MAPPING AND GIS DEVELOPMENT OF LAND USE AND LAND COVER CATEGORIES FOR THE ALBEMARLE-PAMLICO DRAINAGE BASIN

By

Siamak Khorram
and
Heather Cheshire
Computer Graphics Center
North Carolina State University
Raleigh, NC 27695-7106

Karen Siderelis
and
Zsolt Nagy
North Carolina Center for Geographic Information & Analysis
Raleigh, NC 27611

The research on which this report is based was financed in part by the United States Environmental Protection Agency and the North Carolina Department of Environment, Health and Natural Resources, through the Albemarle-Pamlico Estuarine Study.

Project No. 91-08

March, 1992

ACKNOWLEDGEMENTS

We wish to thank the following people and agencies for their assistance in completing this project: Graduate Assistants Jayantha Ediriwickrema, Mike Campbell and Casson Stallings; Systems Programmer Alan Galloway; Systems Manager Scott Ross; GIS Technician Mike Rink and Applications Programmer Bill Miller; visiting scientist Francesca Somma; and personnel from the N.C. Natural Heritage Program. Special thanks to Cecil Frost for his help, which considerably reduced the amount of field work needed. Thanks also to Mr. Glenn Catts and the U.S. Forest Service for providing most of the aerial photography used in the project.

MAPPING AND GIS IMPLEMENTATION OF LAND USE AND LAND COVER CATEGORIES FOR THE ALBEMARLE-PAMLICO DRAINAGE BASIN

ABSTRACT

The Albemarle-Pamlico (A/P) estuarine system in North Carolina is one of the estuaries in the EPA's National Estuary Program. The lack of a current land use/land cover inventory was identified as a critical gap in the A/P Study resource database. At an A/P Study workshop held late in 1987, Landsat Thematic Mapper (TM) digital data were recommended as the most cost effective and practical source for developing an inventory for the 23,000 square mile drainage basin. The Computer Graphics Center (CGC), North Carolina State University, and the North Carolina Center for Geographic Information & Analysis (CGIA) were given responsibility for the development, storage and dissemination of the inventory.

The study area included a portion of Virginia and nearly one-third of North Carolina including almost all of the Tidewater region. CGC had responsibility for analyzing the five Landsat TM scenes needed to cover the area. Digital TM data were converted to a Lambert Conformal Conic projection and classified into 18 land use/land cover classes using a supervised approach. Results of the project included image files in raster format with every pixel classified by land use/land cover category. Classification verification was performed using 1,931 one acre sample sites located on the classified TM imagery and on aerial photography. Class accuracies were 73% or greater for all Level I classes except developed areas which had an accuracy of 46%.

Image data were converted to a format compatible with CGIA's software, filtered using a standard 5X5 mode filter, converted to vector format and integrated with CGIA's database for the A/P drainage basin. Data are georeferenced to the N.C. State Plane Coordinate System and stored as digital ARC/INFO coverages. Land use/land cover data are available from CGIA as map products or in digital format. Final results also include descriptions of methodology and land use/land cover classes as well as classification error matrices for each physiographic province and for the entire study area.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
ABSTRACT	iii
LIST OF FIGURES	vi
LIST OF TABLES	vii
LIST OF ABBREVIATIONS AND ACRONYMS	viii
SUMMARY AND CONCLUSIONS	ix
RECOMMENDATIONS	ix
INTRODUCTION	1
Study Area	1
Selection of Classification Scheme	3
PROJECT MATERIALS	4
Acquisition of TM Data	5
Acquisition of Ancillary Data	6
PRELIMINARY DATA ASSESSMENT	8
Initial File Preparation	9
Tests of Approaches	10
Tests of Bands	11
Assessing Classification Strategies	12
Results of Preliminary Data Assessments	13
PROJECT METHODOLOGY	15
Identification of Spectral Classes and Training Sites	16
Compilation of Training Statistics	17
Image Classification	19
Classification Accuracy Assessment	19
A/P Database Integration	20
PROJECT RESULTS	21
Classification Scheme	22
Classification Verification	26
Data Vectorization and Integration	34

Data Access and Distribution 35

DISCUSSION AND RECOMMENDATIONS 37
Image Analysis 37
Data Use for Future Inventories 39

REFERENCES 41

APPENDICES 43
Appendix I. Study area coverage by county and quadrangle name 43
Appendix II. Land use/land cover classification system for use with remote sensing data 51
Appendix III. Comparison of proposed modified land use and land land cover classification system with A/P Study classification . . . 53

LIST OF FIGURES

		Page
1.	Location of the A/P drainage basin and physiographic provinces	2
2.	Landsat TM coverage of the A/P drainage basin	7
3.	Error matrix for Tidewater	27
4.	Error matrix for Middle and Upper Coastal Plain	28
5.	Error matrix for Piedmont	29
6.	Error matrix for all image data	30
7.	Error matrix for Level I categories	31
8.	U.S. Geological Survey 1:100,000 map coverage of the A/P drainage basin	36

LIST OF TABLES

		Page
1.	Characteristics of Landsat TM data	5
2.	Scene information	6
3.	Scene priorities for ERDAS program 'STITCH'	21
4.	Land use/land cover classes for the Albemarle-Pamlico Estuary Study	23
5.	Classification accuracy estimates	33
6.	Media and copy formats for land use/land cover data	35

ABBREVIATIONS AND ACRONYMS

A/P	Albemarle and Pamlico
ARC/INFO	A geographic information system (ARC) and associated database management package (INFO) marketed by ESRI
CCT	Computer Compatible Tape
CGC	Computer Graphics Center, North Carolina State University
CGIA	Center for Geographic Information and Analysis
DLG	Digital Line Graph
DN	Digital number
EOSAT	Earth Observation Satellite Company
ERDAS	Earth Resources Data Analysis Systems
ESRI	Environmental Systems Research Institute, Inc.
LAS	Land Analysis System
NPDES	National Pollution Discharge Elimination System
P-data	Fully processed TM data (corrected for radiometric and geometric distortions and including necessary header, annotation, and trailer data files
Pixel	Picture element
TIGER	Topologically Integrated Geographic and Referencing System
TIPS	TM Image Processing System
TM	Thematic Mapper
USGS	United States Geological Survey (Dept. of Interior)
VAX	Virtual Address Extension, hardware from Digital Equipment Corp.
VMS	Virtual Memory Sub-system, a computer operating system for Digital
WRS	Worldwide Reference System

SUMMARY AND CONCLUSIONS

Five Landsat Thematic Mapper (TM) scenes covering 97% of the Albemarle-Pamlico estuarine system drainage basin were used to classify land use and land cover. Digital TM data were physiographically stratified, converted to a Lambert Conformal Conic projection and classified into 18 classes using a supervised approach and statistics from TM bands 3, 4 and 5 (red, near infrared and middle infrared). Classification accuracies were determined based on 1,931 verification sample sites. Leaf-off conditions and, near the coast, excessive soil moisture limited differentiation of certain vegetation types particularly within the Tidewater region. Mapping accuracies were relatively low for Urban and Built-up land (46%) and ranged from 73% to 97% for five other Level I categories (Water, Agriculture, Forestland, Wetlands and Barren Land).

Image data were processed and classified into land use and land cover classes at the Computer Graphics Center (CGC) at North Carolina State University and then transferred to the North Carolina Center for Geographic Information & Analysis (CGIA). At the CGIA, image data were filtered using a standard 5x5 mode filter, converted to the ARC/INFO data format and partitioned by USGS 1:100,000 scale map boundaries. Land use/land cover data and products can be obtained from CGIA by USGS 1:100,000 map windows or by county in a variety of formats. Prospective users need to be aware that these data require large amounts of disk storage. Data are georeferenced to the N.C. State Plan Coordinate System, but, because of their derivation, mapping discrepancies may exist between this data layer and data layers derived from different mapping methodologies.

Overall, Landsat TM data appeared to be a good source of information for large area inventories of land use/land cover patterns. The resultant map products provide the level of detail and accuracy required regional/basin-level analyses for management and research needs.

RECOMMENDATIONS

The following recommendations should be considered during use of the current land use/land cover inventory:

1. Data are applicable to inventory and research efforts designed to characterize large geographic areas such as the entire Albemarle-Pamlico estuarine system, groups of counties, or basins, but are not appropriate for site-specific evaluations such as characterization of urban infrastructures.

2. Because of low classification accuracies for developed areas and underestimation of forested wetlands, the estimates of these areas should be considered with great caution. Data on road networks or municipal boundaries can be obtained from alternative sources (USGS DLG files, Bureau of Census TIGER files or CGIA

databases) and can be overlaid with the inventory data to provide quality assurance for developed areas.

3. Users should be aware that data require large amounts of disk storage due to large file sizes. Identification of appropriate hardware needs is recommended before acquisition and manipulation of digital data files.

4. Efficient map production equipment, preferably an electrostatic plotter, is required to produce hard-copy output (a film writer or similar equipment may also be used for photographic output).

5. In order to adequately monitor land use/land cover activities within the A/P basin, an inventory from satellite data should be conducted every five years. The next database should be developed for 1993 conditions.

The following recommendations should be considered for future land use/land cover inventories:

1. Classification schemes should be consistent with the current scheme. Ideally, classification schemes and methodologies should be coordinated with other state, regional or federal mapping efforts to maximize the potential for generating seamless coverages over large areas. Classification schemes developed to meet localized needs or for more detailed analyses should be designed to allow for consistent generalizations.

2. Investigations should continue to evaluate newly developed clustering or classification algorithms and to develop methodologies for change detection.

3. Utilization of multi-temporal data and inclusion of georeferenced data such as soils or topography are expected to improve detail, accuracy, and timeliness of the data.

4. Hardware requirements for efficient data processing, manipulation and output are intensive. Hardware upgrades and expansions which are currently in progress at CGC and CGIA will do much towards improving future operations; however, it is likely that large data volumes and complex analyses will continue to place increasing demands on existing computer systems and additional hardware needs should be anticipated.

5. Successful completion of this and similar projects relies on the expertise and cooperation of specialists in several disciplines including image processing, database management, and spatial data analysis. In addition, it is recommended that future projects include resources to support a specialist in landscape ecology or a related discipline. It is hoped that the advantages of inter-agency, inter-disciplinary cooperation demonstrated in this project can be continued in future efforts.

MAPPING AND GIS IMPLEMENTATION OF LAND USE AND LAND COVER CATEGORIES FOR THE ALBEMARLE-PAMLICO DRAINAGE BASIN

INTRODUCTION

The Albemarle-Pamlico (A/P) system in North Carolina is one of several estuaries which have been intensively studied under the auspices of the U.S. Environmental Protection Agency's National Estuary Program. The ultimate goals of the A/P Study have been to support research efforts aimed at assessing environmental problems facing the estuarine system and to provide basic information needed to formulate management strategies for the area. The lack of a current land use/land cover inventory for the A/P estuary drainage area was identified as a critical gap in the A/P Study resource database. At an A/P Study workshop held late in 1987, Landsat Thematic Mapper (TM) digital data were recommended as the most cost effective and practical source for developing an inventory of the drainage basin. The Computer Graphics Center (CGC) (North Carolina State University) and the North Carolina Center for Geographic Information & Analysis (CGIA) were given responsibility for the development, storage, and dissemination of the inventory. The CGC is a university-wide research unit independent of any department or university college. The mission of CGC is to facilitate and conduct research and training in the fields of remote sensing, image processing, spatial information systems, and database design and management. The CGIA, a receipt funded agency, operates a Geographic Information System (GIS) and serves as the official repository of digital geographic data for the state of North Carolina. CGIA had previously been selected as the data management center for the A/P Data Management Program. A primary responsibility for CGIA is the development and maintenance of the A/P database.

Study Area

The Albemarle-Pamlico estuarine system is one of the largest estuaries in the U.S. It includes five major rivers: Chowan, Roanoke, Alligator, Tar-Pamlico, and Neuse, and many smaller tributaries. The estuarine system actually encompasses Albemarle, Pamlico, Currituck, Roanoke, Croatan, and Core Sounds. The entire drainage basin covers about 23,500 square miles (61,000 square km) of land and water in eastern North Carolina and southeastern Virginia (see Appendix I for a list of counties and quadrangles covered). This figure is based on the A/P drainage boundary in CGIA's database and does not include the upper Roanoke River. In the last 20 years, the A/P system has experienced more frequent and intense occurrences of algal blooms and fish or shellfish infections, increased turbidity, loss of submerged aquatic vegetation, and other evidence of degraded water quality. Population growth and associated increases in the demands placed on resources have resulted in greater pressures on these ecosystems. The estuarine systems are affected not only by activities in the immediate area, but by activities occurring upstream in large portions of the more populated Piedmont.

The A/P drainage basin encompasses portions of both the Coastal Plain and the Piedmont provinces of North Carolina and Virginia (Figure 1). The Coastal Plain is characterized by its low elevation (< 300 feet) and relatively young, unconsolidated sediments. The Coastal Plain

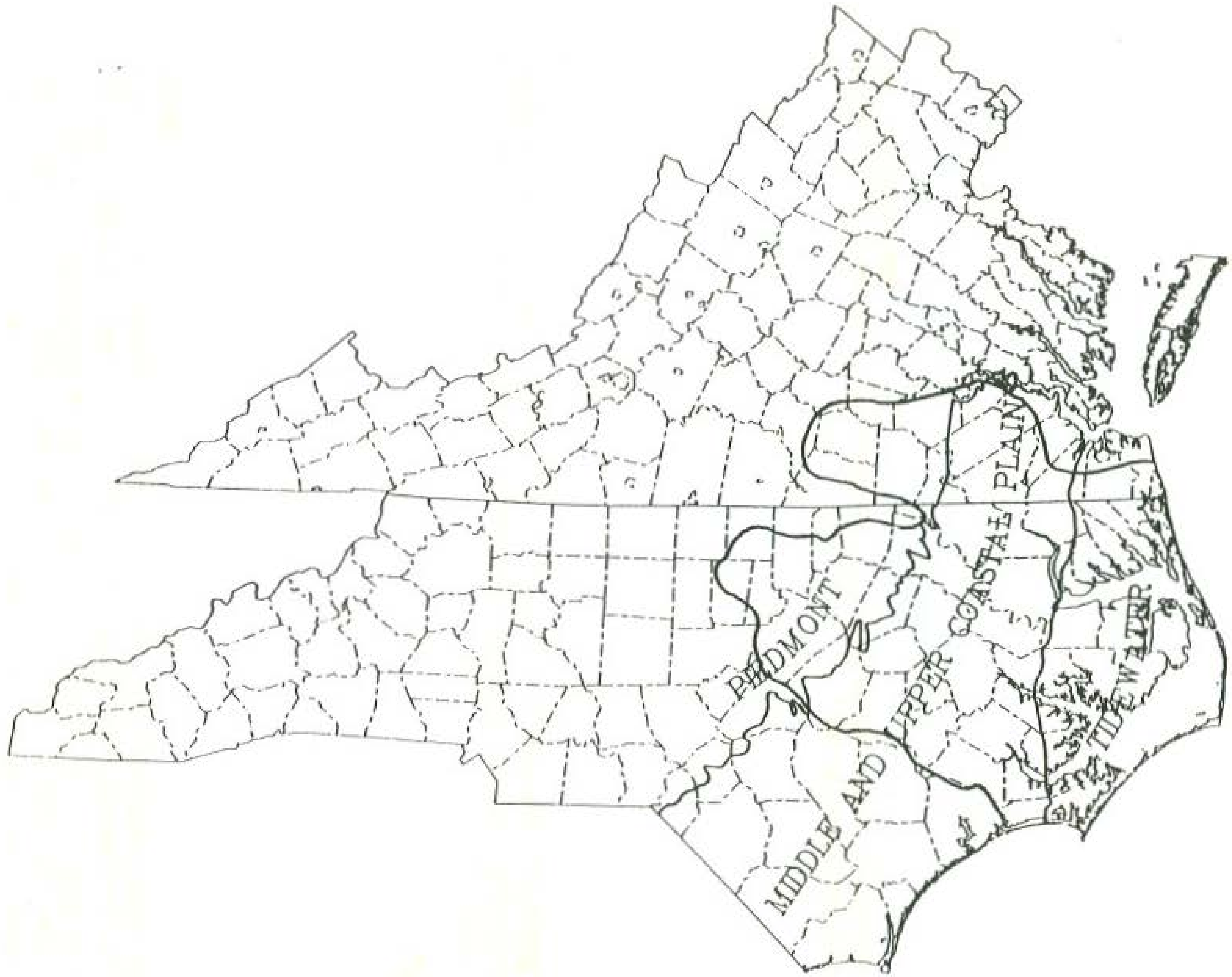

Figure 1. Location of the A/P drainage basin (heavy solid line) and physiographic provinces. Narrower solid lines indicate the Fall Line (between the Piedmont and the Coastal Plain) and the Suffolk Scarp (between the Middle Coastal Plain and Tidewater). Dashed lines are county boundaries.

actually consists of four physiographic provinces, three of which are represented in the A/P drainage area.

