True forest-Map nf (tf_mnf)	Total true forest (Σtf)
Nonforest (nf)	True nf-Map forest (tnf_mf)	True nf-Map nf (tnf_mnf)	Total true nf (Σtnf)
Total	Total map forest (Σmf)	Total map nonforest (Σmnf)	Total
Map marginal proportions	Map proportion forest mpf	Map proportion nf mpnf	

The formulae used to calculate the true probability of interest (forest area), along with its variance estimate, by means of the known map marginal probabilities from a satellite-derived thematic map, are given in Card (1982). An explicit example of how these formulae are used is given in Wynne *et al*. (2000). The true proportion of forest (*pf*) is calculated as follows (refer to Table 1 for notation):

$$pf = mpf * \left(\frac{tf\_mf}{\Sigma mf}\right) + mpnf * \left(\frac{tf\_mnf}{\Sigma mnf}\right) \tag{1}$$

The variance of the percent forest (V*pf*) is calculated as follows (also see Table 1 for notation):

$$Vpf = \frac{\left[mpf - \left(\frac{tf\_mf * mpf}{\Sigma mf} \right) \right] * \left(\frac{tf\_mf * mpf}{\Sigma mf} \right)}{mpf * total} + \frac{\left[mpnf - \left(\frac{tf\_mnf * mpnf}{\Sigma mnf} \right) \right] * \left(\frac{tf\_mnf * mpnf}{\Sigma mnf} \right)}{mpnf * total} . \quad (2)$$

Card (1982) notes that the estimators he derived (used above) are valid regardless of the validation data sample size. However, by examining the latter equation, one can see that the precision of a forest area estimate obtained through satellite image classification is affected by both the classification accuracy (*tf_mf / Σmf* is the proportion of points in the forested map category that were correctly classified) and the sample size of the validation set (*Σmf* + *Σmnf* = *total* points in validation set). Therefore, the choice of a consistently accurate image classification algorithm, as well as the collection of sufficient ground truth data and use of those data in the validation process, are important to achieve the required precision in the variance of forest area estimates.

1.4 Iterative Guided Spectral Class Rejection

Research into various possible satellite image classification methods (for producing thematic maps from which to derive map marginal probabilities) has led to the development of an algorithm called Iterative Guided Spectral Class Rejection (IGSCR; Wayman *et al.*, 2001). IGSCR is a hybrid classifier that combines characteristics of unsupervised and supervised classification algorithms. The first step of IGSCR is to perform an unsupervised classification technique known as ISODATA (Iterative Self-Organizing Data Analysis Technique) on the input image (Fig. 1). Pixel values within user-identified training areas (collected based only upon their informational class) are then extracted from the resulting spectrally clustered image. The pixel data are used to determine which spectral clusters resulting from the ISODATA classification are homogeneous for one informational class. Pure clusters are removed from the image, the signatures of pure spectral clusters are saved and the remaining unclassified pixels are run though the ISODATA algorithm again. After a user-defined number of iterations, the pure signatures accumulated are used in a maximum likelihood classification, which assigns an informational class to each pixel in the image based on its spectral proximity to the pre-identified "pure" spectral classes.

Wayman *et al.* (2001) showed that corrected area estimates obtained from satellite imagery classified via IGSCR for three regions of Virginia were not significantly different (at the

95 % level) from the photo-based estimates. Wayman *et al*. (2001) also showed that IGSCR produces slightly better forest area estimates than those derived from historical Multi-Resolution Land Characteristic Interagency Consortium (MRLC) cover maps (although they were not significantly different at the 95% level). Even though MRLC is an existing program that produces similar results, IGSCR may be the better option in areas where the rapidity of forest change precludes the use of MRLC maps produced on a decadal basis. Furthermore, the land cover definitions used by MRLC do not match FIA definitions (Wayman *et al*., 2001).

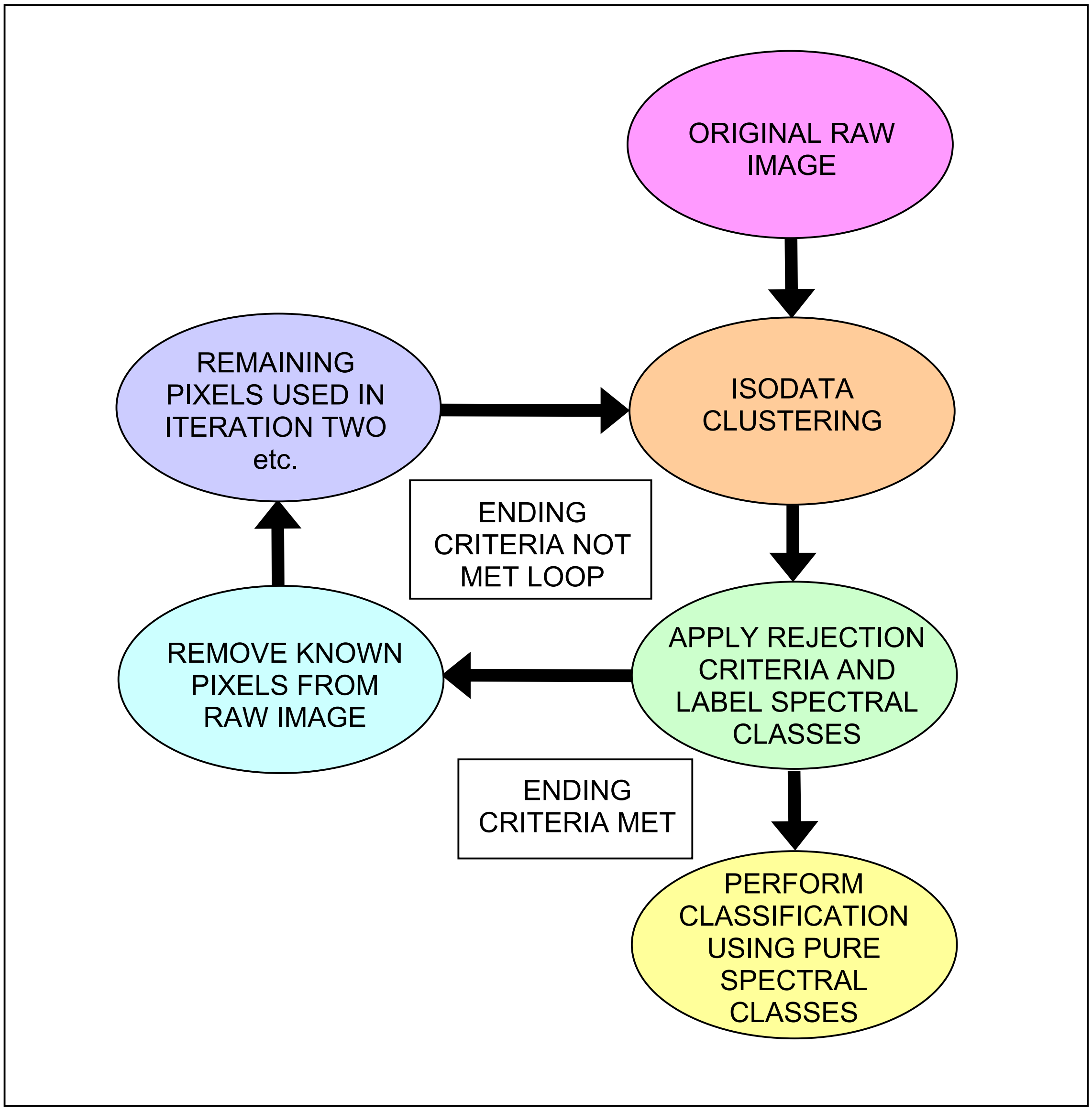


Fig. 1. Flowchart outlining the process of Iterative Guided Spectral Class Rejection (Wayman *et al*., 1999, used with permission).

Wayman *et al*. (2001) have shown the IGSCR method to be rapid, accurate, and repeatable across users, regions and time. However, the implementation of the algorithm up to this point has been quite tedious and prone to user error. This is because the algorithm requires that a number of completely objective and somewhat complicated processes be performed repeatedly. After the input image, training data, and parameters are selected, the remainder of the IGSCR algorithm could be completely automated. This would significantly reduce the analyst time required to perform the process and preclude the possibility of user error. Automation would have the additional benefit of making the IGSCR algorithm more suitable for production-level use by the FIA program. To facilitate the completion of this and other research projects, and for the potential benefits it may offer to the FIA program, I automated the IGSCR algorithm.

1.5 Training data collection

Using current technology and data, it is not possible for satellite image classification to be completely objective. One important stage in the IGSCR algorithm that requires user input is the collection of training data. This process requires the user to select small areas within the image that represent the informational classes of interest (in this case, forest and nonforest). In the IGSCR algorithm, the data from these areas are extracted from a spectrally clustered image and tested to determine which clusters are homogeneous for one informational class. This process ultimately determines the informational class into which each pixel in the image is placed.

Wayman *et al*. (2001) noted that one aspect of the IGSCR process that remained to be standardized was the collection of training data. In their study, data used to train the classifier were gathered with any and all possible means, including the use of hand-drawn areas on maps, on-screen digitization, buffered points and polygons collected with a Global Positioning System (GPS), aerial videography, and even a helicopter to collect polygons of known land use. Given that the FIA program protocol must be objective and repeatable, it is clear that a consistent approach to training data collection is needed.

Preliminary research by Scrivani *et al*. (2001) into possible methods of training data collection for IGSCR showed little difference in classification results between two particular training data protocols used to classify one Landsat TM scene in Virginia. The two training data collection methods that they tested were maplets, digitized over high-resolution digital imagery, and a region-growing approach based on seed pixel initiation at points of known land use. When these two very different approaches to training data collection were tested on one Landsat TM scene in Virginia, despite fewer and less evenly proportioned data than a previous study (Wayman *et al*., 2001), both methods resulted in classification accuracies close to 89 % (Scrivani *et al*., 2001). In addition, the precision of the forest area estimates resulting from the classified images surpassed the FIA standard of 3 percent per million acres of timberland. It was concluded that the seed pixel approach was superior because it did not require ancillary imagery, has the potential for partial automation, and was not as time-consuming as digitizing maplets.

One benefit of using the seed pixel approach initiated at point locations, as was done in the Scrivani *et al*. (2001) study, is that the points can be selected in an objective way, either randomly (from the image or a set of points) or in some other systematic fashion. This reduces some of the subjectivity involved in the collection of training data. For this reason, and those mentioned by Scrivani *et al*. (2001; does not require ancillary imagery, potential for partial automation, not as time-consuming as digitizing maplets), the seed pixel approach to training data collection based on pre-selected points was used in this study.

The pre-selected points employed in the Scrivani *et al*. (2001) study were FIA phase 2 plot center data. These points were removed from the pool of potential validation data to be used as initiation points for training data collection. As was noted earlier, the precision of a forest area estimate obtained through satellite image interpretation is affected by both the classification accuracy and the sample size of the validation set. Ideally, all of the FIA points that were recently visited in the field, including phase 2, phase 3, and other FIA points (where available), would be used in accuracy assessment and forest area estimation. This is not as important in states like Virginia, where the inclusion of supplemental phase 2 plots with the other phase 2 plot data for validation has been shown to produce precision estimates that exceed the national FIA standard of 3 percent per million acres (Scrivani *et al*., 2001). In states where intensification plot data are not collected or where the total percent forest is low, the use of all available points could be critical to meeting the precision standard required of the FIA forest area estimates.

The advantage of using points of known land use is that the analyst has some ground truth information to aid in the collection of training data for the appropriate informational class. Despite this advantage, it is not clear whether or not land use information is necessary to the collection of quality training data via the seed pixel approach for forest/nonforest classifications.

For this reason, I used random non-FIA points selected from the image to determine whether or not random points are acceptable to use in training data collection via the region-growing approach.

An important question that must be addressed is whether or not the use of random sampling from the image is an effective means of sampling the image to be classified. If points are randomly sampled from the image in sufficient number, they will cover the variations in spectral reflectance within the image. However, the number of points required to sufficiently sample the pixels in a satellite image is large, therefore that approach is impractical. This means that there might not be any training data collected in spectrally distinct regions of the image. In that case, the ISODATA clustering portion of the IGSCR process would develop spectral classes for these regions, but they would remain unclassified due to lack of training data. This was determined to be a problem in the classification of water when seed locations were randomly sampled from FIA plots (Scrivani *et al*, 2001). Due to the lack of plot locations in this cover type (very few FIA plots fall in large water bodies), analyst input was required to define training data in areas of the image covered by water during preliminary training data research. However, it is not clear whether this would be a problem if stratified random sampling from the image were used instead of random sampling of the FIA plot locations.

Figs. 2-4 give an example of the type of classification error that might occur if training data are not collected within certain spectral classes due to the use of random sampling. Fig. 2 shows a Landsat TM multispectral image subset overlain with two sets of training data areas of interest (AOIs). The AOIs in Fig. 2(b) are a subset of those in Fig. 2(a). Note that there are no AOIs in Fig. 2(b) in areas of very high reflectance (white or cyan areas). Fig. 3 shows the Landsat TM image subset from Fig. 2 classified via ISODATA into 5 classes. The AOIs shown in black are the ones that are missing from Fig. 2(b). All of these AOIs fall into one spectral class, which is shown in cyan (Fig. 3). If stratified random sampling were used to generate a set of seed pixels with which to collect training data from this image, the AOIs shown in Fig. 3 would be included, and the majority of the image would be classified after performing the IGSCR algorithm with the stacked ISODATA option (Fig. 4(a)). However, if random sampling were used, the spectral class shown in cyan in Fig. 3 might not have been sampled, in which case many pixels would remain unclassified (Fig. 4(b)).

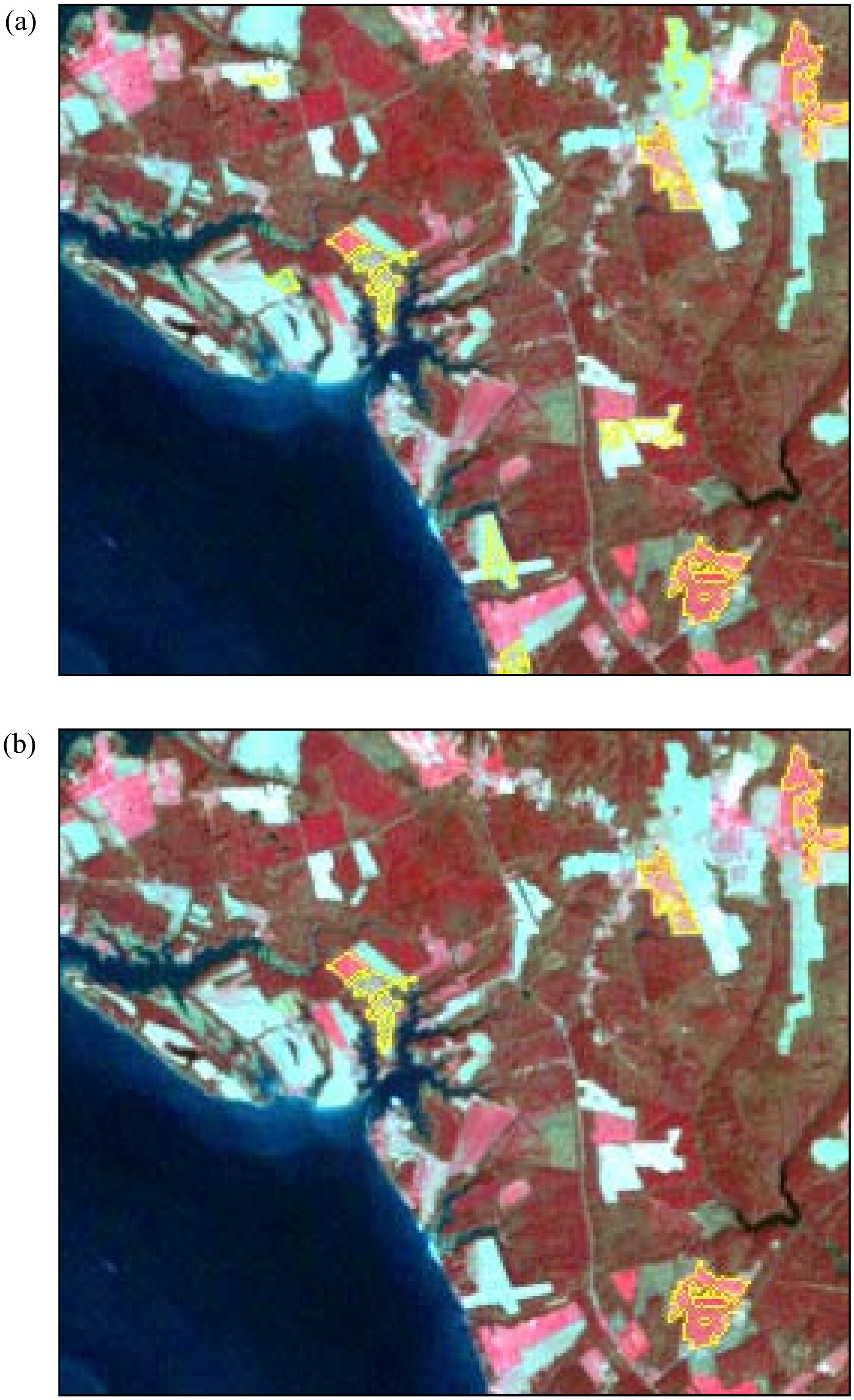


Fig. 2. Landsat TM multispectral image subset showing bands 4, 3 and 2 in red, green and blue, and training data areas of interest (AOIs) in yellow. The AOIs in (b) are a subset of those in (a).

The collection of training data requires a trained image analyst and can be fairly time consuming. These characteristics make training data collection costly, both in time and in money. The goal of this project is to aid in the development of an operational Landsat TM image classification protocol for FIA forest area estimation. Such a protocol must be as inexpensive and time-efficient as possible. This project serves as a means to test, over a wide range of physiographic regions, the use of random sampling to develop seed pixel initiation point sets for training data collection. In addition, the project will serve as the first test of the automated IGSCR program over multiple states using the same parameter values.

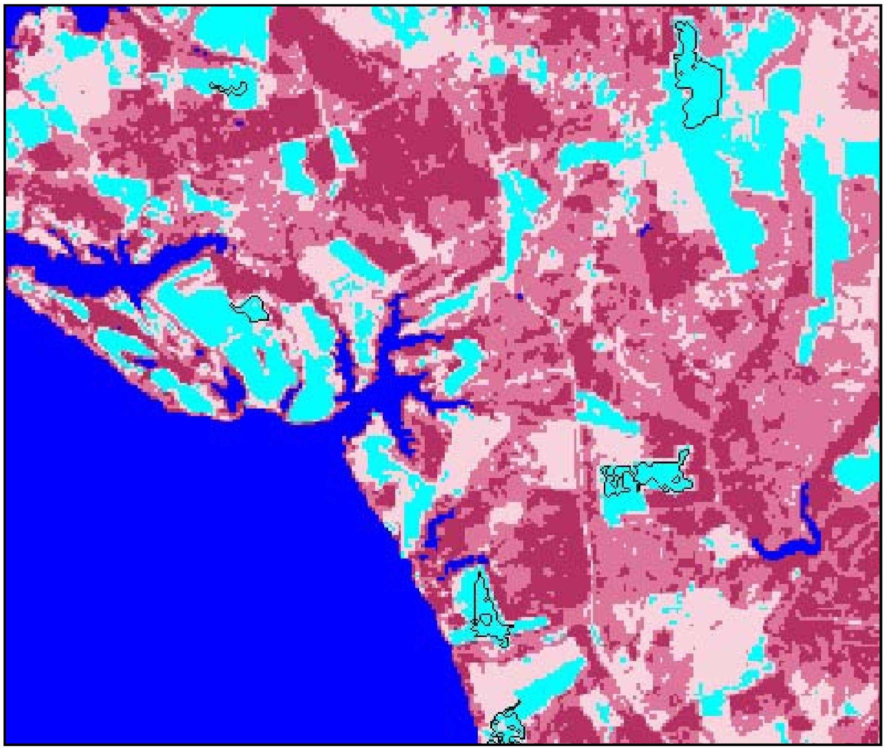


Fig. 3. Landsat TM image from Fig. 2 classified via ISODATA into 5 classes. The AOIs shown in black are the ones that are missing from Fig. 2(b). All of these AOIs fall into one spectral class.

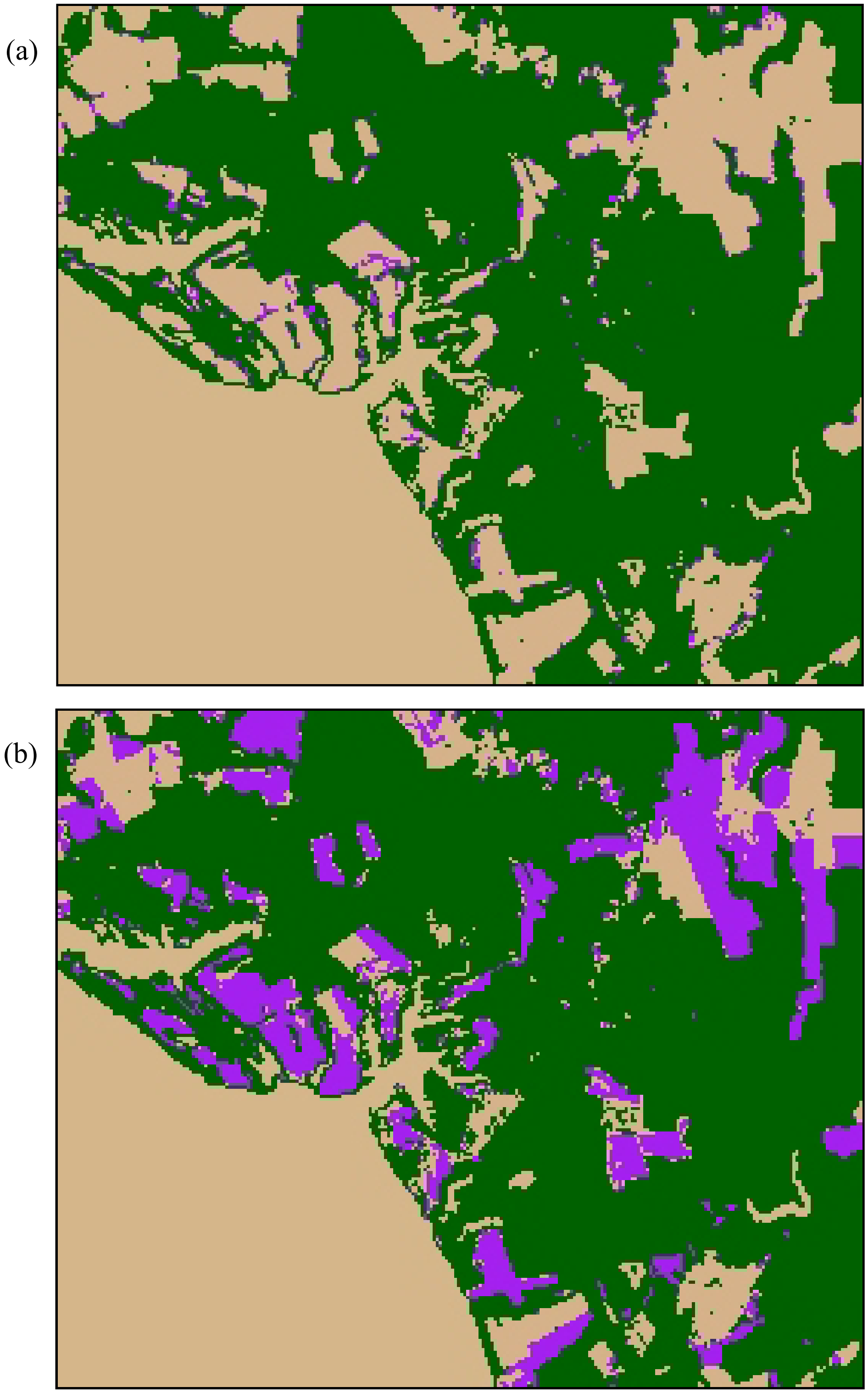


Fig. 4. Landsat TM image from Fig. 2 classified via IGSCR (with the stacked ISODATA option) into forest (green), nonforest (tan) and unclassified (purple) using the same forest and water AOIs for both classifications and the nonforest AOIs in Fig. 2(a) to create (a) and those in Fig. 2(b) to create (b).

1.6 Objectives

The goal of this study is to aid in the development of an operational Landsat TM image classification protocol for FIA forest area estimation through the refinement of the Iterative Guided Spectral Class Rejection protocol. The specific objectives are as follows:

1. Automate the Iterative Guided Spectral Class Rejection (IGSCR) protocol
2. Determine the utility of the following based on the thematic image accuracy and forest area estimates resulting from processing images with IGSCR:
    (a) random points as seed pixel initiation points for training data collection,
    (b) spatial post-processing of the classified image, and
    (c) accuracy set adjustment based on the percent forest at each plot.

Chapter 2
BACKGROUND

2.1 Satellite image classification

Thematic classification of a satellite image is the distillation into categories of the reflectance data recorded in an image. Each pixel in an image is assigned to one category based upon the spectral data it contains. Numerous classification techniques, requiring varied levels of user input and subjectivity, have been developed and tested. Included among these methods are supervised and unsupervised classifications and hybrids thereof, as well as multi-stage iterative classifications. These techniques are outlined in Wayman *et al*. (2001). The ideal satellite image classification algorithm for use in the FIA program should be repeatable, accurate, and require minimal analyst time and input.

2.2 Iterative Guided Spectral Class Rejection

Iterative Guided Spectral Class Rejection (IGSCR) is a hybrid land cover classification algorithm that has been used successfully to achieve accurate binary classifications of Landsat TM imagery (Wayman *et al*., 2001). Research up to this point has shown that IGSCR is a rapid, objective and repeatable classification algorithm. Wayman *et al*. (2001) have also shown that IGSCR is capable of producing classifications that yield forest area estimates that are within the range of precision required by the USDA Forest Service (3 percent per million acres timberland). It is for these reasons that this algorithm is being considered for use in the Forest Inventory and Analysis program.

2.2.1 IGSCR algorithm description

The first step in IGSCR is the collection of training data in each of the informational classes desired in the final classification. Unlike the training data collection involved in a supervised classification, each area collected for use in the IGSCR algorithm includes pixels of one informational class regardless of the spectral class into which the pixels will ultimately be categorized. This removes much of the subjectivity and work involved in a supervised classification resulting from the requirement that each training area must represent a spectral class that is normally distributed and highly separable from other spectral classes. The polygons delineating the training pixels for each informational class are saved in separate files. This means that all forest polygons are stored in one file and all nonforest polygons in another.

After the training data are collected, the image is divided into spectral classes via the ISODATA clustering algorithm, which is commonly used in unsupervised classifications. One difference between an unsupervised classification and IGSCR is that the unsupervised method is followed by user decisions about the informational class to which each spectral class should be assigned. In IGSCR, the training data areas are used to determine the informational classes to which the ISODATA spectral classes should be assigned. Another difference between an unsupervised classification and IGSCR is that in the former, all spectral classes are assigned to an informational class. In the latter, some of the spectral classes might be rejected because they are "impure," that is, they contain data from more than one informational class (Fig. 1).

The pure spectral classes are determined by first extracting the ISODATA class values of the pixels that fall within the training areas for one informational class from the clustered image (Fig. 5, Fig. 6). The extraction process is repeated for each of the informational classes of interest. In a forest/nonforest classification, this means that the extraction process is performed twice, first extracting all of the ISODATA class values for pixels within the forest areas of interest (AOIs), which were stored in one file, and then extracting the values within the nonforest AOIs, which were saved in a separate file.

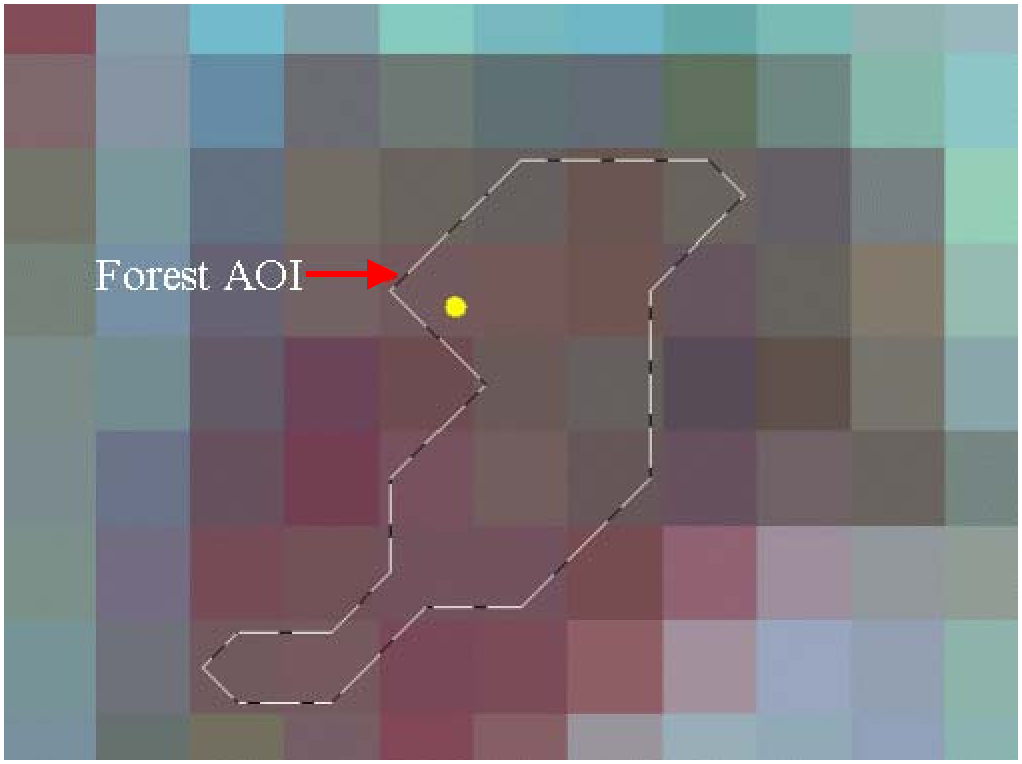
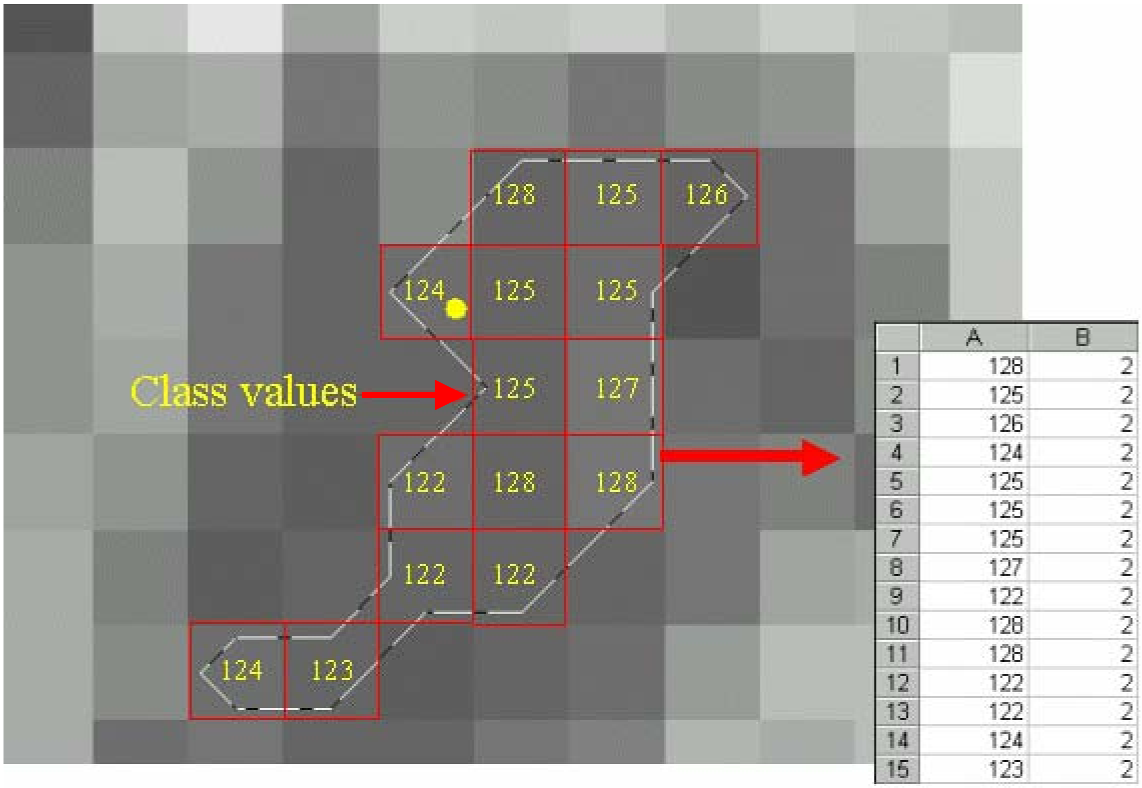


Fig. 5. Forest area of interest (AOI) over a multispectral Landsat TM image.



	A	B
1	128	2
2	125	2
3	126	2
4	124	2
5	125	2
6	125	2
7	125	2
8	127	2
9	122	2
10	128	2
11	128	2
12	122	2
13	122	2
14	124	2
15	123	2

Fig. 6. Forest area of interest (AOI) and corresponding spectral class values of a 150 class ISODATA image. The spectral class values of the pixels within the AOI are extracted and labeled with their informational class, 2 (forest).

Next, using the ISODATA (spectral) class number along with the informational class of each pixel in the training areas, an analysis of the homogeneity of each spectral class is performed. In previous IGSCR research, a class was determined to be homogeneous if 90 percent of the training pixels in a spectral class were of the same informational class and the spectral class contained a minimum of 10 pixels (Wayman *et al.*, 2001). In this study, a test of proportion, adapted for IGSCR by Christine Blinn, is used to determine the homogeneity of each spectral class. Given a user-defined homogeneity threshold, p_0, the minimum number of pixels in a pure class is given by the formula

$$\textbf{count} * (1 - p_0) >= 5$$

where count = total number of pixels in the spectral class within all training areas. If a spectral class does not contain the minimum number of pixels, it is considered to be impure.

The homogeneity of a spectral class is determined by the following test on the proportion of pixels that fall into a given informational class:

if z > z (α) then the class is considered to be pure

where $z = ((p\_hat - p_0 - 0.5) / total) / sqrt(p_0 * (1 - p_0) / total)$
α = the type-I error rate
$z(\alpha)$ = the cutoff point under the standard Gaussian probability density function which is exceeded by a standard Gaussian random variable with probability α
p_hat = majority informational class pixel count / total pixel count (for a spectral class)
p0 = the homogeneity threshold
total = the total pixel count for a given spectral class.

After the purity of each ISODATA class has been determined using the training data, all of the pixels that contain pure spectral classes in the ISODATA image are used to mask the original image values to zero (Fig. 7, Fig. 8, and Fig. 9). The masked image, which contains only pixels from impure classes, is then run through the ISODATA clustering algorithm, beginning the next iteration. The training data are used again to extract pixel spectral class values from the resulting image so that the pure spectral classes can be determined. The clustering, pixel extraction, and pure spectral class determination process is repeated until a user-defined maximum number of iterations has been performed, no more pure classes result, or all of the classes in a given iteration are pure.

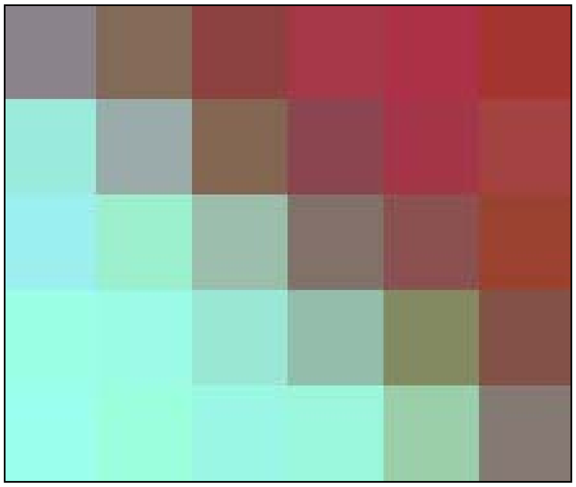
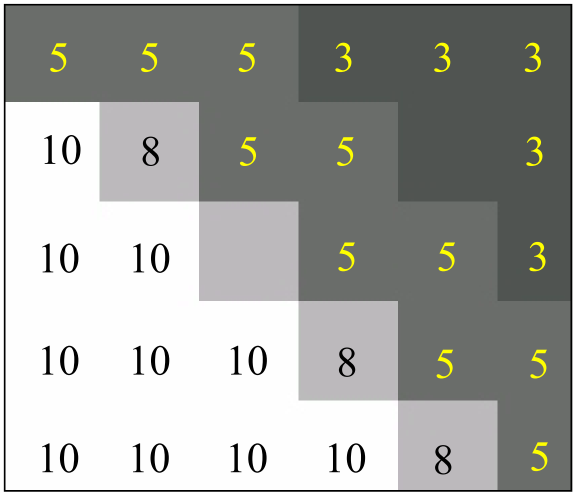
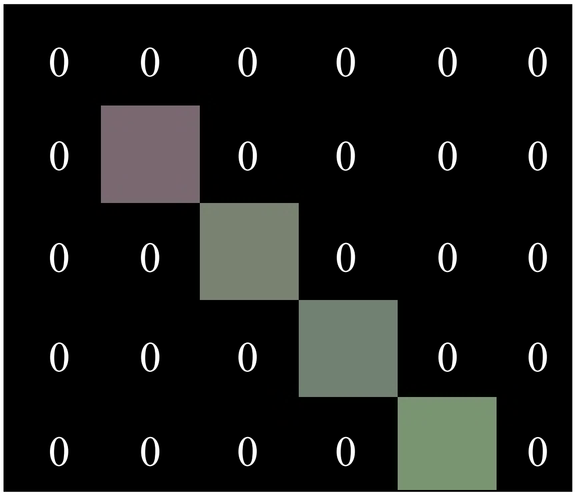


Fig. 7. Subset of a Landsat TM multispectral image.



