DEVELOPMENT OF A CIRCA 2000 LANDCOVER DATABASE FOR THE UNITED STATES

Homer, Collin; Chengquan Huang, Limin Yang, and Bruce Wylie

Raytheon ITSS, USGS/EROS Data Center
Sioux Falls, SD 57198

ABSTRACT

Multi-Resolution Land Characterization 2000 (MRLC 2000) is a second-generation federal consortium to create an updated pool of nation-wide Landsat 7 imagery, and derive a second-generation National Land Cover Database (NLCD 2000). This multi-layer, multisource database will include a suite of 30-meter resolution data that will serve as standardized ingredients for the production of land cover – both nationally and locally. This database will also provide the framework to allow flexibility in developing and applying suites of independent data layers. These nationally standardized independent data layers or components, will be useful not only within the land-cover classification but as data themes for other applications. This database will consist of the following components: (1) normalized tasseled cap (TC) transformations of Landsat 7 imagery for three time periods per scene (early, peak and late), (2) ancillary data layers, including 30m DEM derivatives of slope, aspect and elevation and three STATSCO soil derivatives, (4) image shape and texture information, (5) image derivatives of percent imperviousness and percent tree canopy per-pixel, (6) classified land-cover data derived from the Tassel Capped imagery, ancillary data and derivatives, (7) classification rules and metadata from the land cover classification, allowing future users the potential to modify rules to derive land cover products tailored to their specific local applications. In a pilot study application of the database concept, two mapping zones (Utah and Virginia) were selected for full generation of the above data components. Three derivative layers including, per-pixel imperviousness, per-pixel canopy and land cover were classified from the database. Cross validation accuracies for land cover ranged from 65-82%, and mean absolute error values of 10-15% were reported for percent tree canopy and imperviousness.

I. INTRODUCTION

The USGS Multi-Resolution Land Characteristics (MRLC) Consortium was originally formed in 1993 in order to meet the needs of several federal agencies (USGS, EPA, NOAA, and USFS) for Landsat 5 imagery, and land-cover information. One of the results of this consortium was the completion of a successful mapping of the conterminous United States into the National Land Cover Dataset (NLCD 1992), from circa 1992 Landsat TM at an Anderson et al. (1976) level II thematic detail (Vogelmann et al 2001). The growing need for current Landsat 7 data, land-cover and other geospatial data within the federal government culminated in reforming the MRLC Consortium in 2000 (MRLC 2000)[1]. The MRLC 2000 goals are two-fold. First is a Landsat 7 image acquisition that includes multi-temporal data at a minimum of three dates per path/row (representing different seasonality) for the conterminous United States, Alaska, Hawaii and Puerto Rico. All 8 ETM+ TM bands (including thermal and pan bands) are resampled using cubic convolution into a terrain-corrected Albers Equal Area map projection. The second goal of MRLC 2000 is a value-added database of land cover, called the National Land cover Database 2000 (NLCD 2000) which will be generated across all 50 states and Puerto Rico using both Landsat 7 ETM+ imagery and ancillary data.

This proposed database is designed to overcome the traditional remote sensing classification focus on generation of specialized data products that meet only specific requirements. Historically, products are often developed according to project specific needs, with methods and results not designed to extrapolate to other areas or to crosswalk to different land cover schemes. Consequences of these approaches have resulted in remote sensing datasets and methods that are difficult to compare (spatially and temporally), and have limited flexibility for other uses. This local product focus, historically often a limitation of technology and funding, has restricted the broad-scale development of remote sensing datasets. Additionally, intermediate data layers are often discarded after generation of the final product. These intermediate data layers (such as spectral clusters, ancillary information and training data) provide an untapped potential for flexible application if staged in a related framework. New improvements in remote sensing data availability, hardware capability and software algorithms offer new opportunity in data processing, and methods to increase the utility of remote sensing in building comprehensive more objectively derived databases – not just specialized products.

This database application, (defined as multiple interlinked data layers that are useful either as individual components, or in synergistic groupings), builds upon past database design successes such as the global landcover database (Brown et al. 1999,Loveland et al 1999). NLCD 2000 is planned to accommodate a wide variety of potential users who can tap into the database for both derived land cover products, and other intermediate data layers that will be standardized and consistent for the United States.

II. GUIDING PRINCIPLES

The September 2000 competition of the initial National Landcover Dataset (NLCD 1992) was a tremendous accomplishment (Vogelmann et al 2001). It created a TM pixel scale (30m), data layer with approximately 9 billion pixels. The size of this dataset alone, illustrates the complexity and difficulty of land cover mapping at a national scale. During the five years of mapping required to complete this prototype product, many valuable lessons were learned from both creators and users. This feedback coupled with new MRLC 2000 member requirements were the basis for developing several guiding principles for the follow-on development of NLCD 2000. These guiding principles provided the direction that culminated in this database design. They included the need to: a.) identify land cover products flexible enough for multiple users b.) provide users increased access to intermediate database products and derivatives enabling local application, c.) develop methods that are as objective, consistent and repeatable as possible resulting in standardized land cover products d.) constrain methods to be intuitive, simple, efficient and transferable to others, and e.) ensure design of a second-generation land cover product that maintains reasonable compatibility to NLCD 92.

