Evaluating Classified MODIS Satellite Imagery as a Stratification Tool

Greg C. Liknes, Mark D. Nelson, and Ronald E. McRoberts
USDA Forest Service
1992 Folwell Ave
St. Paul, Minnesota, United States of America 55108
Ph. +1 651 6495192 Fax. +1 651 6495285
E-mail: gliknes@fs.fed.us

Keywords: Stratified estimation, MODIS, spatial resolution

Abstract
The Forest Inventory and Analysis (FIA) program of the USDA Forest Service collects forest attribute data on permanent plots arranged on a hexagonal network across all 50 states and Puerto Rico. Due to budget constraints, sample sizes sufficient to satisfy national FIA precision standards are seldom achieved for most inventory variables unless the estimation process is enhanced with ancillary data. When used to create strata for stratified estimation, satellite imagery can be effective ancillary data. The National Land Cover Dataset (NLCD), a land cover classification based on satellite imagery, has been used to produce substantial increases in the precision of statewide forest inventory estimates.

Because inventories are conducted on an annual basis, it is desirable to create strata using a product that is updated more frequently than the 10-year update cycle of the NLCD. In particular, data from the MODIS sensor are available every 1-2 days, although at a much coarser spatial resolution than the Landsat data used in the creation of the NLCD (250-1000m vs. 30m). In this study, the effectiveness of strata created by classifying MODIS satellite imagery is compared to that of strata created from the NLCD. Results indicate that precision decreases by 0.9 percent per million acres when using a 1-km dataset versus a 30-m dataset.

1. Introduction

The Forest Inventory and Analysis (FIA) program of the US Department of Agriculture (USDA) Forest Service conducts forest inventories of the United States to estimate the area of forest land; the volume, growth, and removal of forest resources; and the health and condition of these resources. The combination of natural variability and budgetary constraints prohibits measurement of a sufficient number of plots to satisfy precision standards (USDA-FS 1970) for estimates of most inventory variables unless the estimation process is enhanced using ancillary data. Stratified estimation (post-sampling stratification) using strata derived from land cover classifications based on satellite imagery can reduce the variances of estimates of forest land area by factors as great as 5.0 (McRoberts et al. 2002b). Specifically, the NLCD, based on 30-m Landsat 5 Thematic Mapper (TM) data, has been successfully used for variance reduction (Hansen and Wendt 2000, McRoberts et al. 2002a).

When creating strata for use in variance reduction of forest attribute estimates, it is desirable to use satellite imagery collected in temporal proximity to the collection of forest attribute data. Because the NLCD is updated approximately every 10-years, there is a mismatch between strata derived from NLCD and most years of forest attribute data collected in an annualized inventory.

At this time it is unknown how long the NLCD will remain a viable data source for FIA stratification in the North Central region. Alternatives to Landsat-based classifications such as NLCD are those based on Moderate-Resolution Imaging Spectroradiometer (MODIS) data. The objective of the study was to investigate the effectiveness of stratified estimation using strata derived from classifications of MODIS satellite imagery.

2. Data

2.1 Satellite-derived data

The NLCD, a digital product of the Multi-Resolution Land Characterization (MRLC) Consortium (Loveland and Shaw 1996), is a land cover map of the conterminous United States consisting of assignment of each 30 m x 30 m pixel to one of 21 land cover classes. The land cover classification was produced by the U.S. Geological Survey and was based on nominal 1992 TM satellite imagery and a variety of ancillary data (Vogelmann et al. 2001). MRLC is currently producing a newer but similar land cover map of the conterminous United States based on nominal 2001 Landsat 7 Enhanced Thematic Mapper Plus (ETM+) satellite imagery and a variety of ancillary data using a decision- tree classifier (Homer and Gallant 2001). Final drafts have been completed for initial prototype mapping zones including Zone 41, which includes sixty percent of the area of Minnesota, USA, and ninety percent of its forest land (Figure 1). Classifications from both 1992 and 2001 NLCD were used in this study.

Figure 1: MRLC Zone 41 mapping zone in Minnesota.

The MODIS instrument resides on the Terra and Aqua satellites, and provides passive detection capability at thirty-six wavelengths of electro-magnetic radiation. These satellites provide complete coverage of the earth every 1-2 days, with Terra passing over the equator southbound in the morning and Aqua passing over the equator northbound in the afternoon. Of the thirty-six spectral bands, the first seven are similar to the bands in TM/ETM+ and are used for the detection of land properties, with the remainder used for atmospheric and oceanic detection (http://modis.gsfc.nasa.gov).

