# Regional Characterization of Land Cover Using Multiple Sources of Data

J.E. Vogelmann, T. Sohl, and S.M. Howard


## Abstract

*Many organizations require accurate intermediate-scale land-cover information for many applications, including modeling nutrient and pesticide runoff, understanding spatial patterns of biodiversity, land-use planning, and policy development. While many techniques have been successfully used to classify land cover in relatively small regions, there are substantial obstacles in applying these methods to large, multiscene regions. The purpose of this study was to generate and evaluate a large region land-cover classification product using a multiple-layer land-characteristics database approach. To derive land-cover information, mosaicked Landsat thematic mapper (TM) scenes were analyzed in conjunction with digital elevation data (and derived slope, aspect, and shaded relief), population census information, Defense Meteorological Satellite Program city lights data, prior land-use and land-cover data, digital line graph data, and National Wetlands Inventory data. Both leaf-on and leaf-off TM data sets were analyzed. The study area was U.S. Federal Region III, which includes the states of Pennsylvania, Virginia, Maryland, Delaware, and West Virginia.*

*The general procedure involved (1) generating mosaics of multiple scenes of leaves-on TM data using histogram equalization methods; (2) clustering mosaics into 100 spectral classes using unsupervised classification; (3) interpreting and labeling spectral classes into approximately 15 land-cover categories (analogous to Anderson Level 1 and 2 classes) using aerial photographs; (4) developing decision-making rules and models using from one to several ancillary data layers to resolve confusion in spectral classes that represented two or more targeted land-cover categories; and (5) incorporating data from other sources (for example, leaf-off TM data and National Wetlands Inventory data) to yield a final land-cover product. Although standard accuracy assessments were not done, a series of consistency checks using available sources of land-cover information were conducted to evaluate the effectiveness of this approach for generating accurate land-cover information for large regions.*


## Introduction

Many agencies, including the U.S. Environmental Protection Agency, the U.S. Forest Service, Bureau of Land Management, National Oceanic and Atmospheric Administration, the U.S. Geological Survey (USGS), state governments, and environmental groups need up-to-date intermediate-scale land-cover data (e.g., spatial resolution of 1 hectare or better). Potential uses for such land-cover data are many and varied, and include assessing ecosystem status and health, modeling nutrient and pesticide runoff, understanding spatial patterns of biodiversity, land-use planning, and developing land management policy. Despite the need for current land-cover data, much of the intermediate-scale spatial land-cover data now available for the United States are outdated and are of questionable accuracy. The most recent intermediate-scale land-cover data set generated for the conterminous United States [land-use and land-cover (LUDA) data] was developed by the USGS (1990) in the 1970s by interpreting high-altitude aerial photographs. Although this data set is probably still adequate for some applications, many land-cover changes have occurred since the data set was compiled. More recently, a land-cover classification for the conterminous United States using 1-km advanced very high resolution radiometer data (Loveland *et al.*, 1991; Brown *et al.*, 1993) was developed for use by the global change research community. However, this data set is spatially too coarse for assessing many of the issues of national concern.

The main objective of this project was to generate a generalized, consistent, seamless, and reasonably accurate land-cover data layer for U.S. Federal Region III, which includes the States of Pennsylvania, Maryland, Delaware, Virginia, and West Virginia. One goal of the study was to create a land-cover data set appropriate for a wide variety of uses. In addition to exploring various methods for efficiently deriving large-area classifications, a major thrust of this project was to evaluate the potential and practicality of generating an intermediate-scale land-cover data set for the conterminous United States.

## Methods

### Study Region

The study region covers more than 30 million hectares in the eastern United States (Plate 1). The majority of the region is densely forested; other features within the region include a variety of types of wetlands, agricultural lands, water, and numerous urban areas. The terrain of the region (Plate 2) ranges from flat (especially along coastal areas in the eastern portion of the region) to mountainous (most notably in the Appalachian Mountain Range in the west). The region contains nine different ecoregions as defined by Omernik (1987).

### Data sources

The primary source of data for this project was leaves-on (summer) Landsat thematic mapper (TM) data acquired in 1991, 1992, and 1993, collected for the Multiresolution Land Characteristics consortium (Loveland and Shaw, 1996). Data sets were de-striped, and terrain corrected using 3-arc-second digital terrain elevation data set (DTED) data and ground con-

Any use of trade, product, or firm names is for descriptive purposes only and does not imply an endorsement by the U.S. Government.

Hughes-STX Corporation, U.S. Geological Survey, EROS Data Center, Sioux Falls, SD 57198 vogel@edcmail.cr.usgs.gov.

Photogrammetric Engineering & Remote Sensing,
Vol. 64, No. 1, January 1998, pp. 45–57.

0099-1112/98/6401-0045$3.00/0

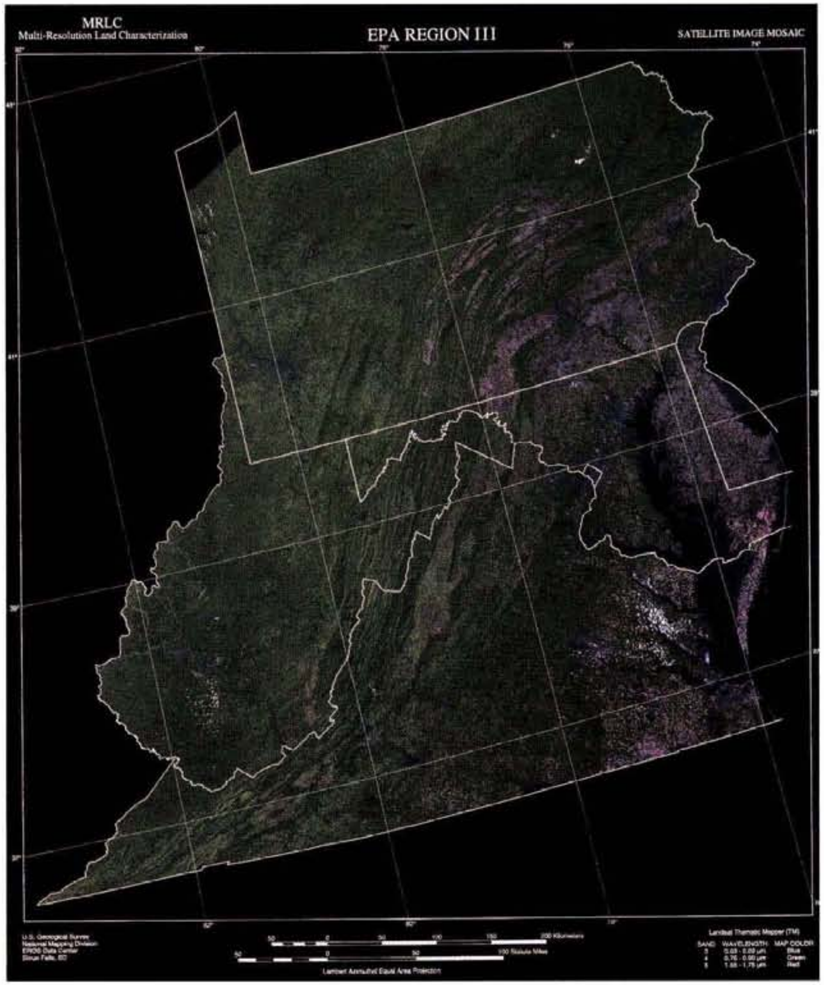


Plate 1. Landsat thematic mapper mosaic of Federal Region III produced using bands 5, 4, and 3 in the order of red, green, and blue. Data represent summer (leaves-on) conditions.

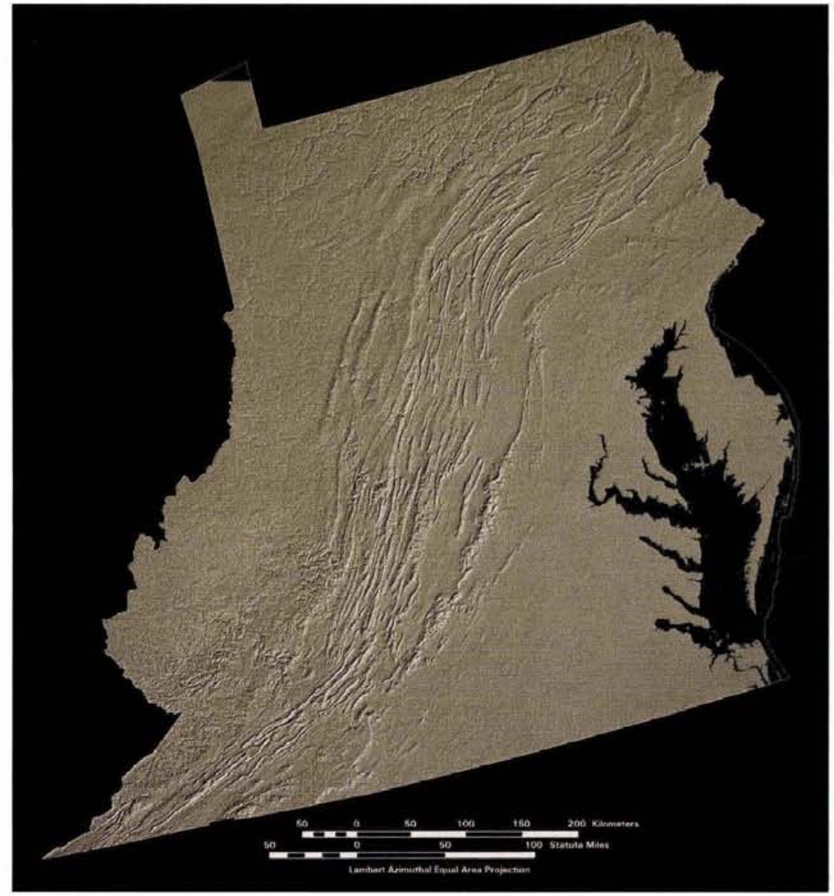


Plate 2. Shaded-relief data set of Federal Region III derived from 3-arc-second Digital Terrain Elevation Data set data.

