SYNERGISTIC USE OF FIA PLOT DATA AND LANDSAT 7 ETM+ IMAGES FOR LARGE AREA FOREST MAPPING*

Chengquan Huang[1], Limin Yang[1], Collin Homer[1], Michael Coan[1], Russell Rykhus[1], Zheng Zhang[1], Bruce Wylie[1], Kent Hegge[1], Zhiliang Zhu[2], Andrew Lister[3], Michael Hoppus[3], Ronald Tymcio[4], Larry DeBlander[4], William Cooke[5], Ronald McRoberts[6], Daniel Wendt[6], and Dale Weyermann[7]

ABSTRACT:--FIA plot data were used to assist classifying forest land cover from Landsat imagery and relevant ancillary data in two regions of the U.S., one around the Chesapeake Bay area and the other around Utah. The overall accuracies for the forest/non-forest classification were over 90% and about 80% in the two regions. The accuracies for deciduous/evergreen/mixed and forest type group classifications were around 80% and 65%, respectively, and were consistent in the two regions. These results suggest that use of FIA plot data together with satellite imagery and relevant ancillary data may substantially improve the efficiency, accuracy and consistency of large area forest land cover mapping.

* To be published in the proceedings of the Thirty-fifth Annual Midwest Forest Mensurationists Meeting and the Third Annual Forest Inventory and Analysis Symposium, October 17-19, 2001, Traverse City, MI.

[1]USGS EROS Data Center, Raytheon, Sioux Falls, SD 57198 USA. [2]USGS EROS Data Center, Sioux Falls, SD 57198 USA. [3]Forest Inventory and Analysis, Northeastern Research Station, 11 Campus Blvd, Suite 200, Newtown Square, PA 19073. [4]Forest Inventory and Analysis, Rocky Mountain Research Station, Ogden, Utah 84401. [5]Forest Inventory and Analysis, Southern Research Station, P. O. Box 928, Starkville, MS 39760. [6]Forest Inventory and Analysis, North Central Research Station, 1992 Folwell Ave., St. Paul, MN 55108. [7]Forest Inventory and Analysis, Pacific Northwest Research Station, 1221 SW Yamhill Suite 200, Portland, OR 97205.

INTRODUCTION

Reliable and updated forest information is required for many scientific and land management applications. Meeting this requirement is of interest to both the Forest Inventory Analysis (FIA) program of the U.S. Forest Service and the Land Cover Characterization (LCC) program of the U.S. Geological Survey (USGS) Earth Resources Observation Systems (EROS) Data Center (EDC). The FIA program has a mandate to collect and report information periodically on status and trends in the nation's forested resources, while the LCC program is mandated to develop a circa 2000 national land cover database through the Multi-Resolution Land Characterization (MRLC) 2000 project. Therefore, it is in the best interest of the government that these two agencies collaborate in mapping the nation's land cover. The current study is the result of an initial collaboration between the two agencies.

Forest land cover information is often derived from remotely sensed images using classification algorithms (e.g. Franklin et al. 1986, Mickelson et al. 1998), many of which require a substantial amount of reference data (Townshend 1992, Hall et al. 1995). Reliable reference data is also required for assessing classification results. One of the challenges to mapping forest land cover over large areas is the lack of adequate reference data. In areas where some reference data sets exist, they may have been collected in different ways and may have different levels of reliability. Such scarcity of reliable reference data and lack of consistency among the available ones often limit the efficiency, consistency and accuracy in deriving forest information from satellite imagery.

The plot data collected through the FIA program is a potentially high quality reference data set for the MRLC 2000 project. FIA plots represent a statistically based sampling of the nation's land. Detailed information on forest status and structure is collected periodically at each plot through intensive field work. With minimal efforts, this data set can be reorganized for use with remotely sensed images. The purpose of the current study is to evaluate the usefulness of the FIA plot data in deriving forest land cover classifications from satellite imagery over large areas, and to test whether using this data set can improve the efficiency, accuracy and consistency in developing the MRLC 2000 national land cover database.

DATA AND PREPROCESSING

Study Area

For mapping efficiency, the conterminous United States was divided into 66 mapping zones for the MRLC project. Two mapping zones: zones 16 and 60 were used in this pilot study (figure 1). Zone 60 represents the eastern coastal environment, covering the Chesapeake Bay area, while zone 16 represents the western arid and less developed landscape, covering Utah and southern Idaho. Figure 1 shows the Landsat paths/rows covered by the two mapping zones.

