CHAPTER 1

THE UNITED STATES 2001 NATIONAL LAND COVER DATABASE

Collin Homer*, Jon Dewitz†, Joyce Fry†, and Nazmul Hossain†

*U.S. Geological Survey (USGS) Center for Earth Resources Observation and Science (EROS)
Sioux Falls, SD 57198

†Science Applications International, Corporation (SAIC), contractor to USGS EROS
Sioux Falls, SD 57198

Key words: *land cover, national database, remote sensing, imperviousness,* and *tree canopy*

INTRODUCTION

Land cover information is required by a broad spectrum of scientific, economic, and governmental applications, and provides essential input to analyze a variety of national issues. Thus, credible consistent national land cover information is increasingly more important. In 1999, new land cover research was implemented to expand and update the United States Geological Survey's National Land Cover Dataset (NLCD) 1992 into a full scale land cover database (with multiple instead of single products), and to produce it across all 50 States and Puerto Rico. This new database called the National Land Cover Database 2001 has been under production for six years. The 2001 refers to the nominal year from which most of the Landsat 5 Thematic Mapper (TM) and Landsat 7 Enhanced Thematic Mapper-plus (ETM+) imagery was acquired. Three products from this database were completed in January 2007, for the conterminous United States, including 16 classes of land cover, percent tree canopy, and percent urban imperviousness.

NLCD 2001 production was funded through an organization called the Multi-Resolution Land Characteristics Consortium (MRLC). This consortium consists of 13 programs across 10 Federal agencies that require land cover data for addressing their agency needs (www.mrlc.gov). MRLC provided the umbrella to coordinate multi-agency NLCD mapping production and funding contributions. In addition to NLCD data, MRLC also offers approximately 8,300 terrain-corrected Thematic Mapper (TM) and Enhanced Thematic Mapper (ETM+) scenes spanning growing season dates from 1982-2007, available for public Web-enabled download from www.mrlc.gov.

METHODS

NLCD 2001 production was accomplished according to protocols outlined in Homer et al. (2004) across 65 mapping zones for the conterminous United States. Production occurred with 12 mapping teams from both the government and private sector. To ensure consistency among teams, products were generated using standardized processes in data preparation, classification, and quality control. NLCD 2001 products were generated from a standardized set of input data layers mosaiced by mapping zone, including multi-season TM and ETM+ imagery centered on a nominal collection year of 2001, and Digital Elevation Model-based derivatives. This standard set of input data layers provided the best available data resources to derive the desired products. All data were geo-registered to the Albers equal-area conic projection grid, and resampled to 30m grid cells.

The land cover classification was accomplished using commercial decision tree (DT) software called See5* (Quinlan 1993). This was applied to standardized input data layers prepared for each mapping zone, and subsequently extrapolated through ERDAS IMAGINE* into classified pixels using customized software. DT is a supervised classification method that relies on large amounts of training data, which were collected from a variety of sources including existing Landsat-based classifications and training data pools, field sampling, and limited on-screen sampling. Training data were used to map all land cover classes except for the four urban classes, which were derived from thresholding of the imperviousness data product. When land cover modeling was completed, the final product was aggregated to a one acre minimum mapping unit (0.4 hectare or five TM pixels) to reduce single pixel

scattering using a "smart eliminate" aggregation algorithm. This algorithm uses eight-corner connectivity from a central pixel to allow non-linear features like roads and streams to remain intact, and accesses a weighting table to allow "smart" decisions on a dissolve protocol. Although every effort was made to maintain consistency in classification between mapping zones during production, some edge matching was required to merge the 65 completed zones. A three kilometer buffer around each zone, (six kilometer overlap was available between mapping zones) served as the interface area for defining the most successful edge-matching boundary.

Imperviousness and tree canopy were classified using commercial regression tree (RT) software called Cubist* (Yang et al. 2002). Training data were generally derived from 1-m resolution Digital Orthoimagery Quarter Quadrangles (DOQQs) that were classified categorically into canopy/non-canopy, or impervious/non-impervious for each 1-m pixel. This training information was then used to derive the RT model, which was subsequently extrapolated across the mapping zone to derive continuous canopy and imperviousness predictions. A masking strategy was then used to further reduce errors of commission over areas with spectrally similar features that proved difficult to discriminate accurately (e.g. shrub and grass areas for canopy and bare agriculture fields for imperviousness). This masking method depended upon other ancillary GIS data layers to help define or eliminate problem areas. Final canopy and imperviousness products were not aggregated with smart eliminate like land cover, but were left in the single pixel format. A three kilometer boundary buffer was also used with these products, (six kilometer overlap was available between mapping zones) to serve as the interface area for defining the most successful edge-matching boundary.