The Tidewater province is a low, nearly level plain bounded on the west by the Suffolk Scarp (an old marine terrace) and composed largely of peninsular tracts divided by broad embayments. Elevations do not exceed 25 feet. Much of the area is subject to flooding from storms or tides and is very poorly drained. The Flatlands (or Middle) Coastal Plain and the Inner (or Upper) Coastal Plain are the two provinces found between the Suffolk Scarp and the Fall Zone, the boundary between the older, more resistant rocks of the Piedmont and the younger, weaker Coastal Plain sediments. As one moves from east to west within these two coastal provinces, elevations gradually increase, drainage improves and soil development is increasingly influenced by sediments derived from the Piedmont. The Upper and Middle Coastal Plain were considered as one province in this project.

The eastern Piedmont is geologically diverse but surficially characterized by a gently rolling topography with well incised streams and rivers. In the small area of Virginia Piedmont, the hills tend to be steeper and the topography somewhat more rugged than that found in North Carolina. Maximum elevations of the A/P drainage basin do not exceed 700 feet.

A great diversity of cover types and land use activities occur within the study area. Vegetation types include marsh grasses and forbes, vines, shrubs, and evergreen and deciduous trees. Forest types range from gum-cypress 'muck' swamps of the Tidewater province to late successional oak-hickory stands found on dry ridges in the Piedmont. Soils may be derived from marine, lagoonal, or fluvial processes and can be sandy, peaty, or clayey. Agricultural and silvicultural activities occur throughout the study area. Though there are only a few cities within the A/P drainage basin which could be characterized as major metropolitan areas, numerous smaller cities and towns can be found throughout the region along with a diversity of associated anthropogenic activities. In general, the drainage basin encompasses a wide assortment of land use and land cover categories. Specific counties, 7.5 minute (1:24,000) U.S. Geological Survey quadrangles and 1:100,000 USGS quadrangles which were included in the study area are listed in Appendix I.

Selection of Classification Scheme

In October, 1988, CGIA (then known as the Land Resource Information Service-LRIS) convened an A/P Land Use and Land Cover Scoping meeting. Participants included federal, state, and local resource managers and university researchers. There were representatives from East Carolina University, North Carolina State University, N.C. Division of Coastal Management, N.C. Division of Environmental Management, N.C. Division of Forest Resources, U.S. Fish and Wildlife Service, the city of New Bern, and Dare County Planning. The purpose of the meeting was to discuss and recommend a classification scheme which would be compatible with research and management needs. Representatives from CGC, N.C. State University, were on hand to provide information on known characteristics, capabilities, and limitations of the Landsat Thematic Mapper with respect to the identification of land use and land cover categories.

The group recommended adopting a classification system which would be compatible with the system used by the U.S. Geological Survey (Anderson et al. 1976, Appendix II). This hierarchical land use/land cover classification system was established to be used nationwide with remote sensor data. At the time of its adoption by the USGS, the majority of remotely sensed data were from airborne camera systems with film formats and land classification was done through visual interpretation. However, the classification scheme was believed to be widely adaptable to computer analyses of digital remotely sensed data. In this system, three to four levels for each of nine major categories were recommended depending on the scale of the imagery available and the amount of detail that could be detected. For example, the Level I category Forest Land (category) is sub-divided into three Level II categories: (41) Deciduous Forest Land, (42) Evergreen Forest Land, and (43) Mixed Forest Land. Within known limitations, the first and second levels of the USGS classification system were used as a basis for the A/P classification since it was desirable, at a minimum, to identify all Level I categories which occur in the A/P drainage area.

The list of potential classes was modified based on prior knowledge of sensor capabilities. In general, land cover is more easily determined than land use. Land cover refers to features or properties - natural or anthropogenic - found on the surface of the ground. Land use, on the other hand, refers to activities occurring on the land. Thus, an area could be covered by grass but it might have an agricultural use (e.g. pasture) or a commercial use (e.g. golf course). The ability to determine land use depends on how well surface features represent activities which are occurring in the area. Determination of land use tends to become even more problematic at more detailed levels of classification. In particular, digital remotely sensed data generally cannot be used to extract many kinds of information on management practices or public use. For example, within Level I, category 1, Urban or Built-up Land, digital spectral data alone could not be used to determine if the land was in residential, commercial or industrial use. It was recommended that other categories of specific interest be investigated to determine if it was feasible to identify them. For example, it would be useful to identify hardwood riverine swamp and Atlantic White Cedar, both of which were in the same Level II category (Forested Wetlands). Further modification of the classification scheme would be based on analysis of digital data. The group also recommended adopting a minimum mapping unit of 5 to 10 acres.

PROJECT MATERIALS

The identification, measurement, and inventory of over 20,000 square miles is a formidable task. The multispectral approach has been shown to be a cost effective and reliable means of gathering data about the earth's surface in a digital format (ASPRS 1983; Khorram et al. 1987). Energy from the sun is reflected or emitted from features on the earth's surface in a characteristic spectral pattern or response. Based on spectral response measured over one or more wavelength ranges (bands) in the electromagnetic spectrum, it is possible to separate and identify different ground cover types. An object or an environmental association's unique spectral response is its signature. It is frequently important to detect and utilize more than one spectral band since different cover types can have the same signature within certain portions of the spectrum.

Landsat Thematic Mapper (TM) multispectral data were selected as the source for this inventory because they offered the potential to produce a broadly consistent database at a spatial, spectral, and temporal resolution that would be useful to resource managers. The Landsat TM is a line scanning device which detects energy in seven spectral intervals or bands including six in the visible and reflected infrared ranges and one in the thermal range (Table 1). Energy detected by the TM is quantized to 256 levels or digital numbers (DNs, zero to 255). The effective ground resolution (or instantaneous field of view) of the TM detectors is approximately one-fifth acre (1/5 ac) in the reflective bands. The thermal band, with its much coarser resolution of about 3.5 acres, is generally used only in geothermal investigations. The single sample unit is referred to as a pixel (picture element). Each full TM scene is 6967 pixels wide (columns) by 5965 pixels high (rows). (For computing purposes, this translates to 6967*5965 = 41,558,155 bytes of data per scene *per band.*) For digital or computer assisted processing, Landsat data are available on magnetic computer compatible tapes (CCTs). Digital image analysis provides a means of more easily incorporating information derived from remotely sensed data into existing spatial databases and of handling data for very large areas.

Table 1. Characteristics of Landsat TM data.

Band Number	Wavelength (μm)	Spectral Region	Resolution (m)
1	.45 - .52	Blue	28.5
2	.52 - .60	Green	28.5
3	.63 - .69	Red	28.5
4	.76 - .90	Near Infrared	28.5
5	1.55 - 1.75	Mid-infrared	28.5
7	2.08 - 2.35	Mid-infrared	28.5
6	10.40 - 12.50	Thermal	120

Landsat 5 TM coverage of any geographical area is repeated every 16 days. The satellite orbit is polar and sun-synchronous; that is, the satellite crosses the equator at about the same local time (approximately 9:30 a.m. standard time) in each north-south transit. Repetitive, sun-synchronous, synoptic coverage at relatively high spatial and spectral resolutions make the Landsat TM a logical and useful source of data for resource inventory and management.

Acquisition of TM Data

Landsat 5 TM digital data were acquired from the Earth Observation Satellite Company (EOSAT) in Lanham, Maryland (Table 2). Major portions of five Landsat scenes were needed to cover all of the A/P drainage basin with the exception of small areas in Johnston, Wayne and Lenoir Counties in North Carolina and Charlotte and Mecklenburg Counties in Virginia. Image corner point coordinates from a TM image search plotted over an outline of the A/P drainage basin were used to determine the best possible coverage with the fewest number of scenes. Missing areas comprise less than 4% of the total basin (Figure 2). Winter imagery was expected to provide the best discrimination between vegetation categories. Cloud cover made the 1989 TM winter scenes unsuitable for use over much of the drainage area, therefore five scenes from winter of 1987 and 1988 were acquired (Table 2). Some clouds and haze still occurred in one scene (Path 14/Row 36, Neuse River estuary and surrounding areas).

Table 2. Scene information.

WRS* Path/Row	Date	Approximate Area of Coverage
14/36	12/05/88	SE section of A/P basin; Neuse River estuary
14/35	11/03/88	NE section of A/P basin; Pamlico River to southern Virginia Beach
15/35	11/24/87	Middle & Upper N.C. Coastal Plain; Suffolk Scarp west to Fall Zone
15/34	01/08/87	Virginia; Suffolk west of Lunenburg Co.
16/35	12/03/88	Western portion of A/P basin; Raleigh

* Landsat Worldwide Reference System

Acquisition of Ancillary Data

National High Altitude Photography (NHAP) aerial photographs were used for selecting training sites and verifying classification accuracies. The NHAP program (now known as the National Aerial Photography Program or NAPP) is a multi-agency Federal program which provides complete quad-centered (USGS 7.5 minute topographic quadrangles) aerial coverage of the U.S. in color infrared photography. NHAP photography of the North Carolina portion of the A/P drainage basin was provided on loan from the U.S. Forest Service. Stereopairs acquired in the spring of 1982 and 1983 at a scale of 1:58000 provided complete east-west coverage of quadrangles and coverage of every other quadrangle north-south. NHAP photographs were not available for the extreme eastern part of North Carolina (Currituck County and part of Dare County, including most of the Outer Banks). Black and white aerial photography acquired in 1981 at a scale of 1:20000 by the USDA Agricultural Stabilization and Conservation Service (ASCS) was used for some of these areas.

Seven and a half minute orthophotography and topographic quadrangles of the Virginia portion of the A/P drainage basin were ordered from the U.S. Geological Survey. State road

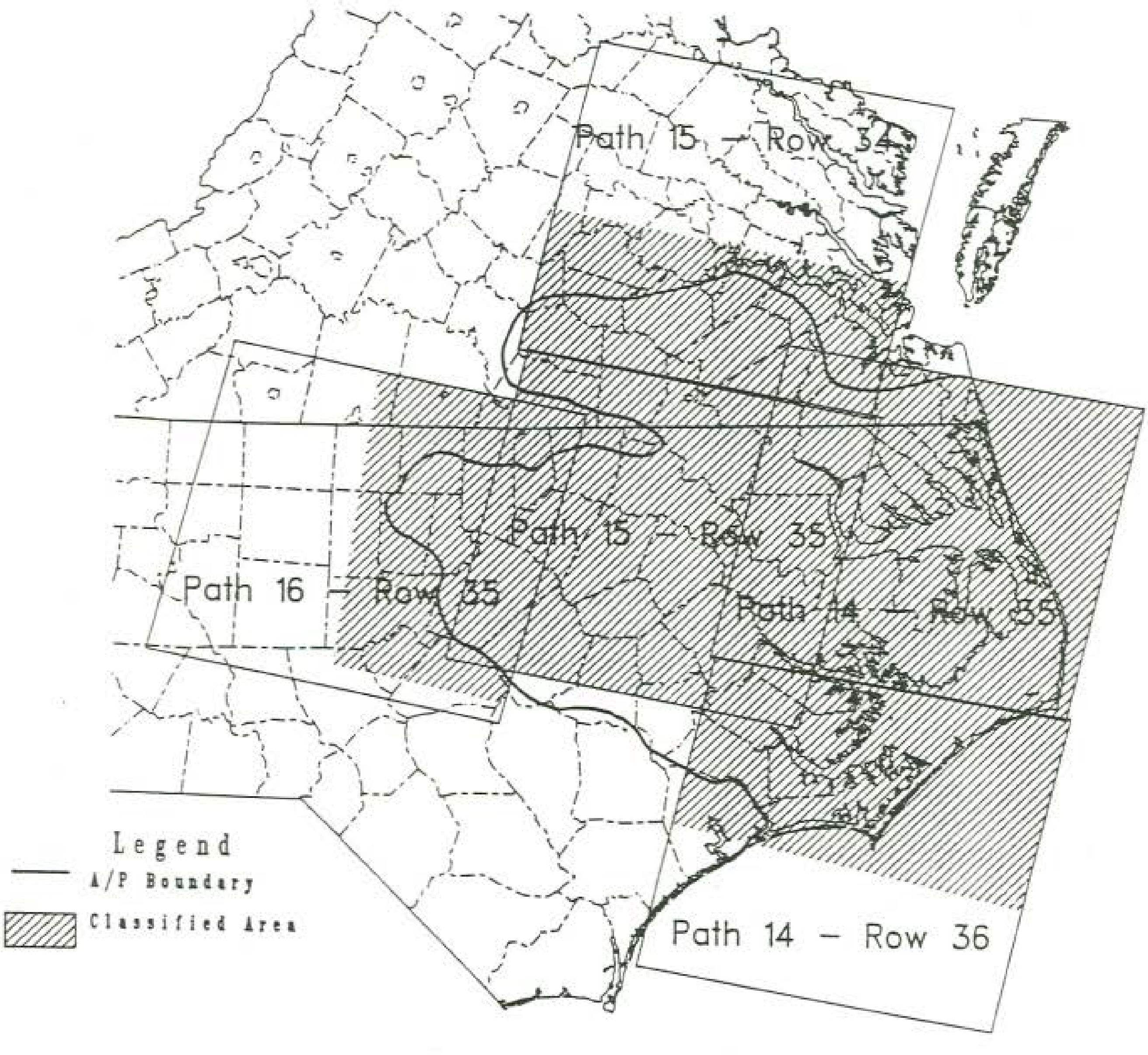

Figure 2. Landsat TM coverage of the A/P drainage basin. Shading indicates areas actually classified by land use or land cover.

maps (N.C. Department of Transportation) were used in the lab as a guide for determining the accessibility of areas for ground visits and were used in the field to record site locations.