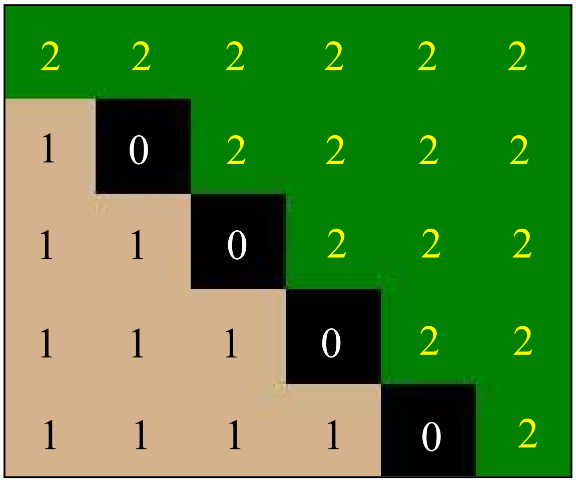
5	5	5	3	3	3
10	8	5	5		3
10	10		5	5	3
10	10	10	8	5	5
10	10	10	10	8	5

Fig. 8. Image subset in Fig. 7 after entire image was classified by ISODATA clustering, with spectral class values shown. Classes 3 and 5 were determined to be pure forest, and 10 pure nonforest, based on homogeneity testing.



0	0	0	0	0	0
0		0	0	0	0
0	0		0	0	0
0	0	0		0	0
0	0	0	0		0

Fig. 9. Masked subset of the Landsat TM image to be used in the second iteration of IGSCR. The pure spectral classes were removed from the raw image by masking the pixels that correspond to pure ISODATA spectral classes (classes 3, 5 and 10 in Fig. 8) to zero.



2	2	2	2	2	2
1	0	2	2	2	2
1	1	0	2	2	2
1	1	1	0	2	2
1	1	1	1	0	2

Fig. 10. The ISODATA image subset (Fig. 8) that was recoded based on the informational class values of the pure classes. Class 10 was recoded to nonforest (1), classes 3 and 5 to forest (2) and class 8, since it was impure, was recoded to zero. This image will be used in creating the stacked ISODATA image.

After one of the stopping criteria has been met, a final classification of the image can be obtained one of four ways: 1) the pure signatures from each iteration can be combined and used in a maximum likelihood algorithm, 2) the ISODATA images can be recoded so that the impure classes are set to zero (Fig. 10) and then added together (in this case, the unclassified pixels remain in a separate, unclassified stratum), 3) a maximum likelihood classification can be performed on the unclassified pixels only and the resulting image added to the recoded, combined ISODATA images, or 4) after performing a complete classification of the image via methods 1 or 3, filters can be used to divide the image into a four-class image indicating forest, nonforest, forest edge and nonforest edge. Hansen and Wendt (2000) at the USDA Forest Service North Central Research Station developed the method of creating the four-class image and models to implement it.

2.2.2 IGSCR parameters

When classifying an image via IGSCR, the user must specify certain parameters. IGSCR begins when the user performs an ISODATA clustering of the input Landsat TM image. The user must choose the number of classes (clusters) to be created, the maximum number of ISODATA iterations to be performed, and the convergence threshold for the clusters formed. In addition, the user must decide whether to initiate the clusters along the principal or diagonal axis of the data, and whether the initial clusters will be scaled automatically or by a user-defined number of standard deviations from the mean of the data. These parameters must be set each time the ISODATA algorithm is performed. Since many IGSCR iterations may be performed, the repeated selection of a specific set of values for these parameters is susceptible to user error. Automation of the IGSCR algorithm eliminates the potential for such errors.

In previous research, the number of ISODATA classes created was changed each iteration as a function of the number of training pixels remaining to be classified. The number of classes, n, to use in the next iteration was obtained by dividing the number of training pixels that were unclassified after the previous iteration by ten, the minimum number of pixels required in a pure spectral class (Wayman *et al.*, 2001). However, this formula was somewhat arbitrarily selected and often resulted in an extremely large number of classes. In this project, the number of spectral classes created in each iteration is held constant. Since pure spectral classes are removed each iteration, holding the number of spectral classes constant results in the division of an ever-smaller data set into an equal number of parts.

Another parameter that must be selected by the user is the maximum number of IGSCR iterations to be performed. In previous research, IGSCR was run until no more pure classes were obtained from the previous iteration. This resulted in as many as 17 iterations being performed (Wayman *et al.*, 2001). Depending upon the number of classes created in each iteration, this large number of iterations could take up to a week to process manually. The automated IGSCR algorithm significantly reduces the time required to classify an image, especially if the user opts to run the classification until no more pure classes are extracted.

The new test of proportion performed in the homogeneity testing phase requires the user to select the type-I error rate and homogeneity threshold to be used in the calculations which determine the purity of spectral classes. Variations in these parameter values, as well as in the number of classes used per iteration, may have the potential to greatly affect the results of the classification. For these reasons, it would be beneficial to perform a factorial analysis of the various combinations of parameter values to determine the optimal values to use. Such an analysis would be extremely time-consuming for the analyst if it were to be performed manually. Now that the IGSCR algorithm is automated, a factorial analysis will be much more feasible. It will also be possible to determine scene-specific optimal parameter values based on the maximization of the accuracy of a set of known points.

After the initial selection of parameter values to be used in the IGSCR algorithm, each step in the process is completely objective. However, when performed manually the process is prone to user error and can be quite time-consuming. Now that the IGSCR algorithm is automated, the analyst can select the input image and parameter values, then leave the process to run to completion. This results in significant time savings and prevents implementation errors.

2.3 Training data collection

As previously mentioned, the collection of training data for IGSCR is unlike the training data collection involved in a supervised classification. This is because each area collected for use in the IGSCR algorithm includes pixels of one informational class regardless of the spectral class into which the pixels will ultimately be categorized. This removes much of the subjectivity and work involved in a supervised classification resulting from the requirement that each training area must represent a spectral class that is normally distributed and highly separable from other spectral classes. Despite these differences, many of the methods developed for collecting supervised training data could be adapted for use in IGSCR.

No matter what classifier is used, it is important that the training data are temporally correlated with the image. If training data are collected long before or after the date of image capture, classification errors may occur due to land cover changes (Richards and Jia, 1999).

2.3.1 Traditional

Traditional methods of training data collection are generally polygon-based. Training data may be gathered via ground surveys, image interpretation, or with the assistance of ancillary data (Richards and Jia, 1999). Common methods include the use of expert knowledge to manually or digitally delineate areas over maps, photographs or digital images.

2.3.2 Non-traditional

A number of alternative methods of generating training data have been developed for use in hybrid classifiers, since the requirements are not as strict for the spectral properties of the regions selected. For example, Bauer *et al*. (1994) defined 88-acre primary sampling units (PSUs), which were stereoscopically interpreted from 35mm color infrared aerial photography to determine cover type, density and size of areas. After the areas were delineated, each PSU was visited on the ground. Since ground truthing for a PSU requires that only a few locations be visited, though sampled more thoroughly, the fieldwork is easier than methods that require random point visits. Other advantages of using PSUs noted by Bauer *et al*. (1994) are variance reduction and collection of proportional area data rather than binary counts.

Maplets are similar to PSUs in that they are delineated areas of a specified extent. Scrivani *et al*. (2001) created maplets via heads-up digitizing over Digital Orthophoto Quarter-Quadrangles (DOQQs). The maplets were then checked for changes after the date of the DOQQs using panchromatic, multi-spectral and pan-sharpened Landsat ETM+ (Enhanced Thematic Mapper Plus) images and edited as necessary. Two disadvantages of using this method operationally in a national forest inventory program are that the process is subjective and analyst-intensive, and high-resolution digital imagery such as DOQQs is not nationally available (Scrivani *et al*., 2001). Stoms (1996) also used maplets, but as validation data for accuracy assessment, not for training.

Wayman *et al*. (2001) tested two innovative, yet expensive, methods of training data collection for use in IGSCR. In one of these methods, a helicopter equipped with real-time Global Positioning System (GPS) equipment was used to collect polygonal areas, called helopolys. This technique allows for collection of data in any terrain, produces data quickly, and has highly accurate positioning. The other novel means of training data collection developed by Wayman *et al*. (2001) was the use of aerial videography to check and edit polygons that were heads-up digitized over orthophotography.

2.3.4 Semi-automated

Buchheim and Lillesand (1989) developed a semi-automated method of extracting training data that used a least-variance approach to “grow” training areas. The collection of training data was based on the coordinates of a manually selected seed pixel, and several user-specified data extraction parameters, which were varied for each region. The growth of a region originates at the seed pixel and moves outward via one of two possible growth strategies: linear or concentric. The linear method selects from among the pixels that are spatially adjacent to the existing training area, and adds the pixel which increases the summed variance of the field the least (or decreases it the most). The concentric growth method adds individual pixels to the training area as it moves outward from the seed pixel in concentric circles. For either method, growth is terminated for an area when either the absolute summed variance, the number of pixels, or the relative-variance-increase ratio has reached the user-defined maximum value.

The method developed by Buchheim and Lillesand (1989) established that a semi-automated region-growing approach can improve analyst efficiency and result in accuracies that are comparable to those attained via other training data collection methods. However, the system did not improve classification accuracy, nor did it employ any spatial, textural or contextual data (Buchheim and Lillesand, 1989) or persistent type-specific information (Bolstad and Lillesand, 1991) to maximize analyst efficiency. To address these issues, Bolstad and Lillesand (1991) also designed a semi-automated training data collection method to define spectral classes for direct use in a maximum likelihood classification. Their method maintained or improved classification accuracy and significantly reduced analyst input.

The system developed by Bolstad and Lillesand (1991) requires that an analyst identify seed pixels in homogeneous features and name them according to their informational class. Within their system, the size of a region is also increased until a growth-terminating threshold is reached. This method is also similar to that developed by Buchheim and Lillesand (1989) in that it evaluates pixels that are spatially adjacent to the seed pixel and adds them to the training set if they meet a set of user-defined criteria. However, the important difference in this method is that pixels are added to a training area based on predefined characteristics of different feature types such as expected size, shape and spectral variability.

The drawback of using the Bolstad and Lillesand (1991) method operationally for the FIA program is the difficulty inherent in developing type-specific region growing characteristics for forest and nonforest areas across the entire United States. The size, shape, and spectral variability of forests in the coastal plain of Virginia differ greatly from those in Minnesota and even those of forests in the mountains of Virginia. Therefore, regions with similar characteristics would have to be delineated. The trained analyst time required to delineate regions in the United States with similar characteristics, determine growth thresholds based on the expected size of an area of a certain land use, and determine the appropriate values for shape parameters, summed variance thresholds, and variance increase ratios specific to those regions diminishes some of the potential time saved by this method.

In addition, a substantial portion of the time reduction resulting from the use of the method described by Bolstad and Lillesand (1991) was due to the partial automation of the training set refinement process and the creation of a reusable color table for use in recoding classified images. The training set refinement required for a traditional, supervised classification is eliminated through the use of hybrid classifiers such as IGSCR, and the recoding process for binary classifications requires trivial effort.

The seed pixel-based region growing approach, which was used by Scrivani *et al*. (2001) and is employed in this study, uses the spectral properties of the region spatially adjacent to the seed to develop training areas in a different manner than either of the previously described methods (Buchheim and Lillesand, 1989; Bolstad and Lillesand, 1991). Rather than using the summed variance, this method uses the spectral Euclidean distance.

In this method, the analyst must locate the seed pixel on the image and choose an initial Euclidean distance value to use in growing the region. Pixels are added to a training area if they are within the user-defined spectral distance from the mean of the pixels already within the area (Scrivani *et al*., 2001). If the area created is completely within the same land use as the seed pixel, the region can be accepted or grown again by increasing the Euclidean distance. If the area includes pixels of a different land use, the Euclidean distance is decreased and a new region created. This process is repeated until the analyst decides that the area is completely within the land use of the seed pixel. Growth can be terminated when a user-defined maximum number of pixels is reached or when the user decides that the area is large enough to represent the spectral variation in the land use in the area surrounding the seed pixel.

Parameter value specification can be a problem in semi-automated training data extraction techniques. In some cases, it may be difficult to determine appropriate values for the parameters without substantial previous experience with their use (Buchheim and Lillesand, 1989). In the methods developed by Buchheim and Lillesand (1989) and Bolstad and Lillesand (1991), the user must specify seed pixel coordinates and parameter values for the entire image prior to creating the areas. If the analyst decides that any of the areas created include pixels of the wrong land use, the entire process must be repeated. The process of selecting appropriate parameter values is therefore fairly detached from the creation of the training areas. With the region-growing approach used in this study, each training area is created individually. Although this adds to the time required to collect training data, it ensures that all of the data collected are of high quality and it alleviates the problem of detached parameter specification.

2.3.5 Automated

McCaffrey and Franklin (1993) described a training site selection method that is almost entirely automated. They developed a computer program that that chooses homogeneous training sites (analogous to the photomorphic area developed in traditional photo interpretation) based only on simple statistics such as the coefficient of variance, the F-statistic and the Student's t-statistic. The program is designed to minimize human bias by using homogeneous sites chosen purely through statistical methods to aid the image analyst in the selection of representative areas for use in a supervised classification. The image is scanned with a 3 x 3 window and performs homogeneity tests based on a user-defined coefficient of variance independently on each band. The pixel in the center of the window is considered to be non-homogeneous and discarded if any of the bands fails the homogeneity test. To develop true class definitions, newly established homogeneous pixels are compared with groups of previously selected homogeneous pixels using the Student's t-test.

The output of the program is a set of homogeneous pixels, which McCaffrey and Franklin (1993) suggest could be used to select sites for field visits that will aid in the development of the classification scheme that will be applied to the image. This is not necessary for forest inventory since the scheme is predetermined by the need for a high quality binary classification into forest and nonforest. Alternatively, McCaffrey and Franklin (1993) point out that some of the points could be used as reference locations to be ground checked and used in an accuracy assessment of the final classification. However, the use of existing forest inventory plots with known land uses as ground truth in assessing the accuracy of images classified for forest area estimation will preclude the need for homogeneous pixels selected through an automated process. Also, heterogeneous validation points are essential to calculating the standard error of the forest area estimates obtained from image classifications.

McCaffrey and Franklin (1993) also propose that the sites could be used in a seed pixel-based classifier such as the ones described above (Buchheim and Lillesand, 1989; Bolstad and Lillesand, 1991; Scrivani *et al*., 2001), or in an unsupervised classification. The research project described herein requires a set of points to initiate a seed pixel-based training data collection process. However, the points selected need not be pixels selected from spectrally homogeneous areas since the training data are not used directly in a supervised classification. The automated training site selection program developed by McCaffrey and Franklin (1993) ultimately amounts to a photomorphic area identifier, which is not what is needed for the collection of training data for use in the IGSCR algorithm.

2.4 Post-processing

Due to the spectral variability inherent in satellite images, classifications sometimes contain individual or small clusters of pixels of one land use within large areas of another land use. In some cases, it may be beneficial to "smooth" the classified data by replacing anomalous pixels with the predominant class value of the spatially adjacent pixels (Lillesand and Kiefer, 2000). Smoothing algorithms use logical operators to determine whether or not to replace a pixel value and, if so, which class should be substituted. One commonly used postclassification technique is a majority filter, in which a moving window of a user-defined size (e.g., 3x3 pixels) is passed over the classified image and the majority class value in the window is assigned to the center pixel in the window.

Post-processing is performed based on the supposition that the aberrant pixels are not correctly classified. However, this is not necessarily true. Since each pixel in the Landsat ETM+ images used in this study represents 900 m^2 (approximately 0.22 acre), it is conceivable that a single nonforest pixel in a large tract of forest may encompass a clearing and a small house, which is not a forest land use. Conversely, a single "forest" pixel is not considered to be a forest based on the FIA definition. In fact, any group of 4 or less "forest" pixels does not meet the FIA definition of a forest, since it takes almost 5 pixels to constitute an acre. This indicates that class-specific post-processing may sometimes be applicable.

Given that the spectral properties and land cover in each image vary, it is not clear whether post-processing is beneficial in all cases, or which method might produce improved results. In particular, it would be of value to determine if post-processing can increase the accuracy of forest/nonforest classifications, thus improving the precision of the derived forest area estimates. This project compares several common post-processing procedures to determine the effects of post-processing on forest area estimates.

2.5 Validation data

After an image has been classified and post-processed, ground truth data are used to assess the accuracy of the thematic map, as well as being used in calculations that determine the true percent forest and estimate its variance. The ground truth data (also referred to as validation data) that were used in this study are FIA phase 2 plot data. The FIA plot design is a cluster of four 1/24 acre plots (Fig. 11) distributed over an area of about 1 acre, which is approximately the size of four TM pixels (Riemann *et al*., 2000). The coordinates of the FIA plot centers were obtained using hand-held GPS data collection units, the accuracy of which is better than 10 m after differential correction (Scrivani *et al*., 2001).

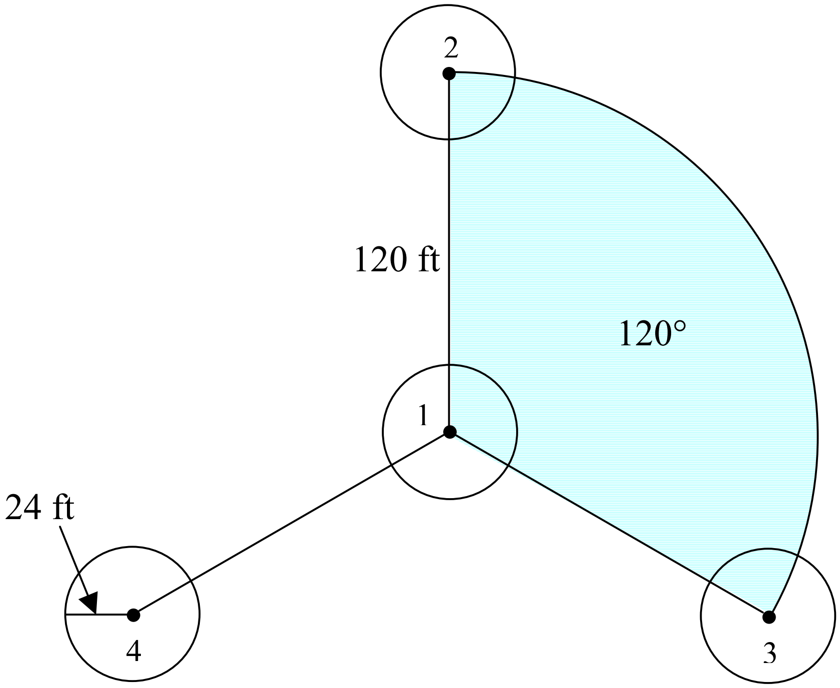


Fig. 11. FIA plot design, showing 4 subplots, their radius and the distance and angle between subplots (modified from USDA-FS, 2001).

The FIA plot dataset is well suited for accuracy assessment because it is designed to be objective, unbiased, consistent and ongoing (Riemann *et al*., 2000). In addition, because the FIA plots are based on a systematic triangular grid (Roesch and Reams, 1999), the accuracy characterization is more statistically informative than one in which no heterogeneous plots are included in the sample (Franco-Lopez, *et al*., 2001).

The greatest benefit of using the FIA data in a study such as this one is that it facilitates the comparison of images from a variety of locations within the United States. Congalton and Green (1999) remark that the misinterpretation of a classification scheme, or use of different definitions when ground data are collected, are the causes of many of the errors reported by accuracy assessments. These types of errors could be compounded when comparing scenes from different areas classified by many analysts. These issues are not a problem with FIA plot data, which are highly uniform and of very high quality (Riemann *et al*., 2000).

Among the data collected at each phase 2 plot are the areas of forest and nonforest land cover on each subplot, which can be summarized as percent forest on the ground. Riemann *et al*. (2000) defined a forested plot as one which contains >50% forest on the ground. After assessing the accuracy of an image using this definition, they determined the type of error that occurred via photo interpretation of the misclassified plots on 1:40,000 photography. The errors were divided into 4 classes: location, resolution, definition and classification. Location errors were misclassifications related to the fact that the ground plot fell within 1 pixel (30 m) of a forest edge. Resolution errors were those occurring in areas where the forest cover was fragmented into patches 1 pixel or smaller in size. Definition errors resulted from the difficulty of trying to classify land use from satellite images that represent ground cover. When the satellite image classification was clearly wrong for no apparent reason, the misclassifications were termed classification errors.

Riemann *et al*. (2000) found that 87 of 452 plots were misclassified. However only 23.5% of the errors were due to classification errors. The majority of errors were location errors (31.5%), and resolution errors (27%) although there were also a limited percentage of definition errors (18%). They conclude that FIA plots can be used in the accuracy assessment of satellite image classifications. However, the results of such an accuracy assessment must be interpreted thoughtfully, giving consideration to the breakdown of potential error sources.

Another interesting study presented in Riemann *et al*. (2000) involved the examination of the homogeneity of the pixels in a 3x3 neighborhood around the plot. The areas were separated into three classes: homogeneous (all 9 pixels were of the same class), intermediate (window primarily homogeneous, but with 1 or 2 pixels from the other class), and mixed (containing 3-6 pixels from each class). When the accuracy of each of these three area types was assessed separately, it was clear that the homogenous areas were far more accurate than the intermediate areas, and the intermediate areas were much more accurate than the mixed areas, with average accuracies for the sub-areas over 4 classified images of about 90, 70, and 60%, respectively. These results indicate that within a classified image, there are local differences in accuracy. The conclusion that must again be drawn is that the accuracy assessment results must be interpreted with care.

The studies performed by Riemann *et al*. (2000) showed that the FIA plot data have the potential to reveal more information about the classified image through the use of variations on a typical accuracy assessment. The plot center land use has been used as the reference information class in some FIA research projects (Scrivani *et al.*, 2001; McRoberts *et al.*, 2002), however, Riemann *et al.*, (2000) defined a forested plot as one which contains >50% forest on the ground. In this study, four validation sets were created for each image, one using the center land use data for all plots and three reduced sets: the plots in which 1) the majority land use agreed with the center land use, 2) 75% or more of the plot agreed with the center land use, and 3) the land use was homogeneous. The validation data were thus refined in an attempt to make them more representative of the predominant land use in the area surrounding their point locations (since a pixel represents land use information in a 30 m^2 area, not at a point). It is not clear how these variations in validation data will affect the overall accuracy or the precision of the resulting forest area estimates, since the number of plots used in each of the validation sets will vary.

Chapter 3
METHODS

3.1 IGSCR algorithm automation

The complete automation of the IGSCR algorithm was accomplished using ERDAS Macro Language (EML) and the ERDAS Imagine C toolkit. EML was mainly used to create a graphical user interface, although it was also used to perform a small number of graphical object-related functions. The ERDAS Imagine C Toolkit allowed access to almost all of the existing functionality of ERDAS Imagine software. Many of the modules available through the ERDAS Imagine user interface, such as unsupervised and supervised classification, were accessed directly through the ERDAS Imagine C toolkit. This obviated the need to rewrite well-designed programs.

3.2 Experimental design for remainder of study

This study was a cooperative effort between the Virginia Tech Department of Forestry, the Virginia Department of Forestry, and the USDA Forest Service North Central Research Station. The Virginia Department of Forestry provided 3 preprocessed Landsat ETM+ images (each of which consisted of 2 or more scenes merged together), and validation data for the project. The USDA Forest Service North Central Research Station provided 1 preprocessed Landsat ETM+ scene, training data collected by trained image analyst, and validation data for the project. I collected the training data for the 3 Virginia images, carried out all of the image processing required for the project and performed the accuracy assessments of the classified images. I performed the analysis of the data.

3.3 Study areas and imagery

A typical Landsat ETM+ image is 183 x 170 km (USGS, 2003), and is named by its path and row in the Landsat Worldwide Reference System (WRS). The Minnesota Landsat ETM+ image is a single full scene, WRS path 28, row 28 (Fig. 12 and Fig. 13; Table 2). The Landsat ETM+ images used in Virginia are not typical Landsat scenes because the images for a certain date in three paths through Virginia were merged into three large images (Fig. 14; Table 2): one for path 15 ("15m"; "m" stands for "merged"), one for 16 ("16m") and one for 17 ("17m"). The merged images were then cut using the Virginia state boundary and the large water bodies in the eastern portion of Virginia were removed from the 15m image using the digital shoreline coverage created by the Virginia Institute of Marine Science as part of their Comprehensive

Coastal Inventory Program (Center for Coastal Resources Management, 1991). As is shown in Fig. 14, there is significant side lap between adjacent scenes.

Table 2. Landsat ETM+ images used in this project, and for each image: the WRS path and row(s), the date of capture, and the total image area.

Image	WRS **Path**/Row(s)	Date	Area (hectares)
15m	**15**/34,35	3/08/2000	3095351
16m	**16**/34,35	2/28/2000	3513834
17m	**17**/33,34,35	4/03/2000	2679556
MN	**28**/28	7/23/1999	3209983

3.3.1 Description

The three images in Virginia cover multiple physiographic provinces, all of which have a northeast to southwest orientation (Fig. 15). The easternmost image, 15m, is almost entirely in the Coastal Plain, with the western portion stretching into the Piedmont (Fig. 14 and Fig. 15). This area is relatively flat, contains two fairly large dammed lakes, and is permeated with rivers running to the Atlantic Ocean. The image contains the highly populated areas of Richmond and Newport News/Hampton, suburban housing, rural and agricultural areas, pine plantations, hardwood and mixed forest, and wetlands, including a portion of the Great Dismal Swamp. The land cover is highly fragmented, with only a few wet or swampy areas appearing to be undisturbed.

The next image to the west, 16m, spans the entire Piedmont region, several Atlantic Rift Basins, includes a large section in the Blue Ridge as well as a portion of the Valley and Ridge province (Fig. 14 and Fig. 15). This image ranges from flat in the east to mountainous in the west, and contains a few fair-sized lakes and waterways. The eastern portion of 16m is similar in land cover to the 15m image, although it becomes gradually less fragmented from southeast to northwest. This is because the Blue Ridge Mountains consist of large expanses of predominantly deciduous forest, and the valleys contain large areas of contiguous agricultural land.

The westernmost image in Virginia, 17m, is mainly in the Valley and Ridge province but its southeastern corner is in the Blue Ridge and its western edge just reaches into the Allegheny Plateau (Fig. 14 and Fig. 15). This image is by far the most mountainous of the three images, which also makes it the least fragmented, since the mountains are mostly covered with forest, and the valleys primarily contain agricultural land. Again, there are a few noteworthy lakes and

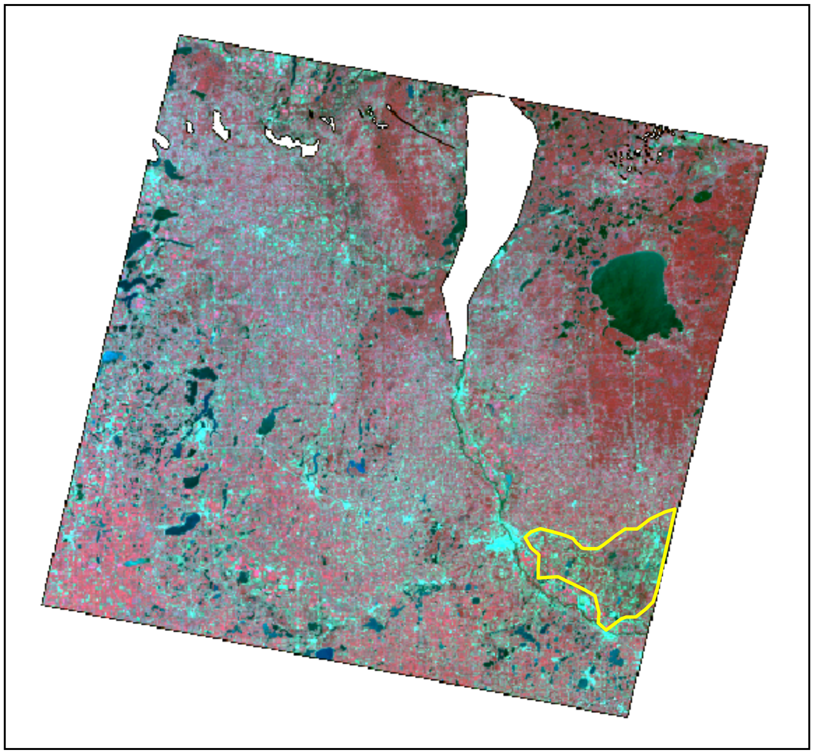


Fig. 12. Landsat ETM+ image of central Minnesota (path 28/row 28) with bands 4, 3, and 2 in red, green and blue, all cloud and cloud shadow areas removed and the Anoka Sand Plain Area in yellow.

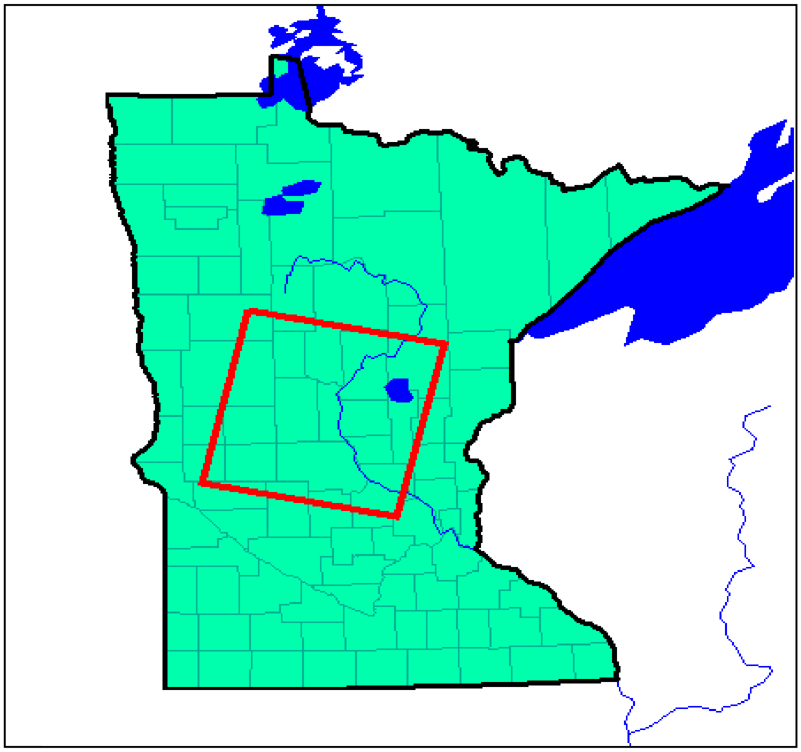


Fig. 13. Map of Minnesota, showing the location of WRS path 28, row 28 as well as county boundaries, major lakes and rivers.

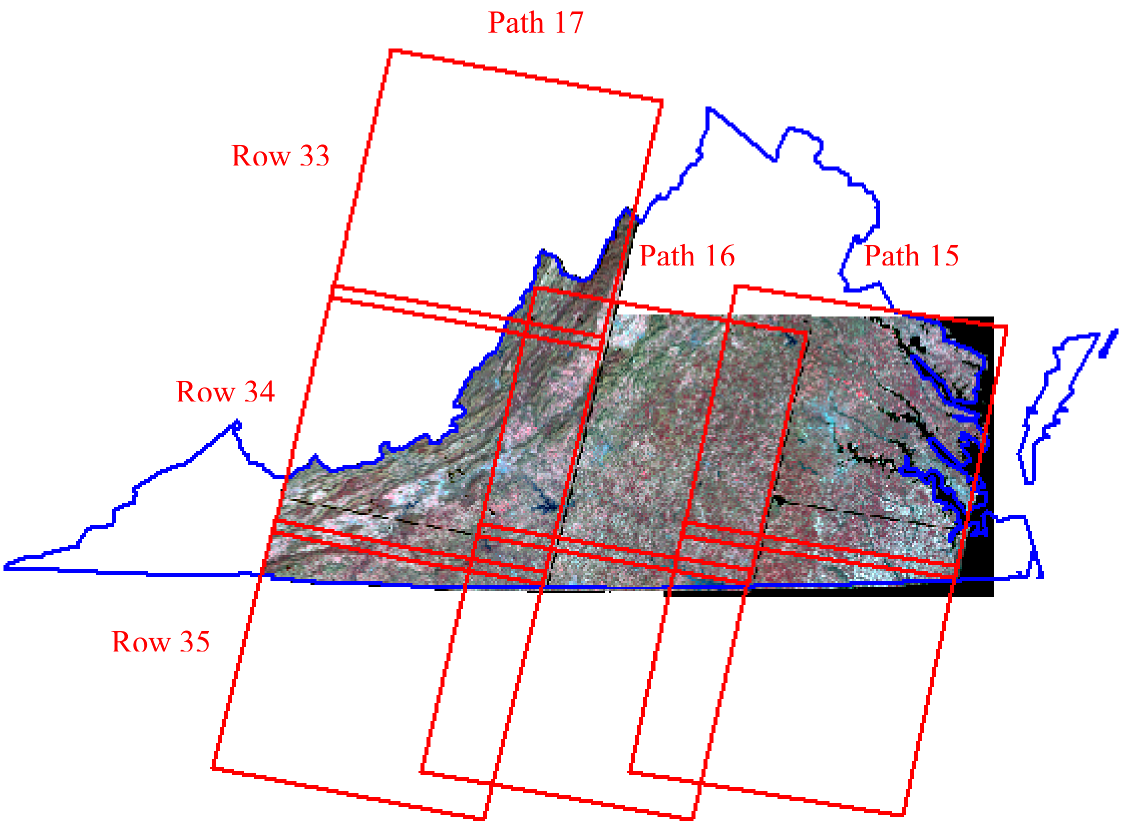


Fig. 14. Three merged Landsat ETM+ images of Virginia (15m, 16m and 17m) with bands 4, 3, and 2 in red, green and blue, all and cloud, cloud shadow and bad data areas removed.

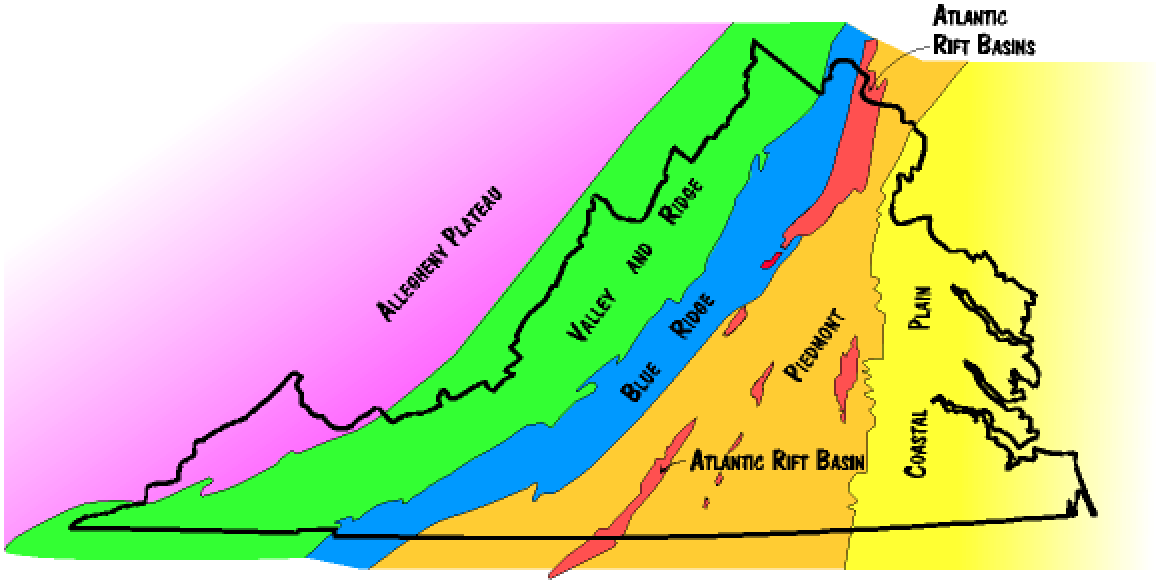


Fig. 15. Physiographic provinces in Virginia (Lynn S. Fichter; http://geollab.jmu.edu/vageol/vahist/PhysProv.html, used with permission).

rivers, but there are almost no wetlands in this image. Other than Roanoke, which is an urban area, there are few cities in this image that are more than minimally populated.

Although there is only one scene in Minnesota, it covers more physiographic regions than the three Virginia scenes combined (Wright, 1972). These are mainly the result of glacial action: moraines, which are areas covered by rocks and debris carried down and deposited by glaciers; drumlins, which are composed of clay molded by glacial action; and part of a till plain, which is clay containing rocks and sand deposited by melting glaciers and ice-sheets (Hawkins and Allen, 1991). In the southeast corner of the image there is an area called the Anoka Sand Plain Area, which is visibly distinct from its surroundings because it contains less agricultural land and more coniferous stands and wetlands (Fig. 12). The area of the Minnesota represented in the image is characterized by numerous lakes and wetlands (hence the name of one region: Sugar Hills Mille Lacs Moraine Area), a small amount of forest, abundant agricultural fields and sparse population.