trol points with a root-mean-square error of less than one pixel (30 metres). Data sets were projected to Lambert Azimuthal coordinates. Additionally, leaf-off TM data sets were analyzed. Although most of the leaves-off data sets were acquired in spring, a few were from late autumn because of the difficulties in acquiring cloud-free TM data during springtime in the eastern United States. In total, 45 TM scenes were used (Table 1).

Other intermediate-scale spatial data were used, including DTED (U.S. Geological Survey, 1993) and derivative DTED products (including slope, aspect, and shaded relief (Plate 2)), population-density data (Plate 3: Bureau of the Census, 1991a; Bureau of the Census, 1991b; Bureau of the Census, 1992; Hitt, 1992), Defense Meteorological Satellite Program city lights data (Plate 4: Elvidge *et al.*, 1997), LUDA data (Plate 5), and National Wetlands Inventory (NWI) data (Plate 6). Political boundary data (county level) derived from Digital Line Graph data (U.S. Geological Survey, 1990) were also used for consistency checks.

#### Classification Procedure

The general procedure was to (1) mosaic the summer TM scenes and classify them using an unsupervised classification algorithm, (2) interpret and label classes into modified National Oceanic and Atmospheric Administration (NOAA) Coastal Change Analysis Program (C-CAP) classes (Klemas *et al.*, 1993; Dobson *et al.*, 1995: see Table 2) using aerial photographs, (3) resolve confused classes using the appropriate ancillary data source(s), and (4) incorporate land-cover information from leaf-off TM data and NWI data to refine and augment the basic classification developed above.

The region was divided into halves for separate analysis. This was done to keep the amount of data reasonable and because scenes from the west half of the region were acquired during late summer and early autumn, whereas scenes from the east half of the region were acquired during early summer. Classification results derived from mosaics comprised of both early summer and late summer scenes might be difficult to interpret and label because of pronounced interscene phenological differences.

For mosaicking purposes, a "master" scene (Homer *et al.*, 1997) was selected, and regions of spatial overlap with adjacent "slave" scenes were used to normalize digital data. From these zones of overlap, histograms of digital values from the slave scenes were adjusted to match the histogram brightness values of the master image on a band-by-band basis. Prior to normalization, areas with clouds and water were masked out, such that normalization was performed solely using digital data from areas dominated by land cover (i.e., the primary focus of this work). Once a slave image was radiometrically matched to the master, it in turn became a master for it's adjacent scenes. While it is recognized that there may be disagreement with the logic of using multiscene mosaics for classification purposes, it should be noted that other investigators have used mosaics successfully to derive large area land-cover data sets in Utah (Edwards *et al.*, 1995; Homer *et al.*, 1997). In the Utah work, however, the techniques employed were different from ours, with the histograms of master-slave images being adjusted by maintaining histogram shape while altering relative position or bias. While we believe that the mosaicking procedure employed in this study was very effective for the purposes the static classification analysis performed, it should be noted that the procedure may not be appropriate for other types of analyses, such as change detection.

Mosaicked scenes were clustered into 100 spectrally distinct classes using the CLUSTER algorithm developed at Los Alamos National Laboratory (Kelly and White, 1993; Benjamin *et al.*, 1996). Classification was accomplished using TM bands 3 (0.63 to 0.69 μm), 4 (0.76 to 0.90 μm), 5 (1.55 to 1.75 μm), and 7 (2.08 to 2.35 μm); for some scenes, bands 1 (0.45 to 0.52 μm) and 2 (0.52 to 0.60 μm) were affected by too much haze to use with confidence. Previous work has indicated that relatively little unique land-cover information is derived by using greater numbers of clusters (Vogelmann *et al.*, 1996), and it was decided that 100 clusters would likely capture most of the regional land-cover variability that could be derived from the leaf-on summer TM data. Clusters were assigned into modified C-CAP classes (Table 2), which are analogous to Anderson level 1 and 2 land cover classes (Anderson *et al.*, 1976), using National High Altitude Photography (NHAP) program aerial photographs as reference information.

Table 1. Multiresolution Land Characteristics Landsat Thematic Mapper Data Sets Used to Develop Federal Region III Data Set.

| Path/Row | Leaf-Off Date | Leaf-On Date |
|---|---|---|
| 14/31 | 05/09/93 | |
| 14/32 | 03/25/89 | 05/20/91 |
| 14/33 | 03/15/91 | 06/10/93 |
| 14/34 | 03/15/91 | 05/04/91 |
| 14/35 | | 06/23/92 |
| 15/31 | 03/31/91 | 06/14/92 |
| 15/32 | 11/14/90 | 06/17/93 |
| 15/33 | 03/16/89 | 05/08/90 |
| 15/34 | 04/11/92 | 06/17/93 |
| 15/35 | | 05/16/93 |
| 16/31 | 03/29/90 | 06/24/93 |
| 16/32 | 03/29/91 | 06/24/93 |
| 16/33 | 04/16/91 | 09/28/93 |
| 16/34 | 04/16/91 | 09/28/93 |
| 16/35 | 03/01/92 | |
| 17/31 | 03/29/88 | 10/02/92 |
| 17/32 | 03/24/86 | 10/02/92 |
| 17/33 | 03/24/86 | 10/02/92 |
| 17/34 | 03/24/86 | 10/02/92 |
| 17/35 | | 11/03/92 |
| 18/31 | 04/22/94 | |
| 18/32 | 04/22/94 | 08/06/92 |
| 18/33 | 04/19/87 | 08/06/92 |
| 18/34 | 11/29/93 | 09/29/94 |
| 18/35 | | 10/25/92 |
| 19/35 | 04/23/92 | 09/30/92 |

Almost invariably, the individual spectral clusters derived from classification were confused between or among two or more of the targeted land-cover classes. Separation of spectral classes into more meaningful land-cover units was accomplished using ancillary data that were rasterized to the same pixel size (30 m) and using the same projection parameters (Lambert Azimuthal) as used for the imagery. Slope, aspect, and shaded relief data sets were derived from the DTED data using standard raster-based image processing software, whereas the NWI, LUDA, and population census block group data layers were obtained by rasterizing and then combining available vector-based coverages.

Briefly, for a given confused spectral class, digital values of the various ancillary data layers were compared (1) to determine which data layers were the most effective for splitting the confused class into the appropriate land-cover units, and (2) to derive the appropriate thresholds for splitting the classes. Models were then developed using one to several data sets to split each confused class into the desired land-cover categories. In this study, the ancillary data layers used for splitting of classes were elevation, slope, aspect, shaded relief, population density, city lights, LUDA, and two TM-derived vegetation indices (normalized difference vegetation index and the TM band 5/4 ratio). It was felt that all data layers chosen characterize certain land-cover features that might