In order to develop applicable solutions, a science team was assembled at USGS EROS Data Center (EDC) in 1999 to review and develop potential methods. Four study sites were picked as

[1] MRLC 2000 Web information and papers are at http://landcover.usgs.gov

locations for methodological development that were representative of the conterminous U.S. (Figure 1). Two sequential Landsat 7 path/rows were selected to represent these sites, which included, Virginia (eastern deciduous forest and agriculture), Nebraska/South Dakota (mid-west crop/prairie/pasture), Utah (Rocky Mountain and Great Basin shrubs/forests and irrigated agriculture) and Oregon (costal forests, agriculture and shrublands). Methods developed in research trials at these study sites were assumed to extrapolate to the conterminous US, and possibly to Alaska, Hawaii and Puerto Rico. Following two years of comprehensive review and research, the database design presented here evolved from efforts to follow these guiding principles. It reflects the primary concern, which was to allow maximum flexibility to future users to derive additional land cover products independent of the NLCD 2000 classification. We anticipate user access to this nation-wide standardized database will foster the exploration, development, application and sharing of land cover information at unprecedented scales.

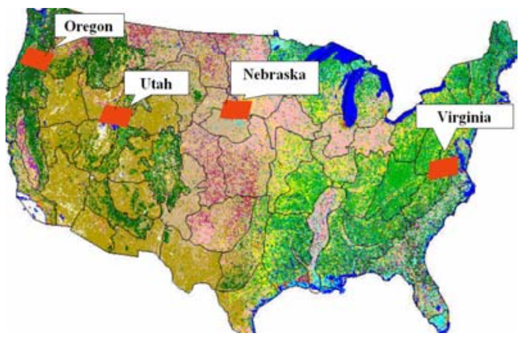
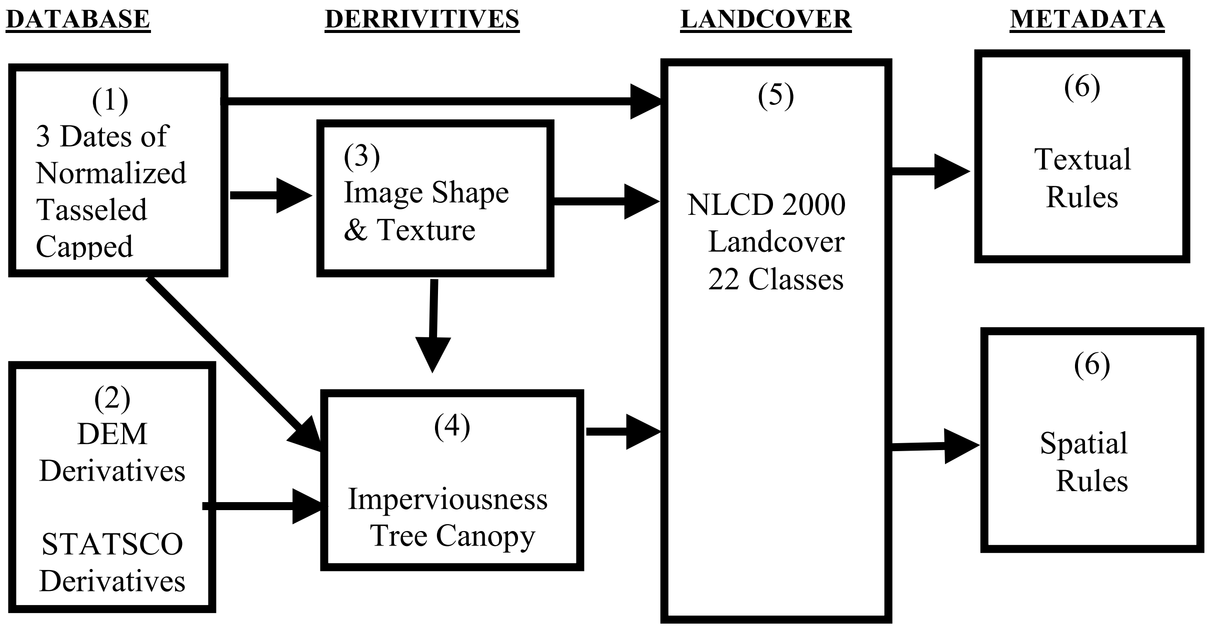


Figure 1. Four mapping strategy study sites

Figure 2 outlines the organization of the database which consists of the following components: (1) normalized tasseled cap (TC) transformations of Landsat 7 imagery for three time periods per scene (early, peak and late), (2) ancillary data layers, including 30m DEM derivatives of slope, aspect and elevation and three STATSCO soil derivatives, (3) image shape and texture information, (4) image derivatives of percent imperviousness and percent tree canopy per-pixel, (5) classified land-cover data derived from the Tassel Capped imagery, ancillary data and derivatives, (6) classification rules and metadata from the land cover classification, allowing future users the potential to modify rules to derive land cover products tailored to their specific local applications.

Figure 2. NLCD 2000 DATABASE MODEL



III. DATABASE

1.) Image Processing

-Selection

The strategy developed for nation-wide Landsat 7 ETM imagery selection was designed to meet the requirements of the land cover database for consistency, while flexible enough to accommodate temporal and data quality requirements. Requirements included three acquisition dates for each path/row covering early, peak and late green-up. The developed method uses a scene selection strategy based on vegetation dynamics of target land cover types over a growing season (Yang et al 2001). It assumes that a distinct seasonal trajectory of land cover dynamics can be identified using multi-temporal coarse-scale remote sensing data, and that this information will provide increased land cover identification capability on finer scale TM data. Scene selection criteria are established using multi-temporal greenness as a surrogate for vegetation phenology. Information on vegetation phenology was derived from the multi-temporal normalized difference vegetation index (NDVI) data of the conterminous U.S. acquired by the advanced very high-resolution radiometer (AVHRR) from 1994-1998.