3.3.2 Rectification

As noted by Wayman *et al*. (2001), the image rectification process is not objective and is, therefore, in need of further research into the development of a standardized methodology. This project does not address the need to develop a consistent means of image rectification. Instead, the typical, somewhat subjective, methodology was employed. The Landsat 7 ETM+ images that were used in this study were pre-processed at the Earth Resources Observation Systems (EROS) Data Center to level 1G, which denotes "a geometrically rectified product free from distortions related to the sensor (e.g., jitter, view angle effects), satellite (e.g., attitude deviations from nominal), and Earth (e.g., rotation, curvature)" with a geodetic error of approximately 30 m (Farr, 1999a). Further rectification was performed at the Virginia Department of Forestry and the Virginia Economic Development Partnership using 30 or more ground control points to reduce the root mean squared error (RMSE) in the images to less than half a pixel (Table 3). No further rectification of the Minnesota image was performed.

Table 3. Rectification information for Virginia images (listed by WRS path/row). Some data were not available (N/A).

Image	Capture date	RMSE (m)			Polynomial order	Resampling method	Registration date	Analyst
		Control points	Check points	All points				
15/33	3/08/2000	7.01	8.27	8.44	2	NN	2/12/2001	Ciminelli
15/34	3/08/2000	N/A	N/A	N/A	1	NN	7/25/2000	Musy
15/35	3/08/2000	6.57	10.56	9.63	1	NN	2/22/2001	Ciminelli
16/34	2/28/2000	11.89	12.43	12.22	1	NN	7/26/2000	Musy
16/35	2/28/2000	12.04	7.93	8.93	2	NN	2/6/2001	Ciminelli
17/33	3/06/2000	N/A	N/A	N/A	2	NN	N/A	Ciminelli
17/34	3/06/2000	14.08	12.32	12.67	1	NN	3/21/2001	Ciminelli
17/35	3/06/2000	20.17	14.36	14.80	1	NN	3/21/2001	Ciminelli

3.3.3 Additional pre-processing

No adjustments to the brightness values in the images, such as haze reduction or atmospheric correction, were performed. The images were manually edited to remove clouds, cloud shadows, bad data lines, and water with high reflectance values. In addition, the feathered edges of the Landsat ETM+ images were removed.

We developed a repeatable method of removing the feathered edges from an image, which we termed "plucking". To "pluck" the image: subset band 1 from the image and use it to mask the image (the recode does not have to be set up since only the zeros in the background will be used to reduce the image); repeat with band 5.

3.4 Training data collection

The training data for input into the IGSCR program consisted of forest and nonforest areas developed by a trained image analyst at 200 random locations within an image using the seed pixel region growing approach.

3.4.1 Analysts

Greg Liknes, of the USDA Forest Service North Central Research Station, and I performed the training data collection for this project. Greg collected training data for the Minnesota image and I collected the data for the three Virginia images.

Greg and I both have a moderate amount of experience with image classification. We have both had practice in performing ISODATA classifications and in the collection and refinement of training data for use in maximum likelihood classifications. Prior to this work, Greg had done little or no hybrid classification, whereas I had some previous experience with IGSCR and other hybrid algorithms. Much of our image classification experience was gained through coursework and projects performed during our graduate studies.

3.4.2 Protocol

A random sample of 200 points per Landsat TM scene were used as seed pixel initiation points to collect training data via the seed pixel-based region growing approach (developed and presented at the January, 2001 IGSCR Training in Blacksburg, Virginia; Appendix B). The random samples were collected using the Create/Add Random Points function in the ERDAS Imagine Accuracy Assessment Module. The random points were selected from within the preprocessed Landsat ETM+ images. The coordinates and spectral information for each of the points used in training are found in Appendices C, D, E and F.

It should be noted that the points in Minnesota were informally geographically stratified. The analyst tried to ensure that each of the counties in the image contained some training data. The initial set of 1,000 points was selected randomly using the ERDAS Imagine Accuracy Assessment Module, however, the analyst performed the selection of points from that random set within each county. Three hundred total training areas were collected in this manner, and 200 of those were selected at random for use in the classification process.

Most of the points selected from the image were visually interpreted by the analyst without the aid of ancillary imagery. Visual interpretation was not excessively difficult since the analyst only had to decide whether a point was forested or not. If the analyst was uncertain of the land use call, the point was dropped and another random point selected. In Minnesota, digital orthophotographs were used to aid in land use interpretation whenever necessary.

3.5 Image processing

All four images were processed using the IGSCR program that was automated as part of this project. The parameters used in running the IGSCR program on each of the four images are listed in Table 4 and Table 5.

Table 4. Options available for IGSCR classification and values selected.

IGSCR Options	
Maximum iterations	15
Type-I error rate	0.05
Homogeneity threshold	0.95
Maximum Likelihood	yes
Edge model	yes
Stacked ISODATA	yes

Table 5. Options for ISODATA portion of IGSCR classification and values selected.

ISODATA Options	
Maximum iterations	100
Number of classes	100
Convergence threshold	0.975
Initialize means along	principal axis
Scaling range	1 std. deviation

3.6 Post-processing

The maximum likelihood images produced by IGSCR were filtered with two 3x3 majority filters, and three clump and eliminate methods in an attempt to improve the agreement of the classification with the land use on the ground.

3.6.1 Majority filters

The majority option in the Neighborhood module of ERDAS Imagine was used to run a 3x3 majority filter on all of the images (ERDAS, 1999). The function was set to include all values in the computation and apply the function at all values. The same function was used to run a 3x3 majority filter that was applied only to the forested pixels. These filters determine the majority class value in a 3x3 window and assign it to the center pixel in the window. Majority filters function as a means to smooth the data.

3.6.2 Clump and eliminate

The Clump function in ERDAS Imagine was used to label each pixel in the images with a value indicating the size of the contiguous cluster of pixels (of one thematic class) that contained the pixel (ERDAS, 1999). This function has 2 options for the way that pixel clusters can be defined: 1) only the adjacent 4 pixels in the cardinal directions are considered to be contiguous with the center pixel, or 2) all 8 neighbors of a given pixel are contiguous with the center pixel. It appeared, upon visual inspection of a subset image that was clustered by each method, that the definition of the neighborhood might significantly affect the resulting filtered image. Therefore, images were clumped with both 4 and 8 contiguous neighbors.

The clumped images were then processed using the Eliminate function in ERDAS Imagine. This function removes from a clumped image any clumps that are smaller than a user-defined number of pixels. It replaces pixels in those clumps with the majority thematic value of the surrounding pixels and recodes the remainder of the image to the original class values. Since the purpose of filtering the images in this way is to produce a classification that more closely matches the FIA forest definition, clumps smaller than 1 acre were removed. One acre is approximately 4.5 pixels, so clusters of 4 pixels or less were removed. To do that, the minimum size was set to 5 in the Eliminate module, since clumps smaller than that size are eliminated.

Since the FIA definition of forest does not include areas less than an acre in size, but non-forested areas less than an acre of size are still considered nonforest, the nonforest pixels that were removed using the clump and eliminate process were replaced using a spatial model which used the original maximum likelihood thematic class value if the pixel was nonforest and the class value from the "eliminated" image otherwise. The accuracy of image that was clumped with 8 neighbors, then had areas of 4 or less pixels removed was also assessed without adding back the nonforest to determine whether or not replacing the nonforest improves the overall accuracy of the image.

3.6.3 Dealing with zeros

All of the filtered images contained more zero pixel values than the unfiltered maximum likelihood images. In order to keep the image area constant between the maximum likelihood images and their filtered derivatives, a model was created to replace the original class value from the maximum likelihood image where the filtered images created new zero pixels. All of the filtered images were processed using this model.

In the 3x3 majority filtered images, sometimes the majority value in a window containing a zero at the center was not zero. This resulted in a class value being present in the filtered image where there was a zero in the original image. To deal with this problem, a second model was created to replace zeros from the original image. This model was run on the 3x3 majority filtered images only, since the 3x3 majority on forest did not filter zero pixels and the clump and eliminate models did not create new zeros.

3.6.4 Abbreviated filter names

In tables and charts, the filters described above will be referred to as follows: 0 = no filter; 3x3maj = 3x3 majority filter on entire image; 3x3on2 = 3x3 majority filter on forest only;

clump1 = clump with 4 neighbors, eliminate areas of 4 pixels or less, and replace nonforest; clump 2 = clump with 8 neighbors, eliminate areas of 4 pixels or less, and replace nonforest; clump 3 = clump with 8 neighbors, eliminate areas of 4 pixels or less, and do not replace nonforest.

3.7 Accuracy assessment

The data used to validate the classifications were FIA phase 2 plot data, which are described in section 2.5.

3.7.1 Validation data

The number of validation points and the total area of the image are critical parameters in the calculation of the precision of forest area estimates. To prevent artificial differences in the resulting precision estimates due to discrepancies in the number of plots used per unit area, the density of plots within each image was standardized. To obtain an equal plot density in each image, the image areas (based on the number of non-zero pixels) were compared and used along with the total available plots in each image to determine which image had the lowest density. The number of plots in the more densely sampled images was then reduced to obtain the number of plots necessary to attain the same density for all images. The Minnesota data set contained the fewest plots per unit area (1.63×10^{-4} plots/acre). The Virginia scenes were reduced by 107, 63, and 37 plots for 15m, 16m and 17m, respectively, to decrease their plot density to that of the Minnesota image.

The FIA data used in this project as validation data were collected over time, since the FIA program currently collects statewide data on a 5-year cycle. The Minnesota data span 3 years (1999-2001) and the Virginia data cover 5 years (1997-2001). In order to improve the temporal agreement between the image dates and the plot dates, the plots that were removed from the Virginia images to equalize the plot density were randomly selected from the plots collected in 1997 (the year that was furthest removed from the image date in 2000).

Four validation sets were created for each image, one using the center land use data and three reduced sets: the plots in which 1) the majority land use agreed with the center land use, 2) 75% or more of the plot agreed with the center land use, and 3) the land use was homogeneous. The following are abbreviated names used hereafter for the validation sets: ctrlu = center land use, 50% = majority land use agreed with center land use, 75% = 75% or more of the plot agreed with center land use, and 100% = the land use was homogeneous.

Only the plot density of the center land use set is equal among the images. The number of points in the other sets were based on the number of plots within the center land use set that also met the plot land cover requirements. The total number of plots used per image in each of the four validation data sets and the percent reduction in plots for each validation set are listed in Table 6.

Table 6. The number of validation points used in each image for each accuracy assessment set and the percent reduction between the center land use sets and the reduced sets.

# of points	Accuracy Assessment Set / **Reduction (in % of ctrlu set)**						
Image	ctrlu	**ctrlu - 50%**	50%	**ctrlu - 75%**	75%	**ctrlu - 100%**	100%
15m	1248	**2.1**	1222	**5.4**	1180	**19.8**	1001
16m	1416	**1.6**	1394	**5.0**	1345	**19.1**	1145
17m	1080	**1.6**	1063	**4.4**	1032	**16.3**	904
MN	1294	**0.9**	1282	**3.2**	1253	**9.8**	1167

3.7.2 Parameters

A contingency table, such as the one shown in Table 1, was created for each image using the validation data in the Accuracy Assessment module in ERDAS Imagine. In the Accuracy Assessment module, the user selects the means by which accuracy point locations are assigned a class value under the heading "Class Value Assignment Options". In this study, we wanted to obtain the class value for the actual pixel in which the accuracy point fell. However, the "Window Size" option (under "Class Value Assignment Options") does not allow the user to select a window smaller than 3x3. To force the module into assigning the center pixel value, under "Window Majority Rule", the "Majority Threshold" option was selected and the threshold was set to 9. Then, in the "No Majority Action" section of the options window, "Use Center Value" was selected. The accuracy of this method was verified using points with known pixel values and various surrounding pixel configurations.

3.7.3 Statistics

The classifications will be compared, both qualitatively and quantitatively, based on their overall classification accuracy, kappa and kappa variance statistics, corrected forest area estimates and the precision of those estimates. The overall classification accuracy is the sum of the correctly classified points (along the diagonal in the contingency matrix) divided by the total number of validation points (Congalton and Green, 1999). The Kappa statistic indicates how the classification compared with what would have resulted from chance alone (Congalton and

Green, 1999). The equation used to estimate Kappa is as follows (from Congalton and Green, 1999):

$$\hat{K} = \frac{n\sum_{i=1}^{k} n_{ii} - \sum_{i=1}^{k} n_{i+}n_{+i}}{n^2 - \sum_{i=1}^{k} n_{i+}n_{+i}}, \tag{3}$$

where k = number of rows in the error matrix, n_{ii} = number of observations in row i, column i (correctly classified points), n_{i+} = number of observations in row i, n_{+i} = number of observations in column i, and n = total number of observations. Using the notation in Table 1, this equation can be written as follows (for a simple forest/nonforest classification):

$$\hat{K} = \frac{Total * \left(tf\_mf + tnf\_mnf\right) - \left[\left(total\ tf * total\ mf\right) + \left(total\ tnf * total\ mnf\right)\right]}{Total^2 - \left[\left(total\ tf * total\ mf\right) + \left(total\ tnf * total\ mnf\right)\right]}. \tag{4}$$

The equation used to estimate the large sample variance of Kappa is as follows (from Congalton and Green, 1999):

$$\hat{\text{var}}(\hat{K}) = \frac{1}{n}\left\{\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2\left(\theta_4 - 4\theta_2^2\right)}{(1-\theta_2)^4}\right\}, \tag{5}$$

where

$$\theta_1 = \frac{1}{n}\sum_{i=1}^{k} n_{ii}, \tag{6}$$

$$\theta_2 = \frac{1}{n^2}\sum_{i=1}^{k} n_{i+}n_{+i}, \tag{7}$$

$$\theta_3 = \frac{1}{n^2}\sum_{i=1}^{k} n_{ii}\left(n_{i+} + n_{+i}\right), \tag{8}$$

$$\theta_4 = \frac{1}{n^3}\sum_{i=1}^{k}\sum_{j=1}^{k} n_{ij}\left(n_{j+} + n_{+i}\right)^2, \tag{9}$$

i refers to map category, and j refers to reference category. The explanation of the notation for k, n_{ii}, n_{i+}, n_{+i}, and n is the same as that given in Equation 3 above.

An estimate of Kappa and its approximate large sample variance are used to calculate a test statistic to test the significance of a single error matrix, and can also be used to calculate the test statistic used to determine whether two independent error matrices are significantly different. The two tests are based on the standard normal deviate, designated with the letter Z. The formulae are as follows (from Congalton and Green, 1999):

$$Z = \frac{\hat{K}_1}{\sqrt{\hat{\mathrm{var}}(\hat{K}_1)}}\text{, and} \tag{10}$$

$$Z = \frac{\left|\hat{K}_1 - \hat{K}_2\right|}{\sqrt{\hat{\mathrm{var}}(\hat{K}_1) + \hat{\mathrm{var}}(\hat{K}_2)}}. \tag{11}$$

The null hypothesis used in testing the significance of a single error matrix is H_0: $K_1 = 0$, and the alternative hypothesis is H_A: $K_1 \neq 0$. The null is rejected if $Z > Z_{\alpha/2,}$ where $\alpha/2$ is the confidence level of a two-tailed Z test and the degrees of freedom are assumed to be ∞ (infinity). The null hypothesis used in the comparison of two independent error matrices is H_0: $(K_1\text{-}K_2) = 0$, and the alternative hypothesis is H_A: $(K_1\text{-}K_2) \neq 0$. H_0 is rejected if $Z > Z_{\alpha/2}$ (Congalton and Green, 1999). In this study, $\alpha = 0.05$, therefore $Z_{\alpha/2} = Z_{0.025} = 1.96$. In these tests, if the Z-score for the comparisons is greater than 1.96, the null hypothesis can be rejected.

3.8 Calculations

The Accuracy Assessment module in ERDAS Imagine performs the calculations needed to obtain the overall accuracy of the classification, and the estimated overall Kappa and also creates a contingency table (also called an error matrix) for the classification. The Accuracy Assessment module does not, however, compute the variance estimate for Kappa, which is needed to calculate the test statistic used in the comparison of two independent error matrices. The information in the error matrix is also needed to calculate the corrected percent forest and variance of percent forest in the image.

The output of the Accuracy Assessment module is a text file containing the statistics listed above and a contingency table. To facilitate the extraction of data for use in further calculations, two macros were written to open text files and copy the pertinent information into a preexisting spreadsheet (Appendices G and H). The automated transfer of information to the spreadsheet using a macro avoided the possibility of errors in transcription and greatly improved the speed at which the data were obtained from the text files.

The corrected forest area estimates for the classifications were obtained by multiplying the percent forest (*pf*) calculated using Equation (1) by the total area within the classified image. The variance of the proportion of forest area (*vpf*) was calculated using Equation (2). These values were used to calculate the precision of the forest area estimates on a percent-per-million-acres-timberland basis using the following equation:

$$\text{Precision} = \sqrt{vpf} * \sqrt{\frac{pf * \text{acres in image}}{1{,}000{,}000}} . \qquad (12)$$

Chapter 4
RESULTS AND DISCUSSION

4.1 IGSCR algorithm automation

I automated the IGSCR algorithm using ERDAS Macro Language (EML) and the ERDAS Developer's Toolkit. EML was used to create the Graphical User Interface (GUI) for the program. The ERDAS Developer's Toolkit was used to support the automation of the IGSCR algorithm (Fig. 16) in C code.

4.1.1 IGSCR program

4.1.1.1 Program interface

The IGSCR program is accessed through a button that was added to the existing ERDAS Imagine Classification menu (Fig. 17). The program takes input through a GUI similar to those used in ERDAS Imagine (Fig. 18, Fig. 19, and Fig. 20). It allows the user to choose the input image, training data and parameter values to be used in the IGSCR algorithm. The parameters which the user must select in the main window (Fig. 18) are the following (with default values in parentheses): type-I error rate (0.05), homogeneity threshold (0.90), maximum number of IGSCR iterations (10), number of ISODATA spectral classes (5).

In the ISODATA Options window (Fig. 19), the following options are available (also with default values in parentheses): axis along which to initialize the means (principal), scaling range (1 standard deviation), maximum ISODATA iterations (100), ISODATA homogeneity threshold (0.975). These options are a subset of the options available for unsupervised classification through the ERDAS Imagine Classification menu. The values of the other parameters sent to the ISODATA function are fixed. If other parameter values (or options) are deemed to be of important in the future, these can be changed (or added) to the program at that time.

The IGSCR Options window (Fig. 20) gives the user the option to output any of five possible final image products: maximum likelihood, stacked ISODATA, stacked ISODATA plus maximum likelihood on unclassified pixels, a four-class image based on the maximum likelihood image, or a four-class image based on the stacked ISODATA plus maximum likelihood image. The user also has the option to save or delete intermediate clustered image outputs and signature files.

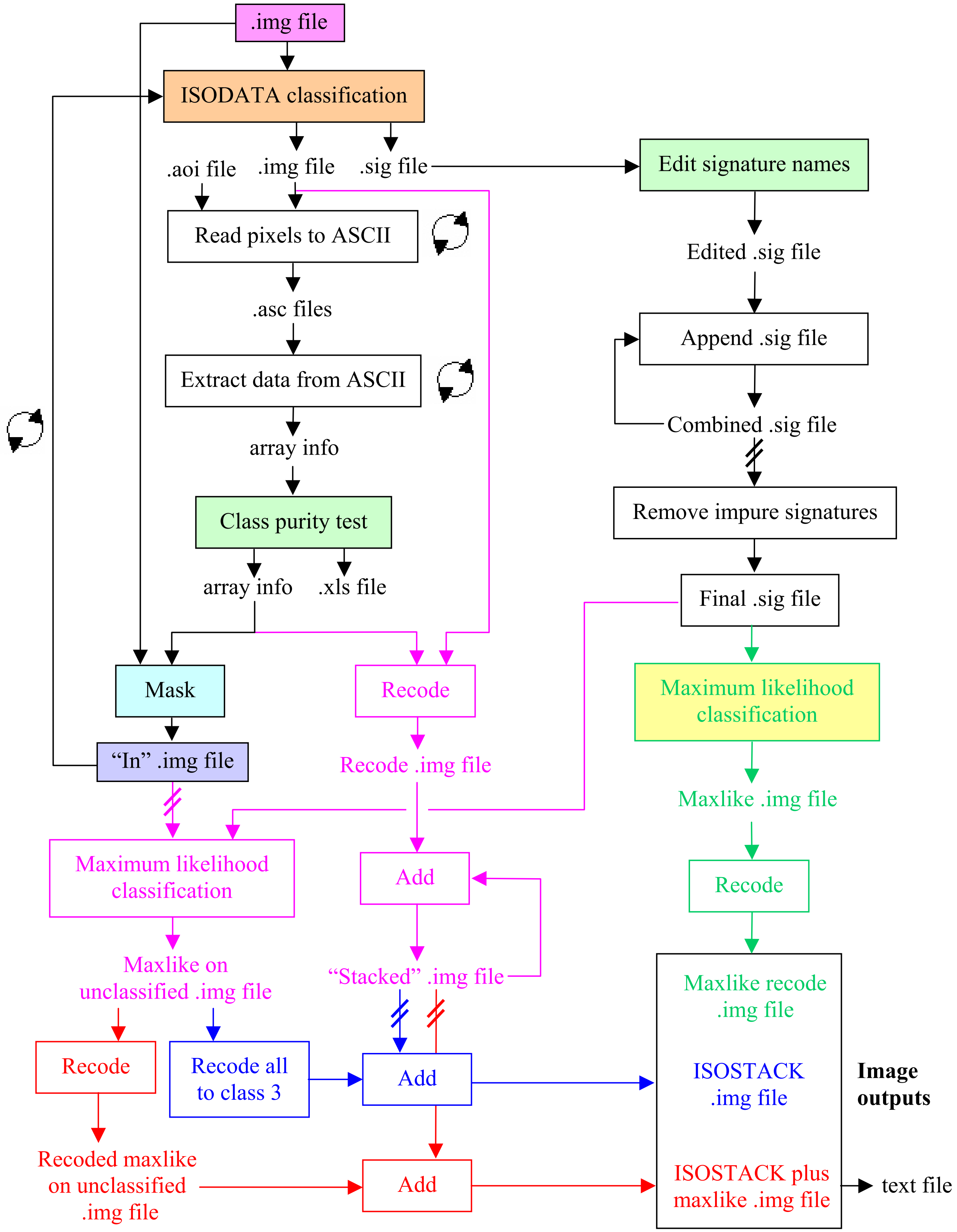


Fig. 16. Iterative Guided Spectral Class Rejection program flow chart, showing main loop (black), maximum likelihood option (green), stacked ISODATA option (blue), stacked ISODATA plus maximum likelihood option (red), and flow if either of the last two options are chosen (magenta).

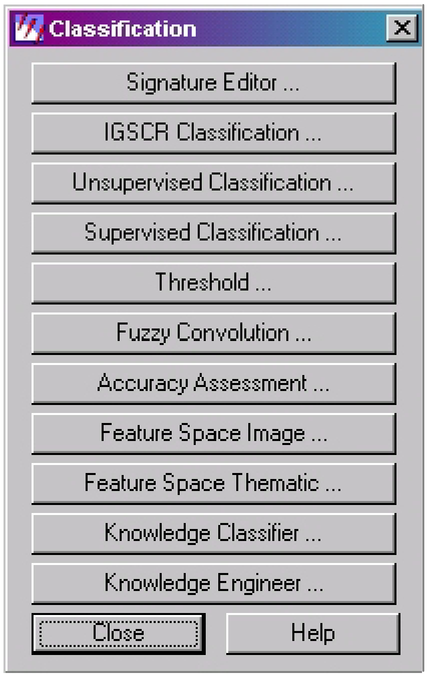
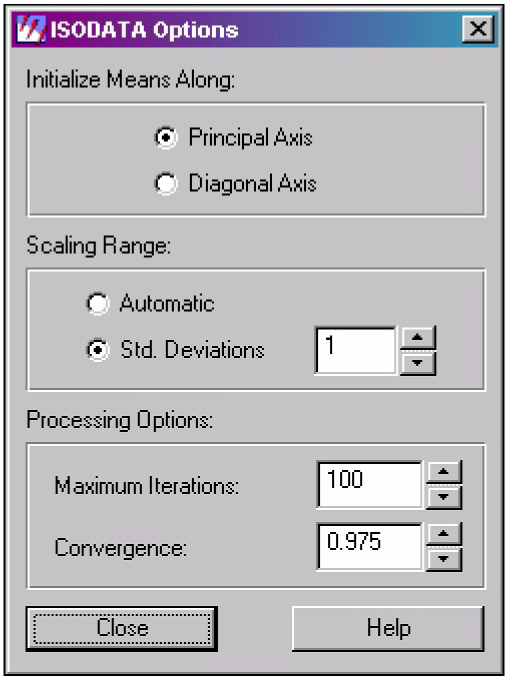


Fig. 17. Edited ERDAS Imagine 8.4 classification menu, with new "IGSCR Classification…" button which opens the IGSCR interface (shown in Fig. 18).

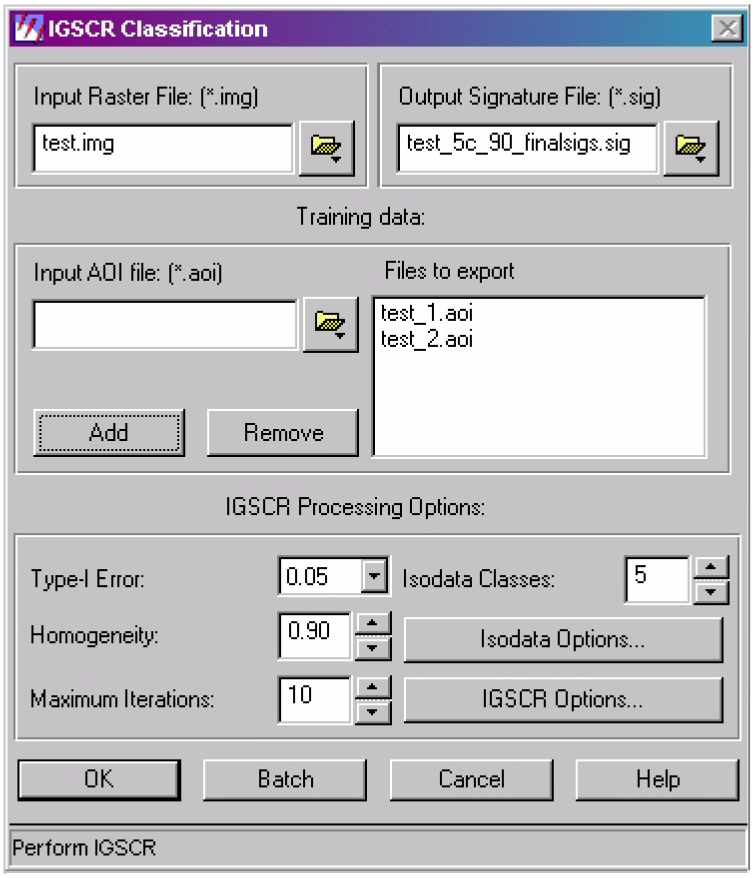


Initialize Means Along:
● Principal Axis
○ Diagonal Axis

Scaling Range:
○ Automatic
● Std. Deviations: 1

Processing Options:
Maximum Iterations: 100
Convergence: 0.975

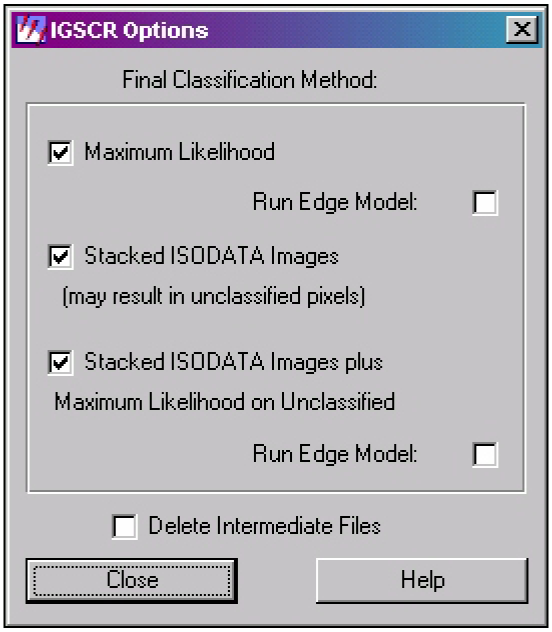
Fig. 19. ISODATA Options window, which is opened by selecting the "ISODATA Options…" button on the IGSCR Classification interface (Fig. 18). Default values are shown.



Input Raster File: (*.img) test.img
Output Signature File: (*.sig) test_5c_90_finalsigs.sig
Input AOI file: (*.aoi)
Files to export:
test_1.aoi
test_2.aoi

Type-I Error: 0.05
Isodata Classes: 5
Homogeneity: 0.90
Maximum Iterations: 10

Fig. 18. IGSCR Classification interface, showing correct format for input image and AOI file names and default output signature file name and parameter values.



Final Classification Method:
☑ Maximum Likelihood
Run Edge Model: ☐
☑ Stacked ISODATA Images
(may result in unclassified pixels)
☑ Stacked ISODATA Images plus Maximum Likelihood on Unclassified
Run Edge Model: ☐
☐ Delete Intermediate Files

Fig. 20. IGSCR Options window, which is opened by selecting the "IGSCR Options…" button on the IGSCR Classification interface (Fig. 18). Default values are shown.

4.1.1.2 Main loop

After the user selects input files and parameter values, then clicks on OK, the EML closes, and the input parameters are sent to the executable. A flow chart outlining the important functions and outputs of the executable is given in Fig. 16. Prior to calling any functions, the IGSCR program creates a structured array large enough to hold data for each of the spectral classes created by the ISODATA function. The array contains variables to hold the information needed to determine the purity of a spectral class: pixel count variables for nonforest, forest, unknown, and a total, p_hat, z, the majority informational class value (nonforest = 1, forest = 2), and the spectral class status (impure = 0, pure nonforest = 1, pure forest = 2, pure class 3 = 3).

Next, the pertinent input parameters are set up and input into the ERDAS classifyisodata job (ISODATA box, Fig. 16). This job produces two outputs, a classified image and a signature file. The classified image, along with one of the user-specified AOI files, is then sent as arguments to the ERDAS pixeltotable job (Read pixels to ASCII). The pixeltotable job is repeated as many times as there are input AOI files (note repeat symbol, ⟳, Fig. 16). Each job produces an ASCII file that contains the coordinates and spectral class values for each of the pixels within the AOIs in the input file.

The heading "Extract data from ACSII" (Fig. 16) refers to a function written in C to extract the spectral class values for each of the pixels that were within the AOIs. Recall that the AOIs each contain data for only one informational class and therefore the same follows for the ASCII files. When the extractData function is called, the name of the ASCII file to be processed and the informational class associated with it are passed into the function. The ASCII file is opened and the pixel spectral classes are read from the file. As each spectral class value is read, the pixel count for the correct informational class within that spectral class and the total pixel count for the spectral class are incremented. This process is repeated for each of the ASCII files.

The array information produced by the extractData function is used to determine spectral class purity (Class purity test, Fig. 16). The function written to do this takes the user-defined critical value and homogeneity threshold as parameters. The purityTest function compares the number of pixels in each of the informational classes for a given spectral class to determine which one is the majority class. Using the number of pixels in the majority informational class, the estimated proportion of the entire spectral class that is in the majority informational class, *p_hat*, is calculated. This value is used in the test of proportion to determine whether or not the spectral class is homogeneous. The status of the spectral class (impure = 0, pure nonforest = 1, pure forest = 2) is assigned based on the outcome of the test of proportion. The array data are printed out to an Excel file after each spectral class has been evaluated.

In the functions that follow, the array information is used again, this time to manipulate the input multispectral Landsat TM and ISODATA images. Prior to performing the Mask and Recode functions (Fig. 16), the status of each of the spectral classes in the array is used to create a lookup table for each of these two functions. Each lookup table contains information indicating the values that will be assigned to each of the pixels in the input image when the function is performed. For the Mask function, the lookup table contains a zero for each of the pure spectral classes and a non-zero value for each of the impure classes. The lookup table for the Recode function is just the opposite, it contains a zero for each of the impure classes and the informational class status (nonforest = 1, forest = 2) for the pure spectral classes.

The Mask function is performed in preparation for the next iteration of the algorithm. It takes as inputs the lookup table and the multispectral Landsat TM image that was used in the ISODATA classification. It produces an image that contains multispectral data in areas that were not pure spectral classes and zeros in areas that were pure. The data in the masked output image, which contains only pixels that fell into impure spectral classes, is then reclustered in the next iteration ISODATA classification. The IGSCR algorithm repeats the main loop, from ISODATA through Mask (Fig. 16), until one of the stopping criteria has been met.

4.1.1.3 Signature file manipulation

In addition to performing image manipulations and data extractions, the IGSCR main loop performs signature file manipulations. First, it edits the signature names (Edit signature names, Fig. 16) so that they contain meaningful information. The results of the pixel purity test (stored in the array of spectral class data) are used to rename each of the spectral signatures as follows:

iteration – value . informational class,

where **iteration** is the iteration that the signature was created via the unsupervised classification algorithm (ISODATA clustering), **value** is the value that the spectral class was assigned in the output ISODATA image for that iteration, and **informational class** is the value of the informational class for which the signature was determined to be homogeneous.

After renaming the signatures, the program appends the most recently created set of renamed signatures to a file containing all of the signatures previously produced during the IGSCR process ("Append .sig file", Fig. 16). All of the renamed signatures created in each iteration are stored in one file. Finally, the combined signature file is edited to remove all of the impure signatures (those with a status of zero). This creates a final signature file that contains only the pure, renamed signatures. Note that this step occurs after the IGSCR main loop has stopped iterating. The arrow pointing to the "Remove impure signatures" box (Fig. 16) has double slash marks through it to indicate that it occurs after iteration has ended.

4.1.1.4 Image product options

If the user chose to perform a maximum likelihood classification using the signature set produced in the IGSCR main loop, then the green part of the IGSCR flow chart in Fig. 16 is executed. A maximum likelihood classification is run, using the edited final signature set, which contains only the pure signatures generated by each iteration of the IGSCR algorithm. A lookup table is generated based on the informational classes stored in the names of the signatures used to create the maximum likelihood image. It assigns an informational class (1 = nonforest, 2 = forest) to each of the spectral classes in the image. The image is then recoded to forest and nonforest using the lookup table that was created.

All of the functions that occur (and products that are created) if the user opted to produce the stacked ISODATA image and/or the stacked ISODATA with maximum likelihood on unclassified pixels image are selected are shown in magenta in Fig. 16. If the user opted to perform only the former option, then the red section of the flow chart (Fig. 16) would be executed following the completion of the magenta portion. If the user chose to perform only the latter option, then the blue segment would be performed. If both options were selected, then both the red and blue branches of the flow chart would be implemented.

The ISODATA image and the recode lookup table are the inputs for the Recode function (magenta "Recode" box, Fig. 16). In this function, the spectral class of each of the pixels in the ISODATA image (Fig. 8) is used to create a lookup table to recode the pixel to the value of the informational class for which it was determined to be pure. The result is an image that contains informational class values only for the pixels from the pure spectral classes found in the iteration that was just completed (Fig. 10). The rest of the image contains zeros where the pixels were impure.

After the recoded image is created, it is combined with recoded images produced in subsequent iterations to produce an image containing the appropriate informational class values for all of the pixels from all iterations that were in pure spectral classes. The images are combined by image addition, i.e., values for each of the pixels in the images being combined are summed (magenta "Add" box, Fig. 16). This can occur because only one recoded image contains a value for any given pixel; impure pixels are recoded to zero in early iterations, then to an informational class value in a given iteration, then are zero in subsequent iterations since they are masked out of the input ISODATA image because they were found to be pure.

The recoding and image addition processes are performed at the end of each iteration of the IGSCR main loop. After iteration has terminated, the final signature file and the last masked image ("In" .img file) created by the main loop are input into the maximum likelihood algorithm (magenta "Maximum likelihood classification" box, Fig. 16; note that the arrow pointing to this box has double slash marks through it to indicate that it occurs after iteration has ended). The last masked image contains the remaining impure pixels after the IGSCR algorithm has stopped iterating. These pixels are classified using all of the signatures produced in the IGSCR main loop. The resulting classified image contains zeros where there were pure pixels, and classified values with class names that correspond to the signature file names used in the maximum likelihood algorithm.

The output image from the maximum likelihood algorithm is used in two different ways. If the user opted to produce the stacked ISODATA image, all of the pixels with values in the maximum likelihood on unclassified image are recoded to the value 3 (red "Recode" box, Fig. 16). The recoded image, which contains zeros for all pure pixels and 3's for all other pixels, is then added to the stacked ISODATA image (red "Add" box, Fig. 16). This produces the final stacked ISODATA image, which contains informational class values for all pure pixels and 3's for all impure pixels. The impure pixels are recoded to 3's to allow the user to differentiate between unclassified pixels and background pixels.

If the user chose to generate the stacked ISODATA with maximum likelihood on unclassified image, all of the pixels with values in the maximum likelihood on unclassified image are recoded to the appropriate informational class value (blue "Recode" box, Fig. 16). The recoded image, which contains zeros for all pure pixels and informational class values for all other pixels, is then added to the stacked ISODATA image (blue "Add" box, Fig. 16). This produces the final stacked ISODATA with maximum likelihood on unclassified image, which contains informational class values for all pixels.

