Completion of the 1990s National Land Cover Data Set for the Conterminous United States from Landsat Thematic Mapper Data and Ancillary Data Sources

BY JAMES E. VOGELMANN, STEPHEN M. HOWARD, LIMIN YANG, CHARLES R. LARSON, BRUCE K. WYLIE, NICK VAN DRIEL

Introduction

General land cover information is required for many environmental, land management, and modeling applications. This information may be derived at a range of spatial scales, which drives the appropriate usage of the data. Spatially coarse land characterization data sets, such as those derived from the Advanced Very High Resolution Radiometer (AVHRR; e.g. Loveland et al., 1991; Brown et al., 1993) are well suited for global analyses. However, such coarse-scale data sets are often of limited value for regional investigations, and are certainly not appropriate for local land use studies. Similarly, fine resolution land cover data sets (e.g. four meter or less spatial resolution) are very appropriate for local land use planning, but are generally inappropriate for regional to global analyses.

In late 2000, the U.S.Geological Survey (USGS) EROS Data Center (EDC) completed the circa 1992 National Land Cover Data set (NLCD). The NLCD, derived from early 1990s Landsat Thematic Mapper (TM) imagery and other sources of digital data, represents an intermediate-scale national land cover data set. The resolution of this data set lends itself to many regional to national scale investigations, including analyses of water quality, ecosystem health, wildlife habitat, land cover assessments and other land management issues. The purpose of this paper is to describe the characteristics and uses of this data set.

One of the goals in the development of the NLCD was to generate a reasonably consistent and seamless 30 meter product for the conterminous United States. The methodology employed to develop the NLCD is analogous to the database approach originally envisioned by Lauer (1986). The early developmental stages of the data set have been described elsewhere (Vogelmann et al. 1998a and b). In addition to the spectral information provided by the TM, ancillary spatial data were used to improve classification results, when appropriate. The NLCD may be considered as a replacement/update to the intermediate-scale Land Use and Land Cover data set (USGS, 1990) developed in the 1970s and early 1980s.

Materials and Methods

TM Data. NLCD is based primarily on 1992 vintage Landsat 5 Thematic Mapper (TM) data that were purchased and pre-processed to a common set of specifications for the Multiresolution Land Characteristics (MRLC) consortium (Loveland and Shaw, 1996). The partner agencies, which included the USGS, the U.S. Environmental Protection Agency (EPA), National Oceanic and Atmospheric Administration (NOAA) and the U.S. Forest Service (USFS), pooled resources to purchase Landsat TM scenes for each of over 400 path/row locations in the lower 48 states. Land cover classification was carried out by several USGS groups at different locations and an outside contractor.

Landsat TM data used to develop the NLCD were terrain-corrected using 3-arc-second digital terrain elevation data (DTED; U.S. Geological Survey, 1993). The TM data were geo-registered to the Albers Equal Area projection grid using ground control points, resulting in a root mean square registration error of less than one pixel (30 m). Two or more TM data sets for each path/row, representing different seasons, were used to generate the NLCD product. Two or more TM scenes representing different times of the growing season (e.g. leaf-on and leaf-off) generally improves upon the quality of land cover information that can be derived as compared to analysis of a single scene.

Ancillary Spatial Data. In addition to TM data, a variety of other intermediate-scale spatial data were used to help develop the NLCD; these included Digital Terrain Elevation Data (DTED) and derivative DTED products (slope, aspect, shaded relief), population density data at the census block level (Bureau of the Census, 1991 a and b; 1992), Land Use and Land Cover data (USGS 1990), and digital National Wetlands Inventory data (NWI; Fish and Wildlife Service, 1996). Other data sets were used to a lesser extent, and included available water capacity and organic carbon (0-40 cm depth) data derived from the State Soil Geographic (STATSGO) Data Base (U.S. Department of Agriculture, 1994), and land cover information derived from various state or national programs. Land cover data from the USGS Biological Resource Division Gap Analysis Program (Scott et al., 1996) were used when available.

Classification System. The NLCD classification system (Table 1) provides a consistent hierarchical approach to defining 21 classes of land cover across the conterminous United States. The classification ap-

Digital	Value Class
11	Open Water
12	Perennial Ice/Snow
21	Low Intensity Residential
22	High Intensity Residential
23	Commercial/Industrial/Transportation
31	Bare/Transitional
32	Quarries/Strip Mines/Gravel Pits
33	Bare Rock/Sand
41	Deciduous Forest
42	Evergreen Forest
43	Mixed Forest
51	Shrubland
61	Orchard/Vineyards/Other
71	Grasslands/Herbaceous
81	Pasture/Hay
82	Row Crops
83	Small Grains
84	Fallow
85	Urban/Recreational Grasses (parks, golf course, cemetery, etc)
91	Woody Wetland
92	Herbaceous Wetlands

Table 1. National Land Cover Data set Classification System.

proach merged existing schemes, including the NOAA Coastal Change Analysis Program (C-CAP) classification protocol (Dobson et al., 1995) and the Anderson system (Anderson et al., 1976). The intent behind the classification system is to enable comparison with other sources of land cover information.

General Classification Methods. Mosaics of leaf-on (i.e. summer) and leaf-off (i.e. spring) imagery were generated for each of 31 regional units (based on political administrative units, contiguity of Landsat scenes and data volume) covering the conterminous United States (Figure 1). Each unit was classified separately using one of several methods. In most cases the general thematic approach was to designate either mosaic (leaf-on or the leaf-off) as the "baseline" data set, and to use that spectral information as the primary source of information from which to derive the classification product. The decision as to which mosaic to use was based on a subjective evaluation of which appeared to be the "best" mosaic in terms of overall data quality and information content. Leaf-off mosaics were chosen as baseline data sets more frequently than leaf-on mosaics. Ancillary spatial information, including spectral data from the image mosaic representing the other season, was used to refine or aid in the labeling process. After development of a first order classification product, a series of recoding operations were performed to fix obvious misclassifications and to further refine the classification.