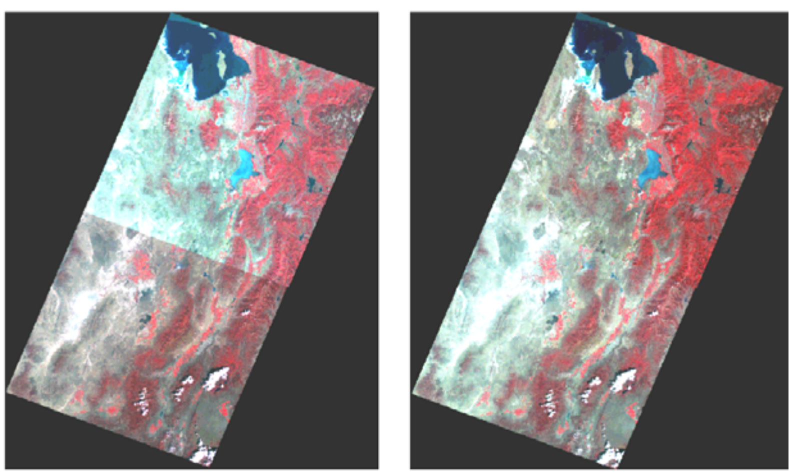
For example, in areas with a single-peak greenness, a scene was chosen at or near time of the peak-greenness (maximum NDVI) of the dominant land cover class, with second and third scenes selected from pre- and post-peak in the leaf-off season. Hence, the first scene choice captures the most rigorous and productive stage of vegetation growth, whereas the second and third scenes choices capturing other vegetative stages and changes over the rest of the growing season. Time of peak-greenness is defined as the biweek period when maximum NDVI occurs for the dominant land cover type(s). A second scene was selected outside the time window of peak greenness, and also outside the dormant season defined by NDVI less than 0.15. The third scene was selected from the remainder of the growing season based on separability among major land cover classes. Overall, this strategy for selecting Landsat 7 imagery based on vegetation phenology and image quality provided a reasonably objective framework to populate a nation-wide image database.

-Preprocessing

One of the challenges to large-scale satellite-based land cover characterization is consistent geometric correction and normalizing noise arising from atmospheric effect, changing illumination geometry, and instrument errors inherent when using multiple frames of imagery (Wharton 1989). Such geometric and radiometric error can hinder the ability to derive land surface information reliably and consistently. For MRLC 2000, images are geometrically corrected using cubic convolution resampling in a single step from Level 0 data to Level 2 terrain correction. Terrain correction is done using the USGS 1-arc second National Elevation Dataset (NED) to improve geo-location accuracy. The selection of cubic convolution as a resampling strategy is based on the superior spatial accuracy it provides over nearest neighbor resampling (Shilen 1979, Park and Schowengerdt 1982). This is of special concern when stacking multiple dates across many path/rows as is the case with NLCD 2000. The 7 visible and infrared bands are resampled to a 30 m spatial resolution; the panchromatic band is resampled to 15 meters and the thermal band to 60 meters.

Great efforts have been made to minimize instrument errors for standard image products of Landsat 7 Enhanced Thematic Mapper Plus (ETM+) (Irish 2000). Noises due to the impact of the atmospheric and illumination geometry can be normalized in several approaches. For MRLC 2000, images are first radiometrically corrected using standard methods at the USGS EROS Data Center to eliminate band bias and gain anomalies (Irish 2000). Secondly images are converted to at-satellite reflectance for the 6 reflective bands and to at-satellite temperature for the thermal band according to Markham and Barker (1986) and the Landsat 7 Science Data Users Handbook (Irish 2000). Considering the tremendous volume of imagery being processed, and uncertainty with algorithms currently available, atmospheric or topographic normalizations are not being used because of their potential to introduce confounding error. Only a first order normalization conversion to at-satellite reflectance is done on clear and near cloud-free ETM+ images. This conversion algorithm is physically based, ready to automate, and does not introduce errors to the data (Huang et al 2002). Tests have shown that this method which normalizes multi-scene noise, coupled with the intelligent scene selection strategy allows assembling of multi-scene datasets without traditional mosaicing (Figure 3), and offers a reasonable pre-processing method for such a large database. This method provides an important first-step to standardizing imagery, but some atmospheric, phenological and topographic noise do remain.

Figure 3. Example of Image Normalization Effect



Before Normalization After Normalization

-Transformation

Potential use of even pieces of a nation-wide three date TM database would require enormous hardware storage capability for a user. Options were explored for optimal ways to distill original resolution TM bands into spectral-efficient transformations without losing important information. Principle Component Analysis (PCA) derivatives were assumed to be the most efficient transformation for compressing spectral information. However, PCA was not considered a viable forum for image compression because if its interpretation difficulty, especially when comparing image to image. Tests and trials using indices such as NDVI, SAVI and LAI and TC transformations were compared against PCA results. Tests showed that TC offered the best potential surrogate to PCA retaining 98% of potential PCA all-band spectral variance information (Bruce Wylie pers. communication). More importantly, TC offers the additional advantage of providing standardized output layers of brightness, greenness and wetness that are linked to scene physical characteristics and comparable from image to image.