For this study, two classifications derived from MODIS were used. The first was based on the Vegetation Continuous Fields (VCF) product, a global dataset consisting of three layers representing percent tree canopy, percent herbaceous cover and percent bare ground cover (Hansen et al. 2002). The dataset is constructed using monthly composites of 500-m data. All seven spectral MODIS land bands are used in the algorithm to determine percent cover. The second was based on the MOD12Q1 product, a 1-km land cover dataset. It is created using a supervised classification process in conjunction with a technique referred to as boosting (http://geography.bu.edu/landcover/userguidelc/lc.html). The classification can be updated every 96-days and is based on the previous twelve months of MODIS data. The classified, 1-km output data used in this study were coded in the International Geosphere-Biosphere Programme (IGBP) scheme (Running et al. 1994).

2.2 FIA Plot Data

For the FIA program of the North Central Research Station (NCRS), FIA field plot data include a variety of plot- and tree-level observations obtained in two phases. In Phase 1, each plot is observed using aerial photography or digital orthophotoquads (DOQs) to determine if trees are growing on it. Plots determined in Phase 1 to be without trees receive no additional observation, and values for all tree-related variables are set to zero. In Phase 2, field crews visit 0.4 ha plots, map land use conditions, and record observations for individual trees that include species, diameter at breast height (DBH) (1.37 m), and height. These plot measurements provide an observation of the portion of the plot that is forest land (Hahn et al. 1995).

For this study, measurements taken during the first four years of the FIA annual inventory in Minnesota (1999-2002) were used. Observations of forest proportion were obtained for 6,635 plots. Of these plots, 3,209 were partially or completely forested and 3,426 were not forested.

3. Methods

3.1 Stratified estimation

Stratified estimation requires (1) assignment of each FIA field plot to a single stratum and (2) calculation of the relative proportion of land area corresponding to each stratum. Plots are assigned to strata based on the strata assignments of their associated pixels. For this study, each plot was assigned to the stratum corresponding to the pixel in which the center of the plot was located. Strata proportions were estimated by counting the number of pixels in each stratum and then calculating the proportions of pixels in strata. These proportions are used as strata weights when calculating stratified estimates of means and variances.

Stratified estimates of forest area were calculated using standard methods (Cochran 1977):

$$\overline{Y} = \sum_{h=1}^{H} w_h \overline{Y}_h \tag{1}$$

and

$$Var(\overline{Y}) = \sum_{h=1}^{H} w_h^2 \frac{\hat{\sigma}_h^2}{n_h} \tag{2}$$

where H denotes the number of strata; w_h is the weight for the h^{th} stratum, calculated as the proportion of pixels assigned to the stratum; n_h is the number of plots assigned to the h^{th} stratum; $\overline{Y}_h$ is the sample mean for the h^{th} stratum, defined as,

$$\overline{Y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{hi} \tag{3}$$

and $\hat{\sigma}_h^2$ is the sample estimate of the stratum variance, defined as,

$$\hat{\sigma}_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(Y_{hi} - \overline{Y}_h\right)^2 . \qquad (4)$$

3.2 Constructing strata

To create strata for the purpose of stratified estimation, NLCD classes were aggregated into forest and non-forest classes. McRoberts et al. (2002a) investigated the utility of 1992 NLCD-derived stratifications for variance reduction purposes and recommended creating four strata using a three-step process: (1) aggregate a subset of NLCD classes into a forest stratum, and the remaining classes into a non-forest stratum; (2) reclassify isolated groups of same-stratum pixels smaller in area than 1 acre into their surrounding class; and (3) create a forest edge stratum from the 2-pixel wide band on the forest side of the forest/non-forest boundary and create a non-forest edge stratum from the 2-pixel wide band on the non-forest side of the forest/non-forest boundary. The first approach to stratification used in this study, denoted NLCD1992-2CLASS, derived only two strata, forest and non-forest, using the older (1992) NLCD. The method used by McRoberts et al. (2002a) for aggregating the 1992 NLCD classes was used in this study.

The second approach to stratification using the newer (2001) NLCD is dentoed NLCD2001-2CLASS. Both NLCD approaches were only processed through step 1 above in order to simplify the comparisons with approaches involving data with much coarser resolution. The 2001 NLCD classes differ slightly from the 1992 NLCD, and aggregation of the 2001 data into forest/non-forest classes is shown in Table 1.

Table 1: Aggregation of 2001 NLCD classes into forest/non-forest strata.

2001 NLCD Class	Strata
Deciduous Forest	Forest
Evergreen Forest	Forest
Mixed Forest	Forest
Short Shrub land	Forest
Woody Wetland	Forest
Open Water	Non-forest
Developed Open Space	Non-forest
Developed Low	Non-forest
Developed Medium	Non-forest
Natural Barren	Non-forest
Herbaceous	Non-forest
Pasture/Hay	Non-forest
Cultivated Crops	Non-forest
Emergent Herbaceous Wetlands	Non-forest

Forest/non-forest strata were created from the VCF data by selecting a percent tree canopy threshold, above which a pixel was identified as forest and below, non-forest. Thresholds were iteratively evaluated until the resulting forest area matched the area estimate for Zone 41 from FIA plot data under the assumption of simple random sampling (SRS). This resulted in a threshold percent tree canopy value of 0.41. This approach will be referred to as VCF-THRESH.