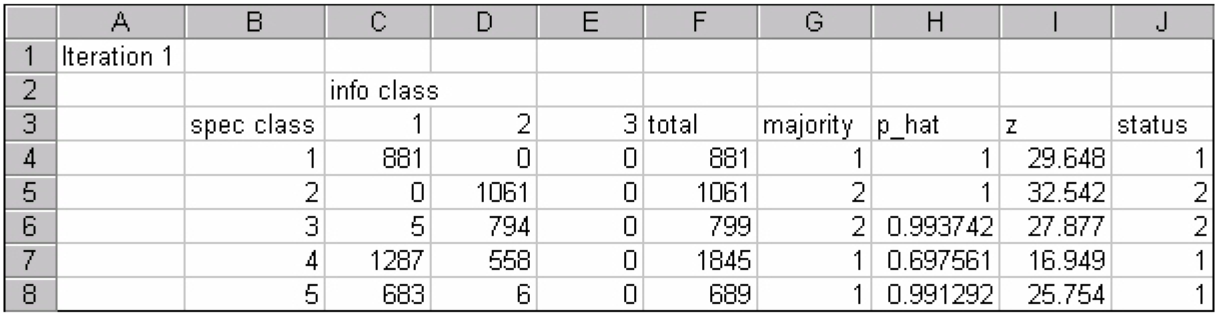
If the user opted to run the edge model on the maximum likelihood image, the recoded maximum likelihood image is run through a model that first determines the distance of each pixel from the forest informational class and then does the same for the nonforest informational class. The model produces two images, each of which contains a 0 if the pixel is within the informational class of interest, a 1 or 2 if the pixel is 1 or 2 pixels away from the class of interest, and a 3 if the pixel is 3 or more pixels away from the informational class of interest.

After the creation of the forest and nonforest distance images, the same model also creates an image containing four classes 1) nonforest, 2) nonforest edge, 3) forest, and 4) forest edge. The function within this model that does this takes as input the original forest/nonforest classification and forest and nonforest distance images. It assigns pixels to the edge category if they were classified as one informational class, but are 1 or 2 pixels away from the other informational class. For example, a pixel that was classified as forest would contain a zero in the forest distance image, but if the nonforest distance image contained a 1 or 2 for that pixel, it would mean that it was 1 or 2 pixels away from nonforest. This pixel would be classified as forest edge in the final model.

If the user opted to run the edge model on the stacked ISODATA with maximum likelihood on unclassified pixels image, the same process is performed as for the edge model on the maximum likelihood image except that the input image is the stacked ISODATA with maximum likelihood on unclassified pixels image.

4.1.2 Program outputs

If the user chooses to delete all intermediate files, the program output is a set of renamed, pure signatures that can be used in a maximum likelihood classifier, along with the image or images that the user selected. Otherwise, the program output consists of all images, the ISODATA iteration 1 signature set, a combined set of all of the signatures that were produced during the IGSCR process with their descriptive names, and a subset of those signatures containing only the pure signatures.



	A	B	C	D	E	F	G	H	I	J
1	Iteration 1									
2			info class							
3		spec class	1	2	3	total	majority	p_hat	z	status
4		1	881	0	0	881	1	1	29.648	1
5		2	0	1061	0	1061	2	1	32.542	2
6		3	5	794	0	799	2	0.993742	27.877	2
7		4	1287	558	0	1845	1	0.697561	16.949	1
8		5	683	6	0	689	1	0.991292	25.754	1

Fig. 21. Example of tab-delineated output file which is created after the iteration 1 purity test.

```
IGSCR parameters

   Input image:      z:/igscr/test_run/test.img
   Output signature set:   z:/igscr/test_run/test_5c_50_finalsigs.sig
   Type-I error:     0.05
   Homogeneity threshold:  0.5
   Maximum iterations:  10

ISODATA parameters

   Number of classes:   5
   Maximum iterations:  100
   Convergence threshold:  0.975
   Initialize means along: principal axis
   Scaling range:    1 std deviation(s)

AOI files used

   AOI file 1:       z:/igscr/test_run/test_1.aoi
   AOI file 2:       z:/igscr/test_run/test_2.aoi

*********************
IGSCR Results

   Stopped after iteration 1 because all of the ISODATA classes were pure
   Total pure classes = 5

The following final classification(s) were performed:

   Maximum Likelihood
                              Count     Percent
            Class 1:          30579     51.3208
            Class 2:          29005     48.6792

   Stacked ISODATA
                              Count     Percent
            Class 1:          30554     51.2789
            Class 2:          29030     48.7211

Output image file(s):

   z:/igscr/test_run/test_5c_50_ml_recode.img
   z:/igscr/test_run/test_5c_50_it1_recode.img
```

Fig. 22. Example of text file output that is created after all of the IGSCR classification products selected by the user have been generated.

In addition to the aforementioned image and signature outputs, an Excel spreadsheet, which presents the spectral and statistical data generated in that iteration, is created for each iteration. An example of the Microsoft Excel spreadsheet created after the iteration 1 purity test is given in Fig. 21. The file includes information pertaining to the spectral classes created as a result of ISODATA clustering: the number of pixels within the AOIs that fell into each spectral class, the total number of pixels in the spectral class, the informational class with the majority of pixels, p_hat- the estimated proportion of pixels in the entire image of a given spectral class that would fall into the majority informational class, z- the test statistic calculated using p_hat, and the class status (0 = impure, 1 = pure nonforest, 2 = pure forest, 3 = pure class 3 and 4 = unclassified).

The program also produces a text file that lists the input parameters used and summarizes the results of the classification (Fig. 22). The program produces a message to let the user know the number of IGSCR iterations performed, why it stopped iterating and the total number of pure classes found. It also accesses the image statistics for each output image and calculates the proportion of pixels that were classified into each of the informational classes. At the end of the file, it lists the names and paths of all if the image products that were created.

4.1.3 Beyond necessity

4.1.3.1 Skip iteration 1 ISODATA

Slight modifications to the program were made in order to facilitate a factorial analysis, which requires the repeated classification of images with a set number of ISODATA classes while other parameters are varied. In this situation, the first iteration of IGSCR would produce an ISODATA image with the given number of classes each time a different set of parameter values are initiated. To improve the efficiency of the program, modifications were made so that the program checks for the presence of the appropriately named ISODATA iteration 1 image and signature file in the input folder. If present, the first ISODATA classification is skipped, and the files are used as the starting point for the remainder of the algorithm. The program will also skip the iteration 1 pixel to ASCII conversion if the ASCII files are present.

4.1.3.2 Set image colors

To further improve the program, a function was added which would change the colors of the recoded images to something appropriate for forest/nonforest classifications. When the color tables of a thematic image are accessed, the values of the red, green and blue color tables for each class can be read or written. In the setColors function, the each of the three color tables associated with a given recoded image are read from the image to temporary color table objects using the ERDAS Imagine toolkit function eimg_ColorTableRead. The colors of the temporary color table objects were set to green for forest, tan for nonforest, and purple for unclassified, then were written out to the image file using the function eimg_ColorTableWrite. Appropriate colors were also selected and set for each of the distance images and the four-class edge image (Fig. 23 -Fig. 26).

4.1.3.3 Add help files

I also wanted to create help files that the user could access through the Graphical User Interface. Fortunately, Imagine 8.5 has a simple function that allows the user to connect to a text or HTML (HyperText Markup Language; often used to create web pages) file. The function that needs to be called is the "helplink" function followed by the path of the html file, for example:

```
helplink ( getenv("IMAGINE_HOME") + "/help/html/igscr/IGSCR_Helpfile.htm");
```

Each of the three windows in the GUI (Fig. 18,Fig. 19, and Fig. 20) has a different help file associated with it, which contains information specific to that window (Appendices K, L, and M, respectively). The help files in Imagine 8.5 are also HTML files, and are organized in such a manner that links to preexisting help files supplied with the Imagine software can be easily accessed from within the IGSCR help files. This improved the integration of information about other classification methods into the IGSCR help files, since the user can simply click on a reference to see the Imagine help files associated with that topic.

Since there were a number of files associated with the program (the EML script, a modified version of an ERDAS EML script, the executable file, the help files, and a spatial model file), I created a document that indicates where each of the files must be placed on the user's computer (read_me.doc; Appendix L). The document also describes how to run the program, by means of an example, and gives some basic guidelines for the user to follow when using the program.

Fig. 23. Subset of a multispectral Landsat TM image, showing band 4 (red), band 3 (green) and band 2 (blue).

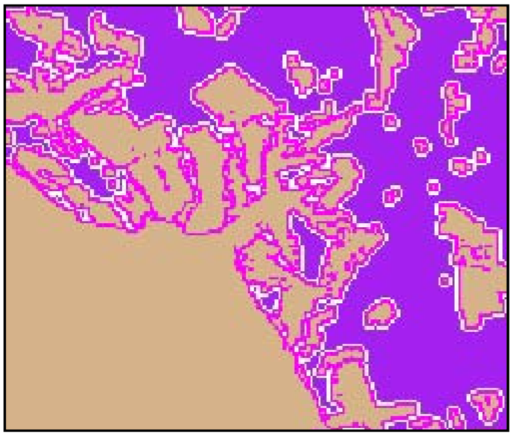


Fig. 24. Landsat TM image in Fig. 23 reclassified to show distance from nonforest. Nonforest (NF) is shown in tan, 1 pixel away from NF in magenta, 2 pixels from NF in white, and 3 or more pixels away from NF in purple.

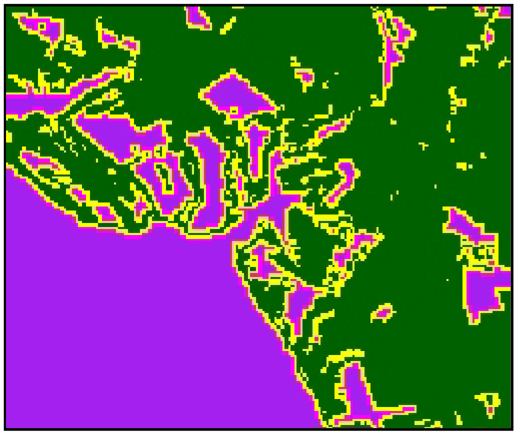


Fig. 25. Landsat TM image in Fig. 23 reclassified to show distance from forest. Forest is shown in dark green, 1 pixel away from forest in yellow, 2 pixels from forest in magenta, and 3 or more pixels from forest in purple.

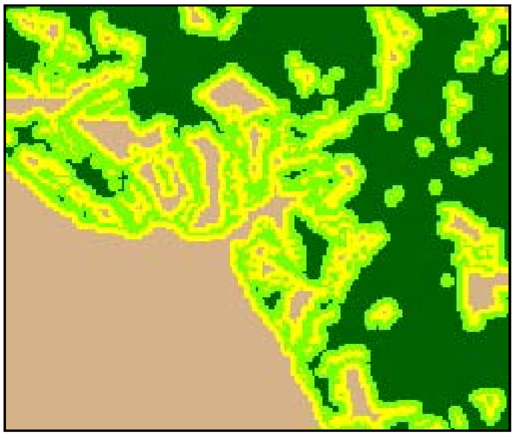


Fig. 26. Landsat TM image in Fig. 23 reclassified to show edges using the nonforest and forest distance images. Nonforest is shown in tan, nonforest edge in yellow, forest is shown in dark green, and forest edge in chartreuse.

4.1.4 Keys to ERDAS programming

Two of the aspects of ERDAS programming which were pivotal to the development of the IGSCR program were also complex enough to warrant describing (in the hope of saving future ERDAS programmers some time).

4.1.4.1 Jobs vs. applications

There are two ways to connect a program written with the ERDAS Developer's Toolkit to an EML script, as an application or as a job. If the C code initializes the EML toolkit package, then the program is called an application. In an application, the EML script is opened from within the C program and parts of the GUI (frameparts) can trigger function calls, known as callback functions, within C programs. When the C program is set up as a job, parameters are passed from the EML script to the C code in command line form, after which there is no further interaction between the EML script and the executable.

The key distinction between running an application and a job is the manner in which preexisting ERDAS Imagine jobs are called from within each of these types of programs. Jobs are ERDAS Imagine functions such as pixeltotable, which is the program that is typically called through the Convert Pixels to ASCII user interface. To call a job from within an EML script, the job command is followed by the name of the executable file, then the parameters to be used in the algorithm, for example

```
job pixeltotable -gui 0 -in "c:/temp/isodata.img" -out "c:/temp/isodata.asc"
   -criteriatype "aoi" -criteriafile "c:/temp/test.aoi" -coord "Map" -s 1 1;
```

While working on early versions of the IGSCR program, this method was used to call the jobs needed to run the algorithm. However, a problem arose due to the fact that in EML it is not possible to run a large number of jobs sequentially; they start as soon as they are called and thus run concurrently. Sequential execution of jobs is a crucial requirement of the IGSCR algorithm, since the output from each job is used as input for the next. Therefore, it was necessary to find a method of calling jobs that would halt the IGSCR process while each job was in progress.

In response to my ERDAS Toolkit listserv request for such a method, Edward Trubia sent a useful code snippet with instructions on how to call a job from within a C program and pause the execution of the program until the job is finished. To do this, after the necessary variables have been declared and the arguments assigned, the following code is needed:

```
job_handle = esmg_ProcessStart(ESMG_PROCESS_JOB, argc, argv, &err);
while (esmg_IsProcessRunning (job_handle, &err)) {
   sleep (500);}
```

This code was the key that made it possible to fully automate the IGSCR algorithm. It checks to see whether the process is running, and if it is, then the "sleep" command is used to suspend the execution of the program for 500 milliseconds. This is the reason that the IGSCR program is set up as a job. After the user selects input files and parameter values, then clicks on OK, the EML closes, input parameters are sent to the executable and the C code runs each process sequentially. One other advantage of creating a job is that the job status and progress can be easily displayed using pre-existing ERDAS Toolkit functions.

4.1.4.2 Spatial Modeler Language

I also learned how to run a model created in spatial modeler using the ERDAS C Toolkit. Prior to the automation of this program, there were three models which were used to create the forest/nonforest edge images. The first model took as input a forest/nonforest recoded image and used it to create an image that contained distances to forest pixels where 0 indicated forest, 1 or 2 meant that the pixel was 1 or 2 pixels from forest, and 3 signified that the pixel was 3 or more pixels from forest. The second model was similar to the first, except that it created an image containing distances to nonforest. The third model combined the forest and nonforest distance images to create an image that showed interior forest and nonforest (forest or nonforest 3 or more pixels away from the other information class) as well as forest edge (forest that was 1 or 2 pixels from nonforest) and nonforest edge (nonforest that was 1 or 2 pixels from forest).

In the ERDAS Spatial Modeler Model Maker interface, under the "Process" menu there is a "Generate Script…" option. I used this option to create a model (.mdl) file for each of the 3 models. Then, I compared the models that I had created with other models that I had already used in my program which were created at ERDAS, such as mask.pmdl. I mimicked the ERDAS model files, modifying them to include all of the functions that were present in the 3 scripts generated from the edge models. Then, I made a function to call the resulting model (edge.mdl, Appendix M) in a manner similar to the way I in which called the other ERDAS models.

4.2 Classification results

After the completion of the IGSCR program, classification of the images for the remainder of this project was greatly facilitated. The results of the image classifications via IGSCR have been divided into two parts: information pertaining to the classification process itself and the accuracy of the images produced by the process. The IGSCR process information is pertinent because this means of classification is still in the developmental stages, and it can also indicate characteristics of the images that accuracy assessment might not reveal. The image accuracy affects the forest area estimates derived from the images and can be used to test the significance of the results.

4.2.1 IGSCR process information

The IGSCR classification produces multiple output files: classified images, signature files, a text document and Excel files for each iteration. The following is a summary of the outputs with comments.

4.2.1.1 Iterations and pure classes

For each image, the number of IGSCR iterations in which pure classes were generated was less than 4 and the total number of pure classes resulting ranged from 48 to 101 (Table 7).

Table 7. Number of pure classes generated in each iteration of IGSCR for each image.

# pure classes	Iteration				
Image	1	2	3	4	All
15m	62	6	2	0	70
16m	72	22	7	0	101
17m	67	6	0	-	73
MN	43	5	0	-	48

These numbers are far lower than in previous IGSCR research (Wayman *et al.*, 2001). This is due to three noteworthy changes in the IGSCR algorithm and parameters. First, the number of classes in each ISODATA iteration remains the same whereas previously it was based on the amount of training data remaining after each iteration. Second, a new test of proportion was used to determine which classes are pure, rather than a simple homogeneity threshold and minimum pixel requirement. Third, in this study the number of ISODATA classes in each iteration was 100, while in the previous study, it ranged from 200 to 1200. Although the overall accuracies in this study are generally higher than those reported by Wayman *et al.* (2001), it is not clear whether or not fewer classes is better, since so many changes to the algorithm have been made. Further research into the effects of varying the parameter values used to perform IGSCR classifications (with the new test of proportion and constant number of classes per iteration incorporated) is required.

4.2.1.2 Training data used

After each iteration of IGSCR, the number of pixels in the image is reduced because the pure spectral classes are removed. This means that the training data areas encompass less information in each iteration, since some of the pixels in the training areas contain zeros. Table 8 lists the total number of training pixels available and those that were unused when iteration ended. Table 9 shows the reductions in the training data used in each iteration.

As can be seen in Table 8, the total amount of training data available for each image was not the same even though the same number of random points was used. Also, the amounts of unused training data were not the same nor were they proportional to the amount available initially. The image with the most training data initially (16m; Table 8) did use the largest proportion of its total available data (Table 9). However, the relationship between the training data that are available and the proportion used is not linear (Fig. 27). The spectral properties of the data undoubtedly play an important role in the determination of whether or not available training data are used to develop pure classes. Additionally, the same training data might be used very differently if the number of classes used in IGSCR were changed.

Table 8. Training data available for use in IGSCR and unused when IGSCR iteration ended.

	Training data					
# of pixels	Available			Unused		
Image	NF	For	Total	NF	For	Total
15m	42990	65295	108285	5698	9918	15616
16m	57184	151631	208815	2877	2493	5370
17m	56350	80344	136694	1546	6287	7833
MN	62518	58034	120552	3951	3492	7443

Table 9. Proportion of initial training data used in each iteration and over all iterations.

	Training data used (%)											
	Iteration 1			Iteration 2			Iteration 3			All Iterations		
Image	NF	For	Total	NF	For	Total	NF	For	Total	NF	For	Total
15m	84.73	81.14	82.57	1.08	3.21	2.36	0.94	0.46	0.65	86.75	84.81	85.58
16m	87.59	96.21	93.85	6.70	1.42	2.87	0.68	0.72	0.71	94.97	98.36	97.43
17m	96.99	89.58	92.64	0.27	2.59	1.63	0.00	0.00	0.00	97.26	92.17	94.27
MN	92.19	93.97	93.05	1.49	0.01	0.78	0.00	0.00	0.00	93.68	93.98	93.83

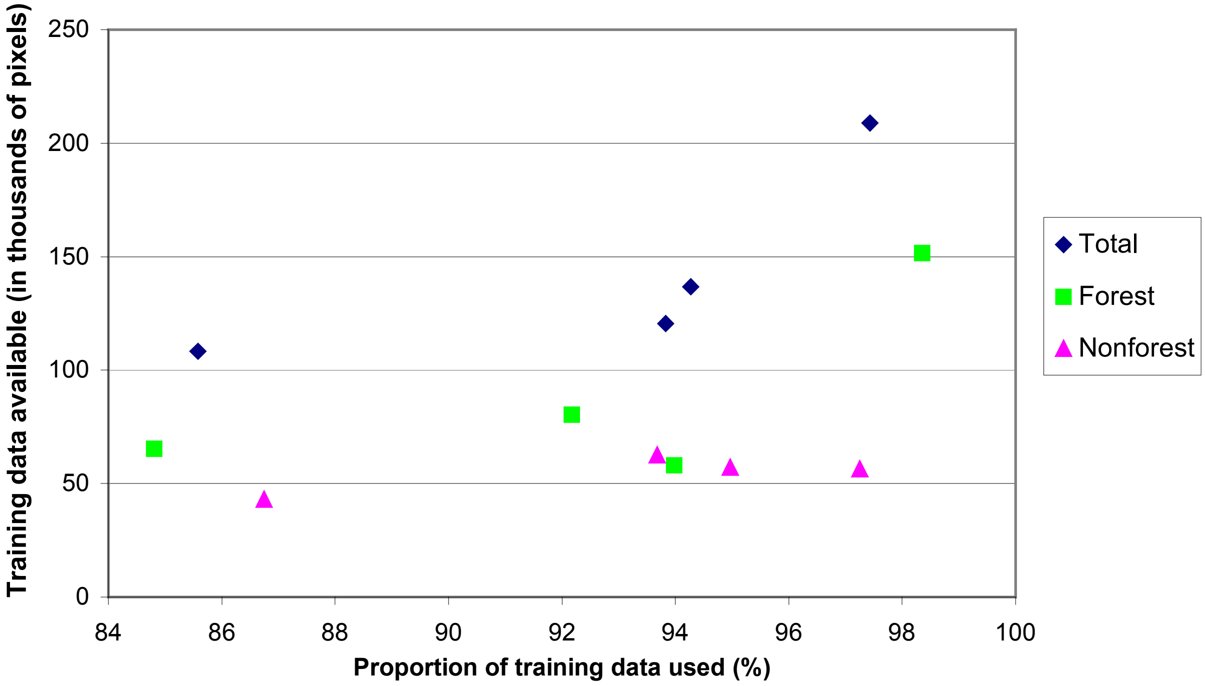


Fig. 27. Relationship between training data available and that which was used in IGSCR.

Another topic related to training data which must be considered is that of spatial autocorrelation between the pixels. Campbell (1981) found that homogeneous land cover categories recorded by the Landsat Multispectral Scanner (MSS; similar to Landsat TM, although of lower spatial, spectral and radiometric resolution; Farr, 1999b) exhibit positive spatial autocorrelation. He states that in order to estimate the characteristics of information classes using a maximum likelihood classifier, the samples of pixels taken in each class must be independent or the estimation will be biased. The training samples used in this project were gathered in clusters of contiguous pixels within a land cover. However, not all of the samples were homogeneous, since IGSCR does not require that the pixels be spectrally similar, just within one land cover.

The conclusions drawn by Campbell (1981) regarding training for a maximum likelihood classifier cannot be directly applied to this study because IGSCR is a hybrid classification. The training fields input into in IGSCR are used to identify the information class of spectral clusters generated via ISODATA clustering, and it is those spectral classes that are used in the maximum likelihood classification. If the only input into a maximum likelihood classification is a set of contiguous, spatially correlated pixels, the estimates of the spectral class means, variances and covariances derived from the training fields may inaccurately represent the characteristics of the land cover categories (Campbell, 1981). Since the spectral classes input into the maximum likelihood portion of IGSCR are derived from spatially independent spectral clusters derived from an ISODATA classification, the contiguity of the training pixels used may not result in significant errors. However, further examination of the relevance of problems arising from spatial autocorrelation should be considered.

The process of training for the IGSCR algorithm seems simple because the areas required must only be within one information class. However, the characteristics that constitute an ideal training data set are not obvious. The relationship between the training data and the number of spectral classes selected is complex. In addition, the decision to use or exclude mixed pixels in training may greatly alter the resulting classification. Finally, the issue of spatial autocorrelation of training data must be considered. It is beyond the scope of this project to attempt to elucidate the qualities of ideal training data for IGSCR, yet the topic is interesting and warrants further investigation.

4.2.1.3 Training data proportions

The training data proportions are derived from the number of pixels in the training areas collected by the image analysts. Despite the fact that training data were collected at the same number of random points in each image, the proportions of forest and nonforest within the samples were not equal across the four images, nor was the proportion of the total image area represented by the data collected (Table 10).

Table 10. For each image, the training data areas, image area and proportion of training data in each information class and in the entire image.

	Area (hectares)				Proportions (%)		
	Training data			Image Total	Training data		
Image name	Nonforest	Forest	Total		Nonforest	Forest	Entire Image
15m	3869	5877	9746	3095351	39.7	60.3	0.31
16m	5147	13647	18793	3513834	27.4	72.6	0.53
17m	5072	7231	12302	2679556	41.2	58.8	0.46
MN	5627	5223	10850	3209983	51.9	48.1	0.34

4.2.1.4 Classified image proportions

The proportions of forest, nonforest and unclassified pixels are automatically calculated for the classified images resulting from IGSCR (Table 11). It is interesting to note that the proportions of forest and nonforest training data collected in the 15m and 16m images (Table 10) are very close to the map marginal proportions in the maximum likelihood image (Table 11) but the proportions are quite different for the 17m and Minnesota images.

IGSCR also produces edge images which are derived from the forest/nonforest images (Table 12). In the edge images, all pixels within the forest category that are 1 or 2 pixels away from a nonforest pixel are labeled forest edge and vice versa for nonforest edge.

The proportion of pixels in the edge categories can be used as an index of fragmentation within the image. The more edge pixels, the more interface between the information classes, therefore the more fragmentation. Table 12 shows that the 15m classification contained the most edge pixels of the four images (61.57%) and 17m had the least (only 37.54%). The fragmented nature of the land cover in 15m was visible in the image prior to classification, as was the contiguity in land cover in 17m (see section 3.3.1). The relationship between fragmentation of the land cover in an image and the accuracy of the resulting classification will be discussed in section 4.2.2.1.

Table 11. Proportions of forest, nonforest and unclassified pixels in the maximum likelihood and stacked ISODATA classifications.

Proportion (%)	Maximum Likelihood		Stacked ISODATA		
Image	Nonforest	Forest	Nonforest	Forest	Unclassified
15m	38.8	61.2	24.0	41.9	34.1
16m	26.4	73.6	19.0	69.2	11.9
17m	29.8	70.3	22.2	55.1	22.8
MN	70.4	29.7	40.4	17.9	41.7

Table 12. Proportions of forest, nonforest, and edge categories in the image created by running the maximum likelihood through an edge model.

Proportion (%)	Edge Model for Maximum Likelihood				
Image	Nonforest	NF edge	Forest	Forest edge	Total edge
15m	13.5	25.3	24.9	36.2	61.6
16m	7.8	18.5	48.5	25.2	43.7
17m	13.0	16.8	49.5	20.8	37.5
MN	38.7	31.7	8.7	21.0	52.7

4.2.2 Accuracy measures

The accuracy of a classified image can be assessed using a variety of measures. The error matrix (Table 1) gives an indication of the overall accuracy of the image based on the number of validation points that were correctly classified. The estimate of Kappa indicates how much better the classification was than what would have resulted from chance.

4.2.2.1 Overall accuracy

Four Landsat ETM+ images were classified via IGSCR, each with 200 training fields initiated at random points within the image. The accuracies were assessed using four variations on a single validation data set for each image (see Section 3.7.1 for a description of the validation sets used). Each of these accuracy sets was used to validate the unfiltered image and five filtered derivatives of the image (see Section 3.6.4 for filters used). The resulting overall accuracies ranged from 81.9 to 95.4% (Fig. 28, Fig. 29, Fig. 30, and Fig. 31).

The differences between the accuracy sets for all images and filters are clearly visible and consistent across all images and filters (Fig. 28, Fig. 29, Fig. 30, and Fig. 31). The center land use set resulted in the lowest accuracies, the set in which at least 50% of the plot was required to agree with the center land use performed slightly better, the set requiring 75% of the plot to agree with the center land use yielded still better results, and the homogeneous set outperformed all of the others.

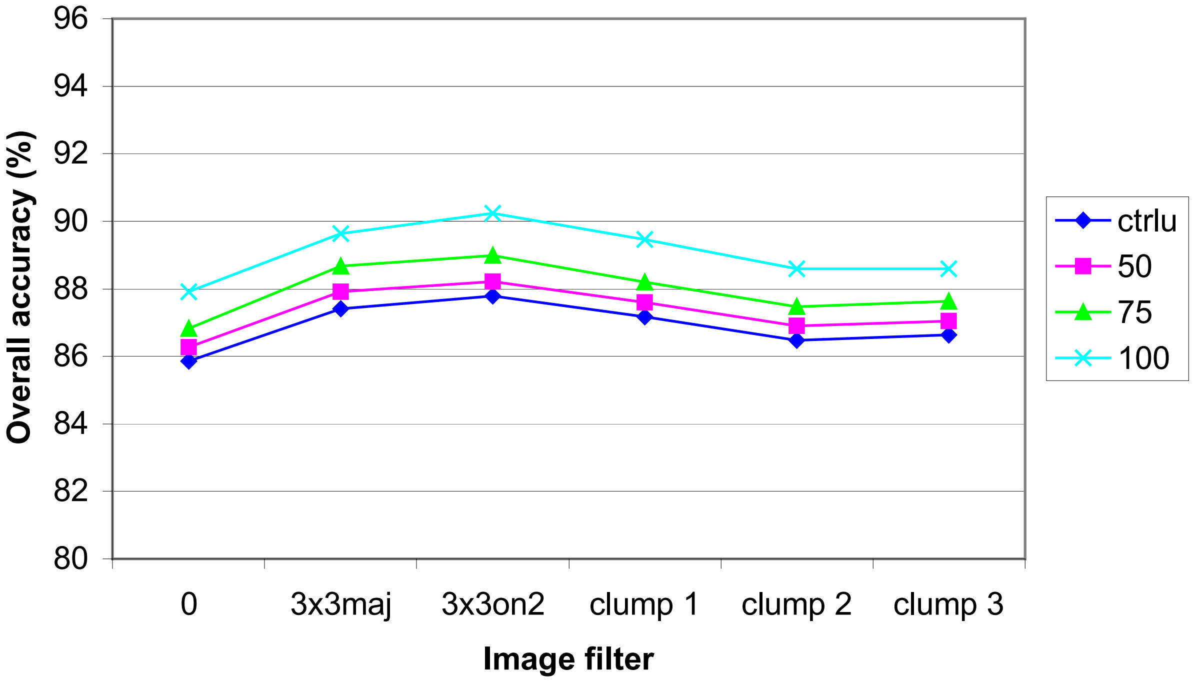


Fig. 28. Overall accuracies for different filters and validation data sets for the Minnesota image.

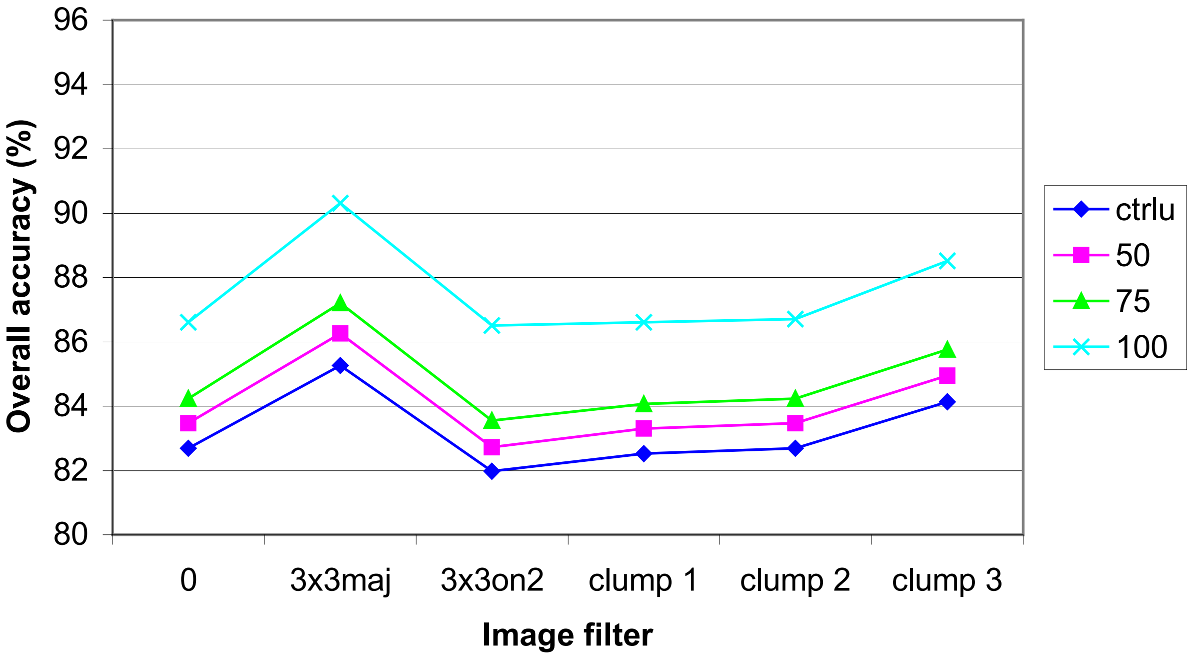


Fig. 29. Overall accuracies for different filters and validation data sets for the 15m image.

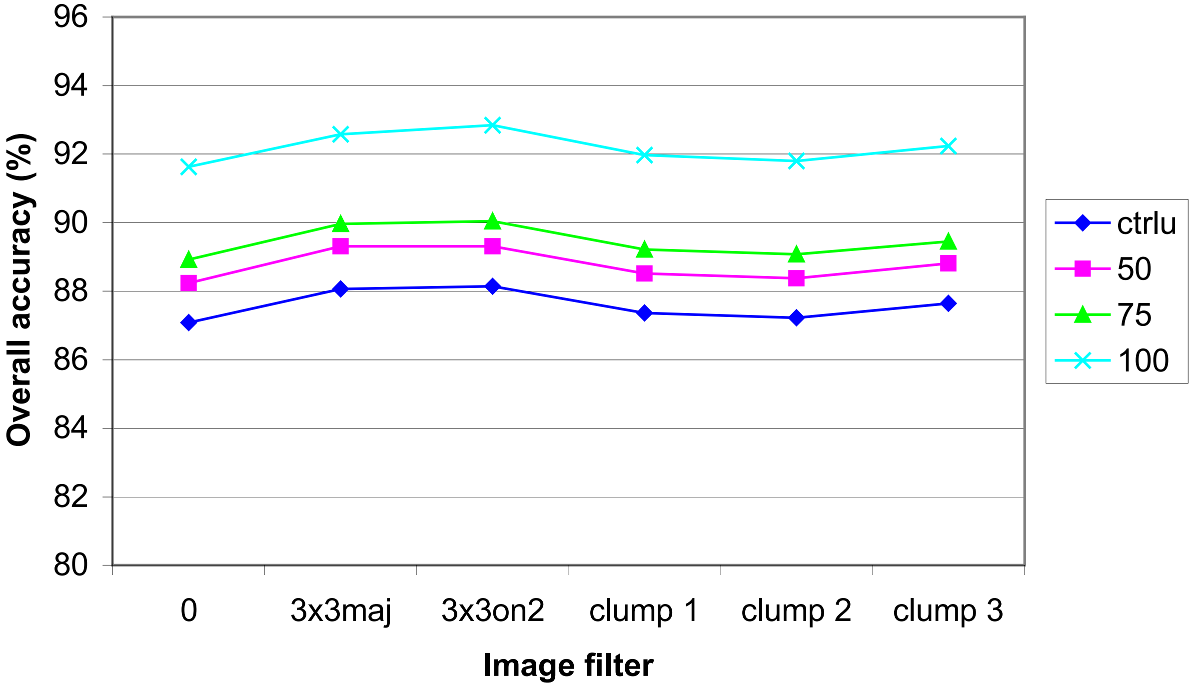


Fig. 30. Overall accuracies for different filters and validation data sets for the 16m image.

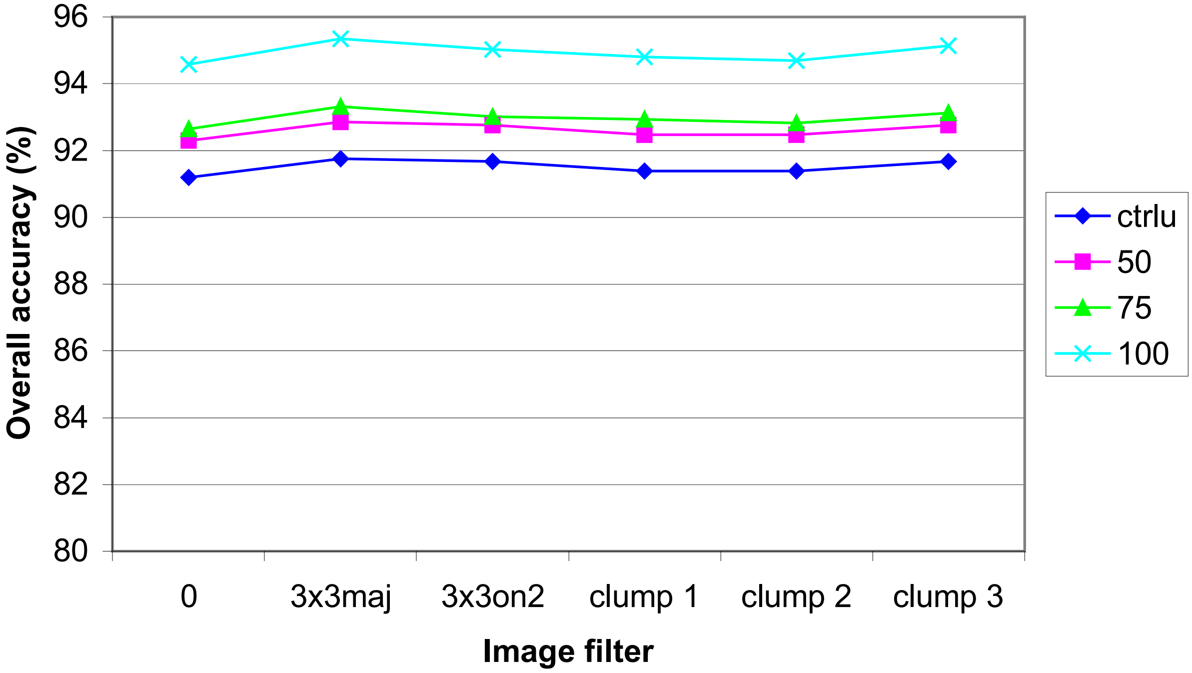


Fig. 31. Overall accuracies for different filters and validation data sets for the 17m image.