A new TC transformation based on Landsat 7 at-satellite reflectance normalized scenes described above was developed from 10 TM scenes representing a variety of landscapes across the United States in both leaf-on and leaf-off seasons (Huang et al 2002). This transformation is most appropriate for regional remote sensing applications where atmospheric correction is not feasible, but greater standardization among multiple images is required. In cases such as this, a DN based transformation would not appropriate for multi-scene comparisons, because values are strongly affected by changing illumination geometry which is normalized here by converting DN to at-satellite reflectance. Applying the ground reflectance factor based transformation directly to at-satellite reflectance images is not appropriate because it may result in unreasonable TC values. The at-satellite reflectance based transformation generally does not have this problem when applied to clear and near cloud-free at-satellite reflectance images. The brightness, greenness and wetness of the derived transformation collectively explained over 97% of the spectral variance of individual scenes used in this study.

2.) Database Development

-Stratification

Originally, NLCD 1992 was mapped in zones determined by EPA administrative boundaries, which were unrelated to the biogeography of land cover and caused difficulties because mosaic boundaries included widely disparate land cover. This experience led to a focus on an improved regional stratification method for NLCD 2000 as a means to stage the database. Because mapping over large landscapes typically involves many satellite scenes, multi-scene mosaicking has often been used to group scenes into a single file for classification. This approach can potentially optimize both classification and edge matching (Homer et al 1997). However, large multi-scene mosaics create a variety of spectral gradients within the file, and files are subsequently useful only as a unit. Spectral gradients typically represent patterns of physiographic, phenologic, solar, atmospheric and instrument influences within and between remotely sensed imagery. The degree to which this variability can be isolated in local context largely determines the success of the classification. A common method of isolating spectral gradients is to stratify landscapes into sub-regions of similar biophysical and spectral characteristics. This process is not new to remote sensing and has been widely used as a method to improve accuracy (Pettinger 1982; White et al. 1995; Lillisand 1996). For example, Bauer et al. (1994) showed overall classification accuracy could potentially be improved by 10 to 15 percent using physiographic regions.

The application of mapping zones as a pre-classification stratification method has shown that carefully defined mapping zones can maximize spectral differentiation, provide a means to facilitate partitioning the workload into logical units, simplify post-classification modeling and improve classification accuracy. The underlying concept of mapping zone delineation is a pre-classification division of the landscape into a finite number of units that represent relative homogeneity with respect to landform, soil, vegetation, spectral reflectance, and image footprints at a project scale that is affordable. There are five general concepts that are useful in defining mapping zones: including economics of size, type of physiography, potential land cover distribution, potential spectral uniformity and edgematching issues.

The development of mapping zones across the conterminous United States included an initial review of project scope, which determined that approximately 60-70 zones would be the appropriate grain size for staging NLCD 2000. Initial mapping zone boundaries were based on Omernik (1987). These boundaries were displayed over two principal data layers, NLCD 1992 and AVHRR normalized greenness maps (NDVI) for modification. These data layers provided a landscape overview of both interpreted land cover and gross vegetation phenology patterns, and provided the context to further refine the initial Omernik boundaries on 1:5,000,000 scale paper maps. Paper map boundaries were subsequently crafted into a digital file by on-screen digitizing with 300m-pixel NLCD 1992 as the background. The next interpretative stage in this process will be to re-draft boundaries over full-resolution TM data to create local interpretation that can be applied at the single pixel scale. It is during this process that mapping zone concepts of physiography, spectral uniformity and edgematching will be fully applied.

The current draft of NLCD 2000 mapping zones is shown in Figure 4. This draft represents only the conceptual scale of the potential boundary lines for the mapping zones. Because the process is iterative, it is anticipated that future updates will be applied as boundaries are interpreted at full TM scales.

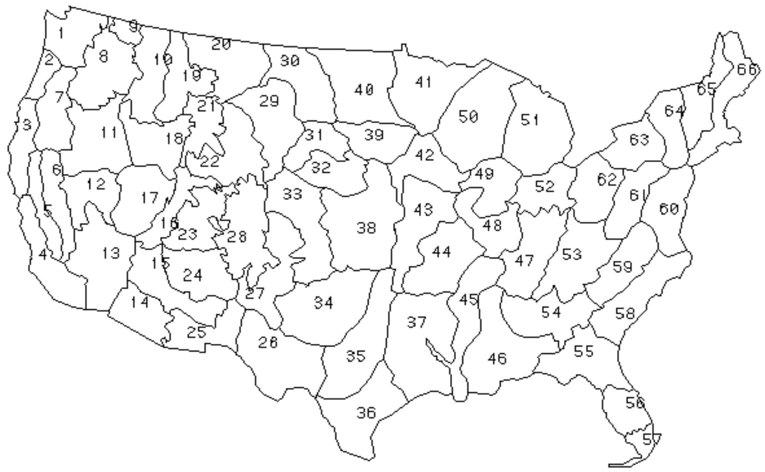


Figure 4. NLCD 2000 Mapping Zones

-Ancillary Data

The use of decision and regression tree algorithms in classification of the database into land cover derivatives allows ancillary data full weighting in the classification process. This highlights the need for consistent, and meaningful ancillary data sources. Based on experience with NLCD 1992, several ancillary data layers have been standardized for use in the NLCD 2000 database. These are derived from two main sources, the National Elevational Database (NED) at 30 meter scale and State Soil Geographic Data Base (STASCO) which is resampled to 30 meter. Besides the NED DEM itself, three DEM derivatives are used including slope, aspect and a positional index. Three ancillary data layers are derived from the STASCO soil data set. They include the Unit for Soil Available Water Capacity (AWC), the Unit for Soil Organic Carbon (OC), and Soil Quality.