Forest/non-forest strata were created from the 1-km MOD12Q1 land cover dataset by aggregating the IGBP scheme into forest and non-forest classes (Table 2).

Table 2: Aggregation of IGBP classes into forest/non-forest strata.

IGBP Class	Strata
Evergreen Needleleaf Forest	Forest
Evergreen Broadleaf Forest	Forest
Deciduous Needleleaf Forest	Forest
Deciduous Broadleaf Forest	Forest
Mixed Forests	Forest
Closed Shrublands	Forest
Woody Savannas	Forest
Water	Non-forest
Open Shrublands	Non-forest
Savannas	Non-forest
Grasslands	Non-forest
Permanent Wetlands	Non-forest
Croplands	Non-forest
Urban and Built-Up	Non-forest
Cropland/Natural Vegetation Mosaic	Non-forest
Snow and Ice	Non-forest
Barren or Sparsely Vegetated	Non-forest

Maps depicting stratification approaches at three different resolutions for a subset of the Zone 41 study area are shown in Figure 2.

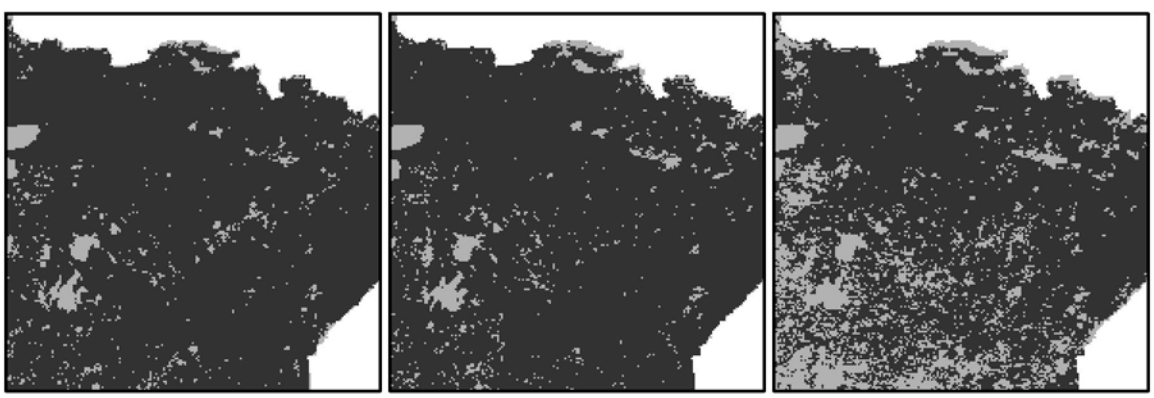

Figure 2: Maps depicting three stratification approaches – NLCD2001-2CLASS (30-m) (left), VCF-THRESH (500-m) (middle), and MOD12Q1 (1-km) (right) – for a portion of the Zone 41 study area. The forest class is dark gray and the non-forest class is light gray.

3.3 Analysis

For comparative purposes, an estimate of forest area and a standard error of that estimate were computed under the assumption of SRS. The SRS approach uses no stratification, calculates the overall sample mean and variance, and is equivalent to stratified estimation using a single stratum.

In addition to obtaining estimates of forest area using SRS and the four stratification approaches, NLCD1992-2CLASS, NLCD2001-2CLASS, VCF-THRESH and MOD12Q1, precision estimates, scaled to a reference of one million acres, were calculated using the national FIA precision guidelines (USDA FS 1970):

$$\text{PREC} = \frac{\left[\text{Var}(\overline{\text{Y}})\right]^{1/2}}{\overline{\text{Y}}}\left[\frac{\overline{\text{Y}}}{1{,}000{,}000}\right]^{1/2} \tag{5}$$

where $\overline{\text{Y}}$ is the estimate of forest area in acres, and Var($\overline{\text{Y}}$) is the variance of the estimate.

4. Results

As shown in Table 3, the SRS approach and each of the four stratification approaches produced a forest area estimate of approximately 15 million acres. The standard error per million acres increased as the pixel size increased or the resolution of the classification decreased (30-m=high resolution, 1000-m=low resolution), approaching the precision from the SRS approach. Although this increase is relatively small (< 1%), this represents an error increase of about 150,000 acres over a 15 million acre area.

Table 3: Stratification approach along with resulting estimates of forest area for Zone 41 and associated precision.