The differences between filters were less noticeable and not entirely consistent across all images (Fig. 28, Fig. 29, Fig. 30, and Fig. 31). The results for the 3x3 majority filter on the 15m image for all accuracy sets deviated greatly from the almost imperceptible differences shown in the other images and among all of the other filters. The 15m image did contain the most fragmented land cover of all of the images. Fig. 32 shows the overall accuracies of the images plotted against the amount of edge in the images. It shows that there is a linear decrease in the overall accuracy of an unfiltered image with an increase in the amount of edge (i.e., the amount of fragmentation) in the image ($R^2 = 0.932$). This indicates that postprocessing of classifications may be more important in highly fragmented areas. The relationship between filtering of images (particularly those with highly fragmented land cover) and classification accuracy requires further investigation.

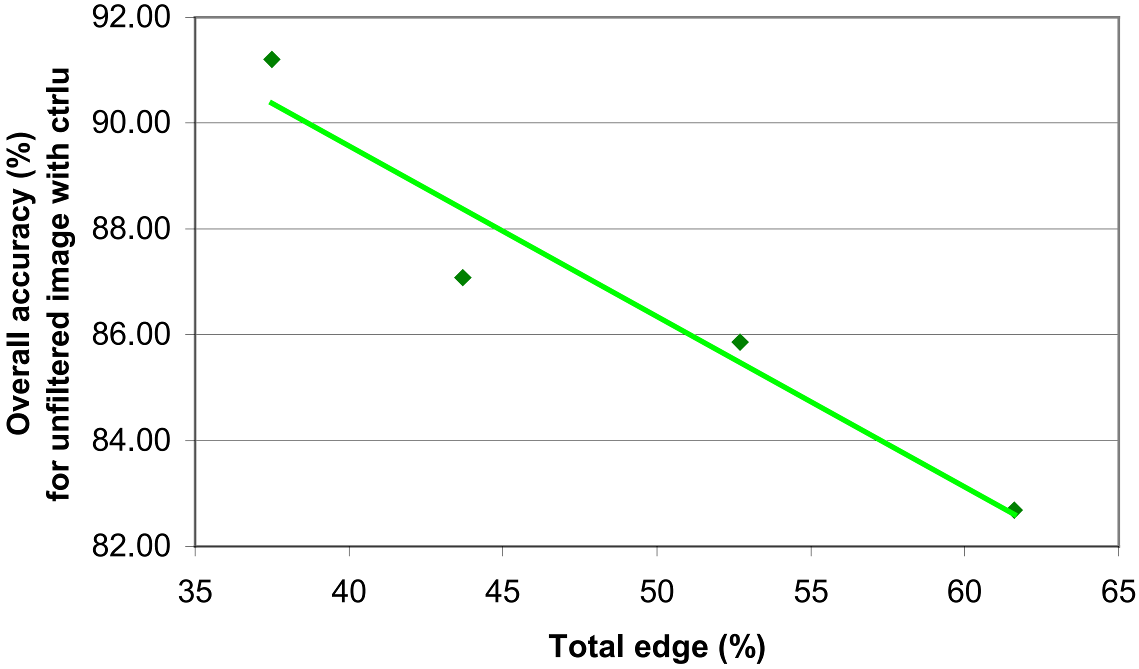


Fig. 32. Relationship between overall accuracy (assessed using the center land use) for 15m, 16m, 17m and MN unfiltered images and total perecent edge in each image.

4.2.2.2 Test statistics

The Kappa statistics for the classifications had a range of 0.6072 to 0.8934, which means that the classifications were all 60-90% better than what would have resulted from chance assignment of each pixel to an information class. The Z-scores for the classifications ranged from 22.01 to 55.73, showing that all error matricies were significant.

The Kappa variances ranged from 0.000257 to 0.000808. The Kappa variances were an order of magnitude lower than those reported by Wayman *et al.* (2001). This can be attributed in part to the fact that the range of overall accuracies was lower in that study than the results presented here. However, the discrepancy is mainly because the number of validation points used in the Wayman *et al*. (2001) study was less than 250 per image, whereas in this study the number of points available for accuracy assessment in each scene ranged from 904 to 1416 (note: the large difference in number of validation points used is partly because the image areas used in this study were much larger than those in the Wayman *et al.* (2001) study).

The pairwise comparisons between the different filters and the unfiltered image were less than what was required to show significance at the 95% level except for that of the 15m 3x3 majority filtered image assessed with the homogeneous validation set (Table 13). There was no filter that consistently improved the accuracy of the classifications assessed using any of the accuracy sets.

The pairwise comparisons among the accuracy sets show that there were significant differences between the homogeneous and the center land use accuracy sets at the 95% level for all of the Virginia images (Table 14). This indicates that the use of a homogeneous validation set will significantly improve the overall accuracy of a classification, which is a somewhat intuitive result, since one would expect that a validation set containing only points in homogeneous areas would result in higher image accuracies than one containing points along edges. The drawback of using only 100% homogeneous plots is that the accuracy of the map in transitional areas is not being assessed. Although the validation sets requiring 50% and 75% homogeneity were not significantly different from the center land use set, the overall accuracies were consistently higher than those for the center land use set. This indicates that some requirement of plot homogeneity is recommended for assessing the accuracy of classifications using FIA plot data if it does not bias the sample.

Table 13. Pairwise comparisons of each filter with the unfiltered image for each image and accuracy set, with significant differences (at the 95% level) in red.

		Filter					
Image	Accuracy set	0	3x3maj	3x3on2	clump 1	clump 2	clump 3
15m	ctrlu	0.0000	1.5561	0.2134	0.0264	0.0325	0.8579
	50%	0.0000	1.7432	0.2146	0.0255	0.0342	0.8843
	75%	0.0000	1.8964	0.1885	0.0309	0.0336	0.9252
	100%	0.0000	2.4544	0.1181	0.0644	0.0909	1.1876
16m	ctrlu	0.0000	0.7367	0.9631	0.2518	0.1257	0.4391
	50%	0.0000	0.8414	1.0056	0.2625	0.1310	0.4598
	75%	0.0000	0.8246	1.0434	0.2721	0.1358	0.4233
	100%	0.0000	0.8206	1.1801	0.3276	0.1632	0.5241
17m	ctrlu	0.0000	0.4443	0.4525	0.1741	0.1637	0.3709
	50%	0.0000	0.4774	0.4790	0.1845	0.1742	0.3982
	75%	0.0000	0.5795	0.3926	0.2711	0.1802	0.4139
	100%	0.0000	0.7334	0.4497	0.2187	0.1089	0.5205
MN	ctrlu	0.0000	1.0669	1.0438	0.7743	0.3704	0.5383
	50%	0.0000	1.1383	1.0752	0.7938	0.3790	0.5483
	75%	0.0000	1.3280	1.2605	0.8152	0.3891	0.5649
	100%	0.0000	1.2341	1.3702	0.9347	0.4058	0.4380

Table 14. Pairwise comparisons of each accuracy set with the center land use set for each image and filtering method, with significant differences (at the 95% level) in red.

		Filter					
Image	Accuracy set	0	3x3maj	3x3on2	clump 1	clump 2	clump 3
15m	ctrlu	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	50%	0.4272	0.6261	0.4332	0.4305	0.4304	0.4602
	75%	0.9419	1.3176	0.9820	0.9426	0.9464	1.0274
	100%	2.5360	3.6312	2.9189	2.6458	2.6082	2.9618
16m	ctrlu	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	50%	0.8652	0.9821	0.9245	0.8803	0.8727	0.8930
	75%	1.4980	1.6102	1.6130	1.5273	1.5126	1.4957
	100%	3.7679	3.9425	4.1247	3.8813	3.8241	3.9099
17m	ctrlu	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	50%	0.8753	0.9130	0.9110	0.8889	0.8883	0.9065
	75%	1.1717	1.3195	1.1275	1.2761	1.1936	1.2244
	100%	3.0195	3.3550	3.0645	3.0836	2.9772	3.2044
MN	ctrlu	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	50%	0.2722	0.3497	0.3031	0.2933	0.2819	0.2854
	75%	0.4373	0.7308	0.6736	0.4959	0.4648	0.4795
	100%	0.1374	0.4453	0.6047	0.4009	0.2199	0.0980

It is not clear from the Z-scores alone whether or not the reduction in the number of points available when using only those that are homogeneous will offset the improved accuracy when calculating the precision of the forest area estimates. This will be discussed in section 4.4.

4.3 Corrected forest proportion and variance

The classified proportion of forest for each image varied by as much as 4% depending upon the filter used (Table 15). The corrected percent forest estimates have only slightly less variation over all filters and accuracy sets (Table 16), and the images with the widest ranges are not the same as those for the uncorrected proportions. Also, the average corrected proportions of forest were much less for all images except 15m, which was much higher than the uncorrected average percent forest. Interestingly, the average corrected proportion of forest for all three scenes in Virginia was approximately the same.

Table 15. Classified proportion of forest in each of the images for each filtering method.

Percent forest	Image			
Filter	15m	16m	17m	MN
0	61.17	73.63	70.25	29.65
3x3maj	62.93	74.34	70.67	28.12
3x3on2	58.59	72.27	68.96	26.10
clump 1	60.39	73.31	69.91	28.27
clump 2	60.81	73.45	70.06	28.72
clump 3	62.10	73.92	70.49	29.07
Min	58.59	72.27	68.96	26.10
Max	62.93	74.34	70.67	29.65
Avg	61.00	73.49	70.06	28.32
Max-Min	4.35	2.07	1.71	3.55

Table 16. Range of corrected proportions of forest for all images, filtering methods and accuracy sets.

Percent forest	Image			
All filters and accuracy sets	15m	16m	17m	MN
Min	67.79	66.34	66.72	20.84
Max	69.49	70.07	68.60	23.60
Avg	68.70	68.08	68.02	22.54
Max-Min	1.69	3.74	1.88	2.76

4.4 Precision of forest area estimates

All of the results met the USDA Forest Service precision requirement of less than 3% per million acres timberland (Fig. 33, Fig. 34, Fig. 35, and Fig. 36). This indicates that the use of 200 random points as training data initiation locations for a region-growing approach is a viable protocol for collecting data for use in the IGSCR algorithm. Also, the improvement in overall accuracies using homogeneous point sets improved the resulting precision estimates despite the reduction in total points used in accuracy assessment. Since the precision was low enough to meet the standard no matter which filters and accuracy sets were used, there is some indication that the protocol used in this study is fairly robust and therefore not greatly affected by variations in validation data or image smoothing technique.

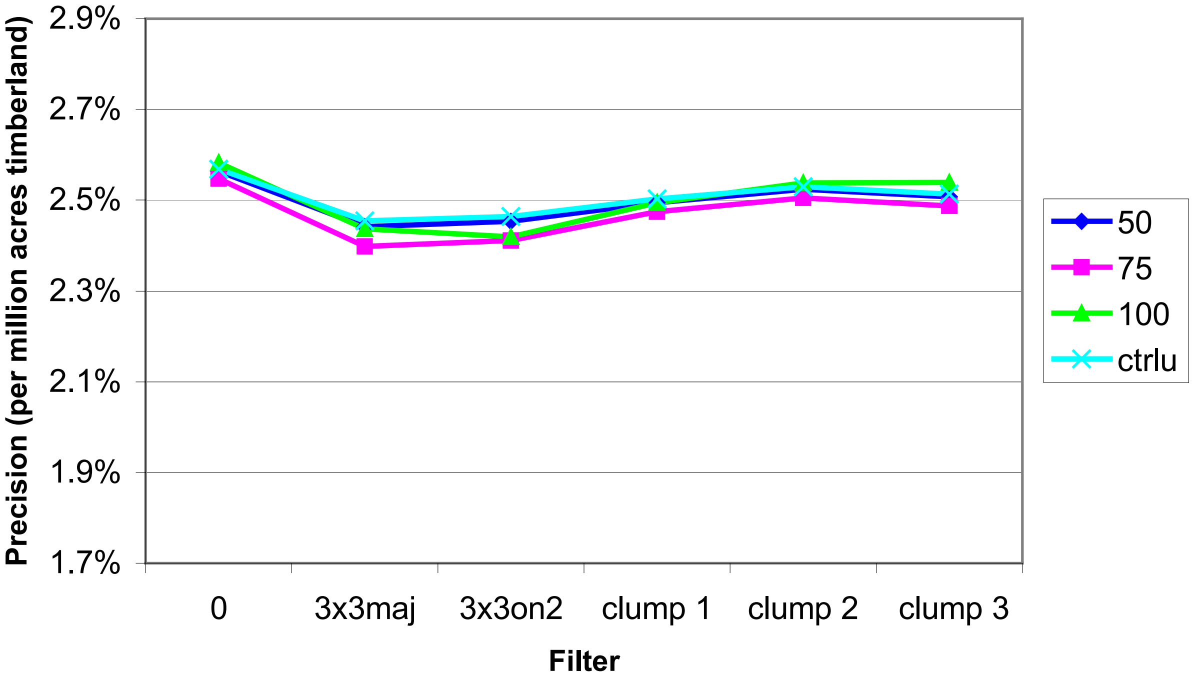


Fig. 33. Precision of forest area estimates for different filters and validation data sets for the Minnesota image.

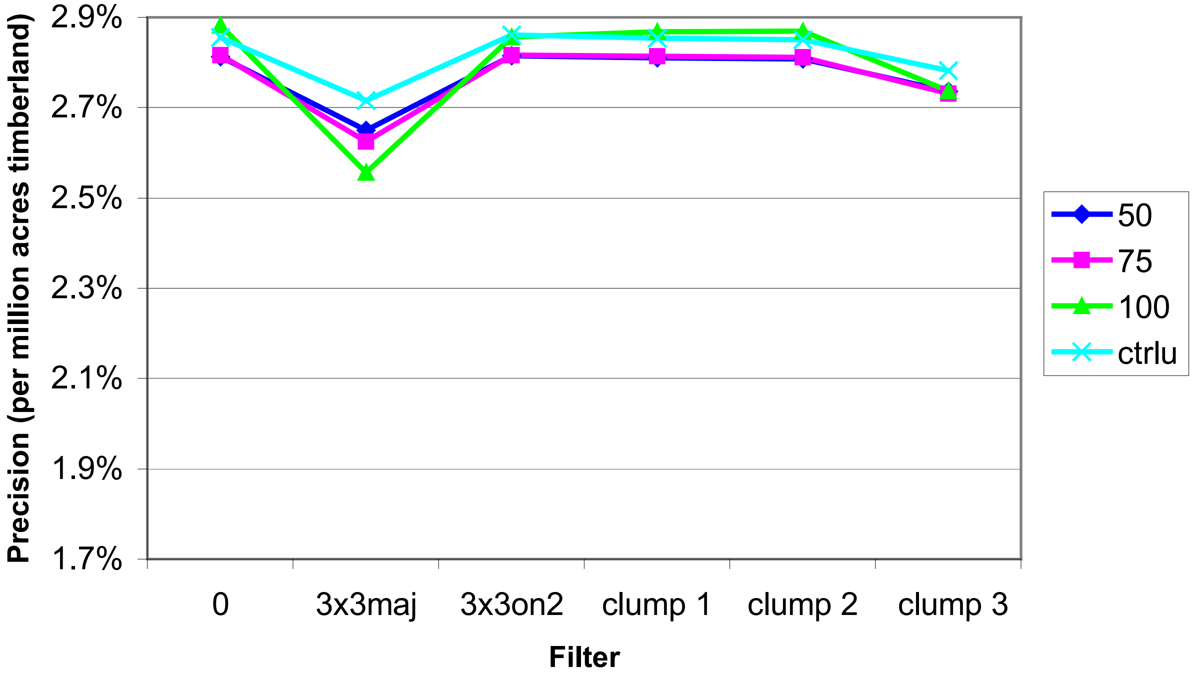


Fig. 34. Precision of forest area estimates for different filters and validation data sets for the 15m image.

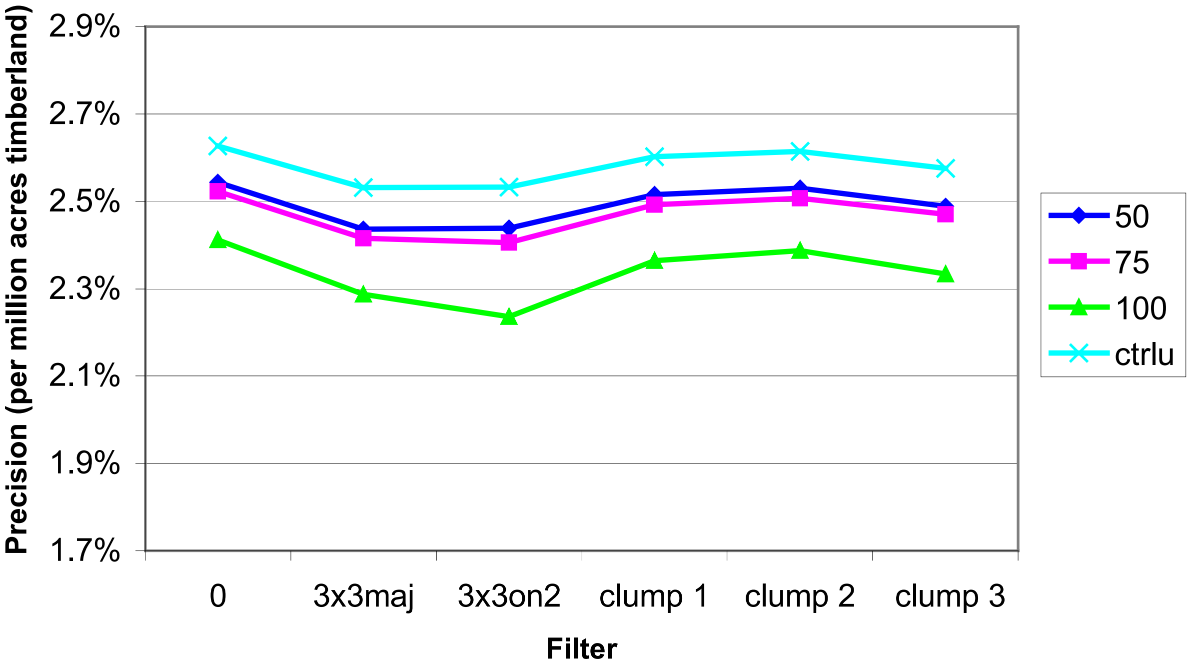


Fig. 35. Precision of forest area estimates for different filters and validation data sets for the 16m image.

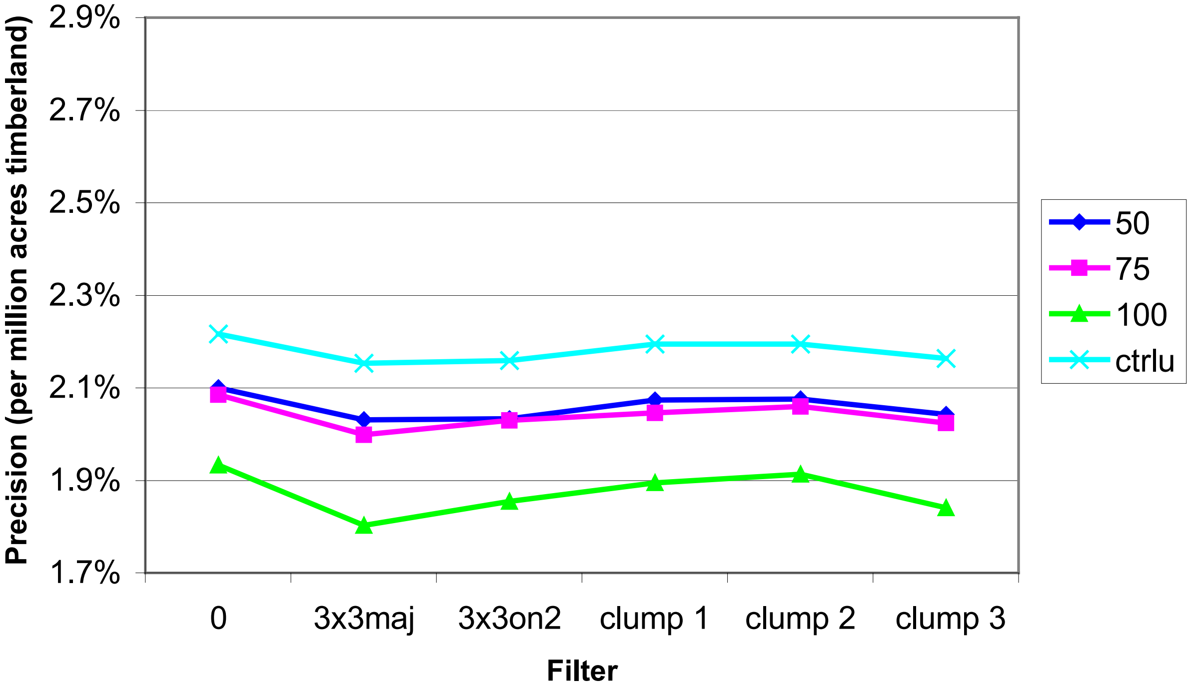


Fig. 36. Precision of forest area estimates for different filters and validation data sets for the 17m image.

Chapter 5

CONCLUSIONS

Prior to the automation of the IGSCR algorithm, the protocol was extremely analyst intensive, prone to error and time-consuming. Although certain calculations within the algorithm had previously been programmed, this was the first effort to successfully integrate the numerous ERDAS Imagine image processing algorithms and functions used in the IGSCR protocol into a fully automated program. The resulting program was used to classify the images in this project and has facilitated research into the optimal parameters to be used in the IGSCR algorithm. The program has also been used operationally by the Virginia Department of Forestry to develop a forest/nonforest map of the entire state of Virginia. We therefore conclude that the automation of the IGSCR algorithm was a worthwhile endeavor and that, if IGSCR classification were selected for use by the USDA Forest Service in the operational classification of Landsat imagery for use in forest area estimation, the program would offer significant time savings and prevent implementation errors.

Although there was a wide range in the classification accuracies resulting from running 4 Landsat ETM+ images through the IGSCR program and employing various filters and accuracy sets (81.9-95.4% overall), all of the results met the USDA Forest Service requirement of less than 3% per million acres timberland (Fig. 33, Fig. 34, Fig. 35, and Fig. 36). The precision of the resulting forest area estimates can be attributed in part to the use of random points, which allowed all available validation data to be used in accuracy assessment. We conclude that random points were acceptable as region-growing initiation locations to collect training data for IGSCR classification of the images used in this study. The use of random points is more repeatable than allowing the user to select locations unsystematically, and therefore is better suited for use in an operational Landsat TM image classification protocol for FIA forest area estimation. It is possible that random selection of points could neglect some spectral classes such as water if the sample number is insufficient. Thus, it is recommended that further investigation into the best possible training data protocol for use with IGSCR be performed.

There were no consistently significant differences in accuracy between the unfiltered image and the various filters (3x3 majority filter on entire image, 3x3 majority filter on forest only, and the 3 clump and eliminate methods; Table 13). Therefore it was concluded that the post-processing of images, at least with the methods and images tested here, is not necessary to obtain sufficiently precise forest area estimates. However, there is some indication that more fragmented images may benefit more from certain smoothing algorithms. Fig. 32 shows that there is a strong inverse correlation ($R^2 = 0.932$) between the proportion of edge in the image (i.e., the fragmentation) and the overall accuracy of the unfiltered image. Since none of the filters decreased the accuracy of the images, it might be advisable to use a filter, such as the 3x3 majority filter, on all images to avoid problems resulting from highly fragmented images. This topic requires further study.

The pairwise comparisons among the accuracy sets show that there were significant differences between the homogeneous and the center land use accuracy sets at the 95% level for all of the Virginia images (Table 14). Although the validation sets requiring 50% and 75% homogeneity were not significantly different from the center land use set, the overall accuracies were consistently higher than those for the center land use set (Fig. 28, Fig. 29, Fig. 30, and Fig. 31). Therefore, incorporation of some level of plot homogeneity is recommended for assessing the accuracy of classifications using FIA plot data if the sample is unbiased. The completely homogeneous validation data do not represent the population of pixels in the image because none of the edge pixels are represented. However, the 50% and 75% homogeneous plots do sample the edges and indicate to which stratum the edge pixels should be assigned. These validation data more accurately reflect the information on the ground than plots where the center land use alone is used, but is not representative of the land use on the rest of the plot. I believe that is the reason why they consistently improved the accuracy of the images, and although the 50% and 75% homogeneous validation sets were not significantly different from the center land use set, they are better suited for use in accuracy assessment than the center land use call alone. However, the optimal means of utilizing the FIA phase 2 plots as validation data without biasing the sample requires further investigation.

This study represents a viable means of producing forest area estimates that meet the precision standards required for the first phase of sampling in the USDA Forest Service Forest Inventory and Analysis forest inventory program. Landsat ETM+ data are a readily available, reasonably priced data source with the spectral and spatial characteristics necessary for forest/nonforest classification. The phase 2 plot data collected as part of the FIA program provide consistent data for accuracy assessment, which are sufficient in number to produce forest area estimates that meet the USDA Forest Service precision requirement. Also, Hansen and Wendt (2000) note that as more inventory data are collected, new strata could be developed which might improve estimates of change over time, and aid in the estimation of growth, removals and mortality. Another benefit of classifying a single image over time via IGSCR is that the training data could be reused, with modifications required only for areas of changed information class. This would greatly reduce the analyst time, which would decrease the cost of producing the classified product considerably.

The goal of this study was to aid in the development of an operational Landsat TM image classification protocol for FIA forest area estimation through the refinement of the Iterative Guided Spectral Class Rejection protocol. Automation of the IGSCR algorithm was successful and the program is now available for use in forest area estimation. The results of the IGSCR classifications performed in this study indicate that the use of random points as seed pixel initiation points for training data collection may be an acceptable and repeatable addition to the IGSCR protocol, given that the training data collected are representative of the spectral characteristics of the image. The various methods of spatial post-processing of the classified images used in this study do not point to the need for addition of such methods to the existing IGSCR protocol, except possibly in areas of significantly fragmented land cover. Finally, if it would not bias the sample, the incorporation of a plot homogeneity requirement to the selection of validation data for accuracy assessment is recommended as an addition to the IGSCR classification protocol based on the improved accuracy of the images assessed with validation sets based on more than just the center land use call.

Chapter 6
FUTURE RESEARCH

The following are recommendations for future research and development related to IGSCR classification for use in forest area estimation:

- Compare the results of this project with results obtained by similar methods in other parts of the United States
- Assess the variability in the training data collected by different users from the same image using the same seed pixel initiation points and validation data
- Determine the utility of stratified random sampling of points for use in training data collection as seed pixel initiation points
- Examine the effects of the relationship between training data and the number of spectral classes selected
- Further automate training data generation via the region-growing method by developing
    - a viewer that zooms to points
    - a program to generate areas of certain Euclidean distances at random points
    - a viewer/selector for AOIs created with specific Euclidean distances
- Investigate the use of slightly modified training data for different image dates (or possibly for adjacent images)
- Use unclassified pixels to inform further training data collection
- Investigate the possible adverse effects of spatial autocorrelation in training data used for IGSCR classification
- Determine whether or not point data alone would suffice in training for IGSCR
- Further investigate the relationship between fragmentation and spatial postprocessing
- Determine the utility of post-processing algorithms not tested in this project
- Attempt to classify into more strata, e.g., hardwood, softwood and nonforest
- Look into the possibility of using the unclassified pixels in the stacked ISODATA image as a separate stratum to reduce the variance of the forest area estimate in images which do not meet the FIA precision requirement
- Determine the optimal number of ISODATA classes, and the best homogeneity threshold and type-I error for use in IGSCR classification.

Chapter 7

LITERATURE CITED

Bauer, M.E., T.E. Burk, A.R. Ek, 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(6):3157-3166.

Buchheim, M.P. and T.M. Lillesand, 1989. Semi-automated training field extraction and analysis for efficient digital image classification, *Photogrammetric Engineering & Remote Sensing*, 55(9)1347-1355.

Campbell, J.B., 1981. Spatial correlation effects upon accuracy of supervised classification of land cover, *Photogrammetric Engineering & Remote Sensing*, 47(3):355-363.

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.

Center for Coastal Resources Management, 1991. 1:24,000 Digital Shoreline Coverage (SHL). Comprehensive Coastal Inventory Program, Virginia Institute of Marine Science, College of William and Mary, Gloucester Point, Virginia, 23062.

Chojnacky, D.C., 1998. Double sampling for stratification: A forest inventory application in the Interior West. Research Paper RMRS-RP-7, Fort Collins, CO: USDA Forest Service, Rocky Mountain Research Station.

Congalton, R., and K. Green, 1999. Assessing the accuracy of remotely sensed data: Principles and practices, 1st ed., New York: CRC Press.

ERDAS, 1999. ERDAS field guide (5th ed.), revised and expanded. Atlanta, GA: ERDAS.

Farr, R., 1999a. Landsat 7 Level 0R, 1R and 1G Data Products Dataset Guide Document. *NASA Landsat 7 Page: Documentation*. http://landsat.gsfc.nasa.gov/guides/LANDSAT-7_dataset.html#9. Access date: April 2, 2003.

Farr, R., 1999b. Comparison of the Landsat missions. *NASA Landsat 7 Page: Documentation*. http://landsat.gsfc.nasa.gov/project/Comparison.html. Access date: April 2, 2003.

Franco-Lopez, H., A.R. Ek, and M.E. Bauer, 2001. Estimation and mapping of forest stand density, volume, and cover type using the *k*-nearest neighbors method, *Remote Sensing of Environment*, 77:251-274.

Frayer, W.E., and G.M. Furnival, 1999. Forest survey sampling designs: a history, *Journal of Forestry*, 97(12):4-10.

Gillespie, J.R., 1999. Rationale for a national annual forest inventory program, *Journal of Forestry*, 97(12):16-20.

Guldin, Richard W., 2000. Remote sensing at the dawn of a new millennium: A Washington, DC Perspective, Keynote remarks presented at the RS2000 Conference, April 13, 2000, Albuquerque, New Mexico.

Hansen, M.H., and D.G. Wendt, 2000. Using classified Landsat Thematic Mapper data for stratification in a statewide forest inventory. In: R.E. McRoberts, G.A. Reams and P.C. Van Deusen (Eds.), *Proceedings of the First Annual Forest Inventory and Analysis Symposium*. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station (General Technical Report NC-213).

Hawkins, J.M. and R. Allen (Eds.), 1991. The Oxford Encyclopedic English Dictionary. New York: Oxford University Press.

Lillesand, T.M., and R.W. Kiefer, 2000. Remote sensing and image interpretation, 4th ed., New York: John Wiley & Sons, Inc.

Lund, H.G., 1999. A 'forest' by any other name..., *Environmental Science and Policy*, 2:125-133.

McCaffrey, T.M. and Franklin, S.E., 1993. Automated training site selection for large-area remote-sensing image analysis, *Computers and Geosciences*, 19(10):1413-1428.

McRoberts, R.E., 1999. Joint Annual Forest Inventory and Monitoring System: The North Central perspective, *Journal of Forestry*, 97(12):27-31.

McRoberts, R.E., Wendt, D.G., Nelson, M.D., Hansen, M.H., 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment*, 81:36-44.

Reams, G. and P.C. Van Deusen, 1999. The southern annual forest inventory system, *Journal of Agricultural, Biological, and Environmental Statistics*, 4(3):108-122.

Richards, J.A. and X. Jia, 1999. Remote Sensing Digital Image Analysis. 3rd ed. New York: Springer-Verlag.

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.

Riemann, R., M. Hoppus and A. Lister. 2000. Using arrays of small ground sample plots to assess the accuracy of Landsat TM-derived forest cover maps. In: *Proceedings,* *Accuracy2000, 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Amsterdam, The Netherlands. July 12-14, 2000.

Roesch, F.A. and G.A. Reams, 1999. Analytical alternatives for an annual inventory system, *Journal of Forestry*, 97(12):33-37.

Scrivani, J.A., R.H. Wynne, C.F. Blinn, and R.F. Musy, 2001. Phase I forest area estimation using Landsat TM and Iterative Guided Spectral Class Rejection: Assessment of possible training data protocols. In: *Proceedings of the 2nd Annual Forest Service FIA Symposium*, Salt Lake City, Utah, October 17-18, 2000. General Technical Report SRS-47. Asheville, NC: USDA Forest Service Southern Research Station, pp. 11-14.

Stoms, D.M., 1996. Validating large-area land cover databases with maplets, *Geocarto International*, 11(2):87-95.

USDA Forest Service (USDA-FS), undated. Fact Sheet Series, Forest Inventory and Analysis Program History. http://fia.fs.fed.us/library/FIA_History_FS.doc. Access date: October 15, 2001.

USDA Forest Service, 1989. Interim Resource Inventory Glossary. File 1900. Washington, DC.

USDA Forest Service, 1992. Forest resource inventories: An overview, Washington, DC.

USDA Forest Service, 2001. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.5, U.S. Department of Agriculture, Forest Service, Washington Office. Internal report. On file with: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, 201 14th St., Washington, DC, 20250.

USGS, 2003. USGS Landsat project. http://landsat7.usgs.gov/outreach.html. Date accessed: April 4, 2003.

Van Deusen, P.C., S.P. Prisley, and A.A. Lucier, 1999. Adopting an annual inventory system: User perspectives, *Journal of Forestry*, 97(12):11-14.

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, *Proceedings of Second International Conference-Geospatial Information in Agriculture and Forestry*, 10-12 January 1000, Lake Buena Vista, Florida, 2:355-362.

Wayman, J.P., R.H. Wynne, J.A. Scrivani, and G.A. Reams, 2001. Landsat TM-based forest area estimation using Iterative Guided Spectral Class Rejection, *Photogrammetric Engineering & Remote Sensing*, 67(10):1155-1166.

Wright, H.E. Jr., 1972. Physiography of Minnesota. In: *Geology of Minnesota:A Centennial Volume*, P.K. Sims and G.B. Morey (Eds.). Minnesota Geological Survey, University of Minnesota, St. Paul, Minnesota. Pp. 561-577.

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.

APPENDIX A
Acronyms and abbreviations

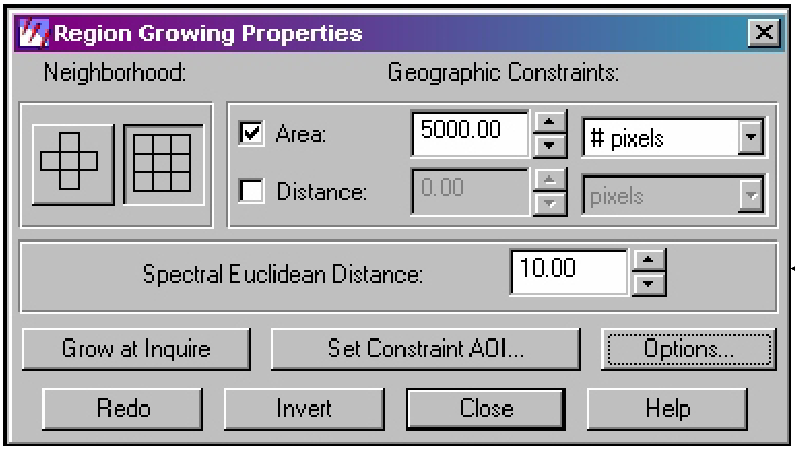
AOI	Area of Interest
ASCII	American Standard Code for Information Interchange
DLL	Dynamic Link Library
DOQQ	Digital Orthophoto Quarter-Quadrangles
EML	ERDAS Macro Language
EuD	Euclidean Distance
EROS	Earth Resources Observation Systems
ETM+	Enhanced Thematic Mapper Plus
FHM	Forest Health Monitoring
FIA	Forest Inventory and Analysis
GPS	Global Positioning System
GUI	Graphical User Interface
HTML	HyperText Markup Language
IGSCR	Iterative Guided Spectral Class Rejection
ISODATA	Iterative Self-Organizing Data Analysis Technique
MRLC	Multi-Resolution Land Characteristic Interagency Consortium
MSS	Multispectral Scanner
P.L.	Public Law
PSU	Primary Sampling Unit
RSB	Remote Sensing Band
TM	Thematic Mapper
USDA	United States Department of Agriculture
USGS	United States Geological Survey
USFS	United States Forest Service
WRS	Worldwide Reference System

APPENDIX B
Training Data Collection- Seed Pixel Approach

1. Open ERDAS Imagine.
2. In the viewer, open the panchromatic image of the 6 county area:
   File|Open|Raster Layer…
   Navigate to and select "**6co_pan.img**".
      Click on the **Raster Options** tab.
      Make sure that **Clear Display** is not checked
      and **Fit to Frame** is checked.

☐ Clear Display
☑ Fit to Frame

      Click OK.
3. In the same viewer, open the multispectral image, "**6co_subset.img**".
4. Still in the same viewer, open the reference point shapefile:
   File|Open|Vector Layer…
   At the bottom of the window that appears, in the **Files of type** dropbox, select **Shapefile** (*.shp).
   Navigate to and select "**6co_sub.shp**".
   Click on the **Vector Options** tab.
   Click on **Use Symbology**.
   Click on **Set…**
   Navigate to CD ROM drive.
   Select "**points.evs**" and click on **OK** (twice, if necessary).
   Make sure that **Clear Display** is not checked.
   Click on **OK.**
5. In the viewer menu, open the AOI tools: **AOI|Tools…**
6. Also in the viewer menu, open the AOI seed properties window: **AOI|Seed Properties...**
7. Change the settings in the **Region Growing Properties** window:
      Set **Neighborhood** button to include the 8 nearest neighbors.
      Set the **Geographic Constraints** to an area of 5000 pixels.
      Set the initial **Spectral Euclidean Distance** to 10.