-Landscape Shape

Shape is one of the primary distinguishing characteristics used when manually interpreting an image. The overall characteristics of an object's boundary along with the size of the object allow many objects to be recognized. This process has proved difficult to automate, with image segmentation the current approach. Image segmentation is the process of separating an image into homogeneous polygons that depict distinct regions on the ground. For this pilot dataset, a hierarchical segmentation was generated at three different scales. Shape methods described here resulted from work completed by Pacific Meridian Resources (now Space Imaging) through Aero-Metric, Inc. under USGS contract 98CRCN1004 to develop this research. ECognition software was used to generate image segments for test study sites using parameters tested on five sample areas per mapping zone, which represented the different types of land use classes found in each site. The segmentation algorithm was run for each test area and visually assessed to determine which set of parameters provided the best results for delineating spectrally unique regions while maintaining a naturally occurring shape of land use classes. Using the segmentation output, four shape measures are calculated including, convexity, compactness, fractal dimension, and form. Fractal dimension is the measure of the polygon's edge roughness and complexity. Convexity is the measure of the deviation of the spatial object from the convex hull of the polygon. Compactness is the measure of the polygon complexity and its deviation from the circularity. Form is the measure of the polygon's boundary roughness. Each of these shape measures are generated, stored and analyzed as an independent files.

-Texture

Texture algorithms are used to measure distinctive spatial and spectral relationships between neighboring pixels and can be helpful in distinguishing land use types that are composed of similar land cover types. Image *texture* is indicated by variance in pixel DN values across space. Texture bands using three Landsat 7 ETM+ bands (bands 1, 7, and 4) acquired during the "Leaf-On" season were calculated for the pilot zones using a standard deviation-based texture measure enhanced by an adaptive 3x3 window filter (Woodcock and Ryherd, 1989). Texture methods described here resulted from work completed by Pacific Meridian Resources (now Space Imaging) through Aero-Metric, Inc. under USGS contract 98CRCN1004.

-Imperviousness

One of the most influential land cover types in urban environments is impervious surfaces – a term that refers to any impenetrable surface such as rooftops, roads and parking lots. Quantification of imperviousness can offer a relative objective measure of urban density and can provide a forum for its classification. For NLCD 2000, imperviousness was chosen as the surrogate for an urban intensity classification in an effort to improve the precision of urban classification from the original NLCD 1992 more subjective based methods.

Quantification of urban impervious surface using remotely sensed data has been the focus of research endeavors using mapping techniques ranging from digital classification, spectral unmixing, artificial neural network (ANN), classification and regression tree (CRT), and integration of remote sensing data with geographic information system (GIS) technology. (Yang et al 2002). For NLCD 2000, a regression tree classification algorithm (Cubist) is used to output a per-pixel estimate of imperviousness in urban areas. This automated image processing procedure quantifies spatial distribution of impervious surfaces as a continuous variable from Landsat 7 Enhanced Thematic Mapper Plus (ETM+) imagery for urban areas. The procedure offers a consistent and repeatable method to characterize urban areas across the nation. The process and procedures of the proposed concept includes several key tasks: 1) delineating potential urban areas, 2) high resolution training data collection, 3) feature selection and initial regression modeling, 4) assessment and modification, and 5) final model (Yang et. al. 2002).

-Tree Canopy

Forest cover, both categories and canopy density, are of great interest to a variety of scientific and land management users. The original NLCD 1992 classification provided 4 forest categories, but made no distinction in forest density. Based on user feedback, a strategy for estimating tree canopy density at a spatial resolution of 30 m was developed for NLCD 2000 (Huang et al 2001A). This strategy is based on empirical relationships between tree canopy density and Landsat data, established using regression tree techniques. One-meter digital orthophoto quadrangles are used to derive reference tree canopy density data needed for calibrating the relationships between canopy density and Landsat spectral data. For NLCD 2000, a regression tree algorithm (Cubist) is used to output a per-pixel estimate of tree canopy in forested areas. This automated image processing procedure quantifies spatial distribution of tree canopy as a continuous variable from Landsat 7 Enhanced Thematic Mapper Plus (ETM+) imagery. The overall approach of the proposed strategy consists of three key steps: deriving reference data from high resolution images, calibrating canopy density models using the derived reference data, and extrapolating the developed models spatially using 30 m resolution images.