During the initial stages of the project, the primary steps for generating the NLCD classification product were: (1) Cluster the baseline mosaic using unsupervised classification. (2) Interpret and label clusters using aerial photographs. (3) Resolve confused clusters by constructing logical or threshold models that utilize appropriate ancillary data. (4) Develop and incorporate information from onscreen digitizing (e.g. quarries, transitional bare areas). As the project progressed, some classification teams individualized the approach on a case-by-case basis. More in-depth methodology for this process has been described elsewhere (Vogelmann et al., 1998 b). Modifications were based upon data quality issues, characteristics of the region being analyzed, and familiarity with other approaches that would facilitate and/or more readily automate the classification process.

In the case of multiscene mosaics, spectral clusters developed from unsupervised clustering can be very complex, and a single cluster may represent many different types of land cover. In these cases, splitting the clusters into meaningful land cover units via modeling based on spectral and ancillary data can be quite difficult. Not only are the thresholds used to make land cover separation important, but the order of threshold implementation can also have substantial effects on the land cover estimates. Determining the optimal set of thresholds and the optimal order of implementation for complex "confused clusters" can be time consuming and difficult. One approach

CONTINUED ON PAGE 652

Figure 1. Regional units used from which TM mosaics were generated and classified.

CONTINUED FROM PAGE 651

taken was to use decision trees to facilitate the modeling process.

Decision trees derive objective, efficient, and ordered thresholds using non-parametric techniques (Friedl and Brodley 1997 and Hansen et al. 1996). Decision trees have been successfully used to derive land cover (Friedl and Brodley 1997) and to identify important data layers or spectral bands (Hansen et al. 1996; Prince and Steininger 1999). For decision tree training, the image clusters were included only as an ancillary data set. Multiple decision trees trained with different data layer set combinations and/or decision tree parameter options produced multiple land cover maps. Majority land cover from the multiple land cover maps for each pixel was used as a preliminary land cover map. Decision trees were also used to help define classification "rules" in an expert system classifier (ERDAS 1998). The expert system allowed quick identification of which rules affected which pixels and quick modification of either rules or rule confidence levels. While it was apparent that the decision trees maintained a high degree of detail in the land cover products, the need for visual inspection and "heads up" on-screen corrections of the preliminary land cover persisted.

In all cases, results were scrutinized in order to ensure comparability of land cover data among regions and to correct obvious classification errors. Edge-matching of adjacent mosaics was then performed to yield a reasonably seamless national-scale product. This was a major task due to seasonal and interpretation differences that invariably resulted in thematic seam lines at the boundaries of the mosaics. As the mosaics were finally pieced together, individual states were extracted using boundaries defined by the 1:100,000 scale Digital Line Graph series. The state files were designated "preliminary" and made available to the user community for review and feedback. Initially, the state land cover files were available to users who contacted the USGS/EDC Land Cover Applications Center to gain access to the data. The intent was to ensure that users understood the "preliminary" nature of the data, as well as its limitations, and to register the user and solicit feedback regarding the utility of the data and any problems that were identified. As feedback was received, it was reviewed and a determination made if an update was required. In many cases, updates were made, and the registered users were apprised of the changes.

Accuracy Assessment Methods. When all the states comprising a Federal Region were finalized, the accuracy assessment (AA) phase was initiated. At this time, the AA for regions 1-4 (Figure 2) are complete, regions 5,7, and 10 are underway, and the remainder are in the planning stages. The accuracy assessment was based on interpretations of 1990 vintage aerial photographs acquired by the National Aerial Photography Program (NAPP). The accuracy assessment of NLCD was achieved with 1) a probability sampling design; 2) a response design for reference data evaluation; and 3) an analysis procedure for estimation of accuracy parameters.

The sampling design incorporated three levels of stratification and a two-stage cluster sampling protocol (Stehman et al., 2000). Each Federal region constituted a stratum and was sampled independently. Within each mapping region, geographic strata were created using 15 x 15 or 30 x 30 minute grid cells, depending on the size of the region. Primary sampling units (PSU) defined by non-overlapping, interior regions of NAPP were delineated within these strata. A single PSU was randomly selected from each grid cell, with all PSU's having an equal probability of being selected. All pixels selected within the first-stage PSU's were stratified by mapped landcover class, and a simple random sample of approximately 80 to 100 pixels was selected for each land-cover class.

To obtain reference land cover class labels, each sample (pixel) was identified on a NAPP aerial photograph. A suite of attributes was collected by photointerpreters, including primary and alternate landcover label (an alternate reference label only provided when appropriate), landcover heterogeneity in the vicinity of the sample unit, and a confidence rating of the photointerpreted landcover label. For a more detailed discussion on the reference data collection and evaluation, refer to Yang et al. (2000) and Zhu et al. (2000). For each mapping region, stratified sampling formulas were applied to estimate the error matrix cell proportions (Stehman and Czaplewski, 1998), and subsequently, the estimates of overall and class-specific user and producer's accuracy (Story and Congalton, 1986). The use of stratified formulas is important because of sampling methods that have been chosen for the project. Accuracy results were computed through weighting the cell proportions by the proportion of each land cover within a given Federal region.

Results and Discussion

NLCD products. The NLCD classification product (Figure 3) provides an overview of the major land cover features for the conterminous United States. Land cover for larger scale areas located in Colorado (Figure 4) provide information regarding the level of detail typical of the NLCD, with the image to the right showing full resolution and detail of the data set.

The equal-area projection of the NLCD allows easy area tabulations

Figure 2. U.S. Federal Regions.