Approach	Resolution	Forest Area (million acres)	Precision (percent per million acres)
SRS	---------	14.820	5.0174
NLCD1992-2CLASS	30-m	15.254	3.2546
NLCD2001-2CLASS	30-m	15.297	3.3464
VCF-THRESH	500-m	14.952	4.0949
MOD12Q1	1000-m	14.840	4.2197

5. Discussion

None of the four stratification maps produced estimates that satisfied the national FIA precision guideline of three percent per million acres for forest area, but all four provided significant reductions from the SRS approach. We did not implement steps two and three of the three-step process recommended by McRoberts et al. (2002a), both of which have been shown to further reduce standard errors of forest land area estimates when using TM-derived stratifications. Although these steps would have been appropriate for the 30-m NLCD-derived stratifications, they appear less appropriate for stratifications derived from larger pixel, e.g., the 500-m VCF-THRESH and the 1000-m MOD12Q1. This is because 500-m and 1000-m pixels already are much larger in area (60 acres and 240 acres, respectively), than the 1-acre minimum area achieved from step 2 with 30-m stratifications. Likewise, the coarser resolution of 500-m and 1000-m pixels does not effectively identify edges and results in numerous mixed pixels containing both forest and non-forest cover, but assigned to a single stratum.

Standard errors of forest land area estimates increased as stratification pixel size increased. Although the 1000-m MOD12Q1 appears least effective for stratification, its precision was only 0.1 percent larger than that of the 500-m VCF-THRESH stratification, despite the fact that we identified an optimal VCF percent tree canopy threshold when creating the 500-m stratification. By identifying multiple thresholds, additional strata could be created for the VCF-THRESH stratification. As McRoberts et al. (2002a) demonstrated for the three-step process, additional strata may concentrate the more variable plot observations into strata with smaller weights, thereby further reducing standard errors. In the case of VCF a threshold of 0.41 differentiates forest from non-forest, but does 100% of such a forested pixel have 41% tree canopy cover or does 41% of the pixel have 100% tree canopy cover, or both? By employing multiple thresholds we could differentiate pixels that are predominantly forest, predominantly non-forest, and one or more categories of uncertainty in between. We suggest additional research for testing the efficacy of multiple strata in moderate- to coarse-resolution stratifications.

Although stratifications derived from coarser spatial resolutions (larger pixels) were not tested, we surmise that increases in standard errors would grow only marginally larger for pixel resolutions larger than 1000-m.

ACKNOWLEDGEMENTS

We thank Geoff Holden of North Central FIA for his work in generating stratified estimates for this study.

REFERENCES

Cochran, W. 1977. Sampling techniques, (3rd ed.). New York, NY: Wiley, 428 p.

Hahn, J.T., C.D. MacLean, S.L. Arner, W.A. Bechtold 1995. Procedures to handle inventory cluster plots that straddle two or more conditions. Forest Science Monograph, 31, 12-25

Hansen, M., R. DeFries, J.R.G. Townshend, R. Sohlberg, C. Dimiceli,, and M. Carrol. 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS, Remote Sensing of Environment, 83, 303-319.

Hansen, M.H., and D.G. Wendt. 2000. Using classified Landsat Thematic Mapper data for stratification in a statewide forest inventory. In: R. E. McRoberts, G. A. Reams, & P. C. Van Deusen (Eds.), Proceedings of the First Annual Forest Inventory and Analysis Symposium. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station (Gen. Tech. Rep. NC-213).

Homer, C.G., and A. Gallant. 2001. Partitioning the conterminous United States in mapping zones for Landsat TM land cover mapping. USGS Draft White Paper.

Loveland, T.L. and D.M. Shaw, 1996. Multiresolution land characterization: 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, MC, (National Biological Service, Moscow, ID), pp. 83-89.

McRoberts, R.E., D.G. Wendt, M.D. Nelson, and M.H. Hansen. 2002a. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. Remote Sensing of Environment, 81, 36-44.

McRoberts, R.E., M.D. Nelson, and D.G. Wendt. 2002b. Stratified estimation of forest area using satellite imagery inventory data, and the k-Nearest Neighbors technique. Remote Sensing of Environment, 82, 457-468.

Running, S.W., T.R. Loveland, and L.L. Pierce. 1994. Vegetation classification logic based on remote sensing for use in global scale biogeochemical models. Ambio, 23, 77-81

USDA Forest Service [USDA-FS]. 1970. Operational procedures. Forest Service Handbook. 4809.11 (Chapter 10, pp. 11-1) Washington, DC. USDA Forest Service.

Vogelman, J. E., S.M. Howard, L. Yang, C.R. Laron, B.K. Wylie, and N. Van Driel. 2001. Completion of the 1990s National Land Cover Data Set for the contemerinous United States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering and Remote Sensing, 67, 650-662.