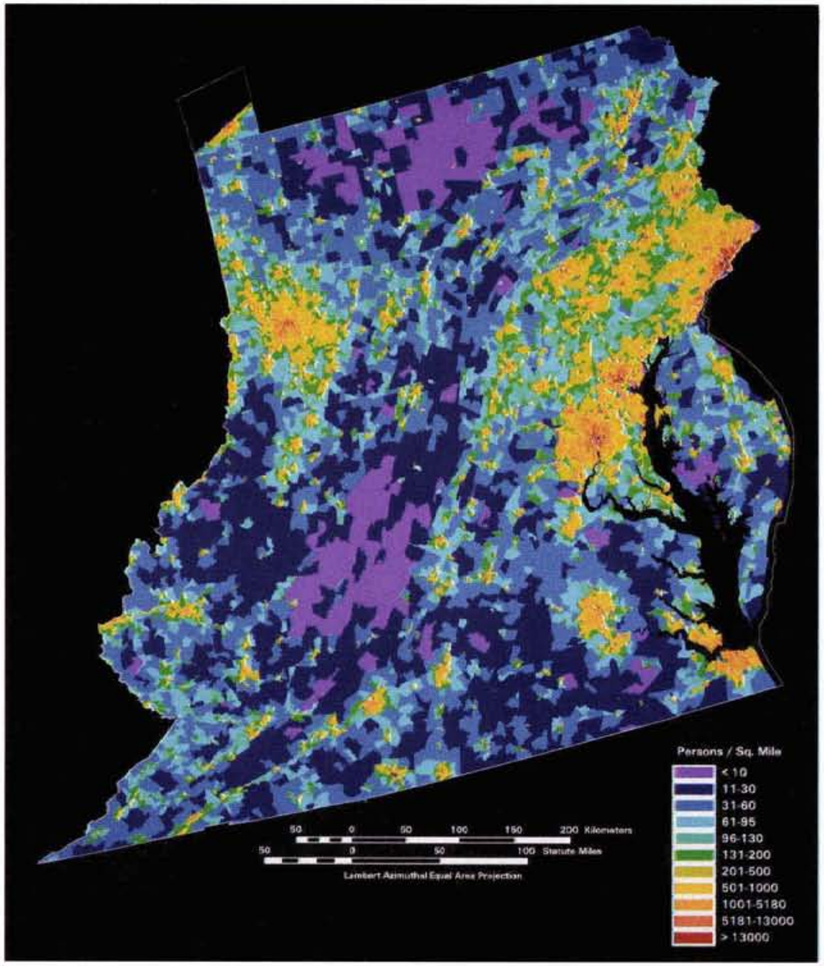


Plate 3. Population block group census data (derived from the 1990 census).

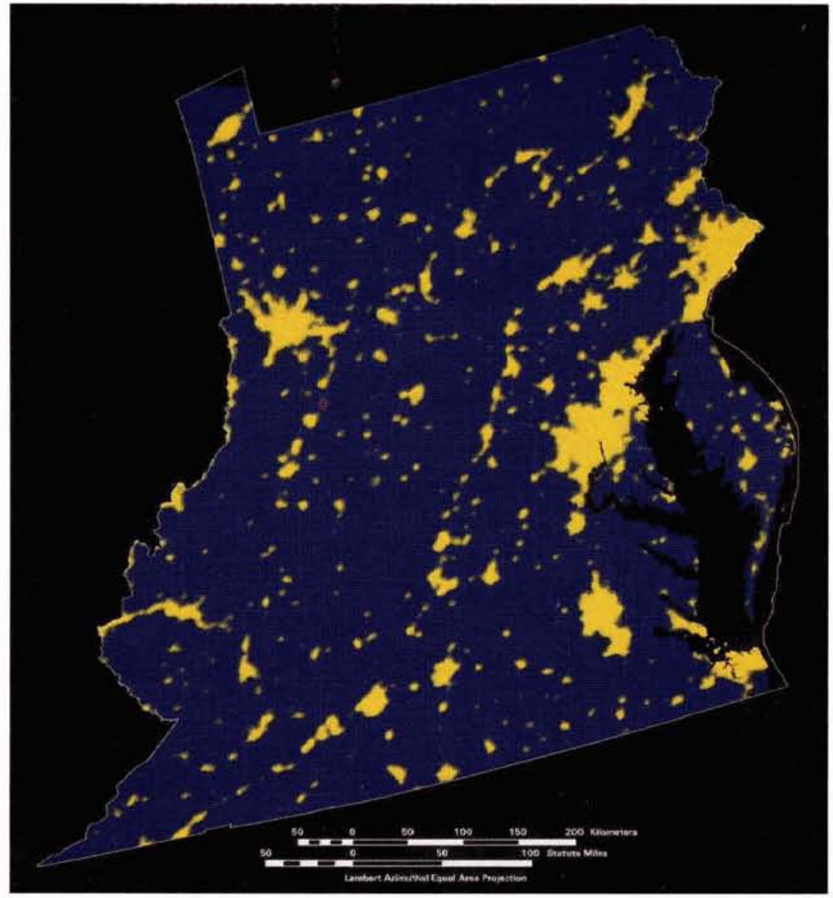


Plate 4. Defense Meteorological Satellite Program city lights data.

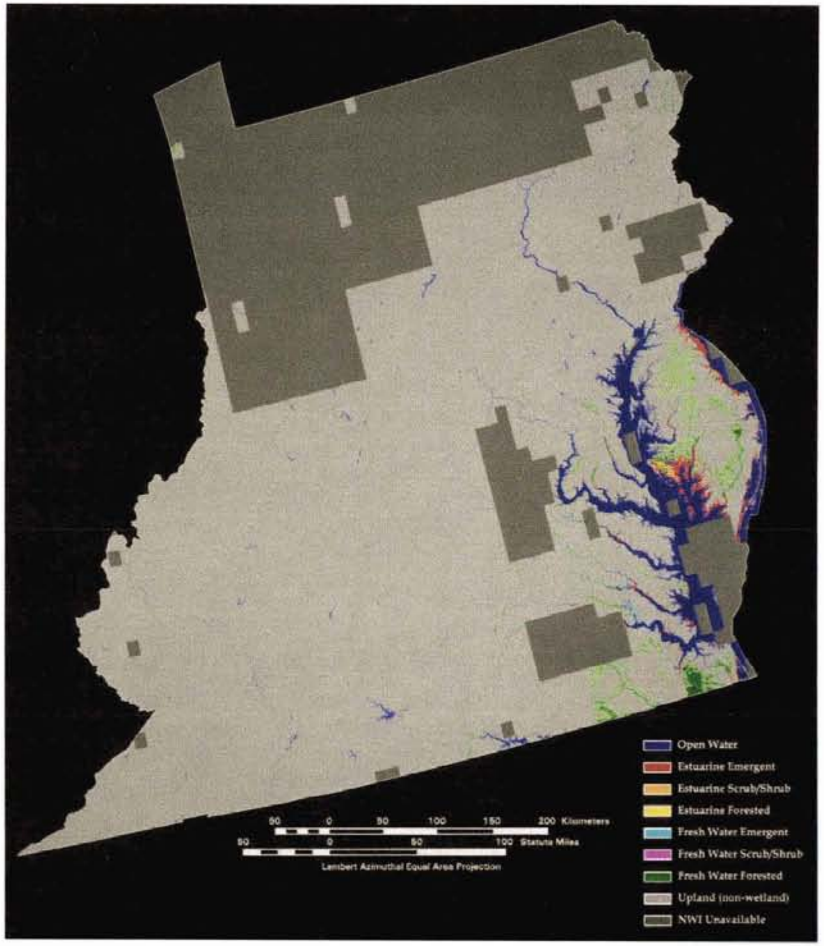


Plate 5. Land-use and land-cover data.

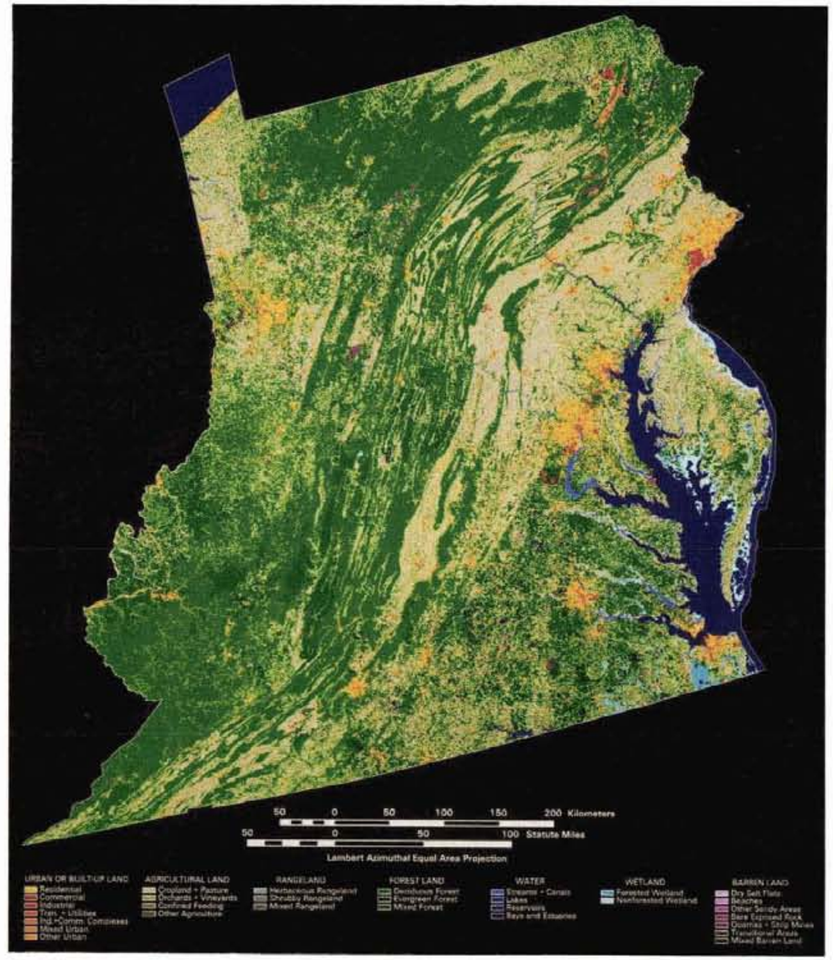


Plate 6. National Wetlands Inventory data, showing where digital wetlands information is available for Federal Region III.