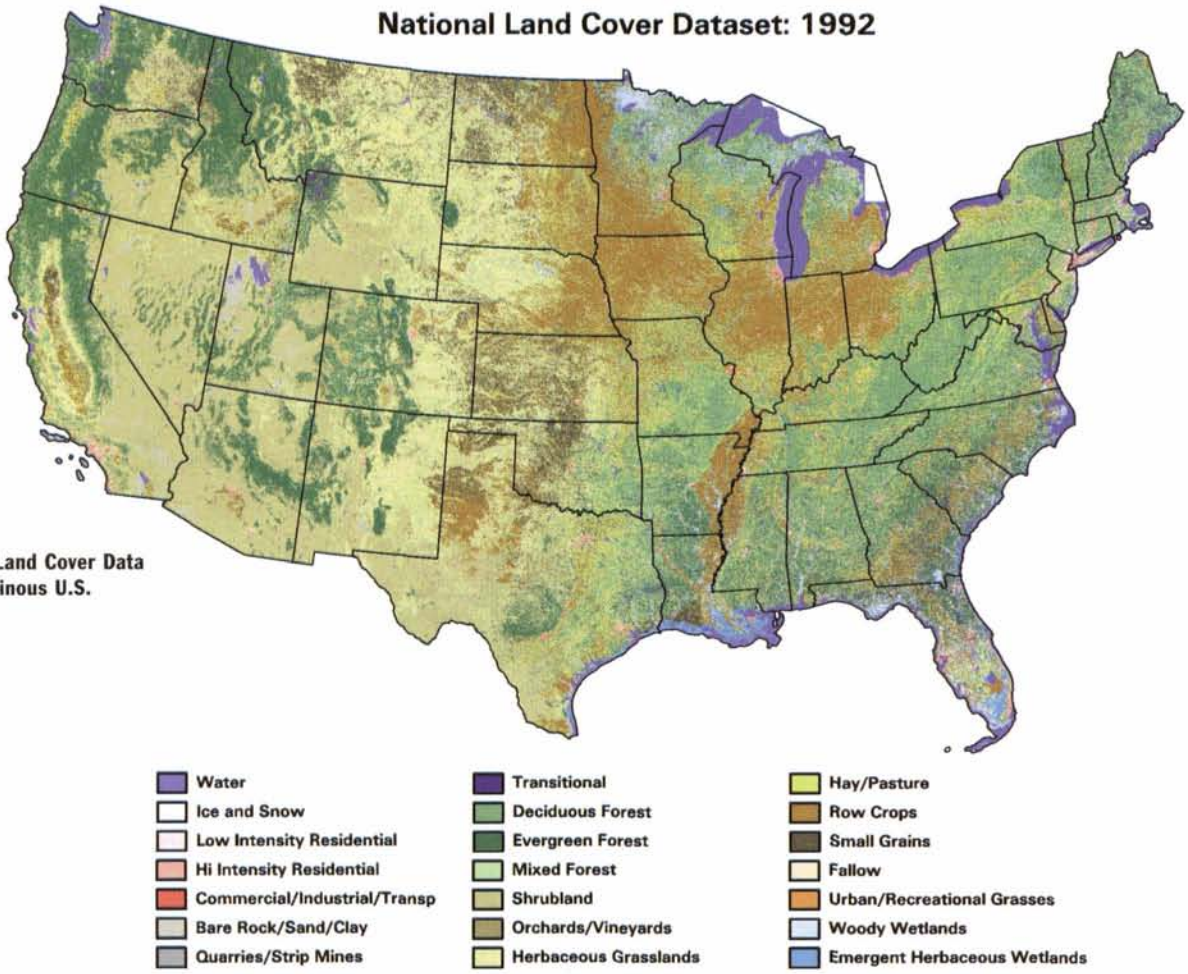


Figure 3. National Land Cover Data set for the conterminous U.S.

Figure 4. (Left) Denver, Colorado. (Right) Downtown Denver Colorado (showing full resolution of the National Land Cover Data set). Refer to figure 3 for color legend.

CONTINUED ON PAGE 654

CONTINUED FROM PAGE 653

of the various land cover classes. Table 2 shows the percentage land cover estimates derived from raw pixel count for each of the ten U.S. Federal Regions of the conterminous U.S. This information is also shown graphically for the conterminous U.S. in Figure 5. At a glance, it can be seen that the four forest classes (deciduous, evergreen, mixed forest and woody wetlands) make up a significant proportion of the NLCD, combining for a total of 32.1% of the area of the conterminous U.S. Agriculture (pasture/hay, row crops, small grains, fallow and orchards/vineyards) makes up about 26.4% of the surface area of the conterminous U.S. Urban classes (low and high intensity residential, commercial/industrial/transportation) account for 2.0 % of the surface area. Land cover area estimates are also available for each of the 48 conterminous U.S. states, and may be obtained at: http://edcwww.cr.usgs.gov/programs/lccp/natllandcover.html.

While many users will want to retain the full 30-m resolution of the NLCD, some users, such as those working at the national scale, will find that there are too many pixels in the original data set for their applications. For these users, files of smaller size representing condensed derivative versions of conterminous U.S. land cover have been created from the NLCD. One of these products simply depicts the dominant land cover class for each square kilometer in the conterminous U.S. A second type of product developed represents percentage land cover type for each 1km unit across the conterminous U.S. There are 21 of these 1 km-resolution data layers, one for each NLCD land cover class. An example showing of percentage deciduous forest is shown in Figure 6. Each individual file of the 1km-resolution data sets is approximately 0.09% the size of the original NLCD data set. Thus, these data layers do not take up huge amounts of computer disk space and are ideal for efforts that require complex computations and manipulations. These data sets are especially appropriate for national-scale efforts in which land cover information is necessary, but not necessarily at the NLCD 30-m resolution. Currently, these data sets are preliminary; more refined versions are anticipated (contact USGS/EDC).

Accuracy. Several rules for defining agreement between map and reference data may be applied given the information collected from NAPP photos and land cover maps. Comparing results across a range ofagreement protocols and data sub-

	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10	TOTAL	% Area
Water	2956	1812	1791	8450	10106	10207	3067	7971	2703	2878	51942	1.7
Perenn. Ice Snow	0	0	0	0	0	0	0	330	17	240	587	0.0
Low Int Res.	2297	2657	2954	7482	4739	3820	1487	1182	3585	1494	31697	1.1
Hi Int Res.	272	777	280	2102	2021	1416	646	154	453	9	8129	0.3
Comm/Ind/Trans	899	695	1297	3179	2864	2842	1887	1145	1698	1043	17549	0.6
Bare Rock	143	34	27	369	74	4579	274	13552	19151	4438	42640	1.4
Quarries/ Mines	46	81	615	475	415	320	97	245	260	29	2584	0.1
Transitional	891	44	842	8948	728	1676	62	2020	134	3815	19162	0.6
Deciduous Forest	19517	20116	60713	87523	72511	47139	29829	13365	3223	6152	360088	12.0
Evergreen Forest	12315	3448	7090	67340	10086	50761	1819	80228	63623	91210	387920	12.9
Mixed Forest	17309	10350	12790	44868	9556	20235	2486	1789	5614	6853	131850	4.4
Shrubland	190	0	0	479	452	128416	1108	121248	219948	74872	546713	18.2
Orch/ Vineyard	16	0	0	1671	4	45	0	8	4216	627	6587	0.2
Grasslands/Herb	0	0	0	5580	2143	128332	79090	203749	38567	21613	479074	15.9
Pasture/Hay	1254	10534	23163	40879	54929	55645	47852	25618	8237	11416	279527	9.3
Row Crops	3412	2976	6795	50137	124217	52888	86524	42888	6262	3711	379811	12.6
Small Grains	0	0	0	657	2262	24941	21515	39679	3619	10325	102998	3.4
Fallow	0	0	0	0	0	260	1575	20186	43	5604	27667	0.9
Urban/Rec Grass	656	472	224	1668	1738	680	778	283	354	115	6967	0.2
Woody Wetlands	2447	1834	2153	37275	25913	11381	2527	1139	193	550	85412	2.8
Emerg/Herb. Wet.	912	452	1058	8000	7686	10179	2983	4879	730	1104	37982	1.3
Totals	65531	56282	121792	377082	332445	555765	285606	581657	382629	248098	3006887	