Landsat Imagery and Ancillary Data

For each Landsat path/row covered by the two mapping zones, Enhanced Thematic Mapper Plus (ETM+) images were acquired in three different dates to capture vegetation dynamics over a growing season and to maximize land cover type separability (Yang et al. 2001a). These images were acquired within the time period between 1999 and 2001, and were selected to minimize the impact of cloud cover and atmospheric effects. The images were geometrically and radiometrically corrected using standard methods at the USGS EROS Data Center (Irish 2000). Terrain correction using the USGS 1-arc second National Elevation Dataset was performed to improve geolocation accuracy. Raw digital numbers were 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 User's Handbook (Irish 2000). All 7 bands were resampled to a 30 m spatial resolution. Tasseled-cap brightness, greenness and wetness were calculated using at-satellite reflectance based coefficients (Huang et al. 2002b).

Ancillary data included the USGS 1-arc second National Elevation Dataset and three derivatives, i.e., slope, aspect and a topographic position index. In addition, three soil attributes, i.e., available water capacity, soil carbon content and a soil quality index, were derived from the State Soil Geographic (STATSGO) Data Base. All ancillary data layers were resampled to a spatial resolution of 30 m.

Reference Data Sets

Through intensive field work, the FIA program provides detailed forest attributes at individual tree, sub-plot and plot levels. Considering the pixel size of the ETM+ imagery and possible geolocation errors of the imagery and FIA plots, only plot level data were deemed appropriate for use with the ETM+ imagery. Therefore, tree level data were summarized to sub-plot level and then to plot level. In zone 60, each plot was labeled with single- or multiple-condition primarily depending on if there were only one or multiple land use/land cover types within the plot area. Multiple-condition plots were excluded to minimize the impact of misregistration errors and other possible inconsistencies between FIA plots and the satellite images. This was not deemed necessary in zone 16 because most of the plots were based on the plot design used prior to 1995, which restricts all plots to single-condition. Each eligible plot was then classified at three levels: forest/non-forest, deciduous/evergreen/mixed, and forest type group. Table 1 lists the number of FIA plots used in this study. Global Positioning System units were used to locate all plots in zone 60 and 147 plots in zone 16. The remaining plots in zone 16 were digitized from aerial photos used in the field. Geolocation errors between the digitized plots and ETM+ images were minimal. A visual check of over 100 random plots digitized from the air photos against corresponding satellite images suggested that less than 10% of the plots had location errors greater than one ETM+ pixel.

Another reference data set available in zone 16 consisted of field data collected by the Fire Science Lab of the Rocky Mountain Research Station of the U.S. Forest Service and the Utah GAP Analysis program of Utah State University. Each field site was classified at two levels: forest/non-forest and deciduous/evergreen/mixed. This was used as an independent data set to evaluate the classification results developed using FIA plot data in mapping zone 16. Although the point location was not based on any statistically rigorous sampling design and the evaluation should not be considered a statistically rigorous accuracy assessment, this independent reference data set should provide useful information on the consistency of FIA plot data and the reliability of derived classifications.

METHODS

Classification Levels

As with the reference data, classification of the ETM+ images and ancillary data was performed at three levels: forest/non-forest, deciduous/evergreen/mixed, and forest type group. A forest/non-forest map is required by FIA to implement a stratified sampling of forested land in order to produce accurate estimates of forest attributes. Deciduous, evergreen and mixed are the main forest categories in the MRLC 2000 classification scheme. Type group information is often required for species conservation planning, fire management and many other applications. Table 2 lists the major forest type groups in the two mapping zones.

Decision Tree Classifier

Many algorithms are available for classifying satellite images (Townshend 1992, Hall et al. 1995), among the most popular of which include the maximum likelihood classifier, neural network classifiers and decision tree classifiers. Decision tree was chosen for this study because it 1) is non-parametric and therefore independent of the distribution of class signature, 2) can handle both continuous and nominal variables, 3) generates interpretable classification rules, and 4) is fast to train and is often as accurate as, sometimes more accurate than, many other classifiers (Hansen et al. 1996, Huang et al. 2002a). The decision tree program used in this study, C5, employs an information gain ratio criterion in tree development and pruning (Quinlan 1993). This program has many advanced features, including boosting and cross-validation.

Boosting

Boosting is a technique for improving classification accuracy (Bauer and Kohavi 1998). With this function, the program develops a sequence of decision trees, each subsequent one trying to fix the misclassification errors in the