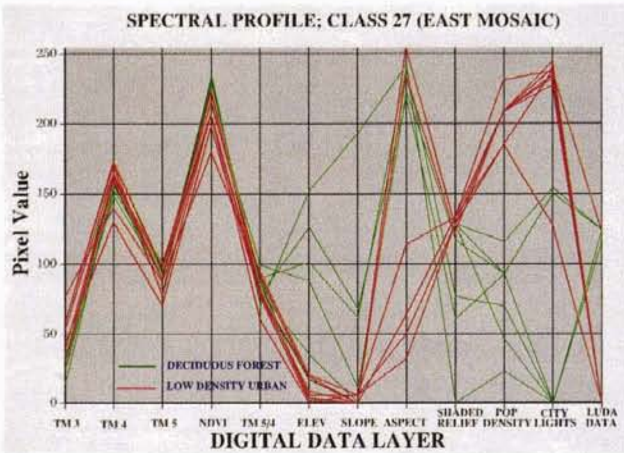


Plate 7. Sample spectral profile illustrating use of ancillary data for splitting a spectral cluster that is spectrally coincident with both deciduous forest and low intensity developed features. In this profile, it can be seen that population density, city lights, slope, elevation, and shaded relief are appropriate for discriminating these two classes.

not be readily obtainable from classification of TM imagery. The two vegetation indices were selected because both have been found to be useful in discriminating various land-cover features in other studies (e.g., Vogelmann and Rock, 1989; Vogelmann, 1990; Loveland *et al.*, 1991; Brown *et al.*, 1993) and, unlike the imagery used for clustering the data, contain information that is largely independent of slope and aspect variables.

Occasionally, we have found that "raw" band data have been useful for model development. For instance, we have found that TM band 3 is occasionally useful for splitting classes. Although we did not use TM band 7 for class splitting operations, subsequent analyses have shown the band also to be sometimes useful for reducing confusion, and we are currently using it in other land-cover mapping projects.

Of the ancillary data layers used, those found to be generally ineffective for separating confused classes included aspect and the two vegetation indices; consequently, these were seldom used in the models. Moderate success was achieved using slope information in model development. The shaded-relief data set, which contains elements of both slope and aspect variables, was found to be much more powerful for eliminating confusion than either slope or aspect data layers used separately. The shaded-relief data set was produced using solar elevation and azimuth values that approximated solar conditions during times of TM data acquisition. This data set was scaled from 1 to 255 such that low digital values represent shaded areas on northwest-facing slopes, and high digital values represent highly illuminated areas on southeast-facing slopes. This data set was especially useful for splitting spectral clusters representing different land-cover classes associated with topographically extreme versus moderate conditions (e.g., spectrally dark pixels on shaded mountain slopes with low shaded-relief digital values representing deciduous forest versus spectrally dark pixels on moderate terrain with medium shaded-relief digital values representing the low intensity developed class).

As an example of the approach employed, a multisource data profile was developed for a spectral cluster in which there was confusion between deciduous forest and low intensity developed land-cover classes (Plate 7). Each line in the illustration represents the profile of individual deciduous or developed pixels. It should be noted that, for graphic representation, values for each ancillary data layer have been scaled between 0 and 255. A linear stretch was used, such that scaled values could be readily converted to original values if desired. As would be expected, substantial differences are not evident between forested and developed pixels for TM bands 3, 4, or 5. Further inspection indicates that slope, shaded relief, population density, and city lights data are all very good for discriminating between the two land-cover categories. Elevation and LUDA data also provide some reasonably consistent separations. Additionally, this graph indicates the appropriate thresholds for each data layer. A model using a series of conditional statements that splits the cluster into two classes (deciduous forest and low intensity developed) was developed and run for a subset of the clustered mosaic. In this particular case, those class-27 pixels with elevation values greater than 100 or slope values greater than 20 or shaded-relief values between 0 and 110 were assigned to the deciduous forest class. The remaining class-27 pixels were assigned into the low intensity developed class if popu-

TABLE 2. LAND-COVER CLASSES AND DEFINITIONS USED IN THIS STUDY.

| Land-Cover Class | Definition |
|---|---|
| Water | All areas of open water, generally with greater than 25 percent cover of water |
| Low Intensity Developed | Approximately 50 to 80 percent construction materials (e.g., asphalt, concrete, building, etc.); often residential development |
| High Intensity Developed | Approximately 80 to 100 percent construction materials; typically low percentage of residential development |
| Hay/Pasture/Grass Lands | Characterized by high percentages of grasses and other herbaceous vegetation that are regularly mowed for hay and/or grazed by livestock; predominantly hay fields but also includes golf courses and city parks. |
| Row Crops | Areas regularly tilled and planted, often on an annual or biennial basis (e.g., corn, cotton, vegetable crops) |
| Conifer (Evergreen) Forest | Conifers making up 70 percent or greater of the forest (area is considered forested if trees cover 40 percent or greater area) |
| Mixed Forest | Both conifers and deciduous trees present, with neither particularly dominant |
| Deciduous Forest | Deciduous trees making up 70 percent or greater of the forest |
| Woody Wetlands | Wetlands with a substantial amount of woody vegetation present (mostly from National Wetlands Inventory) |
| Emergent Wetlands | Wetlands with a substantial amount of herbaceous vegetation present (mostly from National Wetlands Inventory) |
| Bare: Quarry Areas | Includes all quarry areas, including sand/gravel operations; sparse vegetation cover (< 20 percent) |
| Bare: Rock/Sand | Rock or sand; sparse vegetation cover (< 20 percent) |
| Bare: Transitional | Areas of sparse vegetation cover (< 20 percent) that are likely to change or be converted to other land-cover categories in the near future; includes clearcuts |

lation values were greater than 210 and city lights datan 230. Pixels not matching the above criteria were assigned into the developed class if spatially coincident with LUDA urban/residential categories; otherwise, they were assigned into the deciduous forest class. The entire procedure was very empirical, and generally it took several trials and modifications of model parameters using the subset of the mosaicked data set before the class splitting models were considered refined enough to apply to the entire region (determined by visual inspection of model runs).

Fifty-nine and 54 models were generated for east and west halves of the study region, respectively. Generally, from three to seven ancillary data layers were used for each model. After running the models, data were recombined into first-order classification products for each of the two halves of the study region. Data from the leaf-off data were then analyzed with the goal of discerning certain land-cover features not easily discriminated using leaf-on TM data. Classes that were easily defined using leaf-off data included conifer forest and hay/pasture/grass. Both are green in early spring and late autumn, and are readily discernible from each other and from almost all other land-cover categories. Leaves-off scenes were clustered individually using 50 classes with the CLUSTER algorithm described earlier. Spectral clusters that unambiguously corresponded with conifer forest, hay/pastures/grass, and mixed forest classes were identified and recoded, and this information was incorporated into the land-cover classification product. It should be noted that we felt that separate analyses of clustered leaf-on and leaf-off data sets was preferable to clustering and analyzing leaf-on and leaf-off data sets together. This is in part because the analyst can make effective use of seasonally specific phenological information during the labeling process when data sets represent distinct time periods (e.g., leaf-on versus leaf-off). Such information is more difficult to use when the two dates are clustered as a single unit. It should also be noted that previous work (Vogelmann *et al.*, 1997) has indicated that minimal gain in class discrimination is achieved after multitemporal clustering of two TM data sets as opposed to clustering of the two data sets separately. However, it should also be noted that other investigators (Slaymaker *et al.*, 1996) have achieved excellent results after multitemporal clustering of two seasonally distinct TM data sets.

Many bare areas (especially clearcuts and quarries) and wetlands are spectrally similar to other land-cover classes, and consequently are difficult to accurately classify. However, due to spatial characteristics combined with their spectral properties, these areas can often be readily discerned in the TM imagery. Consequently, in this study, quarries, clear cuts, and bare rock/sand classes were obtained by means of on-screen digitizing of the TM images. Similarly, wetlands that were especially large and clearly identifiable from the imagery were digitized where digital NWI data were absent. These digitized data sets were rasterized and re-coded, as were the digital NWI data, into the appropriate land-cover categories, and finally incorporated into the land-cover mosaics. East and west halves of the region were mosaicked, resulting in a land-cover product for the entire region. This product was then inspected in conjunction with the raw imagery, and obvious errors (especially recent residential areas in forested areas classified as row crops) were corrected on a case-by-case basis.