The difference in acquisition or production dates of maps, photos, and satellite coverages created some problems in image interpretation, but it was not economically feasible to obtain up-to-date photo coverage of the entire A/P drainage basin. In many cases, the nature of changes was readily apparent. For example, project participants were well aware of explosive residential and commercial development on the Outer Banks in the last five to eight years (which did not show up in the ASCS photography) and timber harvesting operations accounted for numerous cleared or lightly vegetated areas which previously appeared as mature forest stands (usually pine). Ground truth data were acquired for areas where identification or verification was difficult to determine from existing data sources. Field verification was also necessary for those areas which had undergone recent changes to an unknown use or uses and for assessment of confusion areas.

In addition to existing published data, other researchers and resource managers familiar with localized (or, in some cases, extensive) areas within the A/P basin were frequently consulted for first-hand knowledge of ground conditions. Cooperation from these individuals or agencies not only provided cost effective information to the project but also helped guide selection of the final land use and land cover categories.

PRELIMINARY DATA ASSESSMENT

Radiometrically and geometrically corrected TM computer compatible tapes (TIPS format, corrected or 'P-data') were initially read to verify areas of coverage and scene quality. Prior to analysis, areas known to be outside of the drainage basin were dropped. TM scenes acquired over different areas on different dates are not directly comparable. Variations in environmental conditions as well as small changes in sun angle, satellite attitude or sensor characteristics result in variations in the quantized data (digital numbers). Digital numbers can be converted to radiance values (actual units of energy detected in watts per steradian per square centimeter) or histograms for each scene can be "corrected" in some fashion (e.g. all bands of all scenes could be normalized to the same dynamic range). These procedures will not account for all variations such as those due to atmospheric or other environmental conditions or those due to changes in sun angles. In addition, the signature of a given cover type can vary considerably from one locale to another (Walsh 1987; Karaska et al. 1986). Most of the variation in the spectral properties of a given cover type can be attributed to variations in underlying soil properties, antecedent or indigenous moisture conditions, local topography, or slight differences in climatic regimes. Changes in physiography within the A/P drainage basin as well as the normal variations between scenes were expected to cause variations in the spectral properties recorded by the TM. For these reasons, a decision was reached prior to data analysis to classify each scene and each physiographic province separately.

Initial File Preparation

Windows roughly corresponding to physiographic provinces within each scene were extracted and stored in separate files. When deemed necessary, files were further divided to reduce file sizes. Standard naming conventions were used to simplify file access and to improve record keeping. A total of nine files covering three physiographic areas were thus analyzed: Tidewater (three files); Middle and Upper Coastal Plain (five files); and Piedmont (one file).

Individual files were converted from Space Oblique Mercator to Lambert Conformal Conic projections using a two degree polynomial regression derived from image and state plane coordinates for a series of control points and information encoded on TM CCTs. The Space Oblique Mercator is a cylindrical map projection defined by satellite orbits and the Lambert Conformal Conic is a conic map projection used as a basis for state map series. The geometric transformation was implemented with a cubic convolution interpolation for resampling. Cloudy areas (WRS Path 14/Row 36, only) were masked to a brightness level of zero (0).

Work was initiated on reformatting software so that image files could be transferred from CGC to CGIA. CGC uses the Land Analysis System (LAS) image processing software, developed by NASA Goddard Space Flight Center, running on a VAX super mini-computer under the VMS operating system. (VAX and VMS are registered trademarks of Digital Equipment Corporation.) CGIA operates the state Geographic Information System (GIS) using ARC/INFO software, a high performance GIS produced by Environmental Systems Research Institute, Inc (ESRI). CGIA also uses image processing and GIS software produced by the Earth Resources Data Analysis System, Inc (ERDAS). The ARC/INFO and ERDAS software at CGIA are installed on SUN Microsystems computer workstations (SPARCstations). Though both ARC/INFO and ERDAS software are being used, most databases at CGIA (including the A/P database) are built, stored, and maintained using ARC/INFO. In addition, most databases at CGIA (including the A/P database) are stored using the North Carolina State Plane Coordinate System for referencing

Reformatting software was needed to allow translation from the LAS data format (at CGC) to the ERDAS data format (at CGIA). The software was specifically designed to generate a file that was stripped of both control and header records and then reformatted to an acceptable import format for the ERDAS software. The CGC software also generated new control data such as file size and reference state plane coordinates which were variables that were required to complete the translation at CGIA.

Prior to image classification, tests were conducted to determine (1) which bands or band combinations would provide the best spectral separation between cover types of interest, and (2) what approach (or approaches) would be best to use for image classification. Tests were conducted using a series of small windows (512x512 or 1024x1024 pixels) from the Tidewater region (Currituck County) and from the Piedmont (Wake County).

Test of Approaches

Two basic strategies for land cover classification were tested for their applicability to this project: the supervised and the unsupervised. Within each of these classification strategies, data masking and data conversion techniques were tested as means of improving classification details and/or accuracies. All approaches were based on a minimum-distance-to-means classifier. A minimum distance classifier assigns a pixel to the cluster whose mean falls the shortest distance from it. Euclidean distance is most commonly used as a measure of distance in multi-dimensional space. Previous experience in other land use/land cover classification projects had shown that a minimum distance classifier was less computationally intensive and provided results better than or nearly identical to other classifiers. Processing times were as much as an order of magnitude greater when using a maximum-likelihood supervised approach which uses probabilities to assign unknown pixels to a cluster.

Unsupervised classifications involve the use of clustering algorithms that examine large numbers of pixels and divide them into spectrally distinct clusters based on natural groupings of the pixels in n-dimensional space (where "n" is the number of input bands). The resultant clusters are spectrally distinct but initially unknown in terms of their categorical identity. The analyst must compare classified data to some form of reference data to determine the identity of each of the spectral classes. Using an unsupervised approach, the analyst generally sets a limit on the total number of spectral classes and inputs the bands to be used. Clustering is then totally automated and ends when the specified number of classes is reached. "Guided clustering" is a variation of an unsupervised approach which gives the analyst more options to control the clustering such as deciding which cluster to split at each iteration. Both the totally unsupervised and guided clustering were tested. Statistical summaries of an unsupervised clustering performed on a file window (or image subset) can be used as input to a supervised classification.

In a supervised approach, the analyst chooses the classes or land use/land cover categories that he or she desires and then selects training areas that represent each category. Statistics derived from the training data for each category are then used as a basis for classification. Training areas define categorical spectral response patterns which are used as keys by which unknown pixels are assigned into their appropriate classes. Thus, in the supervised approach, categories are defined and then their spectral separability is determined. In the unsupervised approach, spectral separability is determined and then categories are defined (Avery and Berlin 1985).

In both the supervised and the unsupervised approaches, *spectral class* or *cluster* refers to vectors which describe the inherent grouping of spectral values (digital numbers) in multidimensional space. *Class* or *category* refers to terrestrial types or conditions which may be characterized by one or more clusters.

Data masking involves splitting images such that only a portion of an image is analyzed at one time. That portion of the image which is not analyzed is said to be "masked out" or "turned off" (i.e. pixel values are temporarily set to a known value, usually zero). The criteria for masking can vary. In these tests two criteria were used:

(1) Masking based on an initial unsupervised classification. In this approach, an unsupervised classification is used to separate an image into a few spectral

categories representing some minimal feature separation (e.g. there may be 12 spectral clusters representing three features or environmental associations - vegetation, water and other). Image pixels, in one or more of the original bands, which correspond to locations of spectral clusters for one feature are left "on" while the other pixels are masked out. Classification is then carried out on just one of the basic features identified in the initial classification. This approach permits more detailed analysis within a particular feature and it permits the analyst to select different bands when clustering different feature groups.

(2) Masking based on radiometric value or digital number. In this approach, one band of the original data is used as a source of the mask. A digital number or range of digital numbers are specified to be left "on" while pixels with digital numbers other than those specified are masked out. Additional bands are masked based on the *locations* of the pixels masked in the source band. Because basic features (like vegetation or water) can frequently be separated on the basis of a single band, this approach has almost the same advantages as the first approach and it requires less processing time.

Principal components analysis (PCA) was tested as a means of reducing the amount of data which had to be used in the classification of image data. In principal components analysis new variables (the components) are created from a number of original variables using a simple linear function. The linear function is based on correlations (or covariances) among the original variables and each component generated "explains" a decreasing proportion of the variation amongst the original data. Thus, the first principal component generated from six Landsat TM bands would contain the maximum covariation contained in the original bands and describe, as a linear function, the common element in the bands. (Unfortunately, PCA will not tell the analyst exactly what that common element is in terms of anything other than a mathematical function.) PCA is frequently used as a means of reducing the amount of data which need to be analyzed while retaining the informational content of the original data set. One may, for example, classify TM data using the first two principal components instead of using three or more bands of the spectral data.

Tests of Bands

Since tests of all possible bands or band combinations would involve an impractical number of trials, tests were limited to those bands or band combinations which had been shown to be useful in other land use/land cover projects conducted by researchers at the CGC and elsewhere. In these tests, "band combination" refers to the combining of two or more bands to form a new output variable which is then used as input to the classifier with or without other bands. For example, creating a band ratio (e.g. band 5 divided by band 4) is a common method of combining two bands to form a single new variable.

Use of a single band was immediately excluded from consideration since ample literature was available to indicate that a single band, though often useful for discriminating general categories, would not allow separation of classes to the level of detail required. The thermal band (TM band 6) was excluded because of its poor spatial resolution (Table 1). TM band 1, corresponding to the blue region of the visible spectrum, was not considered in original tests because spectral response in this band tends to be dominated by atmospheric scattering. Of the remaining five TM bands, two or three band (or band combination) classifications were preferred over use of four or five bands because of the exponential increase in processing times associated with increasing the number of input bands.

Two combinations of three bands (TM bands 2, 3, and 4; TM bands 3, 4, and 5) have been used successfully in a number of land use/land cover classifications particularly those classifications which were performed in relatively wet areas (Janssen et al. 1990; Ma and Olson 1989; Trolier and Philipson 1986). From the beginning, these band combinations were considered the most likely to be of value and they were used as a basis of comparison for the other combinations tested. In addition, correlation matrices for TM band responses within training sites were generated as part of the tests of the supervised approach. Correlation matrices were used to evaluate the degree of duplication in band responses within given categories. A high correlation between two bands within a training site would indicate that the bands are providing duplicate spectral information about the land use or land cover category represented by that training site.

Assessing Classification Strategies

Preliminary results from the testing of classification strategies and input bands were assessed based on several criteria. The most obvious requirement was that the approach provide accurate results with sufficient categorical detail to be useful to researchers and resource managers. It would have been too time consuming to perform formal error analyses (including generating error matrices) on all combinations tested, therefore, when necessary, a limited number of sites were compared with ground truth data (aerial or ground photography) to evaluate accuracy and detail. For several combinations of approaches and/or bands, results were visibly inadequate. When testing the use of four input bands versus three input bands, a difference algorithm was used to determine how many pixels actually changed from one class to another and what the direction of change was.

Criteria other than accuracy were considered equally important for project completion. In conducting a study covering such a large area, involving more than one or two analysts, and requiring a number of steps to complete, it was felt that it would be important to maintain an approach which could be kept consistent among areas and among analysts. A consistent approach would be more likely to provide consistent and repeatable results. At the same time, the approach needed to be flexible to the extent that there was such diversity within the land use/land cover categories to be mapped. Numerous studies have been conducted showing a wide range in results due to analyst variability alone. Consistency in approach from one physiographic province to another would also make it easier to achieve consistency of detail and to merge individual scenes into a seamless regional coverage.

Results of Preliminary Data Assessments

The unsupervised approach was not selected as a viable approach primarily because of an inability to achieve consistent results within and between the various physiographic provinces. Clustering algorithms may utilize different criteria for determining when a cluster should be split (or merged, depending on the algorithm used) but many, like the one used here, rely on band variance as a measure of diversity within a cluster. The splitting process is based on natural groupings of pixels. When the spectral variances of the classes are not equal, the unsupervised classifier tends to split classes with high variances into more and more clusters, many of which have very few pixels, while classes with low variances and similar spectral responses remain clumped together. This was particularly problematic in coastal areas where high moisture contents dominated the spectral response in several categories. Drier areas tended to split into greater and greater detail (bare sandy area, sandy area with a tree, sandy area with some grass, sandy area and road, etc.) while major categories like riverine swamps, marshes, and water remained in a single cluster. It was often necessary to generate 100 or more spectral clusters before land use/land cover categories of interest could be adequately differentiated.

Between physiographic provinces, the categories to be identified varied (e.g. certain categories of wetlands which occurred in the Tidewater province were not found in the Piedmont). However, there was considerable variation in the number of clusters required to identify the categories of interest which occurred in all physiographic provinces. Even within provinces, particularly Tidewater, the number of clusters needed to adequately discriminate the categories of interest within a fixed window size varied depending on the characteristics of the areas. Post-clustering identification of categories was also difficult, not only because discrimination of the categories varied, but also because of variations in analysts' knowledge about or familiarity with the areas. Changes in land use/land cover from the time of the aerial photography were difficult to spot on the clustered images.

Rather than helping, guided clustering tended to compound the problem in that it required instantaneous interpretation of the clusters thus relying heavily on the analysts' knowledge and background. Guided clustering also tended to rely on the analyst's ability to precisely locate reference features on the clustered image as the clustering progressed. This is not always easy and is especially difficult when the study area is very diverse. Interactive clustering can be a highly subjective procedure and result in wide differences in output between analysts. Classifier statistics generated from both unsupervised clustering or from guided clustering on different windows within a scene tended to disagree.

These problems can be typical of unsupervised classifications particularly when attempting to apply this approach to very large areas using more than one or two analysts. It was felt that a more structured approach would alleviate many of these problems and provide a more consistent framework for the classification process.

When using the same band combinations, the supervised approach gave visibly better agreement between the image classes and the reference data than either the totally unsupervised or the interactive guided clustering. Classification improvements were particularly noticeable in vegetation classes that had low variations in spectral response within each class but which were spectrally similar (such as pine and pine/hardwood mix).

Several advantages were noted in the supervised approach:

(1) It was easier for analysts, even those unfamiliar with the area, to identify categorical features directly on the aerial photography than to interpret the unsupervised output for identification of those same features;

(2) Changes in land use or land cover which had occurred between the time the reference photographs were taken and the dates the images were collected were generally readily apparent on the TM false color composites and were more likely to be found during location of training sites;

(3) Though the identification of training sites was a lengthy and sometimes tedious procedure, it was easier and less time consuming to modify training statistics as needed than to entirely re-do an unsupervised or guided clustering to generate improved statistics;

(4) More control could be exercised over the identification of features to insure that overlap areas between scenes were classified the same; and,

(5) Results were more consistent between scenes and physiographic provinces.

The supervised approach provided the structure needed to maintain consistency but it also offered flexibility. The number of training sites or spectral classes needed to statistically describe the land use/land cover categories of interest could be varied depending on feature or image characteristics. One analyst could review all training data before they were applied to the individual files both to verify that the training data were complete and to insure that features were consistently defined. It would have been impractical for one analyst to oversee multiple runs using the unsupervised approach and it would have been difficult to maintain commonality in feature definition. When combined with data masking (see below), different band combinations could also be incorporated into the classification procedure.