RESULTS AND DISCUSSION

Sixteen classes of land cover were modeled for the conterminous United States. (Table 1 and Figure 1). Initial land cover product accuracy from cross-validation estimates was generated during classification, with an overall national accuracy of 83.9%. Continuous predictions from 1-100% for both tree canopy and urban imperviousness were also modeled over the conterminous United States, with accuracy estimates derived from cross-validation and reported as an average error estimate. The tree

canopy mapping zone average error estimates ranged within a given zone from to 6% to 17% deviation from prediction, and urban imperviousness average error estimates ranged within a given zone from 4% to 17% deviation from prediction.

NLCD Land Cover Class Digital Code	NLCD Class Name
11	Open Water
12	Perennial Ice/Snow
21	Developed, Open Space
22	Developed, Low Intensity
23	Developed, Medium Intensity
24	Developed, High Intensity
31	Barren Land
41	Deciduous Forest
42	Evergreen Forest
43	Mixed Forest
52	Shrub/Scrub
71	Grassland/Herbaceous
81	Pasture/Hay
82	Cultivated Crops
90	Woody Wetlands
95	Emergent Herbaceous Wetlands

Table 1. Conterminous United States NLCD 2001 land cover legend by digital code

The major value of NLCD 2001 products lie in their ability to provide a complete, consistent coverage of the nation's land cover, and to serve as a resource for regional-to-national scale applications. These types of products are designed to meet land cover requirements over larger areas, and are not designed for local application (e.g., county-level use). However, the ability to modify and customize NLCD 2001 data products for more specific application was accommodated in the original database design. Users not only have three products to synergistically combine, but can also download the original imagery for additional modification or correction if desired.

Now that two eras of national land cover data are available, many users will be tempted to directly compare the two land cover layers as a way to measure land cover change. Users are cautioned that new improvements in mapping methodology, input data, and minor mapping legend modification will confound comparison between NLCD 1992 and NLCD 2001. Direct comparison of these two independently created land cover products is not recommended, because differences in the methodology