Table 2. Land Cover Area (Square Miles) for the 10 Federal Regions of the Lower 48 United States based on raw pixel counts from the National Land Cover Data set 1992.

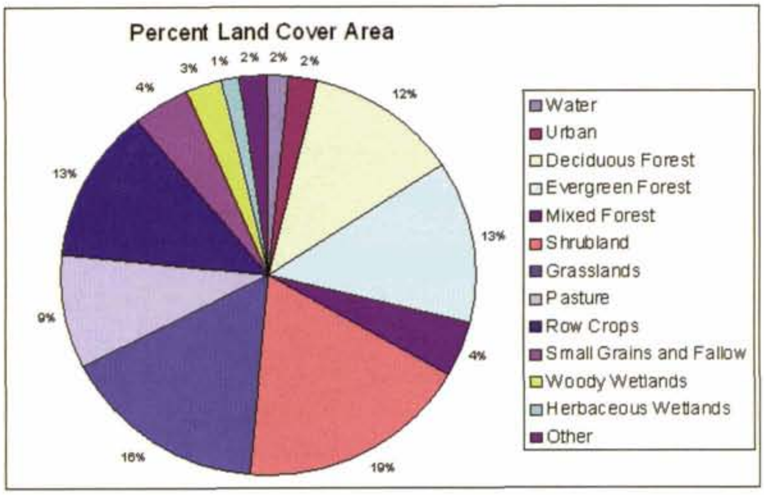


Figure 5a. Percentage land cover area for the conterminous United States based on National Land Cover Data set classification.

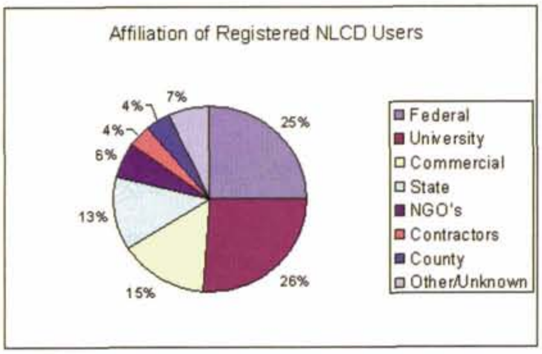


Figure 5b. Affiliation of Registered National Land Cover Data set users.

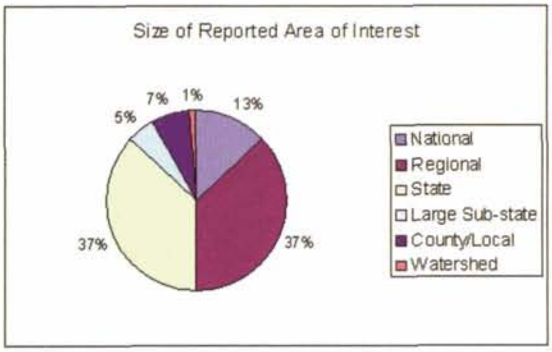


Figure 5c. Size of study area to which National Land Cover Data set are applied.

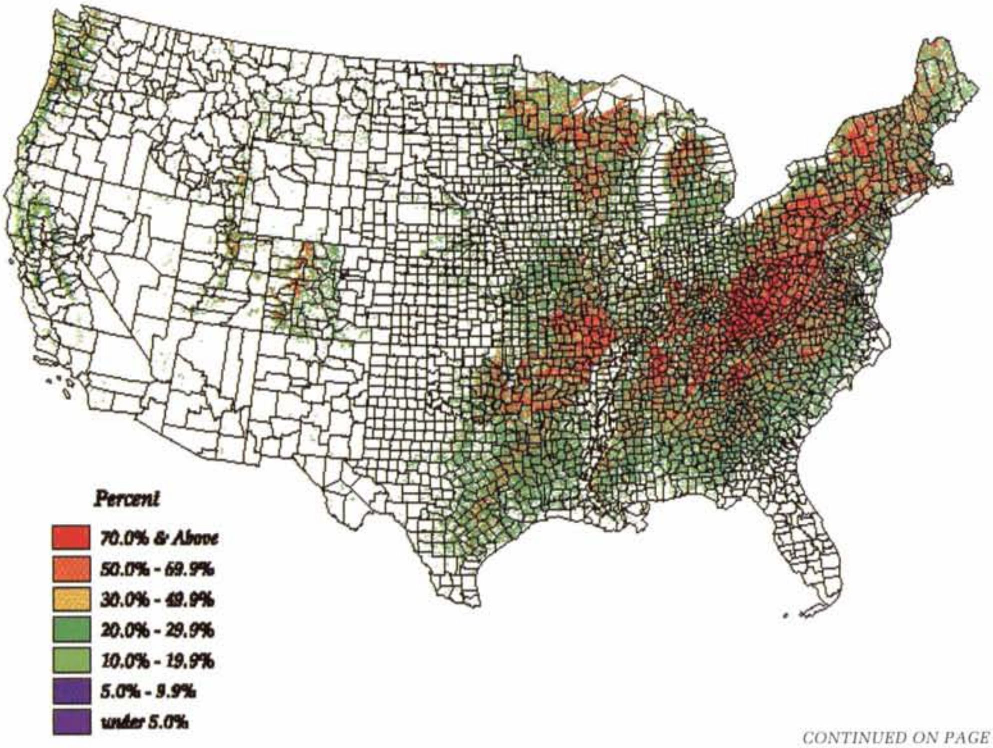
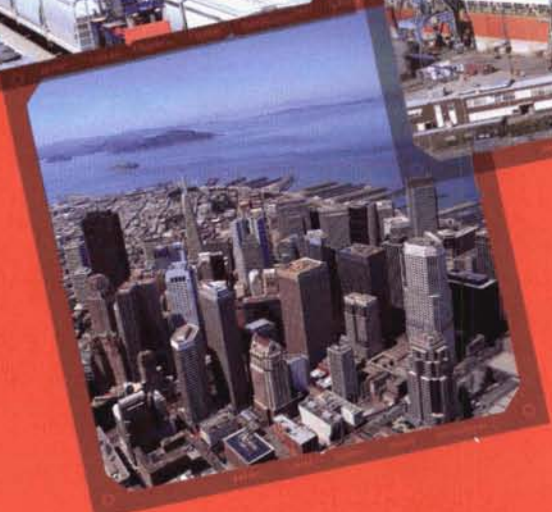