In tests of band combinations, classifications based on two input bands (or band combinations) gave visibly poorer results than the use of either three or four input bands. Input of more than three bands were found to significantly increase processing times with negligible changes in output. Comparison of the number of pixels which actually changed classes when using four or five bands versus three bands (or four) revealed a very low percentage of change (less than 1% to 2% depending on the area).

Data masking proved to be a valuable enhancement to the classification procedure. Using the unsupervised approach, masking and subsequent analysis of individual categorical groups greatly improved classification of individual image windows. However, it was still difficult to maintain consistency in results because of the high diversity in cover types and conditions. Intermediate products (e.g. the initial clustering to differentiate the categorical groups or source for the masking step) often differed considerably even within a scene. In the supervised approach, data masking based on an original single TM band was used to separate areas having high digital numbers (high reflectance) from areas having low digital numbers (low reflectance) prior to application of training statistics. This step appeared to minimize confusion between some categories, particularly those with intermediate reflectances (see PROJECT METHODOLOGY for a more complete discussion of this step).

Principal components analysis had been used previously by researchers at the CGC for data reduction when doing land use/land cover classification for an area around Raleigh.

However, test image windows of the Coastal Plain did not indicate any advantage in using principal components as input to a classification of this region. Over test windows encompassing wetland areas, the first principal component generally accounted for approximately 60% of the variations among TM bands. The remaining variation was distributed fairly evenly among the next four components. A minimum of three components was required to achieve good classification results. Because PCA required numerous additional steps, adding complexity without providing improvements in the classification or reductions in the amount of data needed, this approach was not adopted for use.

Based on statistics from training data, bands two and three were consistently highly correlated as were bands five and six. Inputs which included combinations of these pairs of bands were excluded since high correlations between bands are indicative of redundant information. Band six contained a very low range of spectral values which implied that it contained minimal discriminatory information. This was supported by two-way band plots of training data. Band combinations (5/4 ; 3/4 ; and 2/4) did not provide any improvement in the classification and merely added an extra step to the procedure. Both band ratioing and principal components analysis tend to be scene dependent (Sheffield 1985).

Input of bands 2, 3 and 4 and bands 3, 4 and 5 gave similar results, but, as previously indicated, bands 2 and 3 were highly correlated in all feature categories. Inclusion of band 5 provided better discrimination between vegetation types. Bands 4 and 5 were highly correlated in approximately half the categories indicating that the informational content of these two bands was not redundant in about half the training areas. Band 5 has been shown by other researchers to be particularly useful for vegetation moisture measurement and band 3 was included in the TM design for use in detecting chlorophyll absorption for plant species differentiation. Band 4 has been most frequently used for delineating water bodies. The TM bands 3, 4 and 5 were thus selected as the best input parameters for the classification.

PROJECT METHODOLOGY

Landsat TM image files, each located in a single physiographic region, were converted to a projection compatible with the North Carolina State Plane Coordinate System, which is the coordinate system used at CGIA for the A/P database. Each file was classified separately at the CGC using a supervised approach. The same steps were used to classify each image file but differences in the TM scenes and in regional physiography required different training statistics for each image. Classified images were renumbered to a standard numbering scheme and converted to a format which could be read by the ERDAS image processing software at CGIA. Using ERDAS software, the images were filtered using a 5x5 mode filter and then converted to the ARC/INFO format for final A/P database processing. (Note: An ERDAS to ARC/INFO translator provided in the ARC/INFO software package was used to convert to the ARC/INFO format.) A more complete description follows.

Identification of Spectral Classes and Training Sites

Before beginning preliminary testing of classification approaches, a list of land use/land cover categories to be identified was made using the USGS Anderson classification Level II (refer to Appendix II) as a guide. Initial modifications to the list included: substituting Low-, Medium-, and High-Density Developed for the Anderson Level II classes within Urban and Built-up Land; dropping all Level II classes for Agricultural areas and Water; and, expanding the Wetlands classes to include more specific (Level III) cover types (e.g. Low Marsh and High Marsh in place of Non-forested Wetlands). The Level I categories Rangeland, Tundra, and Perennial Snow or Ice were not used.

Before beginning image analyses, the list was expanded to include anticipated spectral variations for each category. For example, the following descriptions were written for Forestland, Level II classes Pine, Hardwood, and Mixed Pine/Hardwood:

1. Low density pine - mature pine stands with 25-50% crown closure;
2. Medium density pine - mature pine stands with 50-75% crown closure;
3. High density pine - mature pine stands with > 75% crown closure;
4. Young pine - immature pine stands generally less than 10-15 years old with medium to high density crowns above the understory;
5. Bottomland hardwood - hardwood stands occupying topographic lows or in floodplains which are not seasonally flooded or saturated;
6. Riparian hardwood - same as above but with a heavy component of evergreen understory and generally found directly along the banks of streams and rivers;
7. Upland hardwood - hardwood stands found on dry ridges;
8. Other hardwood - hardwood stands that do not fall into the other descriptions;
9. Pine/Hardwood - mixed pine/hardwood with pine comprising 51-75% of the crown density; and,
10. Hardwood/Pine - mixed pine/hardwood with hardwood comprising 51-75% of the crown density.

Prior to beginning location of training sites, over 40 potential spectral variants of 14 initial Piedmont classes had been described. This approach was adopted because of the great variation in spectral properties of certain cover types and land uses found during initial testing. Spectral classes or clusters were intended to represent all unique spectral characterizations of the final classes. Interpretation of Low-, Medium-, and High-Density Developed was based on the amount of impervious surface visible from the vertical perspective. Using this guideline, a residential area, for example, under a nearly closed canopy of trees would be considered "Low-density".

Using every other (east-west) NHAP stereopair, a minimum of three training sites were located on the photography for each expected spectral class. Use of every other east-west stereopair resulted in analysis of one out of every four 7.5 minute quadrangles within most of the A/P drainage basin. Some potential spectral classes were modified during photo interpretation. For example, the anticipated spectral class "Low density pine" was essentially non-existent. Pine "forest" with very low density actually tended to be pines in association with some other land use or land cover, such as landscaping trees in a residential area or scattered pines in a pocosin environment. Additional spectral categories were added as needed, mostly to deal with variations in the spectral characteristics of developed areas (e.g. asphalt vs. concrete). When necessary, ground conditions were verified by field visits or by reference to other existing data sources. An effort was made to distribute the three or more training sites throughout the area covered by each image file. In overlap areas between scenes the same training sites were used for processing both scenes.

Training sites were visually located and delineated on the imagery using a standard false color composite for display (TM bands 2, 3 and 4). Training site polygons encompassed uniform areas and ranged in size from 10 to 75 pixels (about 2-15 acres). Each interactively located training site polygon was uniquely labeled and a record was kept of the label name, the photograph (or orthophoto) on which the site had originally been identified, the class that site represented, and an image line and pixel by which that site could be readily relocated. Statistics were generated from the digital numbers for TM bands 3, 4 and 5 within each polygon.

Duplicative training sites were used to qualitatively assess interpretive and spectral variability within proposed spectral classes. Means and standard deviations of band responses in TM bands 3, 4, and 5 were compared in the three or more training sites within each spectral class. Large differences (> 10 DNs) between site means in one or more bands or high standard deviations (> 5 DNs) within a site were presumed to indicate that a spectral class (or cluster) was being poorly defined. Discrepancies could be attributed to one or more problems:

1. The training site might not represent a homogeneous area;
2. A training site may have been incorrectly interpreted on the photography;
3. A training site may have been located incorrectly on the imagery;
4. A change may have occurred in the area being used as a training site and the change was not detectable on the false color composite; or,
5. One or more of the training sites might represent a new spectral variant of a given class.

Questionable training sites were re-examined to determine the cause of discrepancies. If the reason for the discrepancy could not be determined, the site was dropped. Sites were relocated or their identification corrected when appropriate. If the site was found to be correctly identified but representative of a new spectral variation of a given land use/land cover class, a new spectral class (cluster) was created, similar areas were identified elsewhere in the imagery and additional training sites were delineated for comparison.

Compilation of Training Statistics

After locating and checking the validity of training sites, final cluster statistics were generated for each spectral class. Means and standard deviations were calculated from pixel DNs over all training sites within a given spectral class. That is, each cluster had one mean and standard deviation for each TM band regardless of how many training sites were delineated for that spectral class. Because there were a large number of clusters, statistics files were subdivided to make it easier to examine cluster statistics. The number of clusters varied from 27 to 66 for each image file and each cluster was represented by a minimum of two training sites.

Two-way plots of cluster means and variances were produced for band 3 *vs.* band 4, band 3 *vs.* band 5 and band 4 *vs.* band 5. Two-way plots were preferred because plots representing all three dimensions were difficult to interpret and the perspective could frequently result in hidden clusters. Two-way plots consisted of ellipses representing the variance/covariance information drawn around the cluster means. The x and y distances from the mean to the surrounding ellipse represent the variances of the two bands while the length of the major axis and the angle of the ellipse represents the covariance between the two bands.

Plots were examined to determine cluster separability. Clusters which were represented by other clusters (i.e. the spectral classes had identical or nearly identical means in all bands) were dropped. Training sites for clusters which overlapped in all plots were examined to determine if there were errors or if a class could be better defined. Overlapping ellipses could be indicative of potential misclassifications. When appropriate, clusters which had similar means and a large overlap in variance were combined. For example, pine/hardwood (pine dominant) and hardwood/pine (hardwood dominant) were combined to form a single cluster representing the spectral response of all mixed pine/hardwood stands. Clusters which had overlapping ellipses but which represented a continuum (e.g. low-medium-high density developed or pine-mixed pine/hardwood-hardwood forest) were retained.

Polygonal training sites were inadequate for representing certain training statistics. This occurred when a potential training area was too small or narrow for a polygon to be accurately drawn around it. For example, certain large buildings which had very high reflectances tended to be confused with sand or bare dry soil but were too small to delineate. Roads or runways only 2 or 3 pixels wide were also difficult to delineate as were low-density developed areas (frequently irregularly shaped or covering less than ten pixels). Single pixel values for three to five sites were extracted for each feature which had to be represented (Gong and Howarth 1990). Pixel values for a given feature were compared, averaged and plotted on the two-way plots to assess possible confusion areas. The means were then entered into the statistics file to form a new cluster.

Finally, two-way cluster plots also provided a means for quickly determining if clusters adequately covered the spectral variability within an image. If large "gaps" occurred in cluster statistics it could indicate that some cover type or condition had been missed. Image histograms could be examined to determine the frequency of occurrence of "missing" DNs; however, a visual representation was easier to interpret. Single image bands were displayed in black and white and the missing DNs were turned to a contrasting color (red). A few scattered red pixels indicated the missing DNs were spurious data (class outliers or image "noise") while large blocks of pixels indicated that a land use/land cover class (or spectral variant of a class) had been missed. Additional training sites were delineated as needed and cluster statistics were added to the statistics file.

While examining initial training statistics, it was noted that classes with relatively high means also tended to have higher band variances while those with lower means also had lower variances. This was true for all three TM bands for test areas in both the Piedmont and Tidewater. A discriminant function was used to determine the DN representing the boundary value between "high" and "low" mean DNs. TM band three was used as a source for masking areas in all three bands (bands 3, 4 and 5). This resulted in each input band being split into two parts corresponding to areas with relatively high digital numbers (DNs > 30) and areas with low digital numbers (DNs < 31). Statistical files were also split into two parts for spectral classes or clusters with low DNs and those with high DNs. Clusters with occurrences close to the boundary value were included in both statistics files.

Image Classification

Before classifying a full image, test classifications were performed on two or three test windows (512x512 pixels) corresponding to NHAP photographs and statistics files were edited as needed to add or delete clusters. A k-means minimum distance classifier was used to perform classification of each image file. Input to the classifier was a multiband image file composed of masked bands 3, 4, and 5 and the statistics file. The classifier was run twice - once for areas with high DNs and once for areas with low DNs. Classification was performed by assigning pixels to the cluster they were closest to based on Euclidean distance between pixel band values and cluster mean band values. The number of clusters in any one classification varied from 27 to 66.

After every pixel had been assigned to one of the input clusters, the two output images were renumbered and added together to form a single coverage numbered zero to N, where N was the total number of input clusters for high and low DNs. For most files in the Tidewater and Coastal Plain provinces, confusion between some of the areas to be classified as Agriculture and areas to be classified as Developed were apparent after this step. In one file (Path 14/Row 35, Tidewater) there was also confusion between high marsh and developed areas. Both types of errors occurred because of the similarity in spectral responses of dry agricultural fields, dry matted vegetation and certain building materials such as concrete. In the Piedmont, shadows were classed as water (low DNs) or as border (DNs of zero). Output images from the supervised classification were masked to retain areas corresponding to clusters in which confusion existed. Training statistics were redefined for these areas using polygonal and single pixel values. All bands, including bands one and six, were re-examined in these new training sites to determine if different input bands would be better for differentiating confusion areas. The most appropriate input bands were selected and the minimum distance classifier was re-applied to the selected areas. Output images were renumbered and added back to the data which had been previously masked out.

These coverages were subsequently renumbered to the appropriate final class numbers by assigning one or more clusters to each class found in the area covered by an image file. The number of classes varied from 13 to 20 (including image borders) depending on physiographic province (Table 4, PROJECT RESULTS).

Classification Accuracy Assessment

The upper photograph of every other NHAP stereopair was selected as "ground truth" for the verification. For each photograph a corresponding window was selected from the classified image file, displayed with a magnification factor of two (approximate scale of 1:58000) and photographed. A minimum of one verification sample was taken for every 75,000 pixels in the full image. The actual number of samples taken in any one class was weighted by class occurrences estimated from the image file being sampled. Pixels classed as "Border" or

"Shadow" (Table 4, PROJECT RESULTS) were excluded from the classification verification as nonvalid classes (unverifiable).

One acre sample plots were located on the NHAP photographs using a computer generated overlay of random plots. The same plots were located on prints of the classified image window using a zoom transfer scope to overlay the image data and NHAP photographs. Photo class and image class were recorded on a tally sheet. Sample numbers were recorded on the tally sheets and on the images. When necessary, field verification or other reference data were used to confirm ground truth data. Information from the comparison of image and ground (photo) data was used to construct error matrices or confusion tables. Classification accuracies were calculated for each separate physiographic province and for the entire study area.

A/P Database Integration

A primary objective of this project is to provide investigators, resource managers, government agencies, and other members of the A/P community a current inventory of land use/land cover through the development of land use/land cover data layer. The land use/land cover data layer is one of numerous (more than 65) geographic data layers that are being developed at CGIA for the A/P Database. Some examples of the other layers include hydrography, roads and trails, railroads, citizen water quality monitoring sites, discharge permit sites (NPDES), and county soil surveys.

In order for the full functional potential of the land use/land cover layer to be realized, additional processing beyond image classification by CGC was needed. The data format needed to be compatible with the other data layers in the A/P Database for simple query and display purposes as well as high end spatial analysis. Therefore, image data needed to reside as final data files in the ARC/INFO data format, a vector based GIS. The ARC/INFO software also facilitates data distribution through the use of common standard data exchange formats.