Region Growing Properties

Neighborhood:

Geographic Constraints:

☑ Area: 5000.00 # pixels

☐ Distance: 0.00 pixels

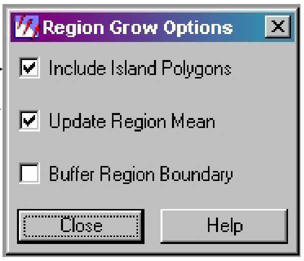
Spectral Euclidean Distance: 10.00

Do not close the Region Growing Properties window.

On the lower right-hand side of the **Region Growing Properties** window, click on **Options…**

Make sure that **Include Island Polygons**
and **Update Region Mean** are checked.
Close the **Region Grow Options** window.

Do not close the Region Growing Properties window



☑ Include Island Polygons
☑ Update Region Mean
☐ Buffer Region Boundary

8. In the viewer menu, select **Vector|Attributes…** A table will appear that contains data pertinent to the reference points in the shapefile. Adjust the column widths so that the following fields are visible: **PLOTID**, **CUR LU**, **SUBPLOT**, and **NO CONDS**.

PLOTID is a unique 8-digit number given to each point which consists of a 2-digit state code (Virginia = 51), a 3-digit county code and a 3-digit plot number.

CUR LU is a two-digit number indicating the land use at the time the plot was visited.

LU Codes: 20 = Timberland
61 = Cropland
62 = Improved Pasture
66 = Other Farmland
67 = Urban and Other
91 = Census Water
92 = Non-Census Water

SUBPLOT is a number between 0 and 400 that indicates the portion of the plot that is forested. For example, a plot with **CUR LU** = 67 and **SUBPLOT** = 200 is half urban and half forest. Note: This number is not given when the plot is a deleted plot that was previously measured and has been revisited.

NO CONDS gives the number of different types of forest, or forested conditions, in the plot. If this field contains a “d”, then it is a deleted plot.

9. Select the first point in the vector attributes table. This will highlight the point in the table and in the viewer in yellow. Zoom in to the area of the point. Examine the area to determine in what land use the point appears to be. Also note the land uses in adjacent areas and whether there appear to be buildings, roads or clearings in forested areas.

Note: There is not an easy way to zoom to selected points in the vector attributes table. Therefore, we suggest that you select a point in the vector attributes table then zoom to the full extent of the image (right-click in the viewer and select **Fit Image to Window**) so that you can see the highlighted point. If you do this in the order that the points appear in the table, it will help you keep track of points that have already been used to create AOIs, since small AOIs are not always visible when zoomed out.

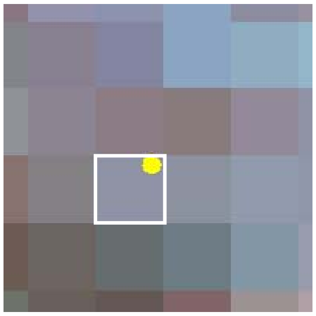


Fig. 1. White box marks a pixel in the multispectral image that contains a point.

10. Zoom in close to the point. You need to zoom in enough to be certain that you can click on the pixel within which the point is contained (see Fig. 1).

11. Click on the seed pixel tool in the AOI tool palette (your cursor should turn into a crosshair).

12. Click inside the pixel that contains the point and wait for the area of interest (AOI) to be created.

13. Zoom out and inspect the area that was created. If the area appears to contain only pixels that are of the same land use as the point, and the area could possibly be increased in size without including pixels of a different land use, increase the **Spectral Euclidean Distance** by **1** or **2** units and click on **Redo** (bottom left button of the **Region Growing Properties** window).

If the area contains any pixels that may be of a different land use or any pixels that appear to be a mixture of two land uses, decrease the **Spectral Euclidean Distance** by **1** unit and click on **Redo**. Edge pixels are any pixels that contain mostly reflectance of the "wrong" land use, e.g., if you are collecting forest and there is a non-forest edge pixel in your AOI.

It is often helpful to check whether any of the pixels in the AOI are mixed by examining the AOI over the panchromatic image (see Fig. 2). To do this, click on **View|Arrange Layers…** in the viewer. In the **Arrange Layers** window that appears, you can drag the multispectral image to the bottom of the list of layers (under the pan image) and click on **Apply**. After you examine the AOI over the panchromatic image, move the pan image to the bottom of the list and click **Apply**.

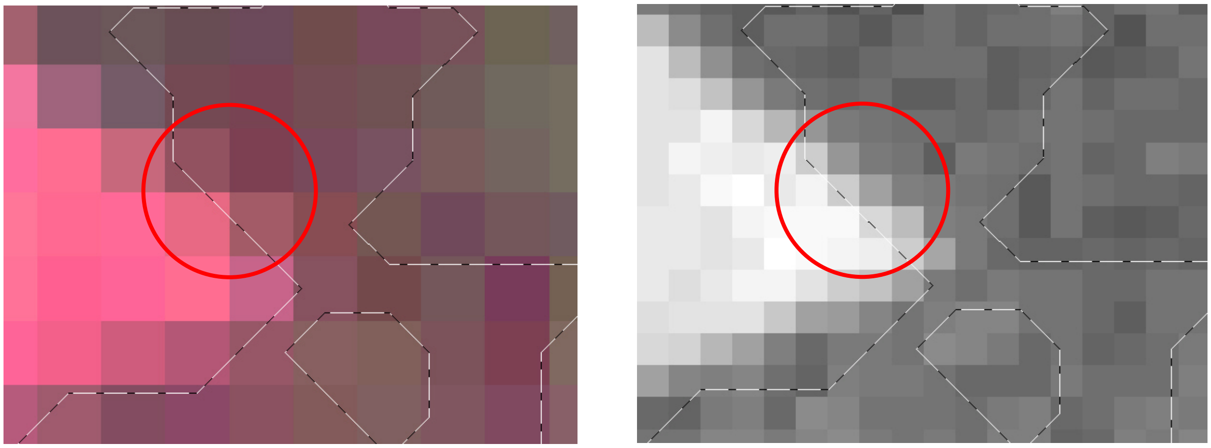


Fig. 2. Area indicated contains mixed pixels on the multispectral image (left) that are more clearly defined as forest edge by viewing the same area of interest on the panchromatic image. (right). The mixed pixels in this AOI can be included since they are considered "forest edge".

Continue to adjust the **Spectral Euclidean Distance** until the area created is as large as possible and contains only pixels that are clearly of the same land use as the pixel that contained the starting point.

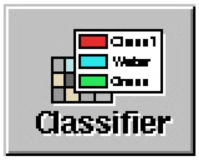
Special Cases

a. If there are clouds or cloud shadows in the area of the point, do not use the point.
b. If the point does not appear to be on a pixel of the land use indicated by **CUR LU**, e.g., if **CUR LU** = 67 and the point is clearly in a pine plantation, do not use the point.
c. If the pixel that contains the point appears to be a mixture of two land uses and seems to be on the edge of the "wrong" land use, you should not use the point.
d. If **SUBPLOT** < 400 and the point appears to be in a homogeneous pixel, initiate a seed pixel at the point. If only pixels of the appropriate land use are collected, use the point.
e. If **CUR LU** = 67 and the point is in a suburban area where there are trees in the pixel which contains the point, do not use the point.
f. If the point is in an area of bad data on your image, don't use the point.

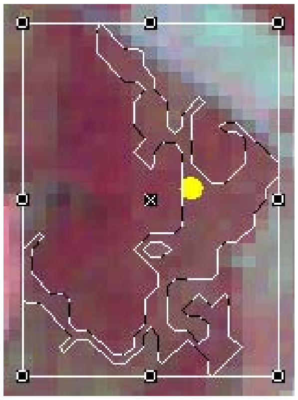
If you choose not to use a point, set the **Spectral Euclidean Distance** to zero and initiate a seed. This will create an area that only contains one pixel. Use this as a placeholder when saving your AOIs to ensure that you have visited each point. When naming the point, include a note that says "don't use".

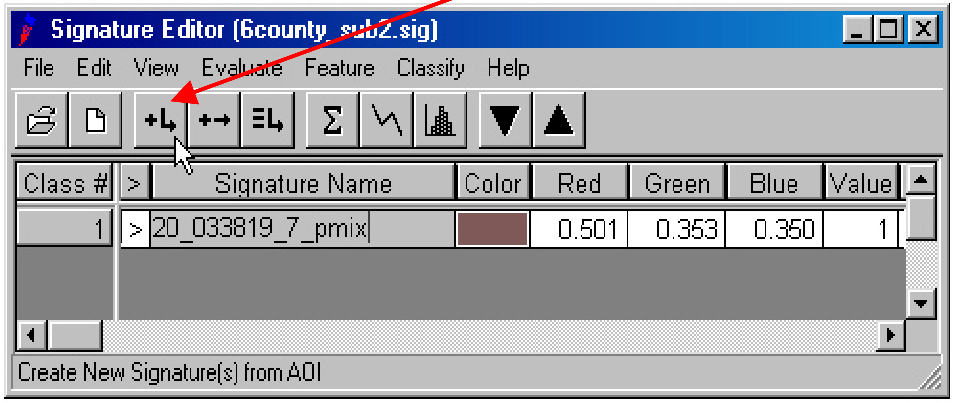
14. To save the AOI, along with information about the point, the **Signature Editor** is used. **<u>Important Note</u>**: the Signature Editor is <u>only</u> used to name and store collections of forest and non-forest pixels. The areas of interest stored in the Signature Editor have no required spectral properties such as normality or separability. The pixels in these areas are used to determine which ISODATA classes are pure. The areas of interest are <u>not</u> spectral classes.

To open the **Signature Editor**, click on the **Classifier** button on the Imagine toolbar, and click on **Signature Editor…** in the **Classification** menu that appears.



15. To add an AOI to your signature file, click on the **Create New Signature(s) from AOI** button in the **Signature Editor** while your AOI is selected.





Class #	>	Signature Name	Color	Red	Green	Blue	Value
1	>	20_033819_7_pmix	🟫	0.501	0.353	0.350	1

Create New Signature(s) from AOI

16. Name your signature as follows: **CUR LU_PLOTID_EuclideanDistance_description**, e.g., **20_035719_7_hw**. The **CUR LU** and **PLOTID** should still be highlighted in the **Attributes** table. The **Euclidean distance** is the Euclidean distance you decided to use to create the AOI.

17. Save your file.
    Select **File|Save As…** from the **Signature Editor** menu bar and navigate to the location where you want to store your file. Name your file "**6co_sub.sig**".
    After naming your file, all other saves can be performed by selecting **File|Save**. It is best to save after each AOI is added to the **Signature Editor**.

18. Create separate forest and non-forest signatures from your subset signature file which contain only the good signatures. This facilitates the creation of AOI layers in the next step.
    When you are finished creating all of the AOIs, select all of the rows that contain good forest signatures by holding down the shift key while left-clicking on the row numbers under **Class #**.
    Then, go to **File|Save As…**
    In the window that appears, click on the **Selected** radio button to save only the selected signatures.
    Save the file.
    Name the file "**6co_for.sig**".

19. Repeat step 18 for the good non-forest signatures and name the file "**6co_nf.sig**".

20. Create forest and non-forest AOIs.
    Open "**6co_subset.img**" in a new viewer.
    Open a new **Signature Editor** by clicking on the **Classifier** button on the Imagine toolbar, and then clicking on **Signature Editor…**
    Open the forest signature file ("**6co_for.sig**").
    Select all rows by right clicking in the column under **Class #** and choosing **Select All**.
    Select **View|Image AOI**. This should make all of your AOIs appear in the viewer.
    In the viewer, select **File|Save|AOI Layer As…**
        Save your forest AOI.
        Name the file "**6co_for.aoi**".
    Click on **View|Arrange Layers…**
        Right-click on the AOI layer and select **Delete Layer and click on Apply**.

21. Repeat step 20 for the non-forest signature file ("**6co_nf.sig**").
    Save the non-forest AOI file as "**6co_nf.aoi**".

22. Since water is often underrepresented in the non-forest plots, water must be addressed separately and added to the non-forest data collected using the plots as initiation points.
    This can be done by initiating seed pixels in user-selected water bodies in an attempt to cover the spectral variability of the water in the scene.
    The feature space image can be used to check whether or not the variability was adequately covered with the selected areas.
    Save the water AOI layer as "**6co_water.aoi**".
    The water AOIs can be added to the other non-forest AOIs by opening both AOI layers in the same viewer. Save all of the non-forest and water AOIs together as "6co_nf.aoi" (overwrite the previous version, which did not include water).

APPENDIX C
Minnesota Image Training Point Information

Projection

Projection type: Albers Conical Equal Area
Spheroid: GRS 80
Datum: NAD 83
1st standard parallel: 29.5 N
2nd standard parallel: 45.5 N
Central meridian: 96 W
Latitude of origin: 23 N
False easting: 0 m
False northing: 0 m

Points

Point	X	Y	LU	EuD	Count	B1	B2	B3	B4	B5	B6
1	111030	2548350	nf	14	323	68	57	45	174	92	40
10	66600	2580000	nf	10	63	69	56	47	139	100	55
16	76200	2505930	nf	15	588	68	50	43	149	103	58
19	110880	2566620	nf	10	42	71	56	45	123	82	42
21	148440	2513460	nf	17	419	68	57	41	156	91	41
24	82020	2550240	nf	14	22	70	61	50	127	107	60
33	140670	2542110	nf	10	33	65	51	38	173	90	38
34	126720	2604330	nf	14	100	69	53	42	134	87	42
44	129360	2516220	nf	15	198	65	53	43	137	92	41
45	95520	2566020	for	12	245	69	55	45	131	75	35
51	22920	2563350	for	14	418	69	51	38	148	74	34
55	83910	2499420	nf	12	10	73	59	49	148	98	53
59	69240	2607210	for	12	221	65	48	36	153	74	32
65	138180	2560620	nf	14	11	75	64	66	126	135	80
68	103440	2530890	nf	10	16	74	66	54	138	115	61
70	144990	2557050	nf	10	40	68	58	46	173	93	42
72	34230	2489100	nf	10	2235	66	47	35	20	10	10
78	147060	2550810	nf	17	169	64	56	45	173	96	42
79	161190	2556210	nf	10	80	62	55	36	163	96	39
83	185190	2612160	for	10	38	63	48	34	158	79	37
86	6960	2518620	nf	20	31	71	60	44	173	87	39
103	215340	2626290	for	12	5000	58	44	31	141	66	30
106	102570	2610390	nf	12	213	68	60	48	142	110	52
108	174720	2461200	for	12	102	60	40	29	47	25	18
111	90720	2651910	nf	10	830	71	59	38	18	10	10
115	94860	2557320	nf	14	37	72	58	50	106	94	55
117	95460	2500440	nf	10	124	66	51	36	162	77	32
126	9540	2533710	nf	14	66	74	71	55	150	96	48
128	185940	2547240	for	10	1058	62	46	32	153	73	30
131	203550	2558370	for	10	5000	59	45	31	156	76	32

Point	X	Y	LU	EuD	Count	B1	B2	B3	B4	B5	B6
139	119460	2525640	nf	14	45	70	57	42	149	100	48
141	179490	2626260	nf	10	84	64	55	40	153	81	34
143	41370	2580270	for	10	4309	64	46	32	144	75	35
153	168990	2549160	nf	12	80	70	60	45	154	114	58
154	128400	2554440	nf	17	64	80	68	63	115	86	52
157	79290	2547390	for	10	21	69	55	42	146	87	41
160	82230	2579280	nf	10	31	64	57	35	200	91	35
167	97320	2510280	nf	14	204	69	53	39	166	84	35
169	74760	2488080	nf	14	1335	68	56	43	129	100	46
170	167970	2597610	nf	10	343	60	36	27	21	11	10
181	30840	2508000	nf	10	355	69	56	40	187	79	35
183	18540	2497230	nf	14	721	69	55	37	225	112	51
185	178920	2460810	for	14	179	62	45	31	145	75	33
186	193440	2586450	for	10	1075	60	46	32	140	74	33
187	64620	2489940	nf	20	11	66	54	40	138	63	32
190	181500	2603220	nf	10	63	63	51	33	156	86	40
191	42870	2546550	nf	10	1144	64	39	28	20	11	10
194	158370	2494830	nf	10	204	69	68	46	27	12	11
197	190140	2608440	for	10	52	58	42	28	161	74	31
200	157380	2620170	nf	12	16	65	50	36	155	79	34
201	129990	2476410	nf	16	24	64	55	38	132	87	46
212	5730	2511270	nf	17	104	68	55	37	153	81	37
214	113610	2544420	nf	12	32	73	66	73	121	106	63
221	111960	2581050	nf	10	5000	64	42	28	19	11	11
227	194130	2612310	for	10	5000	58	42	28	152	69	28
236	74610	2479050	nf	12	42	69	57	42	145	85	39
238	112020	2591520	for	10	69	64	46	31	148	73	27
240	125790	2508300	for	12	229	61	45	33	159	74	29
245	77400	2571660	nf	10	80	71	62	49	170	100	45
249	112050	2472450	nf	14	65	66	55	39	193	98	50
250	29640	2519070	nf	17	126	70	56	37	192	107	49
257	112350	2492250	nf	10	86	66	44	31	22	11	9
258	600	2497080	nf	15	175	72	59	44	173	117	59
260	87330	2558880	nf	14	117	68	60	43	163	87	38
263	167850	2534490	nf	10	54	63	56	38	228	84	31
267	179040	2552220	nf	14	199	63	51	35	176	90	37
275	68730	2557770	nf	14	82	68	57	49	131	88	41
277	192210	2588880	for	10	40	62	53	36	145	85	37
279	166470	2570550	for	10	189	59	45	28	162	70	29
287	30210	2484270	nf	17	192	68	58	43	151	83	37
292	62850	2564220	nf	12	41	71	58	50	109	100	58
294	175380	2559060	for	10	2468	58	43	28	152	71	30
297	137940	2510460	nf	17	24	70	50	40	116	70	37
298	123240	2521020	nf	10	60	65	50	37	166	74	30
310	70440	2476620	nf	12	290	67	54	40	159	81	37

Point	X	Y	LU	EuD	Count	B1	B2	B3	B4	B5	B6
321	156630	2488260	nf	17	70	71	58	45	164	102	52
323	87300	2497500	nf	15	154	69	54	42	159	89	46
330	152340	2569770	nf	17	24	71	59	50	126	117	59
348	44010	2497800	nf	16	93	72	61	48	149	102	50
352	172830	2562840	for	17	5000	59	44	30	138	65	28
355	97590	2629320	nf	12	85	66	58	39	185	98	44
356	99960	2617080	nf	10	57	68	58	42	148	100	46
358	96840	2521440	nf	17	111	74	62	48	178	90	43
367	41460	2599590	nf	10	128	72	61	55	115	90	48
375	148140	2466150	nf	14	20	78	67	66	134	101	66
381	70590	2615220	for	12	754	65	47	35	147	72	30
384	182100	2620560	nf	15	116	69	52	36	184	76	32
391	110040	2595060	nf	14	42	69	60	41	184	90	39
397	176850	2505060	nf	17	8	65	53	43	86	55	34
399	70380	2576310	nf	10	54	71	64	44	163	97	44
400	28110	2548080	nf	15	275	69	55	38	148	94	44
412	82020	2523900	nf	17	18	72	64	47	190	119	54
416	147780	2538000	nf	17	55	65	53	38	157	84	37
419	165120	2488170	nf	10	862	61	39	35	27	7	9
426	106920	2529120	nf	12	31	83	78	92	125	124	82
427	33930	2615310	nf	17	78	73	63	50	106	95	45
439	122520	2493510	nf	12	111	74	65	55	132	99	54
441	66840	2633310	nf	17	78	68	52	40	82	47	26
442	24930	2508420	nf	10	232	68	51	39	194	73	30
448	154170	2577660	for	11	256	64	47	36	144	70	33
453	59670	2487000	nf	12	206	69	55	42	142	104	61
454	43350	2551230	nf	17	9	79	63	68	66	108	80
460	124500	2648670	for	17	126	67	52	39	107	63	32
461	67320	2531100	nf	12	367	66	53	38	201	80	34
463	172920	2599230	for	10	379	63	48	33	153	82	36
467	7530	2513970	nf	10	476	68	51	38	171	76	33
469	106710	2480010	nf	14	1112	70	58	42	187	92	42
471	113520	2652030	nf	17	61	68	55	49	111	103	55
474	170880	2602140	nf	15	6	73	65	51	106	64	33
481	74250	2483250	for	12	524	64	52	36	165	85	38
485	98490	2646030	nf	14	47	71	61	48	140	71	35
496	107910	2627340	for	10	421	64	46	31	146	72	31
497	124140	2489670	nf	17	25	80	74	70	150	139	84
515	77070	2484510	nf	14	486	67	60	38	249	123	57
516	66480	2505180	nf	10	893	69	51	37	183	76	33
519	187470	2538240	for	14	3666	61	45	32	139	69	30
528	98250	2509980	for	12	524	62	45	33	156	69	30
548	125370	2615760	for	17	10	63	46	32	110	57	26
560	107010	2543010	nf	17	82	69	60	39	193	95	41
563	113700	2555910	nf	17	260	70	57	40	162	83	38

Point	X	Y	LU	EuD	Count	B1	B2	B3	B4	B5	B6
569	84120	2627640	for	14	1158	65	49	35	167	84	36
570	72840	2606730	nf	12	99	69	52	45	120	85	46
572	149430	2562420	nf	14	31	72	59	53	120	124	77
575	79350	2558940	nf	15	84	75	67	62	146	129	64
581	39420	2590140	for	10	62	65	49	32	158	74	30
589	169980	2599020	for	12	5000	61	44	29	131	62	27
592	43020	2513190	nf	10	5000	65	47	29	17	10	8
597	116700	2501940	nf	15	84	69	58	42	184	95	44
599	51270	2549220	nf	10	5000	67	48	32	18	10	11
601	70410	2630520	nf	17	56	70	64	46	182	95	43
613	155820	2559600	nf	17	26	69	59	45	156	110	53
618	100800	2524170	nf	17	125	68	56	39	167	78	35
621	187650	2577870	nf	10	5000	63	41	29	19	12	14
623	138240	2577090	nf	17	40	67	56	43	149	91	41
636	72720	2515770	nf	17	503	69	52	39	133	90	49
637	55500	2604480	nf	14	73	73	64	47	177	98	42
646	200730	2617410	nf	10	165	61	45	32	59	37	23
649	103410	2583930	nf	12	12	68	56	37	184	95	41
650	80130	2497800	nf	17	407	71	66	44	200	74	32
662	70530	2498910	nf	17	324	65	51	33	198	104	47
672	160770	2572290	for	11	1218	62	45	33	162	73	29
676	56730	2559390	for	12	32	68	56	40	140	75	34
680	145800	2467290	nf	17	64	67	58	49	134	85	39
682	151590	2487840	for	14	341	61	45	32	138	66	29
684	154890	2469780	for	17	50	63	50	35	118	69	33
686	132390	2535810	for	10	62	61	46	32	157	70	29
693	78930	2555160	nf	14	444	69	59	46	148	85	39
695	29760	2567310	nf	14	14	71	59	42	160	96	44
697	177990	2541630	for	10	875	61	44	30	148	72	30
703	179910	2472690	nf	12	213	66	50	35	201	78	34
710	48330	2596350	for	10	232	67	55	40	127	64	29
713	184920	2626110	for	12	834	63	44	30	140	59	25
719	90990	2636520	for	10	94	64	45	35	110	52	26
725	171630	2557710	for	10	4306	57	45	30	149	76	33
729	119250	2521650	nf	17	12	77	71	68	155	142	79
731	73080	2614530	nf	10	86	72	56	47	125	92	46
736	109320	2481270	nf	10	108	65	51	36	190	75	30
744	40560	2484810	nf	17	283	68	55	42	157	114	63
752	133740	2517120	nf	10	58	68	58	51	134	104	50
761	45330	2557320	nf	10	5000	65	45	29	17	10	7
778	75330	2592990	for	10	14	65	43	31	157	74	30
784	115350	2499390	for	10	61	66	49	35	153	80	37
788	65970	2618340	nf	12	121	68	55	41	150	81	37
792	132270	2474520	nf	14	168	68	59	42	166	108	52
799	191070	2571060	for	11	207	61	50	34	159	83	35

Point	X	Y	LU	EuD	Count	B1	B2	B3	B4	B5	B6
811	86220	2499630	nf	11	96	66	56	38	185	93	40
819	189240	2613360	for	10	23	60	44	28	156	73	28
826	67590	2493060	nf	17	201	71	56	47	129	98	50
843	197040	2628450	for	10	277	60	44	30	160	72	31
864	60570	2481870	nf	10	413	67	56	41	32	11	10
868	114510	2513670	nf	12	68	72	59	56	106	118	64
875	136200	2544870	nf	12	159	70	60	42	158	91	39
877	215040	2606070	for	10	4157	59	46	31	172	85	34
880	82350	2483790	nf	15	4273	62	43	33	31	14	11
885	94560	2545140	nf	10	66	72	62	45	169	101	45
887	73710	2619090	nf	14	97	76	67	48	227	81	39
891	70050	2590050	nf	17	31	73	61	51	113	124	71
892	96240	2604810	for	14	29	69	56	36	162	93	41
898	137580	2516280	nf	17	10	72	62	56	127	108	55
904	101670	2565600	nf	10	36	69	58	45	156	82	37
908	79950	2636250	nf	10	33	76	59	55	114	113	63
919	79470	2636070	nf	12	61	64	49	36	103	63	30
926	54690	2629590	nf	14	128	71	63	42	199	86	36
929	173280	2627220	for	12	459	63	48	34	128	80	34
932	119220	2625900	for	10	94	63	45	33	149	72	33
933	94830	2614830	nf	10	21	66	55	40	142	78	35
937	88560	2510670	nf	14	23	71	64	51	144	107	49
938	90060	2526240	nf	14	20	71	57	42	163	96	43
940	55440	2523570	nf	10	3612	67	50	32	19	11	9
941	87780	2546430	nf	14	10	73	61	55	143	98	52
956	79140	2600490	nf	10	71	70	59	41	157	95	41
957	193170	2525220	for	12	96	60	45	31	146	73	33
959	180720	2574150	nf	10	5000	62	40	29	18	13	12
963	148320	2462250	nf	17	283	65	52	35	197	81	33
966	121020	2469240	nf	10	188	65	52	36	198	81	32
984	28110	2577510	nf	15	41	77	63	51	154	108	53
987	134670	2497380	nf	16	61	69	56	39	172	90	40
989	123030	2532570	nf	15	54	70	60	50	156	102	50
999	120150	2592270	for	17	238	67	47	36	118	64	28
1000	144390	2596260	for	12	743	64	48	33	145	78	33

APPENDIX D
Virginia 15m Image Training Point Information

Projection

Projection type: Lambert Conformal Conic
Spheroid: GRS 80
Datum: NAD 83
1st standard parallel: 37 N
2nd standard parallel: 39.5 N
Central meridian: 79.5 W
Latitude of origin: 36 N
False easting: 0 m
False northing: 0 m

Points

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
1	193060	137390	for	14	388	wet forest	68	44	38	25	37	27
2	252880	144080	for	7	263	hw hazy	76	53	60	50	92	57
3	94030	90440	nf	23	2380	nf	80	65	68	80	116	66
5	114040	76970	nf	17	32	nf edge	75	56	57	56	69	44
7	260470	180680	for	11	200	hazy wet mixed forest	74	52	50	48	70	42
9	184570	242270	for	16	83	wet forest	67	46	42	46	62	38
11	167530	242570	nf	30	1183	nf with edge	87	72	80	83	122	70
12	84370	82400	for	12	60	mixed for	71	53	55	50	91	57
14	131410	119690	for	9	17	wet hw	71	52	56	50	98	59
15	189310	117440	for	17	1380	pine (some wet for)	70	48	42	54	55	30
17	117490	76160	for	13	93	pine mix	65	45	39	48	56	32
18	226390	107900	for	8	36	wet pine mix	74	51	47	49	70	40
19	94240	139910	nf	15	94	nf with edge	79	59	68	66	110	68
21	99400	67520	for	15	181	wet pine	69	48	45	53	67	38
22	103420	134780	nf	23	935	nf	73	57	55	93	94	50
23	153880	207740	for	10	266	hw	73	51	58	50	94	57
24	224680	224870	nf	21	548	nf	79	62	60	78	79	47
25	198070	168710	for	13	65	pine	72	49	46	46	61	38
27	111430	88820	nf	35	10	nf	101	101	125	84	165	130
28	160030	191570	nf	23	309	nf	79	69	73	85	109	70
31	123700	132320	nf	24	555	nf	75	58	53	86	96	51
32	153550	230750	nf	27	1404	nf	86	69	78	72	117	70
34	148420	214310	for	10	3146	hw	72	52	56	47	91	57
35	133780	153080	for	18	278	pine	65	45	35	58	49	25
36	211180	211280	nf	31	544	nf	87	68	74	59	132	87
37	167440	108320	for	12	24	pine	68	48	41	55	61	32
38	153220	64700	for	17	2835	pine	69	45	41	52	69	39
39	199300	199220	for	15	220	hw	76	55	63	54	113	69
41	242710	141410	for	10	18	hw in suburbs (hazy)	80	56	62	50	102	61
42	181150	180560	nf	17	11	nf	78	56	55	44	67	49
43	210190	165050	for	11	2271	wet forest	73	47	41	27	41	29

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
44	218110	224990	nf	17	62	nf	81	63	67	66	92	61
45	207070	153620	for	8	178	wet hw	74	50	51	47	81	46
46	214930	145970	nf	36	2598	nf	86	68	76	55	93	67
48	186730	202940	for	18	70	pine mix	67	49	42	48	63	38
49	109660	107630	nf	31	1323	nf	82	66	82	64	128	78
50	149680	237020	for	7	17	hw	72	51	55	48	86	54
51	117130	208130	for	12	105	hw	71	49	49	43	78	48
52	136660	65960	for	20	568	pine	70	45	44	45	63	36
53	202630	226430	nf	19	872	nf edge	84	62	71	57	117	73
54	245020	184430	for	10	5000	pine	72	51	46	57	58	30
55	173290	211580	nf	26	918	nf	83	65	69	81	106	65
56	135160	88100	for	9	39	hw regrowth	74	54	63	51	113	68
57	121060	185750	for	9	180	hw	73	49	52	44	85	53
60	199690	188210	for	6	60	hw	77	54	55	48	90	58
61	123760	143750	for	11	153	pine	64	45	41	51	62	35
64	152590	173420	nf	18	127	nf	81	63	65	67	108	67
65	177160	231830	nf	32	633	nf	83	69	81	72	125	78
66	112960	152510	nf	21	88	nf	79	62	66	62	120	77
67	129160	186500	for	10	366	hw	71	49	50	41	83	54
70	218470	175820	for	13	214	pine	70	48	39	56	47	24
71	130150	219200	nf	25	1026	nf	81	65	70	66	107	66
72	202600	221330	nf	11	25	nf edge	80	61	66	52	105	71
73	114340	151430	for	14	777	pine	66	47	40	59	59	31
74	181630	130070	for	8	35	hw	75	51	58	50	97	58
75	111130	187670	for	14	18	hw	68	48	46	45	77	49
76	243700	122960	for	12	5000	pine	70	51	42	55	53	29
77	150790	215750	nf	15	216	nf	80	63	71	67	120	72
79	238600	78320	for	12	22	wet hw	74	48	46	31	33	25
80	151660	150500	for	11	47	hw regrowth	83	66	71	52	121	88
81	189730	168800	nf	20	47	nf with edge	91	72	81	50	93	73
82	210340	161540	nf	41	773	nf with edge	92	77	92	67	125	89
84	274840	220820	nf	30	328	nf	87	68	73	62	109	72
85	123340	219380	nf	17	344	nf	78	56	60	64	103	61
86	270310	163070	nf	15	49	nf	76	54	52	62	82	49
87	197440	172310	for	7	203	pine mix	72	51	51	50	80	46
88	156610	202550	for	8	422	hw mix	71	50	47	49	84	52
89	130150	168620	for	11	29	pine	67	45	39	47	58	35
90	263380	134000	nf	28	46	nf	91	70	76	50	82	63
91	265750	93770	nf	28	770	nf	92	74	88	61	132	103
92	177160	100040	nf	48	176	nf	93	82	101	68	146	113
93	240820	131990	for	13	58	wet hw	75	52	50	44	78	48
94	170920	131630	for	13	503	pine	66	49	41	63	56	28
95	123970	207800	for	12	450	hw regrowth	78	59	68	49	112	71
96	197380	159590	for	8	1055	wet hw	73	49	51	42	82	51
97	186370	174230	nf	25	532	nf	75	59	52	83	84	45
99	175450	109010	for	8	475	pine	67	46	41	51	54	30

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
100	111550	228350	for	11	344	hw regrowth	80	58	65	44	116	85
102	190990	94850	for	9	1007	pine mix	68	49	46	52	74	42
103	117040	131270	for	9	104	hw	76	53	62	53	109	65
104	193480	192020	nf	29	1379	nf	82	65	68	68	108	70
105	150700	188390	for	10	35	pine mix	70	49	47	45	71	46
106	154060	126560	for	12	1111	pine	66	45	41	51	61	34
107	155620	112400	nf	29	66	nf with edge	86	71	89	68	135	83
108	154840	197180	for	19	399	pine mix	63	44	36	55	42	24
109	257200	166310	for	11	1660	pine some wet	72	50	44	56	57	32
111	119860	97940	nf	16	300	nf	75	55	59	63	99	55
112	82960	72560	for	11	56	hw	72	48	48	37	81	54
114	155080	234830	for	11	129	pine mix	69	45	40	41	62	40
116	99790	155660	nf	19	321	nf	78	63	71	70	118	73
118	247930	238430	for	14	388	wet pine hazy	76	52	45	51	48	28
120	138460	207650	for	11	372	hw mix hazy	71	50	48	43	76	48
121	188590	92930	for	8	166	pine mix hazy	71	49	45	46	68	41
122	138940	133670	for	9	86	hw mix	68	48	51	48	80	48
124	233590	99650	for	12	418	hw regrowth	81	58	65	47	119	78
125	191200	81920	for	12	988	pine	67	47	39	55	45	24
126	113950	213050	for	6	94	hw	76	53	61	51	103	64
128	162640	236240	for	18	226	pine old?	64	44	36	46	47	28
129	149920	233630	for	10	35	hw	73	52	57	48	98	59
130	169090	178310	nf	12	1707	suburbs	72	52	50	46	70	49
131	201820	218000	for	7	63	hw	74	52	54	52	86	51
134	195190	103040	for	10	222	pine	75	50	46	46	71	41
135	103330	147860	for	10	12	hw	71	49	50	51	94	58
136	176020	189200	for	9	20	hw	71	50	51	47	90	52
138	239410	230360	for	10	150	hw wet hazy	80	52	49	41	59	37
139	262420	112250	nf	25	803	nf	82	65	63	83	86	46
143	223210	196370	for	22	156	hw wet	72	47	41	32	40	29
144	252670	227870	for	8	350	pine mix hazy	77	55	50	47	68	42
146	151870	146180	for	16	2680	pine mix	65	45	37	58	48	26
147	243190	238970	nf	23	1161	nf with edge hazy	86	66	66	64	77	48
148	166630	203870	nf	28	45	nf	75	56	61	55	109	73
152	152500	66320	for	19	194	hw regrowth	77	59	72	63	124	73
154	268660	163970	nf	13	209	nf	78	58	56	62	88	56
155	282550	154790	nf	25	97	nf	82	66	69	71	110	66
156	133060	189920	for	16	126	pine	63	45	39	51	50	28
157	181660	223160	for	7	85	hw	74	52	58	50	94	57
160	92080	127460	for	13	136	hw regrowth	74	53	57	50	98	60
161	126700	117290	for	15	126	pine	68	47	42	51	64	36
162	126700	211310	for	7	320	hw	75	54	64	52	104	68
163	115000	86990	nf	21	1541	nf	79	64	66	71	115	69
164	112570	203510	for	13	409	hw mix wet	75	53	57	47	95	60
165	231160	97460	for	17	5000	wet forest	73	47	43	32	45	31
166	202720	64550	for	6	36	hw wet	74	52	50	40	81	53