Although the TM data sets were mostly cloud-free and of good overall quality, there were a few clouds in several portions of the imagery. Where they occurred, cloud and cloud shadow boundaries were digitized, and data from LUDA were used to fill in these areas for the final classification product. In most cases, clouds and cloud shadows were located in forested areas that appeared to be reasonably stable, and LUDA data were an adequate surrogate for these isolated areas. Also, it should be noted that some of the leaf-off data sets were acquired at times seasonally earlier than optimal.

## Ancillary Data Quality

Numerous studies have shown that the use of ancillary spatial data with satellite-derived spectral cluster information can provide much better land-cover information than the spectral cluster information alone (e.g., Cibula and Nyquist, 1987; Loveland *et al.*, 1991; Franklin, 1994; Harris and Ventura, 1995). Generally, ancillary data have been used to aid in the class labeling procedure, and have been used to split clusters into discrete land-cover classes. Because of the impact of ancillary data quality on the final land-cover products, it is pertinent to provide a brief description of some of the characteristics of these data sources as they relate to the current project.

### DTED and Derivative Products

The digital elevation models used (Plate 2) were the digital terrain elevation data level 1 (DTED-1) products generated by the Defense Mapping Agency. These models cover 1- by 1-degree blocks, and are distributed by the USGS. The majority of the digital elevation models were produced from cartographic and photographic sources (U.S. Geological Survey, 1993). While the elevation data set and the derived shaded-relief data set both appeared reasonable for Region III, it should be noted that major block-specific differences resulting in distinct seamlines were apparent in the derived slope and aspect data sets. These are attributable to differences in the data sources used to derive the digital elevation models, which are highly variable in quality.

### Population Density

The population-density data layer was developed by linking 1990 block-group population census point coverage data with block-group boundaries derived from Topologically Integrated Geographic Encoding and Referencing line files, resulting in a spatial data layer depicting population density on a census block-group by census block-group basis (Plate 3). Spatial resolution of this data layer relates to the size of the individual census block groups, and, thus, is variable and relatively coarse throughout the data set. Nonetheless, the data set is very good at depicting areas of urban development, which are typically very difficult to delineate solely by using spectral classification methods.

### City Lights

Although acquired in a very different manner, the City Lights data set (Plate 4) provides information that has many spatial similarities with the population-density data layer (Plate 3). This data set is described in detail by Elvidge *et al.* (1996). Briefly, the data set was obtained by the U.S. Air Force Defense Meteorological Satellite Program Operational Linescan System using a visible and near-infrared band (0.5 to 0.9 μm) during nighttime. The band signal is intensified using a photomultiplier tube, making it possible to detect faint emission sources (down to $10^{-9}$ watts/cm$^2$/sr/μm). The nominal resolution of the data set used in this study was 2.8 km. This data set was generally used in conjunction with the population-density data layer in modeling/splitting clusters into developed and nondeveloped classes. The data set has advantages over the population-density data layer in that the areas of brightness corresponding to development are not confined to arbitrarily defined political boundaries. City Lights and population-density data sets both tend to overestimate developed classes because of their relatively coarse spatial resolution. We noticed while conducting this study that very small populated areas (i.e., less than 1 km$^2$) were often detectable using the City Lights data set. Thus, this data set is characterized by coarse spatial resolution and very fine radiometric resolution. While the City Lights data set tended to overestimate urbanization, regions characterized by low digital numbers were consistently non-urbanized. Thus, if a pixel from a particular cluster was located in a region with low City Lights digital values, it could be reliably assumed that it did not represent an urban condition. Cluster-splitting models were constructed to take advantage of this logic.

#### LUDA Data

The LUDA data were derived from aerial photographs from the NHAP, NASA, and various special aerial photography projects, usually at scales of less than 1:60,000 (U.S. Geological Survey, 1990). Features corresponding to Anderson Level II land-cover classes were delineated as polygons from the photographs; and minimum mapping units are 4 hectares for some categories (e.g., urban categories, water, quarry, transitional bare, confined feeding operations) and 16 hectares for the others (e.g., forest classes, agricultural classes). Because most of the aerial photographs used to generate the LUDA data are from the 1970s, much of the land-use and land-cover information is out of date. As with many large region land-cover products developed using a number of different photointerpreters, there are some seam lines between adjacent quadrangles in the LUDA data set. Depending on the region and land-cover unit being analyzed, however, much of the LUDA information is still reasonably accurate and useful. For the most part, the models developed in this project incorporated the LUDA data mostly as a "last resort" (typically utilized during the last steps of the models), thereby minimizing the impact of the data set. It should be noted that the LUDA classes represent a mix of land use and land cover; most "natural" landscape units (e.g., wetland and forest classes) represent land cover, whereas those heavily impacted by anthropogenic activities (e.g., urban and agricultural classes) represent land use. The primary focus of our work was land cover, and, thus, the LUDA classes were not always compatible with the our target classes (Table 2).

#### NWI

The NWI data depict location and classification of wetlands as defined by the U.S. Fish and Wildlife Service. The ultimate goal of the NWI program is to generate hardcopy maps and digital NWI data sets for the entire United States on a 7.5-minute quadrangle basis. The data sets are developed by photointerpretation of aerial photographs from various sources, including the National Aerial Photography Program, NHAP, Consolidated Farm Services Agency, NASA, and special project photography (U.S. Fish and Wildlife Service, 1996). The biggest problem in using this data set is that the program is at various stages of completion throughout the United States: some states have no coverage by NWI, other states have just hard copy maps, and some states have digital NWI data as well as map information available. For the purposes of this study, available digital NWI data were rasterized and re-coded into woody wetland, emergent herbaceous wetland, and open water classes and incorporated into the land-cover data set. It should be noted that the data sources used for generating NWI data range from 1971 to 1992, and thus some of the wetlands information from NWI may be outdated. However, in this project, visual comparisons between NWI and TM imagery indicated that this was not a major problem.

It should be recognized that there are many problems in unambiguously delineating wetlands. In a report by Shapiro (1995), comparisons were made between various wetland data sets, including U.S. Fish and Wildlife Service NWI data, U.S. Department of Agriculture Natural Resources Conservation Service Wetland Inventory data, National Oceanic and Atmospheric Administration (NOAA) Coastal Change Analysis Program (C-CAP) data, and State of Maryland Water Resources Administration regulatory wetlands data. These comparisons were made for a county in Maryland. It was noted that there were significant disagreements among the different data sets. The NWI data were conservative in delineating wetlands, and thus errors of commission were noted to be relatively low, whereas levels of omission were relatively high. The heavy reliance on NWI data for discriminating wetlands classes in the current investigation implies that we will have missed a number of the smaller wetlands throughout the region.

## Consistency Checks

While it is recognized that there is much value of assessing accuracies of land-cover classifications, it is difficult to implement the more traditional methods described by Congalton (1991) across especially large regions. In the current project, the land-cover data layer was compared with four sources of data. Together, these sources provide users with useful information regarding the quality of the final product. The data sources used were (1) NHAP program color infrared photographs, (2) a land-cover classification developed for the Chesapeake Bay area by NOAA's Coastal Change Analysis Program(C-CAP) (Hittelman *et al.*, 1994), and (3) 1992 Census of the Agriculture data (Bureau of the Census, 1993) . Comparisons were also made with LUDA data.

Seventy-eight NHAP photographs from throughout the study region were used for comparison with the results from the final land-cover classification. Most of the photographs were acquired during the early 1980s. Areas corresponding to the areas covered by the photographs were subset from the TM imagery as well as from the final classification. Ten samples, each representing a single pixel, were randomly selected from each image subset using ERDAS IMAGINE software, and were located on the appropriate NHAP photographs, photointerpreted, and recorded. Class values representing these same samples were obtained from the final classification and compared with photointerpreted results. For the purposes of this investigation, samples were not used unless they could be clearly located and photointerpreted on the NHAP photographs. When samples were omitted, replacement samples were randomly selected until a total of ten interpreted reference samples per photograph were obtained (780 samples in total).

In addition, a land-cover classification data set developed for the Chesapeake Bay region (Hittelman *et al.*, 1994) was obtained from the NOAA C-CAP (Dobson *et al.*, 1995) and compared with the land-cover data set developed in this study. This data set was generated using 1988 and 1989 Landsat TM data for a four TM scene region (all located within Region III), and incorporated much field data during its development. After reprojecting the NOAA C-CAP data set from Universal Transverse Mercator to Lambert Azimuthal coordinates, tables of class area estimates and spatial coincidence were generated for regions of overlapping classifications.

A third consistency check was done using 1992 Census of Agriculture (Bureau of the Census, 1993). This data set contains statistics of agricultural lands for farms that had sold $1,000 or more of agricultural products, or normally would have sold, during the census year. Data are compiled on a county basis, providing information on total cultivated and (or) managed agricultural land as well as information on individual crops. In this study, spatial representations of Census of Agriculture information were generated for the region and compared with spatial representations of amounts of agricultural land derived from the Region III land-cover product, also summarized and tabulated at the county level. Primary comparisons were made for total amount of agricultural land, total amount of grasslands/pasture/hay, and total amount of row crops.