-Land Cover Classification

There are numerous algorithms for classifying satellite images. Potential methods reviewed for NLCD 2000 included clustering, expert system, neural network and decision tree classifiers. NLCD 1992 classification was based on a several step method of unsupervised clustering, using both pre and post classification stratification with ancillary data, and manual editing to complete the work (Vogelmann et al 2001). For NLCD 2000, a classification method that optimally incorporates many database layers in a single step, with the ability to document this relationship in a rule base was highly desirable. Decision tree classification (De'Atl and Fabricue 2000), was the method chosen for NLCD 2000. Advantages it offers include; 1) it is non-parametric and therefore independent of the distribution of class signature, 2) it can handle both continuous and nominal data, 3) it generates interpretable classification rules, and 4) it is fast to train and often is as accurate as or even slightly more accurate than many other classifiers. Decision trees are a supervised method of classification and require extensive well-balanced training data to perform adequately. The decision tree program used in this case study, C5, employs an information gain ratio method in tree development and pruning (Quinlan 1993), and has many advanced features including boosting and cross-validation.

For NLCD 2000 Decision tree classification potentially offers an efficient robust method to classify huge quantities of information in documentable form. An innovation now being developed allows export of decision tree's generated by the classification into generic rule sets allowing users access to classification parameters. It is envisioned that users can assess both spatial records of decision tree output (similar to spectral clusters) as well as generic rule-set text for review and importation. This comprehensive metadata approach will allow users assess to classification reasoning and will potentially allow local modification of the classification database.

3.) Initial Validation

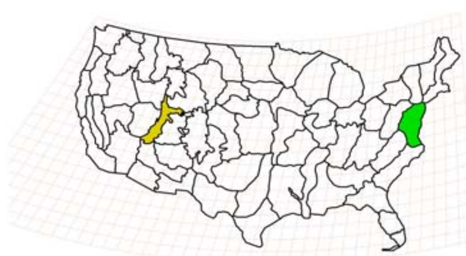
Land cover derivative data layers were initially assessed using cross validation in both regression and decision tree models. Cross-validation is designed to obtain relatively realistic accuracy estimates using a limited number of statistically valid collected reference data samples for both training and accuracy assessment (Michie et al. 1994). For an *N*-fold cross-validation the training data set is divided into *N* subsets. Accuracy estimates are derived by using each subset to evaluate the classification developed using the remaining training samples, and their average value represents the classification developed using all reference samples.

IV. PILOT STUDY

Study Areas

The full database described above was developed in two pilot-mapping zones in the mid-Atlantic region (zone 60) and the western Rocky Mountains (zone 16) (Figure 5). Zone 60 covers costal and inland areas from Staten Island in New York State to Albermarle Sound in North Carolina. This zone includes diverse land use classes from estuarine and coastal areas to agricultural, urban and forested lands. Zone 16 covers the Rocky Mountains of Utah. It extends from the Cache National Forest, located north of Salt Lake City, to Zion National Park in the south. It contains a variety of shrub, forested and agricultural lands.

For each mapping zone, three Enhanced Thematic Mapper Plus (ETM+) images were acquired to capture vegetation dynamics over a growing season and to maximize land cover type separability. These images were selected within the time period between 1999 and 2001 and pre-processed as described above into TC brightness, greenness and wetness bands calculated from at-satellite reflectance based coefficients. Leaf-on images were also processed into three texture bands, and four shape indices.



Ancillary data included the USGS 1-arc second NED and three derivatives including slope, aspect and topographic position index. Three soil attributes, available water capacity, soil carbon and soil quality derived from STATSGO and resampled to 30-meter pixels were also used.

Land cover classification

Landcover was derived from a combination of image and ancillary layers using a C5 decision tree. Input data included 26 bands of spectral and ancillary data in zone 60 and 20 bands in zone 16. Field data for both pilot zones were collected from combined sources. The majority of forested field data was provided for each region through a unique pilot agreement with the USFS Forest and Inventory Assessment Program (FIA). Through intensive fieldwork, the FIA program generates detailed forest sampling information nation-wide at regular intervals which can be used only with strict confidentiality agreements. Incorporating this even sampled data set improved forest mapping considerably, and provided reliable cross-validation estimates (Huang et al 2001B). Other reference data in addition to EDC collected data provided in zone 16, included field data collected by the USFS Fire Science Lab of the

Rocky Mountain Research Station and the Utah GAP Analysis program of Utah State University. In zone 60, field data were also contributed by the state of Delaware.

Imperviousness and Tree Canopy classification

Methods for imperviousness described by Yang et al 2002 and tree canopy methods described by Huang et al. 2001A, were applied in the mapping zones by Earth Satellite Corporation (EarthSat) through Greenhorne & O'Mara Inc. under USGS contract number 010112C0012. In zone 60, 20 DOQ quad subsets were used to generate training data, and in zone 16, 16 subset quads were used. Imperviousness and canopy estimates were developed using the Cubist Regression Tree algorithm. Input data layers included three seasonal TC Landsat 7 images, the leaf-on thermal band, aspect, slope, and soil quality.

Pilot Study Results

Land cover

Results listed here are preliminary, with additional drafts of the classification to follow. A total of 12 classes were mapped in zone 60 using a hierarchical approach that mapped forest classes separate from agriculture and wetland. Cross-validation accuracies for the three forest classes were 74% (SE 1.4%), with the remaining agricultural/wetland classes at 82% (SE 1.1). In zone 16, all 13 land cover classes were mapped concurrently, with cross validation accuracies for all classes at 65% (SE 0.4) (Figure 6).