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
167	224950	243080	nf	14	34	nf	86	66	77	61	115	72
169	119230	97130	nf	26	22	nf	86	77	85	87	117	76
170	113200	220850	for	9	1233	hw	74	52	61	49	102	63
171	197830	239000	for	8	60	hw wet	75	51	54	43	82	53
173	230890	236060	for	9	42	hw hazy	82	57	59	48	84	53
175	108370	73190	nf	19	134	nf	81	64	71	64	117	76
176	120490	76520	nf	23	995	nf	81	63	69	66	118	76
177	151990	109460	for	10	104	hw wet	67	46	44	44	80	48
178	144820	88610	for	7	12	pine	66	45	41	48	56	32
179	234670	93320	for	8	526	hw wet	73	49	43	44	64	38
181	130870	134660	for	7	572	pine	66	47	39	55	61	33
182	153400	212210	for	6	53	hw regrowth	75	56	60	46	112	72
183	141340	195530	for	6	36	hw	72	50	54	48	93	59
185	273940	172610	for	7	42	hw regrowth	80	55	61	50	93	61
186	202210	101810	for	8	448	pine	72	49	44	51	57	34
187	236500	235730	nf	19	1058	nf	89	65	66	63	85	52
188	218740	229100	nf	8	103	water	69	43	35	17	17	12
192	170440	113780	for	6	16	pine near nf	69	48	44	50	75	44
194	204820	229670	nf	17	85	nf	81	62	69	56	105	70
195	246490	138530	nf	24	420	nf	83	65	69	68	99	56
196	145420	66620	for	14	2354	pine	65	45	38	54	52	29
197	139120	132950	for	9	154	hw	71	50	54	48	94	58
200	231010	174920	nf	23	1174	nf	84	64	66	64	114	69
201	189820	183560	nf	18	1718	nf	83	65	65	65	85	62
203	184600	170540	nf	20	717	nf	93	74	82	53	96	73
204	105010	93290	for	7	60	hw	72	51	59	49	99	63
205	142480	194870	for	6	574	hw	71	49	53	45	89	56
208	252940	104840	nf	28	411	nf	90	73	86	66	133	87
210	202510	206090	for	8	53	hw	73	50	51	44	74	48
212	122980	165680	nf	21	58	nf	84	71	82	71	132	84
214	166120	126020	nf	18	264	water	62	41	34	26	23	17
217	147940	144920	nf	8	68	nf edge	65	45	45	49	71	42
218	261460	227960	nf	25	2826	nf	85	64	69	45	88	66
220	229030	122450	for	12	404	hw mix	74	52	52	44	71	46
221	118660	135590	for	13	824	pine	66	45	39	53	50	28
223	141790	90530	for	9	84	hw	72	50	55	51	89	55
224	250810	206450	nf	20	190	nf	90	69	73	56	113	76
225	187180	203030	nf	14	151	nf	78	58	62	61	98	64
230	208750	121820	nf	15	54	nf	77	57	57	41	77	58
231	211600	121910	for	11	3544	pine	70	51	45	51	67	36
232	160960	168200	nf	8	21	nf edge	73	53	53	49	87	57
233	123670	193730	for	10	157	hw	70	47	49	37	75	51
234	101740	133850	for	9	42	hw mix	70	49	51	50	90	55
235	191140	150620	for	7	77	hw regrowth	82	62	75	55	118	75
236	194020	82400	for	13	1420	wet hw	72	48	45	36	58	40
239	210700	218690	for	11	1100	pine	68	49	38	58	45	23

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
240	171430	184430	nf	15	3321	suburbs	83	64	63	65	74	49
242	192460	87530	for	11	124	pine mix	70	44	38	53	47	23
243	154600	181190	nf	29	990	nf	75	57	63	56	109	65
246	213550	65000	nf	39	624	nf	84	63	65	51	88	69
247	186700	64790	for	15	180	pine	67	44	36	51	40	22
249	143500	222620	for	9	176	hw	74	51	58	49	97	59
250	276730	208820	for	7	17	pine mix wet hazy	74	50	46	47	62	36
251	158500	227210	nf	33	151	nf	76	61	72	56	106	70
252	174130	92780	for	16	165	hw mix	69	47	41	60	59	31
254	246160	174740	for	10	215	hw mix	76	52	50	48	76	45
255	241060	233510	for	8	75	hw mix hazy	86	59	63	52	91	58
256	180610	107420	for	9	91	hw	73	49	52	46	87	53
257	228520	186200	for	10	197	hw mix	73	51	50	46	68	39
259	115630	195530	for	10	43	hw mix	70	49	50	49	76	47
261	122860	183110	for	10	69	pine (old?)	65	44	42	43	56	34
262	206470	76910	for	13	123	hw	77	57	56	49	85	52
263	228550	215060	nf	13	41	nf	78	55	58	55	88	49
264	231040	229130	nf	21	39	nf	84	66	71	47	78	63
265	181480	77120	for	6	73	wet hw	70	50	50	43	88	54
266	138610	226700	for	13	513	hw	74	51	58	47	95	58
267	248950	114230	for	13	1465	pine mix	75	50	49	50	69	39
270	195910	206870	for	11	143	wet hw	70	49	45	34	54	40
271	151660	77390	nf	17	14	nf	76	59	58	63	81	50
272	116140	182420	for	7	172	hw	69	46	45	42	73	45
273	221290	106040	nf	32	689	nf	81	67	69	74	90	54
274	196690	101660	for	9	1445	mixed forest	71	50	52	51	77	44
275	268990	123800	nf	18	21	nf	95	68	69	37	56	51
276	189820	159830	for	9	115	hw regrowth	79	59	70	54	114	70
277	143470	162680	nf	21	10	nf	86	71	90	67	132	84
278	134080	213410	for	5	169	hw	74	53	59	49	104	67
279	164080	240590	for	6	163	hw	73	49	55	47	90	55

APPENDIX E
Virginia 16m Image Training Point Information

Projection: same as for 15m image, see Appendix D

Points

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
1	-13011	220635	nf	17	26	water	63	44	43	43	40	30
2	-20301	180015	for	16	796	forest shadow	60	39	38	47	58	39
3	82929	165495	for	20	1523	pine	58	38	40	58	49	31
4	106689	187395	for	10	889	hw mix	63	43	49	60	66	43
5	100839	101925	nf	21	432	field	77	67	85	104	110	70
6	18609	107205	for	19	73	pine mix	56	39	33	70	45	27
7	146109	201315	nf	23	2742	nf	70	58	64	106	100	60
8	77769	236655	for	17	121	hw	64	49	57	71	97	65
9	-6891	199065	for	20	2498	pine mix shadow	61	43	41	69	63	41
10	60279	175275	for	15	82	hw mix	61	41	45	65	69	44
11	95139	128355	for	11	336	hw	61	42	44	59	63	41
12	99639	214065	for	10	937	hw	62	42	46	61	73	47
13	4629	124335	for	20	935	pine	57	39	35	75	51	29
14	39699	219555	for	15	5000	mixed	56	39	40	62	61	37
16	92469	201795	nf	21	1481	nf	68	56	61	104	100	60
17	143259	149295	for	13	5000	hw mix	59	40	43	66	70	43
18	138759	145935	for	15	194	hw mix	60	40	40	61	55	32
19	-41451	144705	nf	17	5000	urban	75	55	60	66	76	55
21	110139	117405	nf	16	16	nf edge	73	58	61	91	79	51
22	10179	175305	for	17	135	hw mix	59	41	44	63	69	43
23	38499	213585	for	21	5000	pine mix shadow	51	33	28	40	31	20
24	13239	211575	for	18	945	hw shadow	63	46	46	68	66	42
25	120099	201465	for	13	669	hw mix	61	45	49	71	83	51
26	-20871	241605	nf	18	119	bare soil	75	63	75	98	97	61
27	131229	198045	for	24	4952	pine	55	39	35	84	44	25
28	156459	244095	for	17	5000	pine mix	58	41	34	71	42	23
29	12339	189075	for	12	487	hw shadow	58	39	40	51	57	37
31	29379	174225	for	13	825	hw	61	44	47	58	76	51
32	48039	133425	for	17	608	hw mix	62	43	43	54	60	40
34	-58581	85335	for	14	367	hw	63	40	45	56	73	47
35	18249	187275	for	15	627	pine mix shadow	55	36	32	60	44	27
36	-48621	126525	for	17	104	hw shadow	62	42	41	69	67	42
37	150609	219105	nf	31	264	nf	81	70	94	97	110	72
38	13059	231195	for	12	35	hw shadow	60	41	42	51	57	40
39	-12021	184905	for	12	5000	hw mix	60	40	37	55	59	37
40	-28371	168975	for	15	15	hw	62	44	47	60	72	47
41	46299	165675	for	13	362	hw	60	43	46	69	76	47
42	31539	77655	for	15	95	hw	68	50	57	77	106	69
43	86349	175125	for	10	28	hw mix	59	42	45	67	73	43
44	106719	209235	for	16	1531	pine mix	56	42	37	74	48	29

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
45	-14871	62235	for	8	24	hw	63	43	51	64	81	54
47	159459	205755	nf	26	529	nf	71	60	68	89	87	54
48	126789	166485	for	14	321	hw mix	59	40	43	57	61	39
49	71229	102465	for	13	892	hw	65	47	56	74	93	60
51	84129	163905	for	12	131	hw	62	42	47	67	79	52
52	104469	120645	for	13	5000	hw	63	48	51	64	79	52
54	-22371	97755	for	14	1782	hw mix	64	45	46	69	85	53
55	142269	156465	for	12	5000	wet hw	59	41	45	53	67	44
56	12819	213135	nf	16	1227	nf	68	56	61	96	102	66
57	135189	139245	nf	20	504	nf	65	52	55	92	73	48
58	67569	82125	for	12	2083	hw	62	45	50	73	91	58
59	-34581	82455	for	13	248	hw	64	45	49	69	83	50
61	-15021	217485	for	14	5000	pine mix shadow	56	39	37	60	55	34
62	166029	236445	nf	26	894	nf	66	51	57	85	88	52
63	95769	120705	for	9	15	hw	59	44	50	69	76	48
64	70119	114885	for	13	755	hw mix	59	40	41	58	58	35
65	62229	78195	for	17	311	hw	61	42	44	58	79	50
66	-19221	225645	for	12	848	hw	62	45	54	71	94	59
67	131709	214485	nf	14	25	nf	67	52	56	94	86	57
68	129549	200625	for	20	1139	pine	57	39	36	74	55	30
70	8799	128235	nf	19	100	nf	68	55	75	85	109	76
71	33429	211875	for	18	5000	pine shadow	52	33	29	45	33	21
72	27309	241005	nf	20	5000	nf	71	57	65	109	124	77
73	73539	221895	for	23	353	hw	58	38	42	52	66	45
74	-20601	129225	for	12	18	wet hw	60	41	40	56	56	38
75	67149	156135	for	20	159	hw	64	45	59	80	91	57
76	-44811	61935	for	17	215	hw shadow	62	40	43	51	61	43
77	55209	121845	for	13	269	hw	61	44	53	67	87	55
78	70299	95895	for	11	447	hw	60	43	43	66	68	43
79	87879	85725	for	19	174	pine	58	40	34	62	48	30
80	52839	62775	for	14	2712	hw mix	59	42	43	60	65	44
81	-53931	82245	for	11	19	hw	64	45	53	68	90	59
82	108969	236385	for	13	120	hw	64	48	59	77	95	63
83	19779	80535	for	10	17	hw	62	48	52	76	98	62
84	-12741	115485	for	11	182	wet pine	61	41	37	63	57	34
85	57939	169365	for	13	137	pine mix	62	41	43	61	62	41
86	129759	227205	for	17	89	hw	60	44	52	76	85	54
87	121449	93615	for	19	151	hw	64	45	47	64	83	52
88	94029	106275	nf	13	38	nf	68	54	55	81	72	47
89	81759	72525	nf	10	5000	water	65	52	43	17	10	11
90	165039	235245	for	15	36	hw	63	44	51	64	73	49
91	50919	185865	for	13	383	hw mix	58	39	39	56	65	41
92	95289	146085	for	17	60	hw mix	60	41	45	64	74	47
93	-7821	62085	for	14	2127	hw clear cut	64	44	51	64	91	59
94	64239	229125	for	13	1350	hw	64	49	59	78	113	74
95	-6681	135225	nf	24	321	nf	76	62	75	101	111	72
96	-14841	190665	for	16	553	hw mix	61	45	50	82	87	51

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
97	90999	200895	nf	19	93	nf	70	58	64	114	103	62
98	88569	90915	for	21	1254	pine	54	37	30	61	36	21
99	-20661	207585	nf	21	158	nf	75	57	65	95	97	62
100	133809	172695	for	11	4396	hw wet	60	41	41	55	58	41
101	-43701	69375	for	12	30	hw wet	60	40	40	52	59	39
102	153009	204135	for	13	1009	pine mix	58	39	39	62	63	37
104	62979	178215	for	16	601	pine	56	41	31	89	43	23
105	84009	187395	for	16	587	hw mix	62	41	45	59	70	47
106	43359	209235	for	14	5000	hw shadow	55	37	36	47	51	33
107	67749	225225	for	15	145	hw	61	42	46	61	76	53
108	142269	162405	for	11	408	hw wet	61	40	41	57	63	40
109	-9411	158925	for	20	488	hw summit	71	56	70	95	134	92
110	41829	229095	nf	18	5000	nf	69	53	58	95	96	61
111	-19641	153015	nf	30	1579	nf	77	64	77	103	115	73
112	-24831	134115	nf	10	2189	water	67	51	41	17	14	9
113	19089	141195	for	16	757	hw mix	63	43	47	62	72	46
114	57819	175785	for	16	1076	pine	56	40	32	63	38	23
115	4299	131505	nf	15	93	nf	71	57	63	88	94	59
116	-57891	72855	for	18	269	hw	62	43	49	74	77	50
117	-7611	181695	for	18	675	hw shadow	61	40	41	51	62	42
118	115839	158505	for	13	1842	hw regrowth	65	46	49	57	82	54
119	77319	183945	for	10	769	hw	62	44	50	64	74	50
120	77859	202815	for	12	174	hw mix	59	42	43	68	66	40
121	115749	211395	for	10	175	hw mix	58	42	47	73	82	52
122	65919	204045	nf	18	420	nf	74	61	63	105	102	70
123	146769	204615	for	15	248	hw regrowth	66	48	53	76	85	52
126	-44211	76125	for	15	47	hw	67	47	52	61	89	57
127	11919	227055	for	19	98	hw sunlit	65	48	65	90	113	70
128	22389	119265	nf	13	50	nf	71	59	66	92	94	65
129	-30171	163275	nf	24	5000	nf	77	66	80	106	113	72
130	-6171	175995	for	20	934	hw shadow	62	42	46	60	69	46
132	-31341	109485	nf	15	23	nf	75	62	71	103	94	61
134	89889	200925	for	13	376	mixed	57	41	38	69	61	38
135	94089	229545	nf	15	22	nf	67	49	55	90	90	56
136	75669	145605	nf	21	1106	nf	68	56	61	98	94	56
137	3699	195885	nf	20	435	nf	66	53	53	112	83	44
138	39549	217035	for	14	1082	mixed	61	47	50	95	92	53
139	134919	127425	for	12	788	hw wet	59	43	41	70	67	42
140	97059	220305	for	13	143	hw	64	46	53	76	88	56
141	339	157395	nf	25	397	nf	70	59	69	103	127	81
142	-46191	103215	nf	15	44	nf	71	55	61	93	87	55
143	-13101	217215	for	30	100	hw	66	49	60	85	115	74
144	79899	208215	nf	27	1982	nf	74	61	74	94	116	77
145	101049	131445	for	22	114	pine	56	38	33	76	41	21
147	36189	157245	for	22	50	hw regrowth	69	51	63	71	116	81
148	4419	199005	nf	21	37	nf	71	54	52	95	75	47
149	80109	211365	nf	15	16	nf	70	57	70	94	115	74

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
150	104889	176925	for	12	532	hw regrowth	62	44	45	57	89	59
151	-3891	161745	for	15	14	hw	65	45	52	63	82	54
152	-5211	68835	nf	22	9	nf	70	57	60	99	96	58
153	119799	163065	for	9	126	pine	58	41	38	72	49	29
154	-39921	81525	for	15	76	hw wet	61	42	44	66	86	54
155	93639	243675	for	10	49	hw wet	59	40	37	48	53	37
156	107859	219255	for	19	299	pine shadow	57	37	36	64	50	32
157	26679	219765	nf	15	269	nf	66	51	53	87	83	50
158	43209	195345	for	9	33	hw mix	60	40	41	52	58	38
159	28929	235635	nf	28	5000	nf	79	68	80	113	129	83
161	4599	218565	for	13	5000	hw mix	62	43	43	58	71	50
162	38979	198645	for	21	447	hw	66	49	58	79	103	69
165	81789	123165	for	11	310	hw	65	50	63	82	105	69
166	42639	222705	nf	23	5000	nf	67	56	62	110	104	64
167	152799	198795	for	11	313	hw mix	64	45	50	67	75	49
168	86379	241245	for	15	300	hw	62	44	54	68	86	54
170	-1071	197295	nf	28	1731	nf	75	65	75	99	106	68
171	143439	232245	for	16	1217	hw regrowth	64	49	55	69	93	64
172	11529	240285	for	12	5000	pine mix shadow	56	40	40	60	57	35
173	15069	127485	for	9	144	hw	62	42	43	65	73	46
174	31449	116295	for	14	127	hw regrowth	62	43	47	65	76	51
176	-16791	120495	for	10	11	hw	63	44	45	58	79	52
177	17019	239775	for	17	5000	hw mix sunlit	65	48	57	90	109	67
178	-26091	112575	for	10	97	hw mix	62	41	47	61	74	48
179	107409	108315	for	14	218	hw wet	58	39	38	54	54	32
180	67989	62055	nf	25	628	nf	63	49	51	98	83	50
181	1149	169605	for	16	63	hw	63	47	54	73	94	64
182	17769	134505	nf	17	100	nf	76	61	70	83	94	66
183	79779	150375	nf	28	49	nf	71	57	65	95	88	45
184	-31611	132765	for	24	883	hw shadow	54	32	26	28	31	21
185	58689	117825	for	14	193	hw regrowth	67	50	59	63	90	63
186	124209	156345	for	10	423	hw	62	42	45	66	72	44
187	-1521	83805	for	13	50	pine	61	43	42	68	61	39
188	-15531	111645	for	16	220	hw regrowth	68	47	55	69	102	66
190	-46311	100245	nf	12	11	nf	78	63	81	92	110	75
194	54279	207465	for	10	312	hw	63	45	53	66	86	60
195	68409	244215	for	10	137	hw	61	44	49	64	90	60
196	97539	161805	nf	22	64	nf	70	57	64	101	95	59
197	53199	73695	for	9	473	mixed wet	62	42	42	60	71	44
198	79569	211335	for	14	179	pine	61	41	40	63	65	42
200	-17511	71775	nf	22	913	nf	71	58	64	99	97	60
203	-38721	81795	for	10	27	hw	64	45	50	68	92	59
204	-13431	169395	for	21	856	hw shadow	58	36	32	34	37	25
205	58719	155865	for	8	71	hw	63	45	49	66	78	52
206	98799	220185	nf	23	44	nf	73	61	76	87	105	64
207	126819	101985	for	12	462	hw mix	61	43	47	64	72	45
208	58359	80535	nf	25	861	nf	69	53	58	91	91	56

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
209	71679	136485	for	11	308	hw mix	59	39	39	61	58	35
213	78129	244125	for	12	391	hw	63	42	48	60	83	58
214	10629	82995	for	18	19	pine mix	62	42	43	64	56	38
216	155409	230055	nf	16	84	nf	71	57	63	88	80	52
217	-24261	87015	for	16	5000	hw shadow	60	43	48	64	70	45
218	129339	228765	for	13	153	pine mix	59	42	44	67	63	37
219	111669	187875	for	11	1689	hw mix	61	40	40	54	62	38
220	-4431	238875	for	23	5000	hw shadow	54	34	30	36	32	22
221	119469	112845	for	8	239	hw wet	59	39	41	56	62	40
222	7089	235395	for	13	5000	pine shadow	56	36	32	52	46	30
223	-14901	134385	nf	17	13	nf	69	54	61	88	107	65
224	117159	200415	for	10	746	hw mix	59	41	47	60	70	43
225	78429	136725	for	11	31	hw wet	58	38	37	66	54	34
226	159399	219945	for	13	168	hw	64	44	52	69	79	49
227	58899	225525	nf	19	46	nf	68	55	57	108	94	59
228	120609	62295	for	15	1731	pine wet	56	38	32	68	44	25
231	91449	158325	for	11	14	hw	63	44	49	70	88	56
232	94029	62235	for	15	231	pine mix	56	36	37	69	51	32
233	61029	213825	for	13	325	hw mix	60	41	42	57	73	48

APPENDIX F
Virginia 17m Image Training Point Information

Projection: same as for 15m image, see Appendix D

Points

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
1	-11632	259670	for	12	62	hw shadow	63	44	41	34	45	31
2	-6862	135710	for	12	28	hw	76	55	65	52	105	67
3	-86722	74240	for	9	60	pine mix	65	46	44	49	73	44
4	-59572	130160	for	11	32	hw sunlit	71	54	69	60	136	87
6	-90142	93380	for	11	503	pine	61	42	38	45	59	34
7	-82882	97670	nf	21	83	nf	72	57	68	67	120	72
8	-26812	105020	for	8	32	hw (in field)	71	48	48	40	70	46
9	-56632	112550	nf	19	276	nf	77	60	67	58	103	68
10	-6052	237650	for	7	96	hw mix shadow	65	43	45	36	72	46
11	-192022	93800	nf	20	2483	nf	77	61	64	79	119	71
12	-38152	117890	nf	23	771	nf	76	61	67	71	111	69
13	-58222	172100	for	8	102	hw mix	65	47	49	47	80	51
14	-143002	100850	for	12	147	hw shadow	62	43	43	34	63	43
16	-113992	123170	nf	17	5000	nf	83	67	72	74	98	68
18	17918	186350	for	17	284	hw	70	53	62	51	115	77
19	-65752	146750	nf	24	262	nf	77	61	69	66	124	79
20	-53932	184940	nf	26	867	nf	76	61	63	85	105	59
22	-3442	186920	nf	22	158	nf	76	56	62	52	95	61
23	-127282	73820	nf	15	45	nf	77	58	58	68	90	55
24	-102952	153800	for	13	477	hw shadow	70	49	51	38	77	56
25	-8542	90980	for	18	34	pine	65	47	41	60	53	30
26	-57862	134030	for	9	148	forest shadow	64	39	36	33	50	32
27	-135472	143390	for	16	36	hw mix edge	68	48	55	51	98	58
28	-73012	171230	nf	17	75	nf	75	58	61	74	105	63
29	-20002	203480	for	16	259	forest shadow	63	41	35	33	48	28
30	-160252	82490	for	11	576	hw shadow	66	43	43	33	60	42
31	18338	257570	for	8	181	pine shadow	62	42	42	47	69	40
32	-36622	210020	for	9	97	hw	69	51	54	48	93	61
33	-35122	214130	for	16	135	hw (near shadow)	68	50	57	49	100	65
37	-143932	121880	nf	19	264	nf	75	61	69	63	116	72
39	-67522	69890	nf	29	81	nf	79	67	81	72	139	88
40	-78772	102620	nf	20	542	nf	73	61	63	86	113	67
41	-95512	145550	for	7	30	hw mix	68	47	48	41	79	52
42	45188	289490	for	14	81	pine mix shadow	62	43	43	45	78	49
43	26978	215420	nf	14	64	nf	75	60	71	58	110	73
44	22388	260420	for	10	1366	pine	60	40	33	47	53	28
45	-106462	141800	nf	16	164	nf	75	60	66	64	104	64
46	-172132	91490	for	18	152	hw mix sunlit	70	53	64	60	119	77
48	17108	252320	nf	18	68	nf	71	58	67	86	119	68
50	-59752	196340	for	9	189	hw sunlit	70	51	60	53	112	74

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
51	-183082	131360	for	21	327	forest shadow	61	36	33	20	28	20
53	-83872	105020	for	20	96	pine	67	46	43	44	62	38
54	34088	282950	for	25	5000	pine shadow	59	35	30	14	33	23
57	-65812	191030	for	18	39	hw mix	69	49	57	52	105	64
58	-183772	78710	for	21	168	hw mix	69	51	61	54	113	73
59	4958	139160	nf	40	1797	nf	67	50	39	121	66	27
60	-126112	144020	for	13	279	hw	69	52	56	51	101	66
61	-39712	146720	nf	16	4168	nf	85	64	65	57	79	54
63	-178582	88910	for	20	2492	hw mix sunlit	72	54	69	59	134	86
64	-126232	139100	for	15	5000	pine	62	43	36	52	57	34
65	-46432	126440	nf	15	323	nf	77	64	75	73	120	73
66	-91492	103430	nf	14	180	nf	79	64	73	68	103	65
67	-180682	77300	for	16	5000	mixed for shadow	62	41	40	33	58	38
70	-45712	174890	for	10	20	pine shadow	64	43	40	41	53	33
71	4868	192230	for	10	18	hw shadow	65	46	45	41	65	42
72	-172	277370	for	12	110	hw	70	53	62	53	115	71
73	-186232	106160	for	10	73	hw shadow	68	45	42	34	61	43
78	-12172	179330	for	8	207	hw	71	54	64	53	117	75
79	25628	256370	for	8	87	hw	68	48	55	49	103	63
80	24788	215300	nf	21	4045	nf	81	68	77	85	118	70
82	-17602	265730	for	10	159	hw mix	64	48	54	43	90	59
83	-142192	65030	nf	23	235	nf	79	69	76	90	127	75
84	-123352	117410	nf	9	72	nf (houses in forest)	65	46	43	46	74	46
85	-123532	113030	for	11	17	pine	62	43	38	43	51	30
86	-18862	261620	for	17	115	hw shadow	66	47	50	39	84	57
87	41378	308960	for	12	523	forest shadow	60	38	35	29	53	35
88	-50392	147860	for	14	227	hw (burned?)	68	46	50	37	92	73
91	5888	225800	for	8	478	forest shadow	63	40	35	32	44	28
92	-88972	82400	for	12	263	pine shadow	59	41	33	45	46	29
93	-158842	130610	for	18	312	hw	67	47	52	58	90	57
94	-167872	90770	for	19	96	hw	70	56	70	60	133	88
96	15008	227390	for	7	158	hw shadow	65	46	48	39	76	49
98	-33772	196280	for	12	128	hw shadow	64	41	41	30	53	34
101	-141982	111350	for	17	273	hw shadow	60	38	33	26	35	23
102	-7732	218450	for	8	610	hw mix	65	45	49	45	83	51
104	-26752	144050	for	7	38	hw	76	55	68	58	120	76
105	-25882	167510	nf	18	1145	nf	68	48	46	46	70	42
108	-17482	189950	for	10	49	pine mix	66	48	51	50	94	57
109	-170272	117830	for	8	672	hw	73	55	68	60	131	86
111	21038	239900	for	8	55	hw shadow	63	40	42	34	63	39
113	34778	254660	for	20	576	forest partial shadow	80	64	72	64	120	76
114	-35212	219260	for	9	182	pine shadow	64	43	39	33	54	36
115	26558	266510	for	22	729	pine shadow	60	40	35	35	48	30
116	-152092	115010	nf	11	18	nf	78	61	71	65	119	74
117	2228	272240	for	26	148	pine mix	61	41	36	43	52	30
118	-41392	134090	for	11	89	hw	72	53	62	54	110	70
119	-33802	236240	for	21	558	hw mix shadow	62	40	40	31	59	42

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
120	-170752	103340	for	14	482	hw mix sunlit	71	51	59	61	108	66
121	-12802	243830	for	11	1446	pine mix sunlit	66	48	47	50	92	55
122	-6532	196580	nf	15	195	nf	80	66	75	72	118	71
123	-11002	210590	for	11	202	hw	65	48	55	45	99	65
124	-13282	208190	for	16	5000	hw mix shadow	61	39	39	31	47	31
125	-132292	111200	for	11	23	hw	72	54	61	55	121	76
127	-53152	206270	for	31	265	pine mix shadow	63	40	37	38	52	33
129	-58552	190520	for	12	5000	pine mix shadow	63	44	48	41	75	49
130	-3232	275210	for	13	474	hw	73	54	64	49	121	82
133	-7912	149180	for	13	26	hw shadow	68	47	50	38	78	53
135	-62242	180980	for	14	32	pine mix	69	49	56	53	103	63
137	1418	272360	for	15	4289	hw mix shadow	66	45	48	40	76	51
138	-66832	167060	for	27	41	pine mix	64	45	46	46	67	42
139	-139612	136670	for	22	99	pine shadow	65	44	44	45	63	41
141	-136522	115070	nf	21	2560	nf	78	65	72	68	125	80
142	45578	310160	nf	11	11	nf	68	50	54	59	93	55
144	46118	293900	for	10	89	hw shadow	62	40	40	31	55	37
145	-96832	132080	nf	19	5000	water	68	42	31	16	15	12
148	-176392	74450	for	17	2171	hw	67	48	54	44	91	63
150	-121222	108740	for	18	77	hw shadow	65	45	45	36	63	43
151	-35962	61040	for	8	26	hw shadow	67	47	46	37	70	45
152	-103912	120710	nf	16	220	nf	76	57	62	51	96	65
153	-104242	147980	for	6	20	hw wet	68	46	46	41	67	45
155	-106192	106010	for	15	506	pine shadow	65	44	41	38	58	37
156	-69802	76700	for	11	25	pine	65	46	41	50	67	38
157	-56272	141980	nf	50	61	nf	86	87	126	70	148	112
158	-35662	177980	for	15	441	forest shadow	59	38	32	30	34	21
159	-4972	228440	for	24	25	pine mix	64	44	45	41	75	46
161	-1702	199790	nf	26	1203	nf	84	71	85	74	134	83
163	-65932	163130	for	23	23	hw shadow	62	42	37	36	50	30
165	-124522	117590	for	33	40	hw	74	54	67	60	127	83
167	-42022	182930	for	9	300	hw mix	69	49	54	50	94	59
168	-54982	176600	for	16	326	hw	73	55	70	62	136	86
170	-9952	238970	for	11	851	hw	71	50	58	48	108	70
172	-37642	163250	nf	20	662	nf	81	66	78	62	119	75
174	-154312	73970	for	11	186	hw	73	54	64	53	117	77
176	19718	259880	for	13	3515	hw mix	65	46	49	49	86	53
177	-193972	89870	for	14	46	hw shadow	67	46	47	33	74	50
178	-3232	173540	for	9	82	forest mix shadow	63	44	43	43	72	45
179	-22822	218120	for	12	1196	hw	70	49	58	49	111	70
180	42608	310160	for	17	248	forest shadow	59	38	34	30	46	32
183	-24622	83000	for	11	200	pine edge	70	49	48	47	77	46
185	11918	169430	for	9	32	forest shadow	66	43	42	33	59	39
186	-34972	142490	nf	13	143	nf suburbs	74	56	55	57	74	50
187	-33982	201950	for	12	1806	hw	70	50	59	53	113	68
189	18998	248120	for	14	5000	forest shadow	61	41	39	32	56	36
190	-15802	224630	for	9	32	forest shadow	65	43	40	37	57	36

Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
191	-6112	160670	for	11	158	hw mix	67	51	55	56	94	58
194	-44002	113060	for	10	98	mixed forest	68	47	46	40	76	49
195	-2632	142430	nf	17	71	nf	78	64	71	76	105	65
196	10208	204710	for	9	72	hw	68	48	49	43	79	52
197	-9232	133310	nf	9	13	nf edge	75	56	57	53	76	48
198	-113872	97820	nf	23	28	nf some water edge	71	55	58	61	90	56
199	-124462	83480	for	14	23	pine edge	63	45	44	48	73	47
200	-191632	90500	nf	12	98	nf	79	62	69	61	117	75
201	-16642	218030	for	12	18	pine	64	45	41	46	61	36
202	-2992	244460	for	7	40	forest shadow	63	43	41	37	64	40
203	-156622	99920	nf	17	5000	nf	79	66	73	73	123	78
205	-155092	135320	for	11	1791	hw sunlit	73	56	69	60	136	92
206	-5062	196430	nf	22	17	nf (edge of pine)	73	55	54	76	84	47
207	-13402	274760	for	10	449	hw hazy?	66	43	44	36	68	44
209	33188	240740	for	7	19	hw edge	69	51	59	47	86	53
211	-31912	95960	for	9	159	hw shadow	68	46	46	38	70	46
212	-123442	94280	for	6	19	hw shadow	66	43	45	38	73	46
213	-103372	76310	nf	22	19	nf	81	67	82	66	131	84
214	-24382	105170	nf	23	3524	nf	79	63	69	68	114	66
215	-103042	68360	for	7	32	pine	64	44	39	41	60	35
216	18338	196940	for	20	353	forest shadow	63	40	39	38	53	34
217	37028	302630	for	31	366	pine mix shadow	60	40	38	37	54	35
219	-49102	152960	for	11	91	hw	72	51	58	54	110	69
220	4688	190100	nf	23	579	nf	76	62	64	83	99	56
221	-68992	137930	for	12	151	mixed forest hazy?	73	51	51	41	70	47
222	-136972	139490	nf	20	97	nf	77	62	69	70	119	74
223	-109162	73910	for	18	15	pine shadow	62	40	35	28	32	23
225	-84142	153440	for	9	71	hw shadow	64	41	38	33	54	35
227	-112792	84020	for	10	22	forest shadow	63	43	41	42	67	41
228	-143602	99710	for	12	151	hw mix sunlit	71	56	66	63	132	81
229	-22702	64970	for	11	37	hw regrowth	72	52	59	46	109	71
230	6848	234740	for	11	552	pine shadow	64	40	37	37	53	34
232	-50932	118880	for	11	85	pine mix	70	50	53	51	89	56
233	-98752	83600	for	15	256	pine shadow	62	40	36	43	56	35
235	-36352	184940	nf	13	18	water	63	42	40	23	33	25
236	8318	236300	for	11	14	forest shadow	62	40	39	32	53	34
238	-60742	128450	for	25	1418	hw mix	69	52	55	57	103	64
239	-55312	202070	for	15	13	hw	67	50	58	47	99	67
240	-138142	85070	for	19	132	pine mix	65	50	52	63	97	58
241	29198	232460	nf	18	5000	nf	76	61	66	65	113	70
242	-8302	216770	for	5	362	hw	63	42	48	38	76	48
243	6128	184280	for	13	32	pine mix	65	46	43	45	69	44
245	-96142	81800	for	14	116	pine	59	42	39	40	51	32
246	-113242	130250	nf	24	396	nf	81	68	77	70	118	74
247	-57682	145430	for	13	211	hw mix	66	48	51	53	90	53
249	-16762	193160	for	13	981	hw shadow	62	40	38	34	49	32
252	-97432	155450	for	13	95	hw shadow	58	37	34	27	37	26

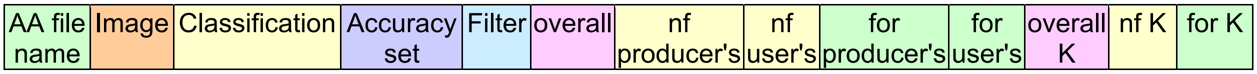
Point	X	Y	LU	EuD	Count	Comments	B1	B2	B3	B4	B5	B6
254	19088	186140	nf	9	36	nf in forest area	67	48	49	48	80	52
255	-102352	150050	nf	15	40	nf	75	55	61	51	94	63
256	45368	292070	for	18	497	hw shadow	61	41	42	38	66	44
257	11798	257390	for	16	58	pine shadow	62	42	40	43	65	39
258	-145552	78320	for	20	33	forest mix shadow	65	43	43	37	67	43
259	-5902	269270	nf	16	300	nf	77	64	75	76	131	83
262	-31822	81560	for	13	842	pine	64	43	37	50	43	26
264	-143542	65060	nf	14	33	nf	70	54	59	68	114	68
266	46958	304310	for	28	1648	pine shadow	58	37	31	29	37	23
267	-93682	157580	for	11	1059	hw	64	45	45	45	73	46
268	30518	229160	nf	19	721	nf	77	66	74	81	118	72
270	47918	299630	for	21	181	pine shadow	60	40	39	39	63	40
273	21338	227630	nf	16	5000	nf	81	67	77	73	118	80
274	-52882	145070	for	15	155	hw shadow	64	41	38	27	42	29
275	-30742	164030	nf	24	1603	nf	77	63	70	77	126	74
276	-59272	200540	for	21	1467	hw	70	50	59	49	115	75
277	-32002	219320	for	10	881	hw shadow	60	40	35	33	45	29
280	35888	266930	nf	26	68	nf	75	64	74	76	117	72
282	-47962	116210	nf	27	448	nf bright pink field	66	50	39	99	66	32

APPENDIX G
Macro to Extract Accuracy Assessment Statistics

This macro extracts the basic statistics from a text file created using the ERDAS Imagine Accuracy Assessment module. This macro was designed for use with text files that contain accuracy assessment information for a binomial classification (in this case forest and nonforest) on a 2-bit image. Accuracy assessment text files for larger images cannot be extracted using this macro.

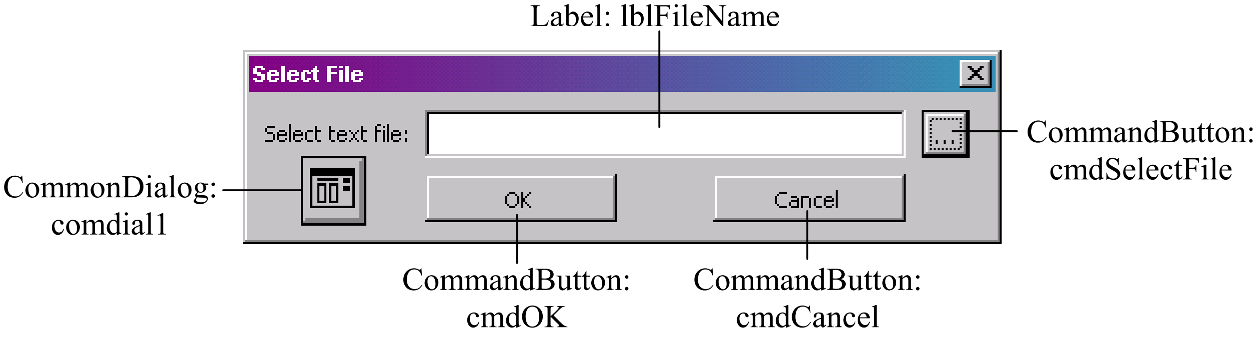
It should also be noted that this macro is designed to parse out the name of the text file that it has been used to open. The text file, therefore, must have a name containing 6 strings separated by underscores. The first part of the name should be the image name. The next two parts of the name are not extracted into the Excel spreadsheet. The last three parts should be the type of classification, the accuracy set used and the type of filter used on the image.