Data layers at CGIA are stored as ARC/INFO coverages (geographic data files). Data layers that require large amounts of data storage are stored as a series of coverages, usually partitioned by a common map window such as the USGS 1:24,000-scale topographic map boundary. These partitions are called tiles. Tiles are usually created to permit more efficient management of the data. The land use/land cover data layer has very large data storage requirements and therefore has been tiled by the USGS 1:100,000-scale map boundary. The map boundaries for each of these maps are delineated every one-half degree latitude and every one degree longitude (Appendix I). A summary of procedures used after image classification at CGC follows.

Header and control information for each classified image in the LAS data format were reformatted at CGC using the reformatting software. The data were copied to 9-track magnetic tapes and transferred to CGIA for data translation. At CGIA, the image data files were loaded onto the computer system and imported into the ERDAS data format. ERDAS software was needed to 1) resolve the discrepancies along the edges of scenes, 2) filter the image data, and 3) act as an intermediate data format to allow translation of the data from the raster based data format in LAS to the vector based data format being used in ARC/INFO.

In order to resolve pixel classification discrepancies along overlapping scene edges, geographic subsets were reselected from the image files and processed using the ERDAS program

'STITCH.' The subsets were at least 25 pixels greater in extent than the USGS 1:100,000-scale map boundaries in order to augment proper filter processing along the USGS map edge. The 'STITCH' program requires that one of the two scenes along an edge be given higher consideration in the pixel re-classification process. Scene priorities for 'STITCH' are identified in Table 3.

Table 3. Scene priorities for ERDAS program 'STITCH.'

SCENE TYPE	SCENARIO 1	SCENARIO 2	SCENARIO 3
Primary Scene	Path 14, Row 35	Path 15, Row 35	Path 16, Row 35
Secondary Scene(s)	Path 14, Row 36 Path 15, Row 35 Path 15, Row 34	Path 14, Row 36 Path 15, Row 34	Path 15, Row 35

The image data were filtered using ERDAS software. Filter mode test plots and reports were generated for the selected images in order to assess and determine the desired filtering mode. Test plots were created for filters at 3x3, 5x5, 7x7, and 9x9 pixel modes and reviewed by staff from CGC and CGIA. A 5x5 pixel filter mode was selected as most desirable and then used on all subsets of the image files. The filtered subset files were then ready to be converted to the ARC/INFO data format for further processing.

The conversion to ARC/INFO was a two step process using two ARC/INFO programs, 'ERDASSVF' and 'GRIDPOLY.' Pixel boundaries were automatically dissolved during this process where classifications of adjacent pixels concurred. In some cases, the dissolve process formed very large polygons that contained numerous single pixel polygons of a different classification within a coverage.

In order to tile the data layer, the ARC/INFO coverages were overlayed with each USGS 1:100,000-scale quadrangle boundary using the ARC/INFO 'CLIP' program. 'CLIP' dropped data outside of each quad boundary (1/2 x 1 degree latitude/longitude). In total, the land use/land cover data layer is comprised of twenty-seven 1/2 x 1 degree latitude/longitude tiled coverages (Figure 8, PROJECT RESULTS; see Appendix I for list of quadrangles).

PROJECT RESULTS

Landsat Thematic Mapper data from fall and winter 1987 and 1988 were used to derive a land use/land cover inventory for 97% of the Albemarle/Pamlico drainage basin using facilities at the Computer Graphics Center (CGC), North Carolina State University. A supervised approach was used to classify the image data. Data were georeferenced to the North Carolina State Plane Coordinate System, filtered and converted to vector format for map and statistics generation.

Information and access to the land use/land cover inventory are available in digital, map, and tabular form on a request basis from the North Carolina Center for Geographic Information

& Analysis (CGIA) through the Albemarle/Pamlico Estuarine Program. The inventory resides as a data layer in the A/P Database at CGIA. In addition, the data are being stored as raw Landsat Thematic Mapper data and as classified ERDAS images. The ERDAS image processing software at CGIA provides the A/P program with the functionality to further maintain, analyze, and update image data for the basin.

Classification Scheme

The final classification scheme (Table 4) was based on initial classes from the USGS classification scheme (Appendix II) and from spectral classes which appeared to be consistently well defined in the image data. Density of developed areas was based on ocular estimates of the amount of exposed concrete or other structures visible from the vertical perspective. Between interpreters, areas with the least vegetation (>85% developed) and the most vegetation (<25% developed) were easiest to identify, consistently. The medium density class had the greatest variability and, therefore, covered a larger range of densities. High-density developed was restricted almost exclusively to major metropolitan areas (Raleigh, Suffolk, etc.) where there was little to no vegetation. Medium-density developed encompassed most transportation corridors which could be resolved in the imagery and areas which had some vegetative cover. Low-density developed areas had more vegetation and fewer structures than the medium-density areas. It is important to note that these classes, based as they are on a vertical view, do not necessarily correlate with the degree of imperviousness of the surface.

Agricultural areas which were in pasture or had some form of grass cover were indistinguishable from other grassy features such as golf courses and large lawns. A decision was made to group grassy areas with the agriculture class because the amount of grassy area in pasture far exceeded the amount of grassy area in other uses. It was anticipated that users of the land use/land cover database could distinguish non-agricultural uses by incorporating ancillary data such as transportation networks and municipal boundaries (see DISCUSSION).

Initially, the class called "Disturbed" was based on training sites established in recent clearcuts (both timber harvest areas and areas cleared for development). Verification of final images revealed that most of the areas classified as Disturbed were in some form of agricultural use. The remainder of the areas were recently cleared areas. The reason for the distinctive spectral characteristics of these areas was not readily apparent but it is suspected that topography tended to be somewhat rougher in these areas (more gullies or uneven terrain) and/or that many of these areas had undergone recent tillage. Time constraints did not permit verification of activities on specific sites at the time of satellite overpass. Bare fields which were in silvicultural use were classed as agricultural since their future use (most had undergone site preparation for planting pine) could not seen in the images.

Forest land was divided into four categories but since the class "Bottomland Hardwood" was restricted to wet areas and areas immediately adjacent to drainages it is more appropriately included in Wetlands in data summaries. The class "Hardwood" is not restricted to upland sites but does occur in drier areas. Wet areas with a closed canopy of pine were classed as "Pine" rather than as a Wetland type. The extent to which this results in underestimation of wetlands is unknown. In the Coastal Plain, many of the closed canopy pine stands were in plantation.

Table 4. Land use/land cover classes for the Albemarle-Pamlico Estuary study.

Class Number	Name	Description
1	BORDER	Areas corresponding to a digital number (spectral value) of zero; includes pixels outside of the classification area and areas obscured by cloud cover.
2	WATER	Lakes, reservoirs, ponds, estuaries and sounds. Also includes streams or rivers wide enough to be resolved by the Thematic Mapper.
3 - 5	LOW, MEDIUM, & HIGH DENSITY DEVELOPED	Residential, commercial and industrial complexes. The three categories correspond roughly to areas where structures and/or pavement cover 25% to 50%, 50% to 85%, and >85%, respectively, of the ground area classified.
6	AGRICULTURE, BARE SOIL AND GRASS	Cropland and pasture, bare and grass covered soils. Includes all land cleared for agricultural or silvicultural activities. Wide transportation corridors (such as interstate highways with grassy medians), beach grasses, golf courses, large athletic fields and other grassy features are also in this class.
7	LOW DENSITY VEGETATION	Areas which have some vegetative cover but are not forested. Fallow fields, cleared areas in early successional stages, and some landscaped residential areas are included in this class. Wide utility corridors (power and communication), some narrow road systems and weed covered spoil piles along drainage ditches also occur in this category.
8	PINE FOREST	Medium and high density conifer stands, predominantly loblolly pine; also includes high pocosins which have a high density (>50% crown closure) of pond pine.
9	BOTTOMLAND HARDWOODS	Hardwood stands found predominantly in the floodplains of streams and rivers. These stands are dominated by deciduous species such as lowland species of maple, black gum, oak, sweetgum, sycamore, birch, elm and ash.
10	HARDWOOD	Hardwood stands found predominantly in upland areas, on gently sloping interstream divides or in drier low lying areas. Stands dominated by oak, hickory, elm and maple.
11	PINE/HARDWOOD	Stands of mixed conifer and deciduous hardwood. Neither pine nor hardwood comprise greater than 75% of the crown density.
12	DISTURBED LAND	Bare fields which have undergone recent disturbance; predominantly agricultural fields and clear cuts but also includes some developed areas such as sites being prepared for construction or around quarries.

Table 4. Land use/land cover classes for the Albemarle-Pamlico Estuary study (cont'd).

Class Number	Name	Description
13	SHADOW AND MIXED PIXELS	In the Piedmont, includes shadows (usually in high density developed areas) and pixels which are a mixture of water and trees (usually bordering lakes and ponds). In the Coastal Plain, many wet areas with organic soils (RIVERINE SWAMP and LOW MARSH) were confused with shadows.
14	RIVERINE SWAMP	Forests occurring along the major Coastal Plain rivers and their tributaries and on sites associated with nearly permanent fresh or brackish water. These riverine swamp forests usually occur in the floodplains of rivers or on wet flats. Dominated by gum-cypress or gum-maple swamps, but also including maple, birch, sycamore, sweetgum and oaks. This class tends to be mixed with SHADOW (Class 13).
15	EVERGREEN HARDWOOD/CONIFER	Dominated by evergreen hardwood shrubs and small trees (magnolias and bay forest); usually found in association with RIVERINE SWAMP or in high pocosins which have a low density of pond or loblolly pines.
16	ATLANTIC WHITE CEDAR	Generally even-aged stands of Atlantic White Cedar which occur on peaty, acidic soils. In areas where drainage channels are bordered by pine forest, the mixed pixel response (black-tannic waters/pine) appear to emulate the response of Atlantic White Cedar.
17	LOW POCOSIN	Predominantly areas with organic soils supporting evergreen and deciduous shrubs, vines, briars and cane. These areas tend to be more poorly drained than areas associated with the EVERGREEN HARDWOOD/CONIFER class (Class 15) and they support fewer tree species.
18	LOW MARSH	Regularly flooded marshes dominated by <u>Spartina</u> (sp. <u>alterniflora</u>), <u>Scirpus</u> and <u>Juncus</u> (cordgrass, bulrushes, and black needlerush). Soils are generally rich in organic matter and remain wet most of the year. This class tends to be mixed with SHADOW (class 13).
19	HIGH MARSH	Generally irregularly flooded marshes dominated by <u>Spartina</u> (sp. <u>cynosuroides</u>), <u>Typha</u> or <u>Phragmites</u> (giant cordgrass, cattail and reeds). In general, these areas are slightly less rich in organic matter but in the fall and winter a thick mat of dead marsh grass may form.
20	SAND	Bare, dry sandy soils. Confined to the Coastal Plain, this class includes sand dunes or bare sandy ridges, and also occurs in agricultural fields which have patches of sandy, well-drained soils.

Extensive drainage networks have been used in these areas to control soil moisture frequently resulting in drier conditions. However, there were also many natural pine stands which have closed canopies. Again, inclusion of additional data, such as topography may clarify the status of areas classed as Pine.

There was considerable variety in the vegetative types found in Wetland areas. As noted earlier, a closed canopy prohibits detection of "wet" versus "dry" surface conditions so many pine stands in the Tidewater region which might be considered Wetland were instead classed as Pine. The class "Evergreen Hardwood/Conifer" was based on training sites located in areas commonly referred to as high pocosin or bay forest and in those pockets of green (in the winter) vegetation found along rivers in the Coastal Plain. The latter areas tend to occur in slightly drier sites (high spots or mineral lenses) found in riverine swamps. Half the sites verified as Evergreen Hardwood/Conifer were, in fact, pine but all were in wet areas and were distinguishable from Pine because of their slightly more open canopies. Several sites classified as Evergreen Hardwood/Conifer had a deciduous overstory (maple-gum-cypress) with a heavy evergreen component in the understory.

Low pocosins are composed largely of shrubs, vines and other low vegetative cover. There was considerable difference in vegetation densities in these areas, but since much of the difference was due to fire or other temporary conditions which were difficult to quantify (some differences were related to the different times of data acquisition), no distinction was made between densities.

The terms "Low Marsh" and "High Marsh" are based more on the heights of characteristic vegetation than on topographic differences, though there is a strong correlation between moisture regime, position relative to water bodies and marsh type. Salinity gradients are not specifically addressed in the classification of marsh types. Species composition of high marsh found in Back Bay on the Outer Banks may differ from the species composition of high marsh found near less saline waters. More detailed resolution of marsh vegetation may require additional, seasonal, coverages or integration of ancillary data such as topography. Atlantic White Cedar was included as a separate class, even though there were large errors associated with this class (refer to RESULTS, Classification Verification) because of the high degree of interest in these stands. If desired, it would be appropriate to aggregate areas classed as Atlantic White Cedar with Evergreen Hardwood/Conifer. These areas are all Level I Wetlands.

Riverine Swamps are wetland areas dominated by hardwood trees (mostly gum and maple) and, in some areas, cypress. These areas tend to have highly organic soils and may have standing water for much of the year. The specific species present can vary depending on proximity to tidal influences or salinity of the water.

The class "Low Density Vegetation" was derived from spectral classes based on training sites established in older clearcuts and fallow agricultural fields. Vegetation in these areas consists primarily of weedy growth and shrubs but sites are drier than Low Pocosin and have fewer evergreen species. Areas of low density vegetation did not fit clearly into any of the Level I categories of the USGS classification (Appendix II) and actually encompassed a number of potential land use categories such as agricultural, silvicultural, developed or low density forested areas. Because these areas were spectrally distinct and did not fit consistently into any one other category, Low Density Vegetation was retained as a separate class (see DISCUSSION AND RECOMMENDATIONS).

The class Shadow/Mixed Pixels was derived from single pixel values in major urban areas where tall buildings cast extensive shadows. Single pixel values were also extracted for the shadowed pixels along the edges of fields bordered by forest stands. Both of these areas had low pixel values which otherwise classified as water.

Classification Verification

A total of 1931 one acre plots were sampled for classification verification. Of these, 697 plots were from the Tidewater area, 806 plots were in the Middle and Upper Coastal Plain, and 428 were in the Piedmont. Error matrices were constructed for each physiographic province separately (Figures 3, 4, and 5) and for the total A/P basin (Figure 6). An additional error matrix was produced to demonstrate class accuracies based on categories analogous to the USGS Level I categories (Figure 7). Diagonal elements of the error matrices are the number of sites correctly classified in the image. The sum of off-diagonal elements in each column indicates the number of sites not identified as being in a particular class (the "omission error"). The sum of off-diagonal elements in each row indicates the number of sites identified as being in a particular class when they, in fact, belonged to a different class (the "commission error"). Error matrices are useful for determining the classes which are most likely to be confused and are sometimes referred to as "Confusion Tables".

Based on information in Figures 3-7, it is apparent that the most common confusion areas are between classes within Level I categories, with the exception of Developed areas. Thirty-seven percent (57 out of 153 sites) of the areas which were classed as Developed were actually in Agriculture (Figure 7). Even though attempts were made through masking and reclassification to separate these classes, it was still likely that an area classed as Developed was, in fact, a highly reflective, bare agricultural field. Also, sites surrounding the areas masked out because of cloud cover tended to classify as Developed. Pixels values around the clouds were probably high due to additional light scattered from the clouds. The influence of cloud cover could have been avoided only by masking out a larger "buffer" area.