The LUDA data (U.S. Geological Survey, 1990) represent the only land-cover data set currently available for the entire conterminous United States. As discussed earlier, this data set was used during the modeling and class splitting components of this project, although an attempt was made to keep the impact of this data set on the final land-cover product to a minimum. Even though the data set was used to produce the final land-cover data set, it was felt that comparative analyses would be helpful to users of the data. For the purposes of this investigation, LUDA data were re-coded into classes that most closely corresponded with the classes of the final land-cover data set, and class area estimates were compared.

## Results and Discussion

### Final Land-Cover Classification

The final land-cover data set (Plate 8) is mostly seamless, and, when compared with the TM three-band composite (Plate 1), appears to be reasonable in terms of general accuracy. Class area estimates for the region (Table 3) indicate that approximately 65 percent of the region is forested, 24 percent is in agriculture, and about 3 percent is developed.

### Comparison with NHAP Photographs

Comparison between the land-cover data set and the reference data obtained from the NHAP photographs (Table 4) shows that there is reasonably good agreement for a number of the classes (e.g., water, low and high intensity developed, deciduous forest, conifer forest, herbaceous wetland), while there is some disagreement for a number of other classes (e.g., row crops, hay/pasture/grass, woody wetland). The overall agreement was 74 percent, and the Kappa coefficient was 66 percent. We wish to emphasize that these numbers only relate to the degree of similarity between the two sources of data, and that they should not be interpreted as accuracy values. This caveat is issued in part because the sample size (780 samples) selected from the NHAP photographs is low given the size of the study area; a much higher number of samples would be required to in order to achieve statistical validity. In addition, there are likely to have been land-cover changes that have taken place between the time that the photographs were acquired (early 1980s) and when the images were acquired (early 1990s), which will decrease quantitative levels of similarity regardless of levels of accuracy. It is difficult to ascertain the impact of land-cover change on these values in this study. Also, it should be noted that some classes, most notably the woody wetland class, are difficult to identify accurately in the NHAP photography due to spatial and temporal scale characteristics of the photographs. Any errors in the reference data will decrease overall agreement and Kappa values. Despite these reservations, we feel that these estimates are acceptable and lend credibility to the approach taken in this study.

One of the largest sources of disagreement between the Region III land-cover data set and NHAP reference data relates to the two agricultural classes (row crops and hay/pasture/grass) and the mixed forest class. In the former case, most of the confusion is between the two agricultural classes, and not between agricultural versus nonagricultural categories. When the two agricultural classes were merged into one class in both land-cover and reference data sets, overall agreement increased to 84 percent, and the Kappa coefficient increased to 78 percent. Some of the discrepancies between the two agricultural classes are related to different cropping patterns during the times of data acquisition. Inter-annual crop rotation is a common agricultural practice throughout much of the area, and because the different data sets being compared represent different years, high levels of agreement between row crop and hay/pasture/grass classes should not be expected. It should also be noted that some of the disagreement is attributable to some of the leaf-off data sets used in the classification procedure. The hay/pasture/grass category was defined using leaf-off TM data sets, based on the principle that grasslands usually green up long before row crops, and that green herbaceous vegetation is markedly spectrally different from other classes at this time of year. However, some of the leaf-off data sets used for this discrimination were acquired too early for optimal separation of these two classes (i.e., before consistent grass green-up).

Accurate delineation of mixed forest as an unambiguous class is problematic in many land-cover mapping studies. This is in part both a definitional problem as well as a photointerpretation problem. In our study, comparisons with NHAP photographs, as well as LUDA comparisons, indicate that it is probably best to regard the mixed forest class simply as a class that can represent any one of the three upland forest land-cover categories.

### C-CAP Classification

Results from the Region III classification compare very favorably with the data set developed by the C-CAP program using 1988-89 Landsat TM data (Table 5; Plate 9). The C-CAP data set was generated for a much smaller area (covered by four TM scenes as opposed to 21 scenes covered in the current effort) and was more field intensive, and thus it is reassuring that the data sets are visually as similar as they are. Some of the differences are due to the 3 by 3 filtering done on the C-CAP product. Results from the Region III land-cover classification were not filtered.

Both C-CAP and Region III land-cover data sets provided similar class area estimates (Table 5). Some of the most pronounced differences related to the two agricultural classes. While the total amount of land in agriculture compared very favorably between the two data sets (27.7 percent versus 30.1 percent for the Region III and C-CAP data sets, respectively), there were marked differences between the types of agricultural land (i.e., hay/pasture/grass versus hay/pasture/grass; row crop versus row crop). The TM data analyzed for the C-CAP project were from 1988-1989, as opposed to the early 1990s for the Region III data set. As noted in the comparisons between NHAP photographs and the Region III product, it is likely that the some of the observed differences relate to changes in agricultural practices between the two time periods.

While the area estimates provided in Table 5 provide some useful comparative information, it should be noted that estimates of spatial coincidence provide more precise information regarding the degree that the data sets match. Spatial coincidence values were also assessed between the C-CAP and Region III data sets. It was noticed that spatial coincidence values were very high for some categories (e.g., water, herbaceous wetland) but were relatively low for others (e.g., woody wetland, mixed forest, conifer forest). In general, spatial coincidence values are especially sensitive to subtle changes in the spatial boundaries of the various classes, and that the results from these types of analyses need to be assessed using caution. This is especially true for classes that are highly dissected and narrow (e.g., low intensity developed), as opposed to classes that tend to be much more blocky and homogeneous (e.g., open water). Slight differences in georeferencing will cause major differences in spa-


Plate 8. Land-cover data set for Federal Region III developed in this investigation.

tial coincidence values for these dissected classes. Additionally, we presume that the filtering of the C-CAP product contributed to some of the differences noted.

#### 1992 Census of the Agriculture

County-wide spatial comparison of percentage "agricultural land" between the Region III land-cover product and Census of the Agriculture statistics indicated similar patterns and, thus, general overall consistencies. As a general rule, the percentage values derived from the Region III product were higher than the values from the Census of the Agriculture. This can be traced in part to the methods applied by the Census of Agriculture to derive their estimates. Only farms with agricultural-related incomes of $1000 or greater annually are used to estimate amount of land in agriculture, and this likely excludes a substantial amount of land that should be categorized as agricultural (e.g., land maintained as fields, but not used for profit). In addition, the classification developed for Region III relates to land cover, whereas Census of Agriculture data relate most closely to land use. This becomes somewhat problematic for such areas as golf courses, city parks, and large residential lawns. These areas are biophysically similar to agricultural classes in that they are fertilized, sprayed, and cut, and were classed as hay/pastures/grass in the Region III product. For the purposes of this comparison, these areas were treated as "agricultural." However, these areas are not considered agricultural by the Census, and represent a source of disparity between the two data sources.

It should be noted that similar spatial representations were generated comparing percentage hay/pasture/grassland values and row crop values separately. The hay/pasture/grassland values compared very favorably, whereas the row crop values did not. In general, the row crop class can be difficult to classify using TM data, and in this project, we suspect that a number of poor quality grassy fields were classified as row crop.

#### LUDA Data

The LUDA data set is the only intermediate-scale source of land cover information currently available for the conterminous United States. We believe that, of the various consistency checks made in this study, comparison with LUDA data would provide the least effective information for evaluating the quality of the Region III data set. Nonetheless, LUDA data have been widely used in the past, and it is worthwhile to provide users with comparative information between the two data sets. Certainly, the LUDA data compared less well with the Region III product than with the C-CAP classification (Plate 9). This is not surprising because of the coarse spatial nature of LUDA data. However, there are some similarities with the Region III classification, at least at a gross level (Table 3).

The Region III product had substantially more woody

Table 3. Class Area Estimates from Area of LUDA and Region III Land-Cover Classification Overlap. For Purposes of Comparison, LUDA Urban Classes Were Merged into Low Intensity Developed (Mostly Residential) and High Intensity Developed (Mostly Commercial/Industrial) Classes.

| | Region III Classification | | LUDA Data Set | |
|---|---|---|---|---|
| Class | Number of Hectares | Percentages | Number of Hectares | Percentage |
| Water | 1,867,717 | 5.7 | 1,728,899 | 5.2 |
| Low Intensity Developed | 827,357 | 2.5 | 1,240,006 | 3.8 |
| High Intensity Developed | 264,574 | 0.8 | 621,790 | 1.9 |
| Hay/Pasture/Grass and Row Crop | 7,916,678 | 24.0 | 9,447,102 | 28.7 |
| Evergreen Forest | 1,809,320 | 5.5 | 1,292,159 | 3.9 |
| Mixed Forest | 3,343,473 | 10.2 | 4,570,696 | 13.9 |
| Deciduous Forest | 15,788,814 | 47.9 | 13,256,777 | 40.3 |
| Woody Wetland | 507,178 | 1.5 | 188,056 | 0.6 |
| Emergent Wetland | 270,577 | 0.8 | 225,259 | 0.7 |
| Bare: Quarries/Gravel Pits | 156,993 | 0.5 | 248,523 | 0.8 |
| Bare: Rock/Sand | 4,370 | 0.0 | 10,657 | 0.0 |
| Bare: Transitional | 178,180 | 0.6 | 105,310 | 0.3 |

wetlands than did the LUDA data set. In the Region III classification, this class was derived in large part from the NWI data, which has been shown to be a conservative data source of wetlands information (Shapiro, 1995). Thus, the LUDA data clearly underestimate the extent of this class.