Figure 6. Zone 16 land cover in 13 classes

Imperviousness and Canopy

Results described here are preliminary, and result from work completed by Earth Satellite Corporation (EarthSat) through Greenhorne & O'Mara Inc. under USGS contract number 010112C0012. Per-pixel imperviousness and tree canopy estimates for each zone were generated and assessed using cross-validation. Tree canopy results from both zones reveal mean absolute errors of approximately 10-15%, with correlation coefficients ranging from .77 in zone 16 to .90 in zone 60 (Figure 7). Imperviousness results from both zones reveal mean absolute errors also of approximately 10-15%, with correlation coefficients ranging from .89 in zone 16 to .91 in zone 60 (Figure 8).

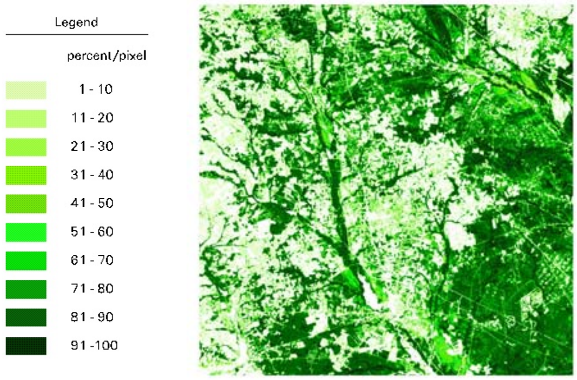


Figure 7. Example of Zone 60 Tree Canopy Estimates

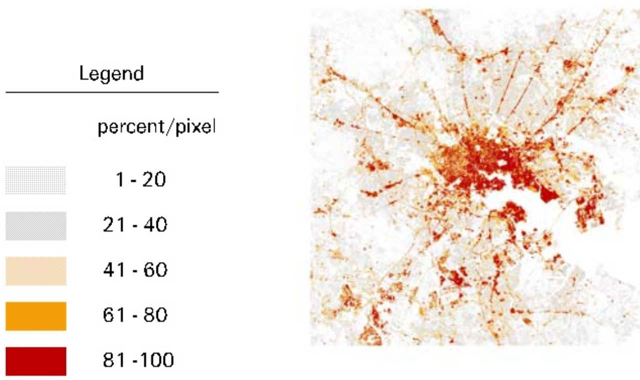


Figure 8. Example of Imperviousness Estimates for Baltimore

V. DISCUSSION

As illustrated by results from the pilot mapping zones, the NLCD 2000 database provides a comprehensive set of data layers that can potentially foster further exploration, development, application and sharing of land cover information by users at national and regional scales. The standardized nature of each data component at reasonable accuracies will allow users the ability to develop data applications that either use layers synergistically or individually. For example, imperviousness can potentially be used not only as a way to classify developed land, but also in water run-off models, green space calculations and urban planning scenarios. The consistency of these data layers will allow direct comparison from place to place, increasing the utility of potential applications.

The database concept will provide users flexible access and interaction with the individual data components and also the land cover products. Spatial and textual metadata generated from land cover product development will allow users the ability to download both database ingredients and rule recipes for local modification. Conceptually, a potential user could modify land cover model parameters directly in any standard software package by manipulating rule-set parameters according to more local information. In this scenario, NLCD 2000 acts as a framework to provide standardized ingredients and a general "recipe" empowering less sophisticated users to generate local value-added land cover without extensive preparation. Further this database could provide a common "language" for users to trade classification methods and results. The production of NLCD 2000 will be implemented in a phased approach using the mapping regions developed by USGS. Full production development is based upon available funding from MRLC 2000 partners and cooperators, but is anticipated to begin in FY 2002, with completion targeted for FY 2005.

AKNOWLEDGEMENTS

This study is made possible in part by the Raytheon Corporation under U.S. Geological Survey contract 1434-CR-97-CN-40274. This paper reflects an enormous team effort with important contributions by many EDC individuals including Mike Coan, Kent Hegge, Tom Loveland, Susan Maxwell, Russel Rykhus, Nick Van Driel, Jim Vogelmann and Jane Zhang. Jim Wickham and Jim Vogelmann authored an initial MRLC 2000 white paper. Dan Civco and Warren Cohen provided some key initial direction. Goshia Bryja, Maria Fiorella, Janet Hoyt and Rick Jones of Space Imaging's Portland office conducted key research on shape and texture (under USGS contract number 010112C0012) and found ways to apply it at the mapping zone scale. David Cunningham, Nate Herold and Greg Koeln of Earthsat also developed a working version of canopy and imperviousness (under USGS contract number 010112C0012 with Greenhorne & O'Mara Inc at the mapping zone scale. Several individuals in the FIA were key in developing an agreement for data namely Bill Cooke of the Southern region, Mike Hoppus of the Northeast region and, Ron Tymcio of the Rocky Mountain region.

REFERENCES

Anderson, J.F., Hardy, E.E., Roach, J.T., Witmer, R.E.: 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964, 28pp.

Bauer, M.E., T.E. Burk, A.R. Ek, P.R. Coppin, S.D. Lime, T.A. Walsh, D.K. Walters, W. Befort, and D. F. Heinzen. 1994. Satellite inventory of Minnesota forest resources. *Photogrammetric Engineering and Remote Sensing* 60(3):287-298.

Brown, J., T. Loveland, D. Ohlen and Z. Zhu. 1999. The global land-cover characteristics data-base: The users' perspective. *Photgrammetric Engineering and Remote Sensing* 65:1069-1074.