Figure 6. Percentage deciduous forest cover as derived from National Land Cover Data set at 1km resolution.

CONTINUED ON PAGE 657



CONTINUED FROM PAGE 655

sets permits evaluation of the reference data quality and more thorough investigation of thematic map accuracy (Congalton and Green 1993, Khorram et al. 1999). In this paper, accuracy results are briefly reported for the first four regions in the eastern United States combined (see Figure 2) by defining agreement as a match between the primary or alternate reference land-cover label and the mode land-cover label in a 3x3 pixel window surrounding the sample. Detailed region specific accuracy assessment results completed thus far can be obtained at: http://edcwww.usgs.gov/programs/lccp/accuracy and also from Yang et al. (2001a).

Overall accuracy for the eastern United States was 81% for Anderson Level I aggregations (i.e. water, urban, barren land, forest, agricultural land, wetland, rangeland; Anderson et al., 1976), and 60% for all classes (analogous to Anderson Level II classes). As expected, a significant source of disagreement between map and reference land-cover labels (approximately 20%) was between classes that aggregate into a single Anderson Level I class. Other sources of disagreement were between the forest and agricultural classes, which accounted for approximately five percent of the error, and between the forested wetland class (91) and the upland forest classes (41, 42, 43).

At Anderson Level II, notably high commission errors occurred for mines (class 32), hay/pasture (class 81) and high intensity residential (class 22). Not surprisingly, high intensity residential was most often confused with the other urban classes (low intensity residential and commercial/industrial). Meanwhile, hay/pasture was most often confused with row crops (class 82). Mines were often confused with transitional, hay/pasture and deciduous forest. Some of these latter discrepancies undoubtedly were related to changing surface conditions that occurred between image and photograph acquisitions.

NLCD Availability, Users and their Applications. The NLCD is available (FTP download or CDROM) from the USGS EROS Data Center website: http://edcwww.cr.usgs.gov/programs/lccp/natllandcover.html Since October, 1999 over 8500 state land cover files have been downloaded from the USGS EROS Data Center FTP site. The files on the FTP site are generic binary files (one file for each state) and are available at no charge to data users. A text file is associated with each binary file, and provides information on map projection, coordinates, and other parameters. Additionally, the states in Regions 1-4 are also available on 8 CDROMs. States in Regions 5-10 will ultimately be available on CDROM. Each CD contains as many contiguous state files as will fit (e.g. all six New England states are available on one CD). A total of 31 CDs are required to hold all the NLCD files. As of December 31, 2000, over 200 CDs (600 files) have been purchased from the following web site: http://edcwww.cr.usgs.gov/programs/lccp/nlcd.jsp.

All users of the NLCD are advised of the intended applications and limitations of the NLCD. The data set was designed and implemented to meet the broad requirements of federal agencies for a consistent national land cover data set. That is, the procedures and available resources dictated that it was not possible or feasible to spend a great deal of effort to ensure the accuracy and integrity of the classification results at local scales. Further, as outlined above, the accuracy assessment is being carried out on the basis of large federal regions and so the accuracy of the NLCD at local levels is unknown. Therefore, users are advised that the NLCD is not recommended for local scale analyses. We recognize that some users will use the data for local purposes due to lack of other available options. These users are advised to scrutinize the NLCD in their study area to verify that there are no major errors in the classification. This assumes that the user has a good understanding of land cover in their study area. Feedback from users confirms this advisory. Some users report very good agreement in their local areas, whereas others report problems.

Since the data have been available, the USGS/EDC Land Cover Applications Center has compiled a database of users and their applications. The goal is to gather user information to better understand the need for land cover data. To date, over 500 people have registered as users of the NLCD. The following information was derived from that database.

Users and Applications of NLCD. As expected, Federal and State agencies along with Universities are the major users of the NLCD, but many, non-governmental organizations (NGOs), commercial and local agencies are also showing interest in the NLCD (see Figure 5). The reported size of study areas to which NLCD data are being applied (Figure 5) indicates that most, but not all users are using NLCD for large area assessments. Approximately 8% of the users are using the data for relatively small areas (county size or smaller).

The NLCD is being used for a wide range of applications (Table 3). While some users appear to push the limits of the data, we have not ascertained whether this group scrutinized the data as recommended and were satisfied with what they found. Many of these users report there is no other source of data.

CONTINUED ON PAGE 658

CONTINUED FROM PAGE 657

Limitations and Reflections:

The NLCD provides excellent regional to national-scale overviews of early 1990s vintage land cover of the United States. During the course of developing the data set, we have become very familiar with some of the strengths and weaknesses of the data set. These come under different headings, and are described below.

Accuracy. While the 80% accuracy level achieved for the NLCD at Anderson Level 1 is a reasonable value considering the scope of the project, the level of accuracy for the Anderson Level 2 type classes (60%) is lower than anticipated. There are a number of variables that affect accuracy levels, and discussion of some of these as they pertain to NLCD is warranted.