Another notable source of disagreement between LUDA and Region III classification data pertains to the urban/developed classes. The sum total of the percentage area of the two developed classes from the Region III data set (1990s vintage) is 3.3 percent, as compared with 5.7 percent for the LUDA data set (1970s vintage). This is a potential source of confusion to those that use both data sets, because area estimates of urban/developed classes would normally be expected to increase rather than decrease over time. In reality, these differences are not related to inaccuracies in either of the two data sets, but merely represent differences in spatial detail and definitions used. The minimum mapping unit for delineating urban classes was 4 hectares for LUDA as opposed to 30 metres for the Region III product, and thus the LUDA data set "misses" many relatively small non-urban features located within predominantly urban settings. In addition, LUDA urban classes are related to land use (e.g., commercial, industrial, residential), whereas the Region III land-cover developed classes (as well as those in the C-CAP product) were defined as percentages of built-up land (Table 2). Thus, a one-to-one correspondence between the urban/developed classes of the two data sets should not be expected.

Other discrepancies between the LUDA data set and the classification results from this study relate to actual land-cover changes that have occurred over the last two decades. This is especially true for some of the bare categories, such as quarry areas and clearcuts (transitional bare), as well as for areas of urban growth. These classes will not exhibit high levels of spatial similarities because of the temporal nature of these categories.

## Conclusions

The approach described in this paper has yielded a very good land-cover classification product for a large region. Although there are some classification errors within the data set (most notably, row crops versus hay/pastures/grasslands), the large-area product appears to have many desirable characteristics (e.g., mostly seamless, and reasonable in terms of accuracy based on visual inspection and consistency checks). Because of the scope of the study, we wish to emphasize that the data set is especially appropriate for regional analyses and applications. Currently, the data set is being used for

Table 4. Consistancy Matrix of NHAP Photograph-Interpreted Points and Corresponding Region III Land-Cover Classification Results. Values Represent Numbers of Observations for Each Pair of Classes. Based on 780 Points.

| | | NHAP Data | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Class | Water | Urban Low Inten | Urban High Inten | Grass | Row Crop | For; Evergreen | For; Mix | For; Dec | Wet-land; Wood | Wet-land Herb. | Bare; Quarries | Bare; Rock/Sand | Bare; Trans |
| Region III Land-Cover Data | Water | 35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| | Urban; Low Intensity | 0 | 47 | 4 | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| | Urban; High Intensity | 0 | 5 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Grass | 0 | 2 | 0 | 41 | 27 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Row Crop | 0 | 2 | 2 | 48 | 67 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 |
| | Forest; Evergreen | 0 | 0 | 0 | 0 | 2 | 41 | 7 | 2 | 1 | 0 | 0 | 0 | 1 |
| | Forest; Mixed | 0 | 2 | 0 | 1 | 0 | 5 | 13 | 23 | 1 | 0 | 0 | 0 | 0 |
| | Forest; Deciduous | 0 | 2 | 0 | 7 | 4 | 11 | 18 | 297 | 2 | 0 | 1 | 0 | 2 |
| | Wetland; Woody | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 6 | 7 | 0 | 0 | 0 | 0 |
| | Wetland; Herbaceous | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 13 | 0 | 0 | 0 |
| | Bare; Quarries | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 |
| | Bare; Rock/sand | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Bare; Transitional | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |

TABLE 5. CLASS AREA ESTIMATES FROM AREA OF C-CAP AND REGION III LAND-COVER CLASSIFICATION OVERLAP (4 TM SCENE REGION).

| | Region III Classification | | C-CAP Classification | |
|---|---|---|---|---|
| Class | Number of Hectares | Percentage | Number of Hectares | Percentage |
| Water | 1,587,363 | 21.1 | 1,550,580 | 20.6 |
| Low Intensity Developed | 331,206 | 4.4 | 315,258 | 4.2 |
| High Intensity Developed | 101,930 | 1.4 | 193,418 | 2.6 |
| Hay/Pasture/Grass | 772,688 | 10.3 | 1,202,697 | 16.0 |
| Row Crop | 1,306,505 | 17.4 | 1,060,509 | 14.1 |
| Evergreen Forest | 500,635 | 6.6 | 320,386 | 4.3 |
| Mixed Forest | 766,728 | 10.2 | 451,181 | 6.0 |
| Deciduous Forest | 1,554,865 | 20.6 | 1,651,772 | 21.9 |
| Woody Wetland | 301,494 | 4.0 | 323,456 | 4.3 |
| Emergent Wetland | 225,216 | 3.0 | 214,244 | 2.8 |
| Bare Soil/Rock/Sand Classes | 17,347 | 0.2 | 7,645 | 0.1 |
| Transitional Bare | 63,660 | 0.8 | — | — |
| Shrub/Scrub | — | — | 231,557 | 3.1 |
| Tidal Flats | — | — | 6,993 | 0.1 |

various activities, including derivation of regional landscape pattern metrics, providing land-use planners with general land-cover area estimates, and for use in pesticide/herbicide runoff models. We recognize, however, that many local scale phenomena may have been missed in such an effort, and that there is no surrogate for more in-depth analyses for obtaining more detailed and precise information relating to localized conditions. For these latter purposes, we believe that the Region III data set may be useful for providing a first-order overview.

The methods described in this study are very empirical. The class-by-class splitting operations employed involve many interactive steps and require numerous decisions by the analyst. While there are certainly other methods of generating land-cover information using multiple sources of data, we found the current approach reasonably efficient, and provided a reasonably consistent land-cover data set. Other techniques, such as regression tree analysis (Michaelsen *et al.*, 1994) and neural nets hold much promise for automating the procedure and decreasing the number of decisions that an analyst needs to make during the course of the work. While certainly meriting exploration, such research was beyond the scope of this investigation.

We view this project as a first step towards the generation of a base-line intermediate-scale land-cover data set for the conterminous United States, and are currently expanding our analyses to other Federal Regions in the eastern United States. We will explore incorporating Gap Analysis (Scott *et al.*, 1996) data when they become available on a state-by-state or region-by-region basis, with the potential of resulting in a product with much more detail for the natural land-cover classes. Ultimately, we wish to use the general approach described here to generate a thematically and spatially consistent national land-cover data set for multiple applications.

## Acknowledgments

This work was performed by the Hughes STX Corporation under U.S. Geological Survey Contract 1434-92-C-40004. We

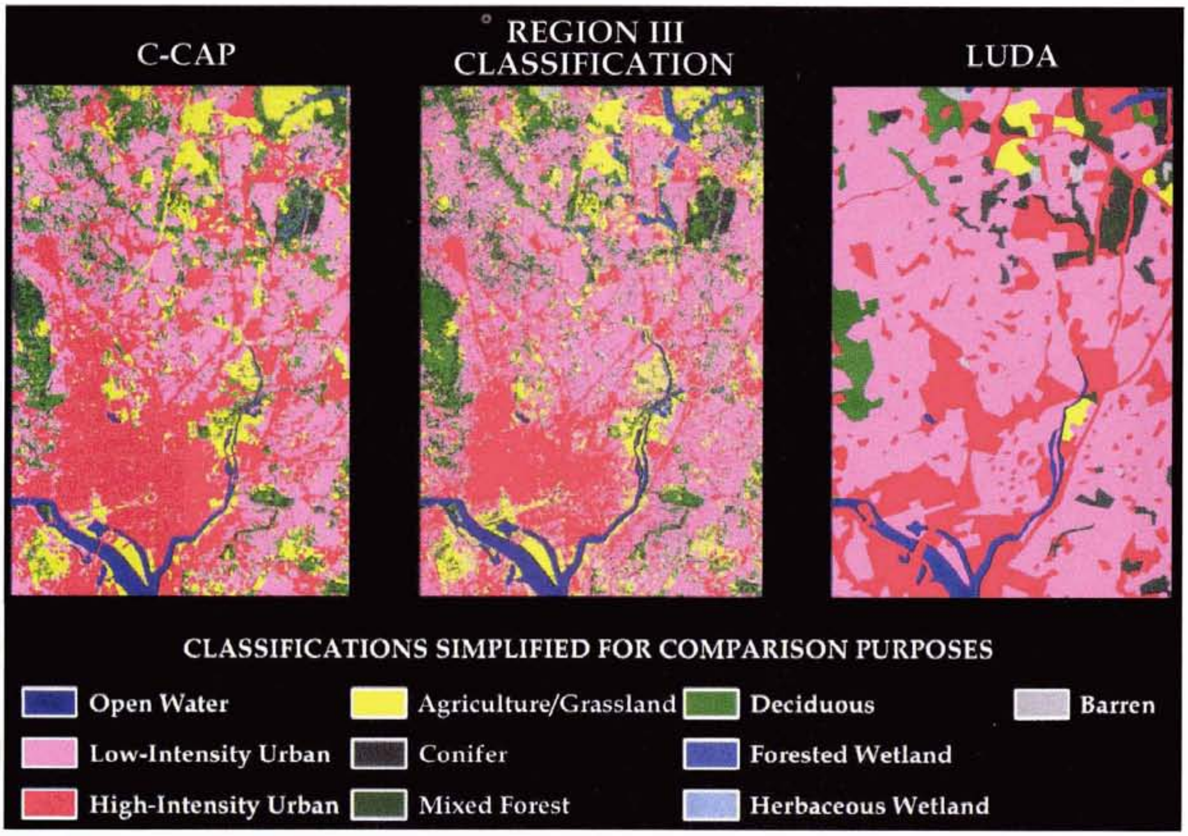


Plate 9. Comparison among Coastal Change Analysis Program data, Region III data developed in this study, and land-use and land-cover data for a portion of the region in the Washington, D.C. area.