This macro needs to run in Microsoft Excel within an open workbook called "extract_data.xls". The following column headings can be inserted into line 1 of sheet1 in the workbook as labels for the data that will be copied:



AA file name	Image	Classification	Accuracy set	Filter	overall	nf producer's	nf user's	for producer's	for user's	overall K	nf K	for K

The macro has a simple interface (a form, called frmMain) to allow the user to select a text file:



The code for the form above is as follows:

```
Private Sub cmdCancel_Click()
    Unload frmMain
End Sub

Private Sub cmdOK_Click()
Dim i As Integer
Dim cell As String
Dim cell_range As String

     'If the caption in the label lblFileName is not blank,
     'open the text file
     If lblFileName.Caption <> "" Then
         Workbooks.OpenText Filename:= _
            comdial1.Filename, Origin:=xlWindows, StartRow:=1, _
            DataType:=xlDelimited, TextQualifier:=xlDoubleQuote, _
            ConsecutiveDelimiter:=True, Tab:=True, Semicolon:=False _
            , Comma:=False, Space:=True, Other:=True, OtherChar:="=", _
            FieldInfo:=Array(Array(1, 1), Array(2, 1), Array(3, 1), _
            Array(4, 1), Array(5, 1), Array(6, 1), Array(7, 1))

        'Activate the spreadsheet to which the data will be copied
        Windows("extract_data.xls").Activate
        'Need to change the upper limit on the counter
        'if more than 999 rows of data are in the spreadsheet!
        For i = 2 To 1000
            'Create a cell name to put the data into
            cell = "A" & Trim(Str(i))
            'Check that the cell is blank
            If range(cell).Value = "" Then
                'If it is, copy data to the spreadsheet
                '(this function is in the module)
                copy_cells i
                'Set width of column A
                Columns("A").EntireColumn.AutoFit
                'Align left the some of the column data
                cell_range = "B2:E" & Trim(Str(i))
                Range(cell_range).Select
                Selection.HorizontalAlignment = xlLeft
                'Select a cell so that whole rows or columns are not selected
                cell = "A" & Trim(Str(i))
                Range(cell).Select
                'Set the counter to 1001 because the data were copied and
                'there is no need to check another row to find a blank row
                i = 1001
            End If
        Next i
    End If

    Unload frmMain
End Sub

Private Sub cmdSelectFile_Click()
    'Display the common dialog box (file selector window)
    comdial1.ShowOpen
```

```
    'If the user selected a file (the file name is not empty)
    If comdial1.Filename <> "" Then
        'Set the caption on the form to the file name
        lblFileName.Caption = comdial1.FileTitle
    Else
        'Otherwise, clear the caption
        lblFileName.Caption = ""
    End If
End Sub
```

The code for the module is as follows:

```
Sub open_file()
    frmMain.Show
End Sub

Sub copy_cells(data_row As Integer)
Dim text_file As String
Dim filename_parts() As String
Dim image As String, classification As String,
Dim aa_set As String, filter As String
Dim overall As String, nfprod As String, nfuser As String
Dim forprod As String, foruser As String
Dim datarow As String
Dim my_cell As String
Dim my_cell_range As String

    'Get the text file name and parse out the image name, classification,
    'accuracy set and filter used
    text_file = frmMain.lblFileName
    filename_parts = Split(text_file, "_")
    image = filename_parts(0)
    classification = filename_parts(3)
    aa_set = filename_parts(4)
    filter = Left(filename_parts(5), Len(filename_parts(5)) - 4)

    'Put the aa file name, the image name, classification, accuracy set
    'and filter used into the spreadsheet
    Windows("extract_data_macros.xls").Activate

    'Convert the row number to a string with no spaces
    datarow = Trim(Str(data_row))

    my_cell = "A" & datarow
    range(my_cell).Value = text_file

    my_cell = "B" & datarow
    range(my_cell).Value = image

    my_cell = "C" & datarow
    range(my_cell).Value = classification

    my_cell = "D" & datarow
    range(my_cell).Value = aa_set

    my_cell = "E" & datarow
```

```
range(my_cell).Value = filter

'Get overall accuracy and paste it
my_cell = "F" & datarow
copy_paste text_file, "D40", "extract_data_macros.xls", my_cell

'Get nf producer's accuracy and paste it
my_cell = "G" & datarow
copy_paste text_file, "G34", "extract_data_macros.xls", my_cell

'Get nf user's accuracy and paste it
my_cell = "H" & datarow
copy_paste text_file, "H34", "extract_data_macros.xls", my_cell

'Get for producer's accuracy and paste it
my_cell = "I" & datarow
copy_paste text_file, "G35", "extract_data_macros.xls", my_cell

'Get for user's accuracy and paste it
my_cell = "J" & datarow
copy_paste text_file, "H35", "extract_data_macros.xls", my_cell

'Get overall K and paste it
my_cell = "K" & datarow
copy_paste text_file, "D49", "extract_data_macros.xls", my_cell

'Get nf K and paste it
my_cell = "L" & datarow
copy_paste text_file, "D57", "extract_data_macros.xls", my_cell

'Get for K and paste it
my_cell = "M" & datarow
copy_paste text_file, "D58", "extract_data_macros.xls", my_cell

'Get the values (with percent signs) for all of the accuracies
my_cell = "F" & datarow
overall = range(my_cell).Value

my_cell = "G" & datarow
nfprod = range(my_cell).Value

my_cell = "H" & datarow
nfuser = range(my_cell).Value

my_cell = "I" & datarow
forprod = range(my_cell).Value

my_cell = "J" & datarow
foruser = range(my_cell).Value

'Multiply the accuracies by 100 to remove the percent signs and
'replace the cell contents with accuracy values with no percent signs
my_cell = "F" & datarow
range(my_cell).Value = overall * 100

my_cell = "G" & datarow
range(my_cell).Value = nfprod * 100
```

```
    my_cell = "H" & datarow
    range(my_cell).Value = nfuser * 100

    my_cell = "I" & datarow
    range(my_cell).Value = forprod * 100

    my_cell = "J" & datarow
    range(my_cell).Value = foruser * 100

    'Reformat the cells containing the accuracies to a number
    'rather than a percent format
    my_cell_range = "F" & datarow & ":" & "J" & _
      datarow
    range(my_cell_range).Select
    Selection.NumberFormat = "0.00"

    'Close the accuracy report
    Windows(text_file).Close
End Sub

Public Sub copy_paste(wkbk1 As String, copy_cell As String, wkbk2 As String,
paste_cell As String)
    Windows(wkbk1).Activate
    range(copy_cell).Select
    Selection.Copy
    Windows(wkbk2).Activate
    range(paste_cell).Select
    ActiveSheet.Paste
End Sub
```

To run the macro, the user must initiate the code in the "open_file" section of the module. That function displays the graphical user interface which allows the user to select a text file to open.

APPENDIX H
Macro to Extract Accuracy Assessment Error Matrix

This macro extracts the error matrix data from a text file created using the ERDAS Imagine Accuracy Assessment module. This macro needs to run in Microsoft Excel within an open workbook called "extract_matrix.xls". The code requires map marginals to be in a sheet in the same workbook. The following column headings can be inserted into line 1 of sheet1 in the workbook as labels for the data that will be copied: AA file name, Image, Classification, Accuracy set, Filter, tf_mf, tf_mnf, tnf_mf, total mf, total mnf, total tf, total tnf, and total points (see Table 1 for an explanation of the abbreviations). See notes in Appendix G for further qualifications regarding the use of this code.

The macro has a simple interface (a form, called frmMain) to allow the user to select a text file. The program runs when the user intiates the code in the open_file function. The frmMain interface and the code to run it is included in Appendix G. The code for the module is as follows:

```
Sub open_file()
   frmMain.Show
End Sub

Sub copy_cells(data_row As Integer)
Dim text_file As String
Dim thiswkbk As String
Dim filename_parts() As String
Dim image As String, classification As String, aa_set As String, filter As String
Dim overall As String, nfprod As String, nfuser As String
Dim forprod As String, foruser As String
Dim my_cell As String
Dim my_cell_range As String
Dim imgfilter As String
Dim my_range As String

'Get the text file name and parse out the image name, classification, accuracy set and filter used
   text_file = frmMain.lblFileName
   thiswkbk = "percent_forest_calcs.xls"
```

```
filename_parts = Split(text_file, "_")
image = filename_parts(0)
classification = filename_parts(3)
aa_set = filename_parts(4)
filter = Left(filename_parts(5), Len(filename_parts(5)) - 4)

'Put the aa file name, the image name, classification, accuracy set and filter used into the
  'spreadsheet
Windows(thiswkbk).Activate

my_cell = "A" & Trim(Str(data_row))
Range(my_cell).Value = text_file

my_cell = "B" & Trim(Str(data_row))
Range(my_cell).Value = image

my_cell = "C" & Trim(Str(data_row))
Range(my_cell).Value = classification

my_cell = "D" & Trim(Str(data_row))
Range(my_cell).Value = aa_set

my_cell = "E" & Trim(Str(data_row))
Range(my_cell).Value = filter

'Get tf_mf and paste it
my_cell = "F" & Trim(Str(data_row))
copy_paste text_file, "F18", thiswkbk, my_cell

'Get tf_mnf and paste it
my_cell = "G" & Trim(Str(data_row))
copy_paste text_file, "F17", thiswkbk, my_cell

'Get tnf_mnf and paste it
my_cell = "H" & Trim(Str(data_row))
copy_paste text_file, "E17", thiswkbk, my_cell

'Get tnf_mf and paste it
my_cell = "I" & Trim(Str(data_row))
copy_paste text_file, "E18", thiswkbk, my_cell

'Get total mf and paste it
my_cell = "J" & Trim(Str(data_row))
copy_paste text_file, "E35", thiswkbk, my_cell

'Get total mnf and paste it
```

```
    my_cell = "K" & Trim(Str(data_row))
    copy_paste text_file, "E34", thiswkbk, my_cell

    'Get total tf and paste it
    my_cell = "L" & Trim(Str(data_row))
    copy_paste text_file, "D35", thiswkbk, my_cell

    'Get total tnf and paste it
    my_cell = "M" & Trim(Str(data_row))
    copy_paste text_file, "D34", thiswkbk, my_cell

    'Get the total number of points and paste it
    my_cell = "P" & Trim(Str(data_row))
    copy_paste text_file, "D38", thiswkbk, my_cell

    'Determine the image for which the proportions are needed
    imgfilter = image & "_" & filter
    my_range = "map_marginals!$A$2:$F$26"

    'Get total map proportion forest and paste it
    my_cell = "N" & Trim(Str(data_row))
    Range(my_cell).Value = Application.WorksheetFunction.VLookup(imgfilter,  _
    Range(my_range), 5, False)
    Range(my_cell).Select
    Selection.NumberFormat = "0.0000"

    'Get total map proportion nonforest and paste it
    my_cell = "O" & Trim(Str(data_row))
    Range(my_cell).Value = Application.WorksheetFunction.VLookup(imgfilter, _
    Range(my_range), 6, False)
    Range(my_cell).Select
    Selection.NumberFormat = "0.0000"

    'Close the accuracy report
    Windows(text_file).Close
End Sub

Public Sub copy_paste(wkbk1 As String, copy_cell As String, wkbk2 As String, _
paste_cell As String)
    Windows(wkbk1).Activate
    Range(copy_cell).Select
    Selection.Copy
    Windows(wkbk2).Activate
    Range(paste_cell).Select
    ActiveSheet.Paste
End Sub
```

APPENDIX I
IGSCR Help File

Iterative Guided Spectral Class Rejection

Select **IGSCR Classification…** from the **Classification** menu to open this dialog.

Iterative Guided Spectral Class Rejection (IGSCR) is a hybrid image classification algorithm which combines the superior qualities of both unsupervised and supervised classifications. It requires as input a six band Landsat TM image and areas of interest (AOIs) selected from within that image to represent the cover types to be classified.

Although IGSCR requires prior knowledge of the cover types in the image, the areas of interest selected do not have to conform to the strict rules which govern the generation of spectral signatures for a supervised classification, namely, they do not need to be highly separable or have a normal distribution of brightness values in each band of the Landsat TM image.

The IGSCR Process

IGSCR first uses the ERDAS IMAGINE **Unsupervised Classification** utility to classify the image using the ISODATA clustering technique. Next, the ISODATA image values are extracted from user identified AOIs using the ERDAS IMAGINE **Pixel to Table** utility. Each AOI file selected by the user should contain only areas within one information class, e.g., one AOI file might contain only forested areas, another the non-forested areas. The Pixel to Table process is performed separately for each of the AOI files. **IGSCR can create images containing either two or three information classes. Therefore 2 or 3 AOIs may be selected.**

Each of the ASCII files produced by the Pixel to Table utility is then processed to determine for every spectral class the number of pixels that fall into each information class. The data obtained are used to determine whether or not each spectral class is "pure". If a spectral signature meets the user-defined **Homogeneity Threshold** at the given **Type I Error** rate, it is considered to be "pure". The homogeneity of a spectral class is determined by the following test on the proportion of pixels that fall into a given information class:

if z > z(α) then the class is considered to be pure

where z = ((p_hat – p0 – 0.5) / Total) / sqrt (p0 * (1 – p0) / Total)
α = the type-I error rate
p_hat = majority information class pixel count / total pixel count (for a given spectral class)
p0 = the homogeneity threshold
Total = the total pixel count for a given spectral class

⇒ Note: Total * (1 – p0) must be >= 5 for a spectral class to be considered pure.

Based on the results of the pixel purity test, a lookup table is created mask the input image. The input image is masked to remove all of the pure classes by setting the pixels within the pure classes to zero. The result is a subset of the input image which contains only pixels that were not considered to be pure in the previous iteration. This masked image is used as input into the next IGSCR iteration ISODATA clustering algorithm.

If the user opted to have the Stacked ISODATA image output, a lookup table is also written to recode the ISODATA image. The ISODATA image is recoded to the appropriate information classes for the pure spectral classes and to zero for spectral classes that were considered to be impure. The result is an image that contains only pure information class data, with zeros in areas that were impure. In subsequent iterations, new pure classes that are found are similarly recoded, then stacked with the previous recodes using simple image addition.

The results of the pixel purity test are also used to rename each of the spectral signatures as follows:

iteration – value . information class

where **iteration** is the iteration that the signature was created via the unsupervised classification algorithm (ISODATA clustering), **value** is the value that the spectral class is assigned in the output ISODATA image for that iteration, and **information class** is the number of the information class for which the signature has been determined to be homogeneous.

IGSCR is an iterative process which accumulates pure signatures developed during each unsupervised classification. Therefore, after each iteration, the renamed signatures are compiled into one file containing all signatures from each iteration (including both pure and impure signatures, named as described above).

IGSCR continues to iterate until either:

- the user-defined maximum number of iterations have been performed, or
- no pure classes are found in a given iteration, or
- all classes in a given iteration are pure.

After any one of the iteration stopping criteria are met, the combined signature file is edited to remove all impure signatures. Then, if the user opted to have a **Maximum Likelihood** image output, the input image and the final signature file are used entered into the ERDAS IMAGINE **Supervised Classification** algorithm. The IGSCR utility then recodes the maximum likelihood output image based on the information classes assigned to the pure signatures.

If the user opted to have a **Stacked ISODATA** output image and unclassified pixels remain when IGSCR stops iterating, those pixels are also run through the maximum likelihood algorithm to obtain a thematic image that can be recoded. The unclassified pixels are then recoded to information class 4 and added to the previously stacked ISODATA recode images. This aids in the discrimination between unclassified data and background pixels (otherwise, both would have a value of zero).

If the user chose to have a **Stacked ISODATA with Maximum Likelihood on Unclassified Pixels** output image, and the unclassified pixels have already been run through the ERDAS IMAGINE **Supervised Classification** algorithm, the resulting maximum likelihood image is simply recoded based on the information classes assigned to the pure signatures and added to the ISODATA images that were stacked prior to the addition of the class 3 pixels. If the **Stacked ISODATA** option was not selected, the unclassified pixels must first be classified using the ERDAS IMAGINE **Supervised Classification** utility, then recoded and stacked.

If the user opted to run the **Edge Model** on the **Maximum Likelihood** image, the recoded maximum likelihood image is run through a model that first determines the distance of each pixel from the forest informational class and then does the same for the nonforest informational class. Two images are produced by the model, each of which contains a 0 if the pixel is within the informational class of interest, a 1 or 2 if the pixel is 1 or 2 pixels away from the class of interest, and a 3 if the pixel is 3 or more pixels away from the informational class of interest.

After the creation of the forest and nonforest distance images, the same model also creates an image containing four classes 1) nonforest, 2) nonforest edge, 3) forest, and 4) forest edge. The function within this model which does this takes as input the original forest/nonforest classification and forest and nonforest distance images. It assigns pixels to the edge category if they were classified as one informational class, but are 1 or 2 pixels away from the other informational class. For example, a pixel that was classified as forest would contain a zero in the forest distance image, but if the nonforest distance image contained a 1 or 2 for that pixel, that would mean that it was 1 or 2 pixels away from nonforest. This pixel would be classified as forest edge in the final model.

If the user opted to run the **Edge Model** on the **Stacked ISODATA with Maximum Likelihood on Unclassified Pixels** image, the same process is performed as for the **Edge Model** on the **Maximum Likelihood** image except that the input image is the Stacked ISODATA with Maximum Likelihood on Unclassified Pixels image.

Important Note: Edge models are not set up to deal with images containing three information classes.

In addition to signature and image outputs, the **IGSCR Classification** utility also produces an Excel file for each iteration listing the spectral classes and the statistics generated to determine class purity and a Rich Text Format document that includes a list of the input parameters, an explanation of the reason that IGSCR stopped iterating, the number of pure classes found, basic statistics for the images produced and a list of the names of the images produced.

SPECIAL FEATURE: In order to improve the efficiency of the IGSCR algorithm in cases where multiple classifications of the same image with varied parameters or areas of interest need to be performed, an additional feature was added to the IGSCR program.

The first iteration of IGSCR involves an ISODATA clustering that depends upon the number of classes selected, but is not affected by any of the other IGSCR parameters. To allow the reuse of the first iteration ISODATA image, the program is set up to check for the presence of that image and its corresponding signature file in the output file folder prior to running the ISODATA clustering algorithm. If found, the first iteration unsupervised classification is not performed.

In addition, if the operator is using the same area of interest (AOI) files, the first iteration ASCII files can also be copied into the output file folder prior to running IGSCR. The program checks for the presence of the first iteration ASCII files, and, if present, skips the first iteration Pixels to Table conversion.

Note 1: To use this feature, the ISODATA image and signature file, containing the same number of ISODATA classes as were specified in the IGSCR interface, must be present in the output file folder selected by the user in the IGSCR interface at the time that the program is initiated.

Note 2: In order to obtain the first iteration ISODATA image, signature file, and ASCII files, **Delete Intermediate Files** in the **IGSCR Options** window must be turned off the first time IGSCR is performed on the image of interest .

Note 3: The files must have the correct name in order to be recognized by the IGSCR program. The names of the first iteration ISODATA image and signature files conform to the following convention: **root of input image name _ number of ISODATA classes + “c_it1.” + extension**. For example, if the input image name was **test.img** and the number of ISODATA classes selected was 100, the first iteration ISODATA image would be called **test_100c_it1.img**.

Back to Top

ISCGR Classification:

Input Raster File: Enter the name of the input image file to classify or click on the File Selector button. The default file extension is .img. Upon selection of the input image, a default name will be created for the output signature file in the same path as the input image file.

Output Signature File: Enter the name of the output signature set file or click on the File Selector button. The default file name is composed of the input image path and root name followed by the number of ISODATA classes selected and the homogeneity threshold for pure classes. The default file extension is .sig.

⇒ The path of the output signature file is used for all of the IGSCR outputs.

Input AOI File: Enter the name of an input AOI file to use in IGSCR or click on the File Selector button. The default file extension is .aoi.

⇒ The AOI file name must conform to following naming convention: **input image name_information class . aoi**, where the information class should be 1, 2 or 3 corresponding to non-forest, forest or unclassified.

For example, if the input image name is **test.img**, the only three AOI names possible are **test_1.aoi**, **test_2.aoi** or **test_3.aoi**.

Add Click to add the selected AOI file to the list of Files to export.

Remove Click to remove the selected AOI file from the list of Files to export.

Files to export This window displays the list of AOI files that will be exported.

Maximum Iterations: Enter the number of maximum number of times that the IGSCR utility should run.

Type-I Error: Select the type-I error rate (α level) to be used in the test on the proportion of pixels in an information class, used to determine whether or not a spectral class meets the homogeneity threshold. The default value is 0.05.

Homogeneity: The homogeneity threshold used to determine whether or not a spectral class is "pure". The default value is 0.90.

Isodata Classes: Enter the number of ISODATA classes to be created in each IGSCR iteration.

ISODATA Options: Use this dialog to specify the ISODATA processing options, including the axis for initializing class means, the scaling range for computing initial class means, the maximum number of iterations and the convergence threshold to be used in the ISODATA clustering algorithm.

IGSCR Options: Use this dialog to specify the IGSCR final classifications desired and whether or not to delete intermediate files.

OK Click to run the IGSCR process with the options selected and close this dialog.

Batch Click to include the IGSCR function in a batch file.

Cancel Click to cancel the IGSCR process and close this dialog.

Help Click to see this help document.

⇒ For information on using the ERDAS IMAGINE graphical interface, see the on-line **IMAGINE Interface** manual

⇒ The IGSCR algorithm was developed by Jared Wayman and Randolph Wynne of the Department of Forestry at Virginia Polytechnic Institute and State University with support from the National Council of the

Paper Industry for Air and Stream Improvement, the United States Forest Service Southern Research Station and the Virginia Department of Forestry.

> Wayman, J.P., R.H. Wynne, J.A. Scrivani, and G.A. Reams, 2001. Landsat TM-based forest area estimation using Iterative Guided Spectral Class Rejection, *Photogrammetric Engineering & Remote Sensing*, 67(10):1155-1166.

⇒ The IGSCR Classification program was written in 2001 by Rebecca Forest Musy during her graduate research under Randolph Wynne at the Department of Forestry at Virginia Polytechnic Institute and State University. This project was supported by the National Council of the Paper Industry for Air and Stream Improvement, the USDA Forest Service North Central and Southern Research Stations, and the Virginia Department of Forestry.

⇒ The edge models used in the program were developed by Mark Hansen and Dan Wendt of the USDA Forest Service North Central Research Station. Note that they recommend clumping and sieving (to 1 acre) a classified image prior to running the edge models. This process was not included in the IGSCR program.

⇒ Special thanks go out to Christine Blinn, my fellow graduate student and partner in crime. Without her experience with and knowledge of the intricacies of the IGSCR algorithm, a thorough testing of this program would not have been possible. Her input and support have been invaluable to me in the development of this program.

Back to Top

APPENDIX J
ISODATA Options Help File

ISODATA Options Dialog

As part of the IGSCR algorithm, the ERDAS IMAGINE **Unsupervised Classification** utility is used to classify the input image using the ISODATA clustering technique. This dialog allows the user to select the axis for initializing class means, the scaling range for computing initial class means, the maximum number of iterations and the convergence threshold to be used in the ISODATA clustering algorithm.

ISODATA Options: Use the following option buttons and number fields to specify the ISODATA processing options.

Initialize Means Along: Specify which axis through the data space to use for initializing class means.

Principal Axis Means are computed to be along the first principal component vector and are evenly distributed within the scaling range for the first principal component. NOTE: This option will cause IGSCR to fail if the input image is a principal components image.

Diagonal Axis Means are computed to be along a diagonal vector and are evenly distributed within the scaling range for each layer.

Scaling Range: Select the scaling range mode for computing initial class means.

Automatic This option will automatically compute the scaling range based on the number of classes requested and the assumption that the data are normally distributed. The more classes requested, the wider the scaling range will be.

Std. Deviations Enter the number of standard deviations to use for scaling. The number of classes requested are distributed evenly within this range along the axis specified above. A higher number of standard deviations initialized more class means in the tails of the distribution.

Maximum Iterations: Enter the number of maximum times that the ISODATA utility should recluster the data. This parameter prevents this utility from running too long, or from potentially getting "stuck" in a cycle without reaching the convergence threshold.

Convergence Threshold: Specify the convergence threshold. The convergence threshold is the maximum percentage of pixels whose cluster assignments can go unchanged between iterations. This threshold prevents the ISODATA utility from running indefinitely.

By specifying a convergence threshold of .95, you would specify that as soon as 95% or more of the pixels stay in the same cluster between one iteration and the next, the utility should stop processing. In other words, as soon as 5% or fewer of the pixels change clusters between iterations, the utility will stop processing.

Cancel Click to accept changes to the ISODATA Options and close the dialog.

Help Click to see this help document.

⇒ For information on using the ERDAS IMAGINE graphical interface, see the on-line **IMAGINE Interface** manual

APPENDIX K
IGSCR Options Help File

IGSCR Options Dialog

The final output of IGSCR includes a set of "pure" signatures, the entire set of signatures produced by the algorithm, as well as any combination of the following three classified image outputs. This dialog also allows the user the choice of deleting intermediate files created during the IGSCR process.

IGSCR Options: Use the following check boxes to specify the IGSCR final classification and file deletion options.

Final Classification Method:

Maximum Likelihood

A Supervised Classification will be performed using the input image file and the final signature set generated by the IGSCR process. The classification is executed without a non-parametric rule and with maximum likelihood as the parametric decision rule. The IGSCR utility also recodes the maximum likelihood output image based on the information classes assigned to the pure signatures.

NOTE: This option will not produce a valid output when the input image is the result of a principal components analysis (PCA), since the signatures developed from the PCA image during IGSCR will be non-invertible.

Edge Model on Maximum Likelihood

If the user opted to run the edge model on the maximum likelihood image, the recoded maximum likelihood image is run through a model that first determines the distance of each pixel from the forest informational class and then does the same for the nonforest informational class. Two images are produced by the model, each of which contains a 0 if the pixel is within the informational class of interest, a 1 or 2 if the pixel is 1 or 2 pixels away from the class of interest, and a 3 if the pixel is 3 or more pixels away from the informational class of interest.

After the creation of the forest and nonforest distance images, the same model also creates an image containing four classes 1) nonforest, 2) nonforest edge, 3) forest, and 4) forest edge. The function within this model which does this takes as input the original forest/nonforest classification and forest and nonforest distance images. It assigns pixels to the edge category if they were classified as one informational class, but are 1 or 2 pixels away from the other informational class. For example, a pixel that was classified as forest would contain a zero in the forest distance image, but if the nonforest distance image contained a 1 or 2 for that pixel, that would mean that it was 1 or 2 pixels away from nonforest. This pixel would be classified as forest edge in the final model.

Stacked ISODATA Images

After each ISODATA classification, the output cluster layer is recoded such that the pure spectral classes are assigned to their respective information classes and the impure classes are set to zero. If only one iteration is performed, this option only recodes the ISODATA output.

If more than one iteration is performed, each of the newly recoded images is added to the previously recoded images by simple image addition. This is possible because each subsequent image is the impure subset of the previous image, therefore newly established information classes fall only in places where previous images had zeros.

If unclassified pixels remain when IGSCR stops iterating, those pixels are recoded to information class 4 and added to the previously stacked ISODATA recode images. This aids in the discrimination between unclassified data and background pixels.

Stacked ISODATA Images plus Maximum Likelihood on Unclassified

The stacked ISODATA classification is performed, except that any remaining unclassified pixels are classified by maximum likelihood using the final signature set generated by the IGSCR process. The resulting maximum likelihood image, which contains only previously unclassified pixels, is then recoded and added to the previously stacked images.

NOTE: The maximum likelihood portion of this option will not produce a valid output when the input image is the result of a principal components analysis (PCA), since the signatures developed from the PCA image during IGSCR will be non-invertible.

Edge Model on Stacked ISODATA Images plus Maximum Likelihood on Unclassified

If the user opted to run the edge model on the Stacked ISODATA Images plus Maximum Likelihood on Unclassified image, the recoded maximum likelihood image is run through a model that first determines the distance of each pixel from the forest informational class and then does the same for the nonforest informational class. Two images are produced by the model, each of which contains a 0 if the pixel is within the informational class of interest, a 1 or 2 if the pixel is 1 or 2 pixels away from the class of interest, and a 3 if the pixel is 3 or more pixels away from the informational class of interest.

After the creation of the forest and nonforest distance images, the same model also creates an image containing four classes 1) nonforest, 2) nonforest edge, 3) forest, and 4) forest edge. The function within this model which does this takes as input the original forest/nonforest classification and forest and nonforest distance images. It assigns pixels to the edge category if they were classified as one informational class, but are 1 or 2 pixels away from the other informational class. For example, a pixel that was classified as forest would contain a zero in the forest distance image, but if the nonforest distance image contained a 1 or 2 for that pixel, that would mean that it was 1 or 2 pixels away from nonforest. This pixel would be classified as forest edge in the final model.

Delete Intermediate Files Select this option to allow the IGSCR process to delete the following intermediate files: the ASCII files, the ISODATA cluster images, the unstacked, recoded ISODATA cluster images, the masked ISODATA images, and the unrecoded maximum likelihood image(s).

With this option selected, the IGSCR process will output the final recoded classification images selected, the final signature file, the combined signature file from all iterations performed prior to removal of impure signatures and informational files about the process.

Close Click to accept changes to the parameters and close IGSCR options dialog.

Help Click to see this help document.

⇒ The edge models used in the program were developed by Mark Hansen and Dan Wendt of the USDA Forest Service North Central Research Station. Note that they recommend clumping and sieving (to 1 acre) a classified image prior to running the edge models. This process was not included in the IGSCR program.

⇒ For information on using the ERDAS IMAGINE graphical interface, see the on-line **IMAGINE Interface** manual

APPENDIX L
Read_me.doc

Setting up the IGSCR program:

1. The computer you are working on must have ERDAS Imagine 8.5 installed.

2. The IGSCR program files (igscr.exe, igscr.eml, classification.eml, edge.mdl) need to be placed in yourdrive:\Users\.imagine850\

3. The "igscr" folder contains the online help files for the program. The "igscr" folder must be placed in yourdrive:\Program Files\Imagine 8.5\help\html\ for the IGSCR help files to be accessible from the user interface.

4. In Imagine, go into Session >Preferences > User Interface and Session and turn off the "Keep Job Status" box. This will prevent you from having to click OK to close the MANY windows that will say "Job finished" after you have run IGSCR.

5. Also in Session > Preferences, go to Image Files (General) and turn off the “Pyramid Layers External” option. The IGSCR program is set up to deal with images that have internal pyramid layers. If you choose to keep the pyramid layers external to your image files, intermediate .rrd files will not be deleted.

6. In Session > Properties, go to the Environment tab and click on Personal Defaults. This will make Imagine look in the Users\.imagine850 folder for the IGSCR files.

6. After you save these preferences, you must reopen Imagine to get them to work.

The IGSCR interface can be accessed by clicking on the Classification button in the Imagine main toolbar. If the above files have been installed correctly, the Classification menu should contain a button that says “IGSCR Classification …”

Once the IGSCR interface is opened, the first thing you need to do is choose your image file. The final signature file name will get automatically created and put into the same folder as the input image. You can change the final signature file name and path.

NOTE: All of the output files will be put into the same path as the output signature file.

NOTE: As the homogeneity threshold and number of ISODATA classes are changed, the name of the final signature file will also change. Each time the name changes, ERDAS stores the new signature file name in the list of recent files. To avoid filling your list of recent files with variations on one file name, it is recommended that you either type in changes to these two parameters or choose the input file after changing these parameters.

The next thing you need to do is select your AOIs. The AOI files must be named with the root name of the image file then "_" and the number of the info class. For example, if the image name is "test.img", the AOIs should be named"test_1.aoi" and "test_2.aoi".

After selecting the AOI files, the IGSCR and ISODATA options can be changed. For help, please refer to the help files.

I included a copy of a small image subset (test.img) and the associated AOIs (test_1.aoi and test_2.aoi) so that you can do a quick test before running any longer tests. You can put the files in any folder. The AOI files need not be in the same place as the input image or the output files. If you leave the parameters on the interface set to the defaults, IGSCR will stop after 4 iterations since no more pure classes get extracted. It should only take 3-5 minutes.

APPENDIX M
Edge.mdl

```
#
set window union;
#
# set area of interest for the model
#
set aoi none;
#
# declarations
#
if (argcount != 4) { quit; }

integer raster in1 file old input arg1;
integer raster out1 file new output useall thematic bin direct default 4 bit
unsigned integer arg2;
integer raster out2 file new output useall thematic bin direct default 4 bit
unsigned integer arg3;
integer raster out3 file new output ignore 0 thematic bin direct default 4
bit unsigned integer arg4;
#
# function definitions
#
# note: the search function takes as input a thematic image, the distance to
search and the class from
# which the distances are being found
#
out1 = SEARCH ( in1 , 2 , 2 ) ;
out2 = SEARCH ( in1 , 2 , 1 ) ;
out3 = CONDITIONAL {(in1==1 && out2==0 && out1==3) 1,
   (in1==1 && out2==0 && out1<3) 2,
   (in1==2 && out2==3 && out1==0) 3,
   (in1==2 && out2<3 && out1==0) 4};
quit;
```

Rebecca Forest Musy

OBJECTIVE

To apply and further my knowledge of remote sensing and to contribute to the development and advancement of that science as it applies to environmental research.

WORK EXPERIENCE

Dates Employed	Job Title, Location	Responsibilities
Jan-Sept 2002	**Remote Sensing Specialist,** Questerra, LLC	Independent contractor while completing Master's research at Virginia Tech; Developed computer programs to process satellite imagery. Working with little direction, dealt with the manipulation and storage of extremely large data sets, contributing to broad commercial availability of GIS data.
Jan 2000-Present	**Research Assistant/ Teaching Assistant**, Virginia Tech	Researched, developed, and implemented spatial analyses involving GIS and remotely sensed data; assisted in writing grant proposals; gave presentations; taught photogrammetry, ArcView and ArcInfo to forestry students.
Nov 1998-Jan 2000	**Forestry Assistant**, Virginia Department of Forestry	Corrected Global Positioning Systems data, managed the GPS database, provided training and technical support for GPS software and hardware; generated and analyzed GIS data; assisted in field data collection.
Jan-May 1998	**Biology Lab Instructor**, University of Florida	Planned and presented weekly general biology labs; supervised lab activities; prepared, administered and graded quizzes, practical exams, and lab reports.

EDUCATIONAL PREPARATION

High School Diploma	St. John's International School	Waterloo, Belgium	1993
B.S., Zoology, cum laude	University of Florida	Gainesville, Florida	1997
M.S., Forestry	Virginia Polytechnic Institute and State University	Blacksburg, Virginia	2003
Additional Courses:	Probability, University of Virginia		1999
	Visual Basic Programming, Piedmont Virginia Community College		1999

HONORS AND AWARDS

1995	Marlin Perkins Mutual of Omaha's Wild Kingdom Scholarship
1995	Tropical Ecology Scholarship (for bat research in Panama)
1997	Phi Beta Kappa, Senior inductee
1997	Golden Key National Honor Society
2000	Elected Chapter President of the American Society of Photogrammetry and Remote Sensing
2002	First place in Virginia Tech Remote Sensing and GIS Research Symposium student poster competition
2003	First place in Natural and Biological Sciences category at Virginia Tech Graduate Student Assembly Research Symposium poster competition
2003	Awarded conference admission and travel fees to attend the American Society of Photogrammetry and Remote Sensing (ASPRS) conference in Anchorage, AK (as well as a 1 year membership in ASPRS) by ASPRS and the International Geographic Information Foundation

SKILLS

Trilingual: English, French, Spanish
Programming languages: MapObjects, Visual Basic, C++, ERDAS Macro Language, ERDAS C Toolkit
Software Experience: ArcGIS, ArcView, ArcInfo, ENVI, ERDAS Imagine, MS Office Suite, SAS, WinZip, WS-FTP

RELEVANT COURSEWORK

Remote Sensing of Natural Resources, GIS Applications in Natural Resources, Analysis of Spatial Data, Algorithms in GIS, Natural Resources Research Procedures, Seminar in Remote Sensing & GIS, Statistics in Research, Integrated Principles of Biology (2 sections with labs), General Ecology, Vertebrate Zoology

PUBLICATIONS

Scrivani, J.A., R.H. Wynne, C.F. Blinn, and R.F. Musy, 2001. Phase I forest area estimation using Landsat TM and Iterative Guided Spectral Class Rejection: Assessment of possible training data protocols. In: *Proceedings of the 2nd Annual Forest Service FIA Symposium*, Salt Lake City, Utah, October 17-18, 2000. General Technical Report SRS-47. Asheville, NC: USDA Forest Service Southern Research Station, pp. 11-14.

Musy, R.F., 2003. Refinement of Automated Forest Area Estimation via Iterative Guided Spectral Class Rejection. MS Thesis. Virginia Polytechnic Institute and State University, Blacksburg, VA, 137 pp.

PRESENTATIONS AND POSTERS

January 2001	Virginia Tech Iterative Guided Spectral Class Rejection Training, Blacksburg, VA Co-Instructor and presenter of IGSCR methods and training data collection
March 2001	Virginia Tech Third Annual Remote Sensing and GIS Research Symposium, Roanoke, VA Presentation: Towards More Objective Training Data Collection for Land Use Classification
March 2002	Virginia Tech Fourth Annual Remote Sensing and GIS Research Symposium, Roanoke, VA Poster: Automation of Iterative Guided Spectral Class Rejection
April 2002	XXII FIG (International Federation of Surveyors) International Congress, Washington, D.C. Presentation: Training Data for Satellite-Based Forest Area Estimation
April 2003	Virginia Tech Graduate Student Assembly Research Symposium, Blacksburg, VA Poster: Automation of Iterative Guided Spectral Class Rejection
May 2003	American Society of Photogrammetry and Remote Sensing 2003 Annual Conference, Anchorage, AK Poster: Automation of Iterative Guided Spectral Class Rejection