One source of Anderson level 2-type error relates to the level of class distinctness. Some groups of classes fall on a continuum, grading from one class to another (both in the field as well as spectrally). The forest classes and residential classes are good examples of this as they grade from one class to another with no definite demarcation between them (e.g. deciduous to mixed to evergreen; low intensity to high intensity residential). While the definitions of these classes include percentage thresholds as cutoff values (e.g. low intensity residential is characterized by having 0-75% built-up material, and high intensity residential has 75-100% built-up material), the classes are not particularly well-defined, especially near the threshold value. Yet unless the accuracy values are thoroughly scrutinized, a point that is labeled as "high-intensity residential" that has 72% built-up material will simply be a misclassification contributing to error.

Land use planning
Land cover change assessment
Landscape pattern metric analyses
Wildlife habitat evaluation
Forestry mapping
Agricultural applications
Atmospheric dispersion modeling
Assessing wind energy resources
Environmental Impact Statements
Siting telecommunications transmission towers (modeling radio signal propagation/coverage)
Disease vectors (lyme disease, LaCrosse Encephalitis, i.e. deer (deer tick) and mosquito habitats)
Hydrological/Watershed modeling, impervious surfaces
Water quality (agricultural chemical runoff, erosion)
Fire Fuels/ Fire Risk assessment
Human Health risks (toxic spills, pesticide exposure, military installation environs)
Hazards assessment (wind, earthquake, hurricane storm surge)
Oil spill contingency planning
Urban Studies/Sprawl
Visualization (cockpit/air-flight simulations)

Table 3. Reported applications of NLCD data.

Another source of error relates to interpretability and quality of reference material used in the accuracy assessment. In this investigation, AA was accomplished using NAPP photography. It is noteworthy that the forested wetland class had a relatively high error based upon the AA, yet the original source for much of the forested wetlands information was NWI. Wetlands for NWI were originally delineated by interpretation of air photos. The NWI data in turn were relied upon heavily to delineate wetlands for NLCD. Yet the air photo validation of forested wetlands of NLCD indicated high levels of "error." The discrepancies resulted from (1) differences in photointerpreters, and/ or (2) the fact that the photographic materials used for NLCD were not ideal for discriminating forested wetlands (whereas the photos used for NWI were better). In either case, it is apparent that there is error associated with the photointerpretation phase of the accuracy assessment that adds to classification "error." Quality of air photos, which includes year and season of photo acquisition, may be a major factor influencing the accuracy assessment error for other classes as well, especially those that have a propensity to change intra- and interannually (e.g. agricultural crops).

It is important that users of the NLCD carefully evaluate the accuracy values in the context of their particular applications. Those users who do not need Anderson Level 2 type class information may wish to combine classes into an Anderson Level 1-type of a product, and thus work with land cover data with higher accuracies that will enhance the integrity of their results. Conversely, users working with the NLCD data in spatial modeling capacities may wish to keep the data in the 21-class form. These users may recognize that while there is error associated with the classification product, there may also be high levels of uncertainly associated with the models that they are running, and that some information for given classes is better than none.

It should also be kept in mind that scaling issues affect classification accuracies, with a general trend of increasing accuracies with coarser levels of spatial aggregations of pixels. For instance, higher levels of error are associated with these error assessments conducted at the single pixel level than at the 3x3 pixel level. Many users do not need the NLCD at the level of spatial detail provided, and will find it beneficial to spatially aggregate to the scale appropriate for their application. While we do not know the levels of accuracy associated with the 1 km NLCD products, we are reasonably certain that the accuracy levels of these data sets will be high enough for many national-scale applications.

Flexibility. Users often desire more land cover classes than NLCD, but they need to recognize that increasing the number of classes tends to exponentially increase the level of effort in creating the land cover classification product. In addition, higher numbers of classes generally result in higher levels of error. In spite of these observations, we recognize that there is a need within the user community for products that provide more thematic information than is currently provided by NLCD. Certainly there is more land cover information that can be gleaned from the Landsat data and ancillary data sets than provided by NLCD. Standard classifications such as those done to generate NLCD tend to generalize and simplify the landscape into a set number of discrete units, and much potentially useful information tends to be lost in the process.

One of the goals with future work will be to generate a land characteristics database at 30 m resolution that provides users with more capabilities towards tailoring products that will fit with their applications. Techniques are currently being developed for this, and will be implemented during "MRLC 2000" when a new National Land Cover Database will be generated for the entire United States using 2000 vintage Landsat data. One of the goals of this research will be to provide more quantitative information for selected continuous variables (e.g. canopy closure, impermeable surfaces). In addition, it is anticipated that users will have access to more intermediate data layers (i.e. pre-classification derivative products) that will be useful for deriving specific land cover variables appropriate for their applications that are not included in the final NLCD data sets.

Quality of Source Data. High quality land cover products can be derived only when input data are also of high quality. Generally, the Landsat 5 TM data acquired by the MRLC were of high quality from the standpoint of cloud and haze coverage. However, many of the data sets were sub-optimal in terms of seasonality. Many land cover features are much easier to map accurately using remotely sensed data sets from the appropriate time of year. For instance, hay/pasture is easier to discriminate from other types of land cover in early spring rather than mid-summer. While early spring and midsummer TM data sets were targeted for developing the NLCD, the data sets from the optimal time periods were often not available due to cloud contamination. In addition, it was often difficult to ascertain what "optimal" time periods were for each scene. Thus, during the scene-selection phase of the project, there were two confounding issues: (1) lack of data from the appropriate time(s) of year and (2) lack of knowledge as to the time frames of optimal land cover separability for each scene. Determination of the date to be used was generally subjective, and was based on our somewhat limited understanding of phenological patterns and opinion as to what time periods would give us the best land cover separability.