thank the following people for their many contributions throughout the course of this study: Tom Loveland, Jess Brown, Paul Seevers, Brenda Jones, Don Ohlen, Kent Hegge, Luke Hieb, Paul Severson, Pete Campbell, and Denice Shaw.

## References


Anderson, J.F., 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. Geological Survey Professional Paper 964, 28 p.

Benjamin, S., J.M. White, D. Argiro, and K. Lowell., 1996. Land cover mapping with Spectrum, *Gap Analysis: A Landscape Approach to Biodiversity Planning* (J.M. Scott, T. Tear, and F. Davis, editors), Proceedings of the ASPRS/GAP Symposium, National Biological Service, Moscow, Idaho, pp. 279–288.

Brown, J.F., T.R. Loveland, J.W. Merchant, B.C. Reed, and D.O. Ohlen, 1993. Using multisource data in global land cover characterization: Concepts, requirements, and methods, *Photogrammetric Engineering & Remote Sensing*, 59(6):977–987.

Bureau of the Census, 1991a. *Census of Population and Housing, 1990*, public law 94–171 data (United States) (machine readable data files), The U.S. Bureau of the Census (producer and distributor), Washington, D.C.

———, 1991b. *Census of Population and Housing, 1990*, public law 94-171 data, on-line documentation (United States), The U.S. Bureau of the Census, Washington, D.C.

———, 1992. *TIGER/Line Files* (machine readable data files), The Bureau of the Census (producer and distributor), Washington, D.C.

———, 1993. *Census of the Agriculture, Final County Files* (machine-readable data file), 1992, The Bureau of the Census (producer and distributer), Washington, D.C.

Cibula, W.G., and M.O. Nyquist, 1987. Use of topographic and climatological models in a geographical data base to improve Landsat MSS classification for Olympic National Park, *Photogrammetric Engineering & Remote Sensing*, 53:67–75.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37:35–46.

Dobson, J.E., E.A. Bright, R.L. Ferguson, D.W. Field, L.L. Wood, K.D. Haddad, H. Iredale III, J.R. Jensen, V.V.Klemas, R.J. Orth, and J.P. Thomas, 1995. *NOAA Coastal Change Analysis Program (C-CAP): Guidance for Regional Implementation*, NOAA Technical Report NMFS 123, U.S. Department of Commerce, Seattle, Washington.

Edwards, T.C., Jr., C.G. Homer, S.D. Bassett, A. Falconer, R.D. Ramsey, and D.W. Wight, 1995. *Utah Gap Analysis: An Environmental Information System*, Final Project Report 95-1, Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, Utah, 60 p.

Elvidge, C.D., K.E. Baugh, E.A. Kihn, H.W. Kroehl, and E.R. Davis, 1997. Mapping city light with nighttime data from the DMSP Operational Linescan System, *Photogrammetric Engineering & Remote Sensing*, 63(6):727–734.

Franklin, S.E., 1994. Discrimination of subalpine forest species and canopy density using digital CASI, SPOT PLA and Landsat TM data, *Photogrammetric Engineering & Remote Sensing*, 60:1233–1241.

Harris, P.M., and S.J. Ventura, 1995. The integration of geographic data with remotely sensed imagery to improve classification in an urban area, *Photogrammetric Engineering & Remote Sensing*, 61:993–998.

Hitt, K.J., 1992. *Digital Map File of 1990 Population and Housing Data for the United States, Edition 1.2* (digital data file), U.S. Geological Survey, Reston, Virginia.

Hittelman, A. M., H. Iredale, M. Hollinger, and D. Field, 1994. *Coastal Change Analysis Project, Chesapeake Bay Region (1984–1989), CD-ROM User's Manual* (and accompanying machine-readable data files), U.S. Department of Commerce, 22 p.

Homer, C.G., D. Ramsey, T.C. Edwards, Jr., and A. Falconer, 1997. Landscape cover-type mapping and modeling using a multi-scene Thematic Mapper mosaic, *Photogrammetric Engineering & Remote Sensing*, 63:59–67.

Klemas, V., J.E. Dobson, R.L. Ferguson, and K.D. Haddad, 1993. A coastal land cover classification system for the NOAA Coast Watch Change Analysis Project, *J. Coastal Res.*, 9:862–872.

Loveland, T.R., J.W. Merchant, D.O. Ohlen, and J.F. Brown, 1991. Development of a land-cover characteristics database for the conterminous U.S., *Photogrammetric Engineering & Remote Sensing*, 57:1453–1463.

Loveland, T.R., and D.M. Shaw, 1996. Multiresolution land characterization: Building collaborative partnerships, *Gap Analysis: A Landscape Approach to Biodiversity Planning* (J.M. Scott, T. Tear, and F. Davis, editors), Proceedings of the ASPRS/GAP Symposium, Charlotte, North Carolina, National Biological Service, Moscow, Idaho, pp. 83–89.

Michaelsen, J., D.S. Schimel, M.A. Friedl, F.W. Davis, and R.C. Dubayah, 1994. Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys, *Journal of Vegetation Science*, 5:673–686.

Kelly, P.M., and J.M. White, 1993. Preprocessing remotely sensed data for efficient analysis and classification, *Applications of Artificial Intelligence 1993, Knowledge-Based Systems in Aerospace and Industry, Proceedings of SPIE*, pp. 24–30.

Omernik, J.M., 1987. Ecoregions of the conterminous United States, *Annals of the Association of American Geographers*, 77:118–125.

Scott, J.M., T.H. Tear, and F.W. Davis (editors), 1996. *Gap Analysis. A Landscape Approach to Biodiversity Planning*, Proceedings of the ASPRS/GAP Symposium, National Biological Service, Moscow, Idaho, 320 p.

Shapiro, C., 1995. *Coordination and Integration of Wetland Data for Status and Trends and Inventory Estimates*, Federal Geographic Data Committee Technical Report 2, U.S. Geological Survey, 210 p.

Slaymaker, D.M., K. M. L. Jones, C. R. Griffin, and J. T. Finn, 1996. Mapping deciduous forests in southern New England using aerial videography and hyperclustered multi-temporal Landsat TM imagery, *Gap Analysis: A Landscape Approach to Biodiversity Planning* (J.M. Scott, T. Tear, and F. Davis, editors), Proceedings of the ASPRS/GAP Symposium, National Biological Service, Moscow, Idaho, pp. 87–101.

U.S. Fish and Wildlife Service, 1996. *National Wetlands Inventory (NWI) Metadata*, U.S. Fish and Wildlife Service, National Wetlands Inventory, St. Petersburg, Florida.

U.S. Geological Survey, 1990. *Land Use and Land Cover Digital Data from 1:250,000- and 1:1,000,000-Scale Maps, Data Users Guide 4*, U.S. Geological Survey, Department of the Interior, Reston, Virginia, 33 p.

———, 1993. *US GeoData Digital Elevation Models, Data Users Guide 5*, U.S. Geological Survey, Department of the Interior, Reston, Virginia, 51 p.

Vogelmann, J.E., 1990. Comparison between two vegetation indices for measuring different types of forest damage in the northeastern United States, *International Journal of Remote Sensing*, 11: 2281–2297.

Vogelmann, J.E., and B.N. Rock, 1988. Assessing forest damage in high-elevation coniferous forests in Vermont and New Hampshire using Thematic Mapper data, *Remote Sensing of Environment*, 24:227–246.

Vogelmann, J.E., P.M. Seevers, and M. Oimoen, 1996. Effects of selected variables for discriminating land cover: Multiseasonal data, different clustering algorithms, and varying numbers of clusters, *Proceedings of the Pecora 13 Symposium*, Sioux Falls, South Dakota, 20–22 August, in press.


(Received 14 August 1996; accepted 09 December 1996; revised 12 March 1997)