Disturbed areas are listed separately (Figures 3-6) but, as previously noted, these areas are more likely to be agricultural areas (24 sites) than they are to be clearcuts or areas cleared for potential development (20 sites) (Figure 7). In the Middle and Upper Coastal Plain, Mixed Pine/Hardwood stands were most likely to be confused with Evergreen Hardwood/Conifer (11 sites, Figure 4). These classes are similar in that they both may be composed of a mixture of pine and hardwoods (with Evergreen Hardwood/Conifer consisting of species more characteristic of wet sites) and their distinction apparently becomes more confused in this transitional area between the very wet Tidewater province and the drier Piedmont.

Very few areas were classified as Sand and nearly half of these turned out to be well-drained, bare agricultural fields (4 out of 11 sites). This was not considered an error since the condition (sandy) was correctly identified. Three sites identified as Sand were actually sparsely vegetated with a sandy surface (Figures 3 and 6). Five sites identified as Low Density Vegetation turned out to be in agricultural use (Figure 6). No attempt was made to determine if those areas were actually idle fields at the time of satellite overpass.

No attempt was made to verify pixels classified as Shadows/Mixed Pixels though the class name was modified (from just Shadows) based on observations of class occurrences during the

GROUND TRUTH

IMAGE CLASS

CLASS NAME	#	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18	19	20	Total IC
Water	2	215					1						6							222
Low Dev	3		2	3		2														7
Med Dev	4	2		5		4							1							12
High Dev	5				3	3														6
Agriculture	6					129							2				1			132
Low Veg	7					2	18						1							21
Pine	8		1					69		1	3		2	2			1			79
Bottom Hdwd	9							1	17											18
Hardwood	10									5										5
Mixed	11							1			12		1			2				16
Disturbed	12					3						7								10
Riverine	14							1			2		35	2			2			42
Evergreen	15								1	1	1		3	31		1				38
White Cedar	16							2	1						2					5
Low Pocosin	17					1		3			1			1		24				30
Low Marsh	18	1				1		1					1	1		1	26			32
High Marsh	19						1											10		11
Sand	20						3												8	11
Total GT		218	3	8	3	145	23	78	19	7	19	7	52	37	2	28	30	10	8	697

Figure 3. Error matrix for Tidewater.

GROUND TRUTH

IMAGE CLASS

CLASS NAME	#	2	3	4	5	6	7	8	9	10	11	12	14	15	Total IC
Water	2	20											1		21
Low Dev	3		5			8			1			1			15
Med Dev	4		2	2		7									11
High Dev	5					7									7
Agriculture	6		5		3	184	2			2					196
Low Veg	7					2	38			1			2		43
Pine	8	1						96	3		1		1		102
Bottom Hdwd	9							2	48	2	6	2	2	4	66
Hardwood	10			1			4		9	93	5	1			113
Mixed	11							21	2	5	66			4	98
Disturbed	12			1		20	4					10			35
Riverine	14					1		1	2		1	1	25		31
Evergreen	15								3	1	11		10	43	68
Total GT		21	12	4	3	229	48	120	68	104	90	15	41	51	806

Figure 4. Error matrix for Middle and Upper Coastal Plain.

GROUND TRUTH

IMAGE CLASS

CLASS NAME	#	2	3	4	5	6	7	8	9	10	11	12	Total IC
Water	2	9											9
Low Dev	3		13			5		3		3	6	1	31
Med Dev	4		1	23		13		1		3	5	1	47
High Dev	5			4	8	5							17
Agriculture	6		3	5		83		1		2			94
Low Veg	7					1	25						26
Pine	8		1					49		1			51
Bottom Hdwd	9			1					41	1	1		44
Hardwood	10		1			1				65	3		70
Mixed	11							2			33		35
Disturbed	12					1						3	4
Total GT		9	19	33	8	109	25	56	41	75	48	5	428

Figure 5. Error matrix for Piedmont.

GROUND TRUTH

IMAGE CLASS

CLASS NAME	#	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18	19	20	Total IC
Water	2	244					1						7							252
Low Dev	3		20	3		15		3	1	3	6	2								53
Med Dev	4	2	3	30		24		1		3	5	1	1							70
High Dev	5			4	11	15														30
Agriculture	6		8	5	3	396	2	1		4			2				1			422
Low Veg	7					5	81			1			3							90
Pine	8	1	2					214	3	2	4		3	2			1			232
Bottom Hdwd	9			1				3	106	3	7	2	2	4						128
Hardwood	10		1	1		1	4		9	163	8	1								188
Mixed	11							24	2	5	111		1	4		2				149
Disturbed	12			1		24	4					20								49
Riverine	14					1		2	2		3	1	60	2			2			73
Evergreen	15								4	2	12		13	74[a]		1				106
White Cedar	16							2	1						2					5
Low Pocosin	17					1		3			1			1		24				30
Low Marsh	18	1				1		1					1	1		1	26			32
High Marsh	19						1											10		11
Sand	20						3												8[b]	11
Total GT		248	34	45	14	483	96	254	128	186	157	27	93	88	2	28	30	10	8	1931

a. Includes 22 points which were actually Pine (all in wet areas).
b. Includes 4 points which were actually bare sandy agricultural fields.

Figure 6. Error matrix for all image data.

GROUND TRUTH

IMAGE CLASS

Level I Class	Class Name	#	2	3, 4, 5	6, 12	7	8, 10, 11	9, 14, 15, 16, 17, 18, 19	20	Total IC
Water	Water	2	244			1		7		252
Urban or Built-up Land	Low Dev Med Dev High Dev	3 4 5	2	71	57		21	2		153
Agriculture	Agriculture Disturbed	6 12		17	440	6	5	3		471
Shrub/Scrub	Low Veg	7			5	81	1	3		90
Forest Land	Pine Hardwood Mixed	8 10 11	1	4	2	4	531	27		569
Wetland	Botton Hdwd Riverine Evergreen White Cedar Low Pocosin Low Marsh High Marsh	9 14 15 16 17 18 19	1	1	6	1	39	337		385
Barren Land	Sand	20				3			8	11
Total GT			248	93	510	96	597	379	8	1931

Figure 7. Error matrix for Level I categories.

verification procedures. Pixels which were a mixture of water or very wet ground and deciduous (leaf-off) vegetation or shoreline tended to be assigned to this class. In the Piedmont, this class occurred in metropolitan areas (building shadows) and along the borders of ponds and lakes (mixed water/shoreline). In the tidewater, low marsh and riverine swamp (wet ground/deciduous vegetation) tended to fall in this class.

Based on the error matrices, there were no obvious differences between physiographic provinces in the types of errors which occurred. Slightly more errors occurred in the identification of Developed areas in the Middle and Upper Coastal Plain than in the other two provinces. This may be because better drainage and fewer organic soils were more likely to be present which increased the likelihood of confusing these highly reflective areas with the reflection from concrete or other structural surfaces. It is also possible that conditions were drier at the time these images were acquired and the small sizes of developed areas tended to increase the likelihood of confusion. Sand was not included as a potential class in this region since it was considered to be confined to the Tidewater area. Subsequent attempts to distinguish bare, well-drained fields from developed areas were not successful.

Two types of accuracy assessments were calculated for each province, for the total drainage basin and for each Level I and Level II class (Table 5). The first of these is referred to as "producer's accuracy." It is the probability that an area which is in class N has been correctly identified as being in class N. This accuracy is indicative of possible errors of omission as it defines the number of verification sites which were actually "found" in the classification. The second accuracy, "user's accuracy," is the probability that an area which has been classified as N actually is in class N. This is indicative of errors of commission as it defines the number of verification sites "committed" to the correct class. Using the class Atlantic White Cedar as an example - Producer's accuracy would address the following question: If you were in a stand of Atlantic White Cedar, what is the probability that that stand was classified as Atlantic White Cedar on the imagery? User's accuracy would answer the question: If you were to go to an area identified on the imagery as Atlantic White Cedar, what is the probability that you will find Atlantic White Cedar on that site? Both types of accuracy assessments are useful depending on one's use of the data and can be indicative of possible limitations or qualifications in the information.

Standard errors for Level I categories (Table 5) are indicative of the amount of variation associated with each estimate of a class accuracy at the 95% level of confidence. They can be used to determine a confidence interval about an estimated class accuracy. For example, 95 times out of 100, water will be classified with a user's accuracy of 94.89% to 99.11% (97% plus or minus 2.11%; Table 5, Level I, column B). When the number of sample sites is low, the degree of confidence in an estimate tends to decrease (i.e. the standard error tends to increase and the confidence interval widens). A class accuracy cannot be greater than 100%; however, when no errors are found during the classification procedure, no estimate of error can be determined (e.g. producer's accuracy for the class Sand; Table 5, Level I, column A).

Water was clearly the most likely class to be correctly identified and Developed areas the least likely. Level I producer's vs. user's accuracies for Developed areas indicate that an area which is developed is likely to be correctly identified 76% of the time but that only 46% of those areas identified as Developed were actually in that class. Similarly, all of the verification sites which actually were Atlantic White Cedar were correctly identified as such, but only 40% of the

Table 5. Classification accuracy estimates. (Standard errors were calculated for Level I categories using a 95% confidence level.)

Level I Class	Class Name & Number	Tidewater A*	Tidewater B**	Mid & Upper Coastal Plain A*	Mid & Upper Coastal Plain B**	Piedmont A*	Piedmont B**	Level II Total A*	Level II Total B**	Level I Total A*	Level I Total B**
Water	WATER/2	99	97	95	95	100	100	99	97	99 (±1.24) (max. 100%)	97 (±2.11)
Urban or Built-up Land	LOW DEV/3	67	29	42	33	68	42	59	38	76 (±7.35)	46 (±7.90)
	MED DEV/4	63	42	50	18	70	49	67	43		
	HIGH DEV/5	100	50	0	0	100	47	79	37		
Agriculture	[a]AGRICULTURE/6	89	98	80	94	76	88	82	94	86 (±3.01)	93 (±2.30)
	DISTURBED/12	100	70	67	28	60	75	74	41		
Shrub/Scrub	LOW VEG/7	78	86	79	88	100	96	84	90	84 (±6.21)	90 (±5.25)
Forest Land	[b]PINE/8	88	87	80	94	88	96	84	92	89 (±2.51)	93 (±2.10)
	HARDWOOD/10	71	100	89	82	87	93	88	87		
	MIXED/11	63	75	73	67	69	94	71	75		
Wetland	BOTTOM HDWD/9	89	94	71	73	100	93	83	83	89 (±3.15)	88 (±3.25)
	RIVERINE/14	67	83	61	81	-	-	65	82		
	[c]EVERGREEN/15	84	82	84	63	-	-	84	70		
	WHITE CEDAR/16	100	40	-	-	-	-	100	40		
	LOW POCOSIN/17	86	80	-	-	-	-	86	80		
	LOW MARSH/18	87	81	-	-	-	-	87	84		
	HIGH MARSH/19	100	91	-	-	-	-	100	91		
Barren Land	[d]SAND/20	100	73	-	-	-	-	100	73	100 (-)	73 (±29.5)

A* - Percent probability that an area which is actually in class N has been classified as class N on the image; "Producer's accuracy"

B**- Percent probability that an area which has been classified as class N on the image actually is class N; "User's accuracy"

a. Does not include areas classified as SAND.
b. Does not include areas classified as EVERGREEN HARDWOOD/CONIFER.
c. Includes areas which were actually pine stands in wet areas.
d. Includes areas which were actually bare agricultural fields.

sites classified as Atlantic White Cedar turned out to be cedar. (The amount of area classified as Atlantic White Cedar represented a very small percentage of the drainage basin and the effects of these misclassifications on over-all accuracy would be minimal.) The same type of problem was inherent in the class Sand as many areas which would be more appropriately classified as something else were classified as Sand but sandy areas were very likely to be classified as such.

Other than those noted, no differences in producer's vs. user's errors were obvious between physiographic provinces or within Level I or Level II categories. Some apparent differences were more a function of sample size rather than any persistent bias. In general, accuracies are very good for all classes except Developed areas.

Data Vectorization and Integration

Most activities associated with vectorization and integration of the land use/land cover image data were straightforward albeit lengthy processes. Vectorizing image data (converting data from LAS format to ARC/INFO format) involved eight steps. The image data were: 1) reformatted at CGC; 2) copied to tape, physically transferred to CGIA and loaded from tape; 3) generated as ERDAS files; 4) resolved for pixel confusion along overlapping scene boundaries; 5) geographically subset; 6) filtered; 7) translated to ARC/INFO (vector) format; and, finally, 8) clipped at 1/2 x 1 degree latitude/longitude and by county boundaries. Data integrity measures were required after each step in order to assure that data handling were appropriately addressed. These measures consisted mostly of data review through visual display and classification summary reporting.

Image and vector files were immense in data size. As an example, the Cape Hatteras 1/2 x 1 degree window, which has only partial scene coverage and has mostly open water classification, contains more than 4,000 polygons, requires 1.6 megabytes (mb) of storage in ARC/INFO coverage format, approximately 2.6 mb in ARC/INFO EXPORT format, and approximately 2.9 mb in DLG format. The land use/land cover data layer in EXPORT format will require more than one gigabyte (one thousand million bytes) of disk storage. The size of the files had a direct affect on overall processing and plotting times for each image or subset thereof. Inadequate disk storage capacity resulted in geo-processing periodic failures, especially when the processes involved generation of large transparent, temporary data files. Lengthy processing and plotting times have been a primary reason for delays in image vectorization and subsequent generation of summary reports.

Vertical feature integration of the data layer with other data in the A/P database such as hydrography was considered early in the project but removed from the procedures after further consideration. Vertical integration will allow users to produce large scale hydrographic map overlays onto a land use/land cover base without showing mapping discrepancies in common features from both layers, such as shoreline. However, resolving the discrepancies would alter the original placement of common features that were based on unique and separate mapping methodologies. With original feature placement preserved, the land use/land cover data layer remains genuine and therefore may better serve the A/P user community at large. Vertical integration is deferred for action to a project request basis.

Data Access and Distribution

Land use/land cover data and products can be obtained at cost from CGIA. The data are available by USGS 1:100,000-scale map window (Figure 8) or by county. Lists of these windows and counties are also provided in Appendix I.

Standard hardcopy products consist of acre summary reports by USGS map window or by county and land use/land cover maps by county. Map production occurs at CGIA using pen plotters. Generation of finished quality plots depicting data from this layer had required from three to twelve hours for each plot. Excessive plotting time has resulted in delays during the data review phases and final map generation phases of this project. Installation of an electrostatic plotter at CGIA is expected later this year. The electrostatic plotter will permit production of the maps in less time and in greater volume.

Data can be acquired using typical data distribution formats such as ARC/INFO EXPORT and Digital Line Graph (DLG). Other output formats supported by ARC/INFO software will also be supported on request. The intermediate ERDAS data files are available for users with special requirements.