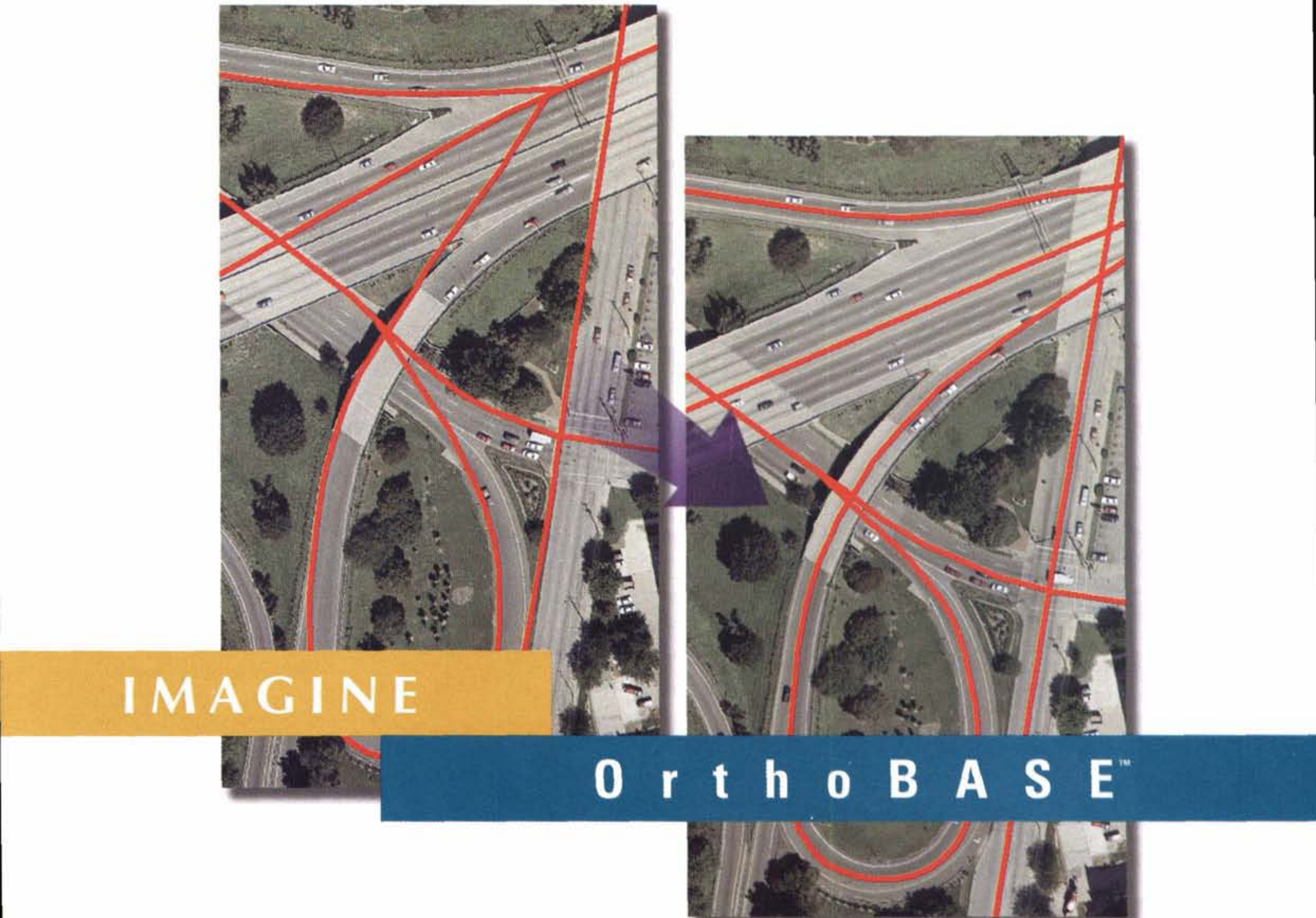
With the potential of using both Landsat 5 and 7 data for the next national mapping effort, the issue of clouds adversely affecting data availability may not be as great a problem as during the first effort. The two satellites are offset by eight days, and thus there will be twice as many opportunities to acquire cloud-free data throughout the growing season as during the first mapping effort. While the primary source of data for land cover mapping will be from Landsat 7 ETM+ data, it should be possible to augment the database with Landsat 5 TM data when necessary. Both TM and ETM+ data have been shown to provide similar information from an applications standpoint (Vogelmann et al., 2001), and with some exceptions and modifications, it should be possible to use the two sources of data interchangeably.

The other issue, lack of knowledge as to the time frames of optimal land cover separability, has been the focus of research since the completion of NLCD. In brief, it has been determined that use of the high temporal resolution data provided by AVHRR in conjunction with NLCD-derived land cover information provides useful information regarding optimal time periods for Landsat data acquisitions. Briefly, seasonal AVHRR-derived Normalized Difference Vegetation Index (NVDI) data were obtained for the dominant NLCD-derived land cover types for each path and row or mapping zone. Plots of AVHRR-derived NDVI versus date were developed for each major type of land cover for each path/row or mapping zone. Comparisons of the plots were then done to determine the three best dates for separating as many of the dominant types of land cover for each path/row as possible. Further details are provided in Yang et al. (2001b). This information is being used to target dates for scene acquisition for the next mapping effort. It is anticipated that three scenes, representing different seasons, will be acquired for each scene for MRLC 2000.

Conclusions

The completion of NLCD has provided a 1990s vintage land cover

CONTINUED ON PAGE 661



(Your Solution.)

ERDAS®
geographic imaging made simple™

Get an **ortho-accurate GIS** with IMAGINE OrthoBASE™, the revolutionary product by ERDAS® that **bridges the gap between digital photogrammetry and GIS**. This powerful product **dramatically reduces the cost and time** associated with triangulating and orthorectifying aerial photography, satellite imagery, and digital and video camera imagery. And, because ERDAS is **committed to building long-term relationships** with its clients, technical support, training, custom applications development, and project consulting services are all available to **make sure your project is a success.** ERDAS has your solution. Call today or visit **www.erdas.com/orthobase**.