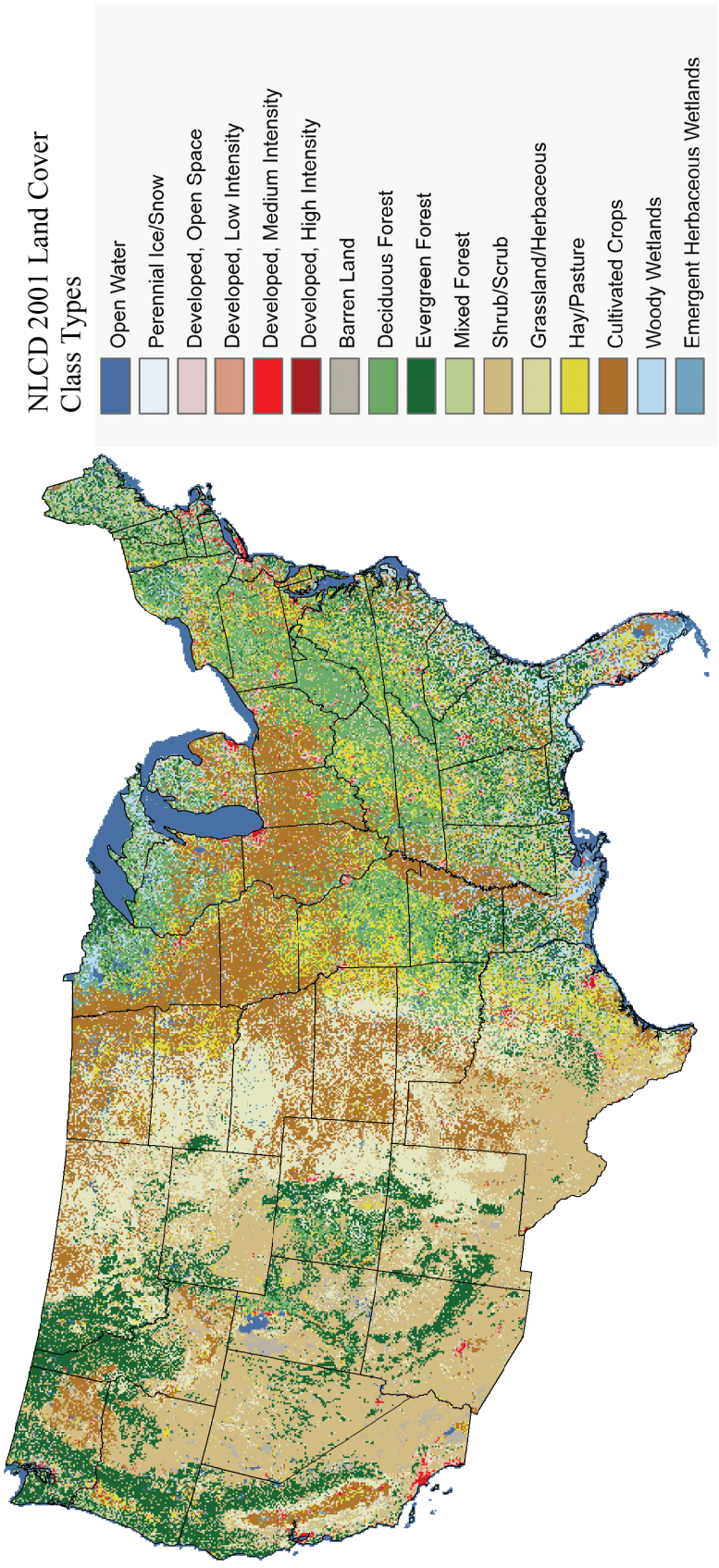

Figure 1. NLCD 2001 conterminous United States land cover product.

used to produce the two products will overwhelm true differences due to land cover change. However, an NLCD "bridge product" to aid land cover change analysis between the two eras will soon be available at the Anderson Level I thematic scale (Fry et al. in prep), and is scheduled for release for the conterminous United States by the end of 2007.

All NLCD 2001 products and mapping tools are available via Web-enabled file download from the MRLC Consortium website (www.mrlc.gov) with options for both dynamic download (user-defined download areas) and FTP download by zonal groupings. In most cases, files are available in several formats, and mapping zone metadata are supplied with all downloads. All NLCD 2001 product sets are distributed at 30m resolution in the NLCD standard NAD 83, Albers equal-area conic projection.

NLCD 2001 data for Alaska, Hawaii, and Puerto Rico will be completed by August 2008, which will then represent the first compilation of nationwide land cover ever produced at 30m resolution. NLCD 2001 will then provide a comprehensive land cover resource for the entire United States, and support hundreds of applications that require this scale of information. Future updates of NLCD 2001 are now being prototyped to ensure that land cover information stays current, and that land cover change is quantified and analyzed to broaden the utility of this information.

REFERENCES

Fry, J., M. Coan, C. Homer, D. Meyer and J. Wickham. 2008. Completion of the National Land Cover Database 1992-2001 Retrofit Change Product. USGS Open File Report. In preparation.

Homer, C., C. Huang, L. Yang, B. Wylie and M. Coan. 2004. Development of a 2001 National Landcover Database for the United States. *Photogrammetric Engineering and Remote Sensing*, 70 (7): 829-840.

Quinlan, J.R. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann: San Mateo, California.

Yang, L., C. Huang, C. Homer, B. Wylie, and M. Coan. 2002. An Approach for Mapping Large-area Impervious Surfaces: Synergistic use of Landsat 7 ETM+ and High Spatial Resolution Imagery. *Canadian Journal of Remote Sensing* 29 (2): 230-240.

AKNOWLEDGMENTS

This research reflects an enormous team effort across many organizations and individuals. Special acknowledgment is made for the support of the individuals and agencies of the MRLC Consortium, the many NLCD 2001 Federal mapping teams from the USGS, the National Oceanic and Atmospheric Administration (NOAA), the U.S. Department of Agriculture, Forest Service, and the private-sector mapping teams of SAIC, MDA Federal, and Sanborn who performed work under several USGS and NOAA contracts.

Endnote

***** The use of any trade, product, or firm name is for descriptive purposes only and does not imply endorsement by the U.S. Government.