De 'Ath, Glenn and Katharina E. Fabricus. 2000. Classification and Regression Trees: A powerful yet simple technique for ecological data analysis. *The Ecological Society of America*. Vol. 81. No. 11. November 2000. pp 3178-3192.

Homer, C. G., and A. Gallant, 2001.Partitioning the Conterminous United States into Mapping Zones for Landsat TM Land Cover Mapping", USGS Draft White Paper, available at http://landcover.usgs.gov.

Homer, C.G., R.D. Ramsey, T.C. Edwards, Jr., and A. Falconer. 1997. Land cover-type modeling using multi-scene Thematic Mapper mosaic. *Photgrammetric Engineering and Remote Sensing* 63:59-67.

Huang, C., Wylie, B., Homer, C., Yang, L., and Zylstra, G. 2002. Derivation of a Tasseled cap transformation based on Landsat 7 at-satellite reflectance. International Journal of Remote Sensing. In press.

Huang, C., Yang, L., Wylie, B., and Homer, C. 2001A. A strategy for estimating tree canopy density using Landsat 7 ETM+ and high resolution images over large areas. in: Third International Conference on Geospatial Information in Agriculture and Forestry; November 5-7, 2001; Denver, Colorado. CD-ROM, 1 disk.

Huang, C., Yang, L., Homer, C., Coan, M., Rykhus, R., Zhang, Z., Wylie, B., Hegge, K., Zhu, Z., Lister, A., Hoppus, M., Tymcio, R., DeBlander, L., Cooke, W., McRoberts, R., Wendt, D., and Weyermann, a. D. 2001B. Synergistic use of fia plot data and landsat 7 etm+ images for large area forest mapping. in: The Thirty-fifth Annual Midwest Forest Mensurationists and the Third Annual FIA Symposium; October 17 - 19, 2001; Traverse City, MI.

Irish, R. R., 2000, Landsat 7 science data user's handbook, Report 430-15-01-003-0, National Aeronautics and Space Administration, http://ltpwww.gsfc.nasa.gov/ IAS/handbook/handbook_toc.html.

Lillisand, T. M. 1996. A Protocol for Satellite-Based Land Cover Classification in the Upper Midwest. *Gap Analysis: A Landscape Approach to Biodiversity Planning*, editors J. Michael Scott, Timothy H. Tear and Frank W. Davis. ASPRS, 320 pp.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, J, Yang, L., and Merchant, J.W., 2000. Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data. International Journal of Remote Sensing, v. 21, no. 6/7, p. 1,303-1,330.

Hall, F. G., Townshend, J. R., and Engman, E. T., 1995, Status of remote sensing algorithms for estimation of land surface state parameters, *Remote Sensing of Environment*, **51**: 138-156.

Markham, B.L. and Barker, J.L. 1986. Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures, *EOSAT Landsat Technical Notes*, 1:3-8.

Michie, D., Spiegelhalter, D. J., and Taylor, C. C. (Editors), 1994, *Machine learning, neural and statistical classification* (New York: Ellis Horwood), 289 pp.

Omernik, J.M. 1987 "Ecoregions of the Conterminous United States." Map (scale 1:7,500,000). Annals of the Association of American Geographers 77(1): 118-125.

Park, Stephen K. and Robert A. Schowengerdt. 1982. Image Reconstruction by Parametric Cubic Convolution. Computer Vision, Graphics and Image Processing 23:258-272.

Pettinger, L.R. 1982. Digital Classification of Landsat Data for Vegetation and Land-Cover Mapping in the Blackfoot River Watershed, Southeastern Idaho.: U. S. Geological Survey Professional Paper

Quinlan, J. R., 1993, *C4.5 programs for machine learning* (San Mateo, California: Morgan Kaufmann Publishers).

Shilen, Seymour. 1979. Geometric Correction, Registration and Resampling of Landsat Imagery. Canadian Journal of Remote Sensing, Vol. 5, No. 1.

Vogelmann, J.E., S.M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and J. N. Van Driel, 2001, Completion of the 1990's National Land Cover Data Set for the conterminous United States, Photogrammetric Engineering and Remote Sensing 67:650-662.

Wharton, S. W., 1989, Knowledge-based spectral classification of remotely sensed image data, In *Theory and applications of optical remote sensing*, edited by G. Asrar (New York: John Wiley & Sons), 548-577

White, J.D., G.C. Kroh, C. Glenn, and J.E. Pinder 3rd. 1995. Forest mapping at Lassen Volcanic National Park, California, using Landsat TM data and a geographical information system. Photogrammetric Engineering and Remote Sensing v 61 Mar P. 299-305.

Woodcock, C.E. and S.L. Ryheard, 1989. Generation of Texture Images using Adaptive Windows, Technical Papers, 55th Annual Meeting ASPRS, 2:11-22.

Yang, L., Collin Homer, Kent Hegge, Chengquan Huang, Bruce Wylie and Bradley Reed, 2001, A Landsat 7 Scene Selection Strategy for a National Land Cover Database. Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium, Sydney, Australia, CD-ROM, 1 Disc.

Yang, L, Chengquan Huang, Collin Homer, Bruce Wylie and Mike Coan. 2002. An approach for mapping impervious surfaces using multi-source and mult-sensor datasets, Manuscript in preparation, available at http://landcover.usgs.gov.