ERDAS Worldwide Headquarters
Toll Free: +1 877-GO ERDAS (463-7327)
Phone: +1 404-248-9000
Fax: +1 404-248-9400
Email: info@erdas.com

ERDAS in Europe, Africa, Middle East
Phone: +44 1223-881774
Fax: +44 1223-880160
Email: eame@erdas.com

All Other Worldwide Inquiries
Phone: +1 404-248-9000
Fax: +1 404-248-9400
Email: info@erdas.com

CONTINUED FROM PAGE 659

data set for the conterminous United States. In addition to being more current, this product is spatially improved over the Land Use and Land Cover data set (USGS, 1990). Since being released, the usage level of this data set has been high, with many investigators using the data set for a multitude of applications. This underscores the importance of the data set, and also implies that updated data sets need to be developed on a regular basis. Not only does the user community require current land cover, but they also require accurate and detailed information that can potentially be used for monitoring purposes. While the data set provides an excellent overview of the nation's land cover patterns, it needs to be emphasized that this was the first of its kind for the conterminous United States, and that expectations for the quality of such a data set need to be tempered with reality. While we believe that this data set will fit the needs of many who work at the regional to national scales, we do not believe that the data set is particularly appropriate for local applications without modification. Conversely, the data may serve as a first-order land cover data set from which more refined data sets can be developed at the local level. We are cognizant of the need for more land cover information than provided by NLCD, and are currently working on methods to accommodate this need for the next national mapping effort. We will continue to refine methods to create accurate land cover projects, which will include continued research and development activities.

. . . .

Acknowledgments. This work was performed in part by the Raytheon Corporation under U.S. Geological Survey Contract 1434-CR-97-40274.

Authors

James E. Vogelmann[1], Stephen M. Howard[1], Limin Yang[1], Charles R. Larson[1], Bruce K. Wylie[1] and Nick Van Driel[2]

[1]Raytheon ITSS, EROS Data Center, Sioux Falls, SD 57198

[2]U.S. Geological Survey, EROS Data Center, Sioux Falls, SD 57198

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 pp.

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 and Remote Sensing*, 59:977—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.

Bureau of the Census, 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.

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

Congalton, R. and K. Green, 1993. A practical look at sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing*, 59:641-644.

Dobson, J.E., E.A. Bright, R.L. Ferguson, D.W. Field, L.L. Wood, K.D. Haddad, H. Iredale, 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.

ERDAS, 1998. ERDAS Imagine Expert Classifier. ERDAS, Atlanta, GA.

Friedl, M.A. and C.E. Brodley, 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61:399-409.

Hansen, M., R. Dubayah and R. Defries, 1996. Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, 17:1075-1081.

Khorram, S., G. Biging, D. Colby, R. Congalton, J. Dobson, R. Ferguson, M. Goodchild, J. Jensen, and T. Mace, 1999. *Accuracy Assessment of Remote Sensing-derived Change Detection*. Monograph. American Society of Photogrammetry and Remote Sensing, Bethesda, MD, 64 pp.

Lauer, D.T, 1986. Applications of Landsat data and the data base approach. *Photogrammetric Engineering and Remote Sensing*, 52: 1193-1199.

Loveland, T.L., J.W. Merchant, D.O. Ohlen, and J.F. Brown, 1991. Development of a landcover characteristics database for the conterminous U.S. *Photogrammetric Engineeering and Remote Sensing*, 57:1,453-1,463.

Loveland, T.L. and D. M. Shaw, 1996. Multiresolution land chracterization: building collaborative partnerships, In *Gap Analysis: A Landscape Approach to Biodiversity Planning* (eds. J.M. Scott, T. Tear, and F. Davis), Proceedings of the ASPRS/GAP Symposium, Charlotte, NC, (National Biological Service, Moscow, ID), pp. 83-89.

Prince, S.D. and M.K. Steininger, 1999. Biophysical stratification of the Amazon basin. *Global Change Biology*, 5:1-22.

Scott, J.M., T.H. Tear, and F.W. Davis (eds.), 1996. *Gap Analysis. A Landscape Approach to Biodiversity Planning*, American Society for Photogrammetry and Remote Sensing, Bethesda, MD, 320 pp.

Stehman, S.V. and R.L. Czaplewski, 1998. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment*, 64:331-344.

Stehman, S.V., J.D. Wickham, L. Yang, and J.H. Smith, 2000. Assessing the accuracy of large-area land cover maps: Experiences from the Multi-resolution Land-cover Characteristics (MRLC) Project. *Pro-*

CONTINUED ON PAGE 662

CONTINUED FROM PAGE 661

ceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Delft University Press, The Netherlands, 601-608.

Story, M. and R.G. Congalton, 1986. Accuracy assessment: A user's perspective, *Photogrammetric Engineering and Remote Sensing*, 52:397-399.

U.S. Department of Agriculture, 1994. State Soil Geographic (STATSGO) Data Base, Data Use Information, *United States Department of Agriculture Miscellaneous Publication Number 1492*.

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 User's Guide 4*, Reston, Virginia, Department of the Interior, U.S. Geological Survey, 33 pp.

U.S. Geological Survey, 1993. U.S. GeoData digital elevation models, *Data User's Guide 5*, Reston, Virginia, Department of the Interior, U.S. Geological Survey, 51 pp.

Vogelmann, J.E., D. Helder, R.A. Morfitt, M.J. Choate, J.W. Merchant, and H. Bulley, 2001. Effects of Landsat 5 TM and Landsat 7 ETM+ radiometric and geometric calibrations and corrections for landscape characterization, *Remote Sensing of Environment*, in press.

Vogelmann, J.E., T.L. Sohl, and S.M. Howard, 1998. Regional characterization of land cover using multiple sources of data, *Photogrammetric Engineering and Remote Sensing*, 64:45-57.

Vogelmann, J.E., T.L. Sohl, P. V. Campbell, and D.M. Shaw, 1998. Regional land cover characterization using Landsat Thematic Mapper data and ancillary data sources. *Environmental Monitoring and Assessment*, 451:415-428.

Yang, L., S.V. Stehman, J.D. Wickham, J.H. Smith, and N.J. Van Driel, 2000. Thematic validation of land cover data of the eastern United States using aerial photography: feasibility and challenges. *Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, pp. 747-754.

Yang, L., S. Stehman, J. Smith, and J. Wickman, 2001a. Thematic Accuracy of MRLC Land Cover for the Eastern United States, *Remote Sensing of Environment*, in press.

Yang, L., C. Homer, K. Hegge, B. Wylie, and B. Reed, 2001b. A Landsat 7 scene selection strategy for national land cover and land use characterization. In prep.

Zhu, Z, L. Yang, S.V. Stehman, and R.L. Czaplewski, 2000. Accuracy assessment for the U.S. Geological Survey regional land cover mapping program: New York and New Jersey Region, *Photogrammetric Engineering and Remote Sensing*, 66:1425-1435.