Data are transferable on 3.5" diskette, 1/4" data cartridge, or 9-track tape using DOS COPY, UNIX TAR, and ARC/INFO TAPEWRITE copy formats. Table 6 illustrates the possible combinations of media and copy formats available for each data type. Prospective users are cautioned that land use/land cover data require large amounts of disk storage. Prior to acquiring the data, users should determine appropriate disk and media storage requirements for their systems. An average USGS 1:100,000-scale land use/land cover data file contains approximately 66,453 polygons and requires approximately 26.7 megabytes for the ARC/INFO coverage format, 44.2 megabytes for the ARC/INFO EXPORT format, and 49.0 megabytes for the DLG format.

Table 6. Media and copy formats for land use/land cover data.

DATA	MEDIA			
Format	3.5" Diskette (1.4mb)	Data Cartridge	9-TRACK (1600/6250)	
TEXT	DOS COPY	TAR	DD	TAR
COVERAGE	NOT APPLICABLE	TAR	TAR	
EXPORT	DOS COPY	TAR	TAPEWRITE	TAR

DD - Data Dump Command/UNIX
TAR - Tape Archive Retrieval/UNIX
TAPEWRITE - ARC/INFO Dump/Load

SPECIAL NOTE: Distribution on 5.25" Diskette is available only at low density (360K) with DOS COPY dump format.

Direct on-line access to the data layer (for users directly linked to the CGIA computer system network) is available through special arrangements with CGIA. Due to data storage requirements, the land use/land cover data layer is currently not available on-line for the entire

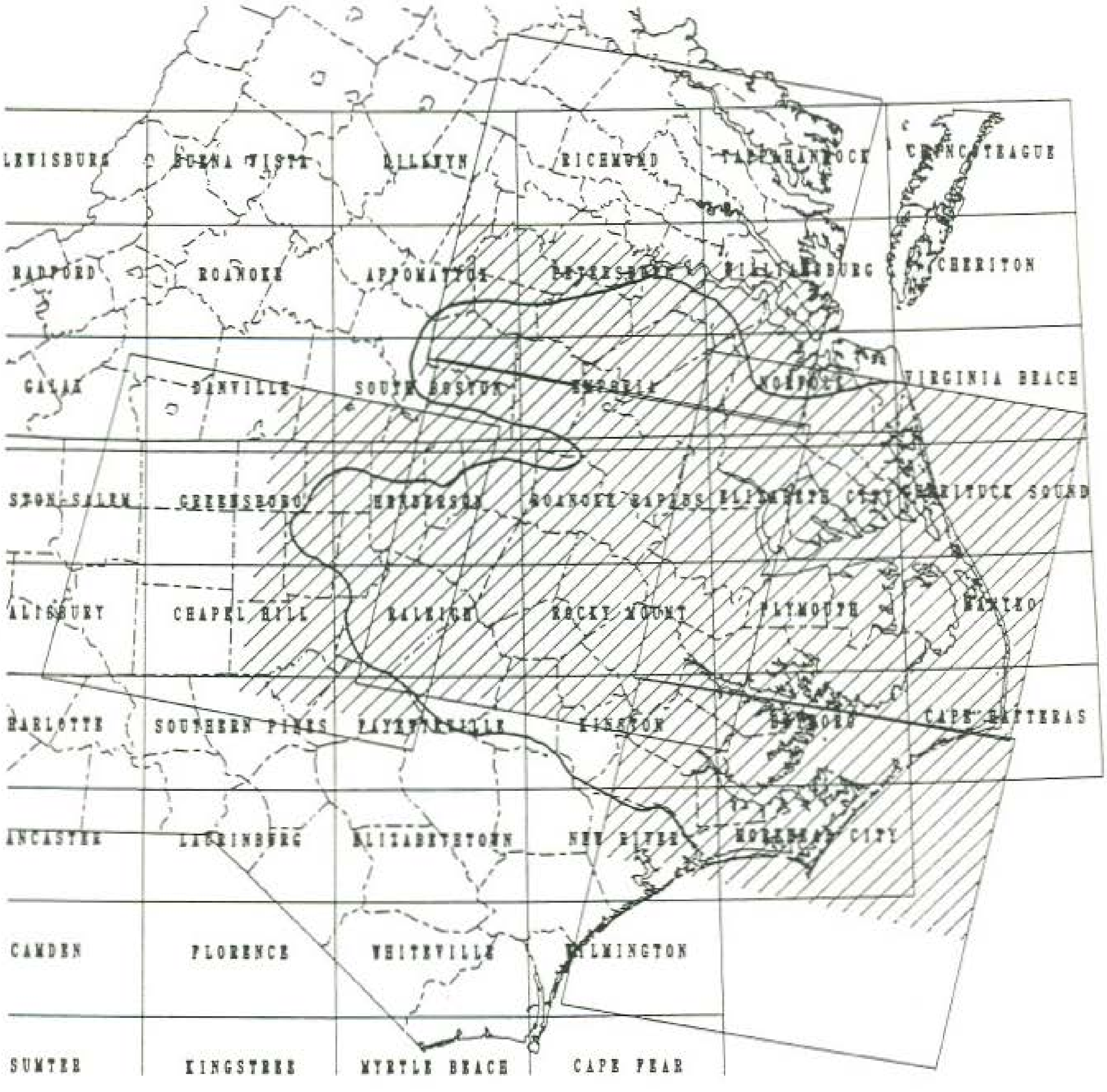

Figure 8. U.S. Geological Survey 1:100,000 map coverage of the A/P drainage basin (with Landsat scenes overlain). Shaded area indicates actual extent of land use/land cover data.

A/P Study area. Only areas that are specifically requested for access are loaded onto the computer system. Expanded disk storage capacity on the state GIS computer network is expected later this year and will allow on-line storage of the layer for the entire A/P basin.

For those without direct access to a geographic information system, CGIA offers geo-processing services on a cost recovery basis. Specialized map products, reports, and analysis will be provided by CGIA staff through typical project arrangements.

DISCUSSION AND RECOMMENDATIONS

Image Analyses

Land use and land cover classes used in this project were based on the USGS hierarchical classification scheme for use with remote sensor data (Appendix II). This scheme was intended as a standard for visual interpretation, and not all classes (or all Levels) are directly applicable to digital analyses. Digital spectral data, for instance, cannot be used to determine land use *per se*, but they do provide information on land cover, including man-made structures, from which land use can be inferred.

On February 12, 1991, a meeting was held in Washington D.C. to define a wetlands/uplands classification scheme for the National Oceanographic and Atmospheric Administration (NOAA) Habitat Mapping Program. Participants included representatives from several federal agencies including NOAA, the Environmental Protection Agency, U.S. Army Corps of Engineers, U.S. Geological Survey, U.S. Fish and Wildlife Service, U.S.D.A. Soil Conservation Service, the National Park Service, and Oak Ridge National Laboratory, and representatives from state agencies or universities from Virginia, Delaware, Maryland, Florida and North Carolina. This was one of a series of meetings to be held nationwide as part of NOAA's efforts to coordinate federally funded wetlands and uplands mapping projects.

The classification scheme adopted at this meeting is also based on the USGS classification system but contains a few modifications and caveats (Appendix III). Most notably, it was recommended that rangeland be replaced by Grasslands (Herbaceous) principally because the land use implications for the term "rangeland" could be misleading, and that the category Shrub/Scrub be added (analogous to Low Density Vegetation in the A/P classification). Level II classes are a combination of classes from the USGS system and wetlands classes used by the U.S. Fish and Wildlife Service in their National Wetlands Inventory maps (Cowardin et al. 1979). Project participants acknowledged that more Level II classes could be added to meet regional needs and that much of the Level II information would not be obtainable from TM data but would come from ancillary sources.

In general, the final classes used in the A/P satellite study have good agreement with the proposed national standard classification. The most notable exception is the classification of Grasslands as a separate category using Level II categories to distinguish land use. In this classification, one proposal for Level II categories for Grasslands separates Unmanaged Grassland (naturally occurring grasses and forbes) and Managed Grassland (e.g. pasture, yards, golf courses, etc.). Distinguishing unmanaged vs. managed grassland as well as alternative or subsequent levels of detail still requires incorporation of ancillary data. In NOAA's prototype land use/land cover classification focused on the Chesapeake Bay (Dobson and Bright 1991), agriculture and grass were combined in a single class as in the A/P study primarily because of inability to distinguish grassy agricultural fields from grassy areas in some other use. However, participants in NOAA's meetings have recommended retaining these as separate categories and using ancillary data to distinguish land use.

The proposed classification is more appropriate for digital analyses than the original USGS classification because the Level I classes are based more on cover types that may be distinguished from spectral data. Flexibility in the type, source and level of detail of additional information is provided in succeeding levels.

The approach used to identify land use and land cover categories on the TM data has been termed a supervised approach. However, it was, in essence, a combination of the supervised and unsupervised approaches because an effort was made to identify spectral categories first. This approach provides better training statistics particularly when a given class of interest exhibits highly variable spectral characteristics; but, this approach requires that an analyst have some *a priori* knowledge of cover conditions as well as knowledge of the likely impacts of environmental conditions on the spectral properties of the classes. Subjectivity in the approach could be minimized by development of more automated techniques for selecting training sites. One promising approach has utilized digital soils data in conjunction with an unsupervised classification to reduce the subjectivity of site selection. Sample sites for training and ground truth are located in areas represented by unique land cover/soil combinations in a stratified manner based on their proportional contribution to the area being analyzed (Warren et al. 1990). When combined with stratification based on physiography, this type of approach could make it easier to take into account the effects of soil type and moisture conditions on spectral signatures and may provide more information on spectral variability within land cover classes.

Physiographic stratification greatly improved both the ease and accuracy of image classification. Results may have been further enhanced, particularly in the Middle and Upper Coastal Plains, by stratifying using more detailed physiographic data. More research would be needed to determine the best approach for this geographic area, or for other areas, and to determine the types of ancillary data needed. In mountainous terrain, for example, topographic data would be needed to stratify image data by aspect in order to account for differences in insolation (Justice et al. 1981). As computational resources become faster and digital databases improve, it becomes more practical to utilize a larger number of data layers or more sophisticated clustering and classification algorithms which may provide improvements in classification accuracies.

The season in which data were collected greatly affects the ability to categorize ground features. Winter imagery was good for distinguishing general forest types (deciduous vs. coniferous) and bare fields but did not help in distinguishing highly reflective bare soil from concrete. Multi-temporal coverage would have been of value in clearing up confusion between agricultural use and development and may have helped improve classification accuracies within wetlands classes. Wet areas were relatively easy to distinguish from drier areas in the winter imagery, but supplemental spring or fall coverages could be used to improve differentiation of vegetation types within the wet areas. In some sites within the Tidewater region (e.g. Dismal Swamp and Currituck County), deciduous riverine swamps with a high density of evergreen trees or shrubs in the understory tended to misclassify as pine or pine-hardwood. Multi-temporal coverage would have helped alleviate most of these types of errors. Summer coverages could be used to reduce the influence of shadows. Also in the Tidewater region, shadows, water and wet organic soils such as those found in low marshes, riverine swamps and some agricultural fields, tended to be confused. As previously noted, much of the area classified as Shadow/Mixed pixels in the near coast areas are actually low marsh or swamp. Multi-temporal coverage would also provide additional opportunities for classifying areas which may be obscured by cloud cover or haze in any one scene.

The following recommendations should be considered for the image processing aspect of future land use/land cover mapping projects which utilize Landsat TM or other digital spectral data:

1. Utilize multi-temporal data sets to improve detail, accuracy and timeliness of the data;
2. Continue to investigate the use of newly developed clustering or classification algorithms which may improve class discrimination;
3. Expand research efforts to support the inclusion of other georeferenced data which can be used for image stratification, such as topography, soils, or more detailed physiography;
4. Coordinate classification schemes and methodologies with other state or regional mapping efforts to maximize the potential for generating seamless coverages over larger areas; and
5. Include support for a full time ecologist/botanist with field experience in identifying and classifying vegetative communities.

Data Use for Future Inventories

The land use/land cover inventory is highly recommended for use on region-wide applications. The data can be used to inventory, map, and characterize geographic areas such as the entire Albemarle-Pamlico estuarine system, groups of counties, basins, and similar areas. Applied in analyses with other layers in the A/P database, the data can be utilized to carry out A/P research projects such as the development of models that determine non-point source pollutant loads. With other land use/land cover inventories, the data can further be used to monitor land use/land cover status and trends in the A/P area.

In order to regularly monitor and research land use/land cover activities adequately, an inventory from satellite data is recommended for the Albemarle-Pamlico estuarine basin every five years. The classification scheme for future inventories should be consistent with the current scheme. In order to augment proper trends analysis, future inventories that are more detailed in classification and resolution should be designed to allow for generalizations which are consistent with this inventory. It is hoped that future inventories will be extended into other river basins that are in or affect North Carolina.

Hardware limitations apparent during this project should be overcome on future classification projects. Ample amounts of disk storage space and memory should be available for data processing. Efforts involving the A/P area (five scenes) using ERDAS and ARC/INFO software require at least one and one-half gigabytes of storage to efficiently conduct operations.

Efficient map production equipment is also a requirement. Maps that depict land use/land cover data for entire counties typically require two or three hours to produce using a pen plotter. A fifty percent failure rate associated with most pen plots increases the overall time to complete each map to four or five hours. A map of the entire A/P area would require almost twelve hours of plotting time. Alternative map production equipment is needed to assist with the quality control process and final map production. An electrostatic plotter would produce a county map in less than fifteen minutes. This plotter would provide additional colors and polygon shade fill capabilities which are necessary on detailed land use/land cover maps. An electrostatic plotter would also allow efficient generation of multiple copies of each map.

In addition to hardware expansions (currently in progress at CGC and CGIA), future projects should employ an adequate number of trained full-time staff who are available for classification of satellite data and data integration on the GIS. The expertise and cooperation exhibited by staff at CGC and CGIA was highly advantageous to the success of this project. It is hoped that similar efforts with these agencies can be arranged to continue classification of land use/land cover data on a regular basis.

REFERENCES

American Society of Photogrammetry. 1983. Manual of Remote Sensing, 2nd Edition. D.S. Simonett and J.E. Estes, Editors, 2724 p.

Anderson, J.R., E.E. Hardy, J.T. Roach and R.E. Witmer. 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. U.S. Dept. of Interior, Geological Survey Professional Paper 964, 28 p.

Avery, T.E. and G.L. Berlin. 1985. Interpretation of Aerial Photographs, 4th Edition. Burgess Publishing Co., Minneapolis, MN, 554 p.

Cowardin, L.M., V. Carter, F.C. Golet and E.T. LaRoe. 1979. Classification of Wetlands and Deepwater Habitats of the United States. U.S. Dept. of Interior, Fish and Wildlife Service, Office of Biological Services, FWS/OBS-79/31.

Gong, P. and P.J. Howarth. 1990. "An Assessment of Some Factors Influencing Multispectral Land-Cover Classification." Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 5, pp. 597-603.

Janssen, L.L.F., M.N. Jaarsma and E.T.M. van der Linden. 1990. "Integrating Topographic Data with Remote Sensing for Land-Cover Classification." Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 11, pp. 1503-1506.

Justice, C.O., S.W. Wharton and B.N. Holben. 1981. "Application of Digital Terrain Data to Quantify and Reduce the Topographic Effect on Landsat Data." International Journal of Remote Sensing, 2(2), pp. 213-221.

Karaska, M.S., S.J. Walsh and D.R. Butler. 1986. "Impact of Environmental Variables on Spectral Signatures Acquired by the Landsat Thematic Mapper." International Journal of Remote Sensing, 7(12), pp. 1653-1667.

Khorram, S., J.A. Brockhaus and H.M. Cheshire. 1987. "Comparison of Landsat MSS and TM Data for Urban Land-Use Classification." IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-25, No. 2, pp. 238-243.

Ma, Z. and C.E. Olson, Jr. 1989. "A Measurement of Spectral Overlap Among Cover Types." Photogrammetric Engineering and Remote Sensing, Vol. 55, No. 10, pp. 1441-1444.

Sheffield, C. 1985. "Selecting Band Combinations from Multispectral Data." Photogrammetric Engineering and Remote Sensing, Vol. 51, No. 6, pp. 681-687.

Trolier, L.J. and W.R. Philipson. 1986. "Visual Analysis of Landsat Thematic Mapper Images for Hydrologic Land Use and Cover." Photogrammetric Engineering and Remote Sensing, Vol. 52, No. 9, pp. 1531-1538.

Walsh, S.J. 1987. "Variability of Landsat MSS Spectral Response of Forests in Relation to Stand and Site Characteristics." International Journal of Remote Sensing, 8(9), pp. 1289-1299.

Warren, S.D., M.O. Johnson, W.D. Goran and V.E. Diersing. 1990. "An Automated, Objective Procedure for Selecting Representative Field Sample Sites." Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 3, pp. 333-335.

APPENDIX I

Study area coverage by county and USGS quadrangle name

COUNTY COVERAGE

North Carolina Counties

(42 counties including 19 counties with partial coverage)

Beaufort
Bertie
Caswell
Camden
*Carteret
*Chatham
Chowan
Craven
Currituck
Dare
*Duplin
*Durham
Edgecombe
Franklin
Gates
*Granville
Greene
*Halifax
*Harnett
Hertford
Hyde
*Johnston
*Jones
*Lenoir
Martin
Nash
*Northampton
*Onslow
*Orange
Pamlico
Pasquotank
Perquimans
*Person
Pitt
*Sampson
Tyrrell
*Vance
*Wake
*Warren
Washington
*Wayne
Wilson

Virginia Cities/Counties

(18 cities/counties including 10 with partial coverage)

*Brunswick County
*Charlotte County
Chesapeake City
*Dinwiddie County
Greensville County
*Isle of Wight County
*Lunenburg County
*Mecklenburg County
Norfolk City
*Nottoway County
Petersburg County
Portsmouth City
*Prince Edward County
Southampton County
Suffolk City
*Surry County
Sussex County
*Virginia Beach City

* Indicates partial coverage

U.S. Geological Survey 7.5 Minute (1:24000) Quadrangles

North Carolina Quadrangles
(350 quadrangles including 82 with partial coverage)

Afton
Ahoskie
Albemarle Sound
*Albertson
*Angier
Apex
Arapahoe
Askin
*Atlantic
Aulander
Aurelian Springs
Aurora
Ayden
Bailey
Barco
*Barley
Bath
Bayboro
Bayleaf
*Beaufort
Belhaven
*Benson
*Berea
Blounts Bay
Bluff Point
Boones Crossroads
Boykins
Broad Creek
Buffalo City
Bunn E
Bunn W
Bunyan
*Buxton
Caldwell
Camden Point
*Cape Hatteras
*Cape Lookout
Cary
Castalia
Catfish Lake
*Cedar Grove
Center Hill
Centerville
Chapanoke
*Chapel Hill
Cherry Point
Claresville
Clayton
*Coats
Coinjock
*Cokesbury
Colerain
Columbia E
Columbia W
Comfort
Conetoe
Conway
Corapeake
Core Creek
*Corolla
Cove City
Creedmoor
Creeds
Creswell
Creswell SE
Currituck
Darlington
Davis
Dawson Crossroads
Deer Run
*Dobbersville S
Dover
Drake
Draughn
*Dunn
East Lake SE
Edenhouse
Edenton
Edmondson
Edward
*Efland
Elizabeth City
Elm City
Enfield
Engelhard E
Engelhard NE
Engelhard NW
Engelhard W
Ernul
Essex
Fairfield
Fairfield NE
Fairfield NW
Falkland
Falling Creek
Farmlife
Farmville
Flowers
Fort Barnwell
Fort Landing
Fountain
Four Oaks
Four Oaks NE
Franklinton
Freemont
Frying Pan
*Fuquay-Varina
Galatia

* Indicates partial coverage

U.S. Geological Survey 7.5 Minute (1:24000) Quadrangles

North Carolina Quadrangles
(cont'd)

Garner
Garnersville
*Gasburg
Gates
Gatesville
Gold Sand
*Grantham
*Green Island
*Green Level
Greenville NE
Greenville NW
Greenville SE
Greenville SW
Grifton
Grimesland
Grissom
Hackney
Hadnot Creek
Halifax
Hamilton
*Harkers Island
Harrellsville
Hartsease
Harvey Neck
*Hatteras
Havelock
*Henderson
Hertford
*Hillsborough
Hobbsville
Hobgood
Hoke
Hollister
Hookerton
*Horsepen Point
*Howard Reef
*Hubert
*Hurdle Mills
Inez
Ingleside
Jackson
*Jacksonville NE
*Jacksonville NW
Jamesville
*Jarvisburg
Jason
Jasper
Jones Bay
Justice
Kelford
Kenly E
Kenly W
Kinston
Kittrell
*Kitty Hawk
Knightdale
*Knotts Island
La Grange
Lake Drummond
Lake Drummond SE
Lake Michie
Lake Wheeler
Lamps Corner
Leggetts Crossroads
Leonards Point
Little Fishing Point
*Little Kinnakeet
*Littleton
Long Bay
Long Shoal Point
Louisburg
Lowland
Lucama
Lynchs Corner
*Macon
Manns Harbor
*Mansfield
*Manteo
Margaretsville
*Martin Point
Masontown
*Maysville
Merchants Millpond
Merrimon
Merry Hill
*Middleburg
Middlesex
Middletown
Middletown Anchorage
Mintonsville
Moriah
*Mossey Islands
Moyock
*Mt Olive
Murfreesboro
NE Durham
NW Durham
NE Goldsboro
NW Goldsboro
Nashville
New Bern
*New Hill
New Holland
New Lake
New Lake NW
New Lake SE
Newport
*Newton Grove N

* Indicates partial coverage

U.S. Geological Survey 7.5 Minute (1:24000) Quadrangles

North Carolina Quadrangles
(cont'd)

*Newton Grove S
Nixonton
Norfleet
North Bay
Oak City
*Ocracoke
Old Ford
Old Sparta
*Olive Hill
*Oregon Inlet
Oriental
*Oxford
Palmyra
Pamlico Beach
Pamlico Point
Pamlico Sound
Pantego
Pasquotank
*Pea Island
*Peacocks Crossroads
Phillips Crossroads
Pike Road
Pinetops
Pinetown
*Pink Hill
Plymouth E
Plymouth W
Point Harbor
*Point of Marsh
Pollocksville
Ponzer
Portsmouth
*Potters Hill
Powellsville
Powhatan
Princeton

Pungo Lake
Quitsna
Raleigh E
Raleigh W
Ransomville
Red Oak
Reelsboro
Republican
Rich Square
*Richlands
*Ridgeville
Ringwood
Riverdale
Rivermont
*Roanoke Island NE
*Roanoke Rapids
Robersonville E
Robersonville W
Rocky Mt
*Rodanthe
Rolesville
Roper N
Roper S
Rougemont
*Roxboro
SE Durham
*SW Durham
SE Goldsboro
SW Goldsboro
*Salter Path
Saratoga
*Satterwhite
Scotia
Scotland Neck
Scranton
Selma

*Seven Springs
Shiloh
*Skippers
Snow Hill
South Creek
*South Hill SE
South Mills
South River
Speed
Spring Hope
Stancils Chapel
Stantonsburg
*Stella
Stem
Stevenson Point
*Stovall
Stumpy Point
*Styron Bay
Sunbeam
Sunbury
Swanquarter
*Swansboro
Tarboro
*Thelma
Timberlake
*Townsville
Trenton
*Triple Springs
Union
Upper Broad Creek
*Valentines
Valhalla
Vanceboro
Vandemere
*Vicksboro
Wade Point

* Indicates partial coverage

U.S. Geological Survey 7.5 Minute (1:24000) Quadrangles

North Carolina Quadrangles
(cont'd)

*Wainwright Island
Wake Forest
Walstonburg
Wanchese
*Warrenton
Washington
Weeksville
Weldon
Westover
Whaleyville
Whitakers
*Williams
Williamston
Williston
Wilmar
Wilson
Wilton
Windsor N
Windsor S
Winstead Crossroads
Winton
Woodard
Woodland
Woodville
Yeopim River
Zebulon

* Indicates partial coverage

U.S. Geological Survey 7.5 Minute (1:24000) Quadrangles

Virginia Quadrangles
(79 Quadrangles including 36 with partial coverage)

Adams Grove
Alberta
Ante
*Baskerville
*Blackstone East
*Blackstone West
*Boydton
*Buckhorn
Capron
Carson
Chase City
Cherry Hill
*Church Road
*Claremont
*Clarksville North
Courtland
*Crewe West
*Crewe East
Danieltown
*Darvills
*Deep Creek
*Dendron
*Dewitt
Dinwiddie
*Disputanta South
Disputanta North
*Drakes Branch
Drewryville
Emporia
*Eureka
*Fentress
Forksville
Franklin
Ft Mitchell
*Green Bay
*Hebron
*Holland
Ivor
Jarrett
Kenbridge West
Kenbridge East
*Keysville
La Crosse
Lake Drummond NW
Lawrenceville
Littleton
Lunenburg
Manry
Mc Kenney
*Meherrin
*North Bay
North View
*Petersburg
*Pleasant Ridge
Powellton
*Prince George
Purdy
*Raynor
*Rubermont
*Runnymede
*Savedge
Sebrell
Sedley
Smoky Ordinary
*South Hill
Stony Creek
Suffolk
Sussex
*Sutherland
Templeton
Vicksville
Warfield
Waverly
White Plains
Wightman
*Windsor
*Wylliesburg
Yale
*Zuni

* Indicates partial coverage

U.S. Geological Survey 1:100,000 Quadrangles

North Carolina and Virginia

*Appomattox
Bayboro
Cape Hatteras
*Chapel Hill
Currituck Sound
*Danville
Dillwyn
Elizabeth City
Emporia
*Fayetteville
*Greensboro
Henderson
*Kinston
Manteo
Morehead City
*New River
*Norfolk
*Petersburg
Plymouth
Raleigh
Richmond
Roanoke Rapids
Rocky Mount
*South Boston
Southern Pines
*Virginia Beach
*Williamsburg

* Indicates partial coverage

APPENDIX II

Land use/land cover classification system
for use with remote sensing data

Land Use/Land Cover Classification System for use with Remote Sensor Data
Adopted from USGS Professional Paper 964

Level I		Level II	
1.	Urban or Built-up Land	11.	Residential
		12.	Commercial and Services
		13.	Industrial
		14.	Transportation, Communications, and Utilities
		15.	Industrial and Commercial Complexes
		16.	Mixed Urban or Built-up Land
		17.	Other Urban or Built-up Land
2.	Agricultural Land	21.	Cropland and Pasture
		22.	Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas
		23.	Confined Feeding Operations
		24.	Other Agricultural Land
3.	Rangeland	31.	Herbaceous Rangeland
		32.	Shrub and Brush Rangeland
		33.	Mixed Rangeland
4.	Forest Land	41.	Deciduous Forest Land
		42.	Evergreen Forest Land
		43.	Mixed Forest Land
5.	Water	51.	Streams and Canals
		52.	Lakes
		53.	Reservoirs
		54.	Bays and Estuaries
6.	Wetland	61.	Forest Wetland
		62.	Nonforested Wetland
7.	Barren Land	71.	Dry Salt Flats
		72.	Beaches
		73.	Sandy Areas other than Beaches
		74.	Bare Exposed Rock
		75.	Strip Mines, Quarries, and Gravel Pits
		76.	Transitional Areas
		77.	Mixed Barren Land
8.	Tundra	81.	Shrub and Brush Tundra
		82.	Herbaceous Tundra
		83.	Bare Ground Tundra
		84.	Wet Tundra
		85.	Mixed Tundra
9.	Perennial Snow or Ice	91.	Perennial Snowfields
		92.	Glaciers

APPENDIX III

Comparison of proposed modified land use and land cover classification system for use with remote sensor data with A/P study classification. The modified system has been proposed for use in a national wetlands/uplands classification effort to be coordinated by the NOAA Habitat Program. Level II classes in the A/P study which are directly comparable to proposed classes have the standard Level II numerical designation in parentheses following the A/P class name.

Modified Anderson		A/P Study	
Level I	Level II	Level I	Level II
1. Urban or Built-up Land	11. Residential 12. Commercial and Services 13. Industrial 14. Transportation, Communications, and Utilities 15. Industrial and Commercial Complexes 16. Mixed Urban or Built-up Land 17. Other Urban or Built-up Land (Anderson et al.)	Urban or Built-up Land	Low Density Developed Medium Density Developed High Density Developed
2. Agricultural	21. Cropland 22. Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas 23. Confined Feeding Operations 24. Other Agriculutural Land (Modified Anderson et al.)	Agriculture/Grasslands	Agriculture/Grass (21) Disturbed
3. Grassland (Herbaceous)	31. Herbaceous Grassland		
4. Forest Land	41. Deciduous Forest Land 42. Evergreen Forest Land 43. Mixed Forest Land (Anderson et al.)	Forest Land	Hardwood (41) Pine (42) Mixed P/H (43)
5. Scrub/Shrub	51. Deciduous Scrub/Shrub 52. Evergreen Scrub/Shrub 53. Mixed Scrub/Shrub (New Classes)	Shrub/Scrub	Low Density Vegetation

Modified Anderson		A/P Study	
Level I	Level II	Level I	Level II
6. Water (Deepwater)	61. Marine 62. Estuarine 63. Lacustrine 64. Riverine (Cowardin et al.)	Water	Water
7. Wetland	71. Estuarine Intertidal Herbaceous 72. Estuarine Intertidal Woody 73. Estuarine Intertidal Non-Vegetated 74. Palustrine Forested 75. Palustrine Scrub/Shrub 76. Palustrine Emergent 77. Palustrine Non-Vegetated (Cowardin et al.)	Wetland	Bottomland Hardwood (74) Riverine Swamp (72 and 74) Evergreen Hardwood/Conifer (74) Atlantic White Cedar (74) Low Pocosin (similar to 75) Low Marsh (71) High Marsh (71)
8. Barren Land	81. Dry Salt Flats 82. Beaches 83. Sandy Areas other than Beaches 84. Bare Exposed Rock 85. Strip Mines, Quarries, and Gravel Pits 86. Transitional Areas 87. Mixed Barren Land (Anderson et al.)	Barren Land	Sand (82 and 83)
9. Tundra	91. Shrub and Brush Tundra 92. Herbaceous Tundra 93. Bare Ground Tundra 94. Wet Tundra 95. Mixed Tundra (Anderson et al.)		
10. Perennial Snow or Ice	101. Perennial Snowfields 102. Glaciers (Anderson et al.)		N/A