The Utility of the Cropland Data Layer for Forest Inventory and Analysis

Greg C. Liknes[1], Mark D. Nelson[2], Dale D. Gormanson[3], and Mark Hansen[2]

Abstract.—The Forest Service, U.S. Department of Agriculture's (USDA's) Northern Research Station Forest Inventory and Analysis program (NRS-FIA) uses digital land cover products derived from remotely sensed imagery, such as the National Land Cover Dataset (NLCD), for the purpose of variance reduction via postsampling stratification. The update cycle of the NLCD product is infrequent; NLCD 2001 was the first update since the release of NLCD 1992, and was not yet fully completed as of late 2006. Consequently, FIA field data collected as recently as 2005 are being poststratified with land cover data collected more than a decade before. In addition, NRS-FIA has performed its own land cover classification of remotely sensed imagery for use in the stratification of some States, a time-consuming process. Alternative sources of information need to be evaluated both to eliminate the temporal mismatch between land cover data and FIA plot information and to reduce the amount of analyst time required to perform the stratification process.

The USDA's National Agricultural Statistics Service, in conjunction with the Foreign Agricultural Service, produces the Cropland Data Layer (CDL) using satellite imagery. The product, updated yearly for multiple States, includes detailed classification of crop type and also a "Woodland" category. In this study, the CDL was compared to the NLCD 1992 data set for Wisconsin. This comparison included two components: (1) county-level, pixel-derived estimates of forest land area relative to each other and to plot-based FIA estimates and (2) variance reduction produced by each layer when used in postsampling stratification of FIA plots for estimating primary FIA attributes (forest area, number of trees, timber volume, and tree biomass). Results indicate poor agreement between CDL pixel-based area estimates and FIA plot-based estimates; however, when used for poststratification, the CDL produces similar estimates of primary FIA attributes to those of the NLCD at the State level and higher relative efficiency at the FIA survey-unit level.

Introduction

The Forest Service's Forest Inventory and Analysis (FIA) program makes estimates of the area of forest land across all 50 States and Puerto Rico. In addition, the volume, growth, removal, and health of forest resources are assessed. Estimates are achieved via data collected on a consistent, nationwide sampling framework (Bechtold and Patterson 2005). The FIA program at the Northern Research Station (NRS-FIA) uses digital land cover products derived from remotely sensed imagery, such as the National Land Cover Dataset (NLCD), for the purpose of reducing the variance of estimates via postsampling stratification. The update cycle of the NLCD product is infrequent. NLCD 2001 (Homer *et al.* 2004) was the first update since the release of NLCD 1992 (Vogelman *et al.* 2001), and was not yet fully completed as of late 2006. Consequently, FIA field data collected as recently as 2005 are being poststratified with land cover data collected more than a decade before. In addition, NRS-FIA has performed its own land cover classification of remotely sensed imagery for use in the stratification of some States, a time-consuming process. Alternative sources of information need to be evaluated both to eliminate the temporal

[1] Research Physical Scientist, U.S. Department of Agriculture (USDA) Forest Service, Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.
[2] Research Forester, USDA Forest Service, Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.
[3] Supervisory Forester, USDA Forest Service, Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

mismatch between land cover data and FIA plot information and to reduce the amount of analyst time required to perform the stratification process.

The goal of this study was to compare another consistent land cover data set to NLCD and to evaluate the utility of a data product from another U.S. Department of Agriculture (USDA) agency for FIA. The Cropland Data Layer (CDL) (Craig 2001) was compared to the NLCD 1992 data set for Wisconsin. This comparison included two components (1) county-level, pixel-derived estimates of forest land area relative to each other and to plot-based FIA estimates and (2) variance reduction produced by each layer when used in postsampling stratification of FIA plots for estimating primary FIA attributes (forest area, number of trees, timber volume, and tree biomass).

Data

National Land Cover Dataset

The NLCD, a digital product of the Multi-Resolution Land Characterization Consortium (Loveland and Shaw 1996), is a land cover map of the conterminous United States consisting of the assignment of each 30-m x 30-m pixel to 1 of 21 land cover classes, 4 of which are assumed to be sufficiently similar to FIA's definition of forest to be considered forest for this study. The land cover classification was produced by the U.S. Geological Survey and was based on nominal 1992 Landsat Thematic Mapper (TM) satellite imagery and a variety of ancillary data. NLCD 1992 data are freely available for each of the 48 conterminous United States (Vogelman *et al.* 2001).

Cropland Data Layer

The CDL (Craig 2001, Mueller and Ozga 2002) is a digital product produced as part of a cooperative venture between two USDA agencies, the National Agricultural Statistics Service and the Foreign Agricultural Service. The product, updated yearly for multiple States, includes detailed classification of crop type and also a broad "Woodland" category, which includes woods and wooded pastures. The primary purpose of the CDL is to monitor cropland and other agricultural lands; the identification of woodlands is a side product of the classification effort. The layer used in this study is the Wisconsin 2005 CDL, which is based on multiple dates of Landsat TM imagery from that same year. As with the NLCD, CDL pixels are 30 m x 30 m.

FIA Plot Data

For this study, measurements taken during the first FIA annual inventory in Wisconsin (2000–04) were used. Plot-level data on forest area, number of trees, timber volume, and tree biomass were obtained for 12,885 plots. Of these plots, 6,478 were partially or completely forested and 6,407 were not forested. County-level land area estimates based on plot data were obtained from the FIA Mapmaker program (Miles 2001).

Methods

NLCD/CDL Processing

The 21 land cover classes of the NLCD 1992 were collapsed into forest and nonforest categories, and a raster filtering process was used to remove forest/nonforest pixel groupings smaller than 1 acre. This filtering process was done to conform to the FIA definition of forest land, which requires that groups of trees occupy a minimum of 1 acre. Three forest classes (deciduous forest, evergreen forest, and mixed forest); shrublands; woody wetlands; and lands in transition to forest were included in the aggregate forest class. Next, an edge category, 2 pixels in width, was created along both sides of the forest/nonforest interface. This procedure resulted in a total of four categories, or strata: forest, forest edge, nonforest, and nonforest edge.

CDL categories were also collapsed into forest and nonforest classes and filtered to remove groupings smaller than 1 acre. The Woodland category was the sole member of the forest stratum; all other categories became nonforest, with the exception of clouds. Cloud pixels accounted for less than 0.1 percent of the total CDL area and were recoded to either forest or nonforest class by applying a series of successively finer majority filters to assign unclassified cloud pixels the majority land cover value surrounding them. As with the NLCD, 2-pixel edge strata were created, resulting in a total of four strata. No CDL data were available for small portions of Wisconsin (e.g., a few islands in Lake Michigan beyond the Door County peninsula), and therefore the NLCD was clipped to the smaller extent of

the CDL to create equivalent extents for each stratification layer.

Pixel-Based Area Estimation

Estimates of forest land area were compared between FIA and CDL, FIA and NLCD, and CDL and NLCD data. County estimates were compared using percent differences and area-weighted root mean square deviation (RMSD). Percent differences for each pair were defined by equations 1, 2, and 3.

$$\frac{\left(CDL - FIA\right)}{FIA} \quad (1)$$

$$\frac{\left(NLCD - FIA\right)}{FIA} \quad (2)$$

$$\frac{\left(CDL - NLCD\right)}{NLCD} \quad (3)$$

where CDL and NLCD are the pixel-based estimates of forest area and FIA is the plot-based estimate of forest area. The RMSD was modified from Häme *et al.* (2001), equation (4).

$$RMSD_{rv} = \sqrt{\sum_i \frac{a_i}{A}\left(\hat{p}_{ir} - \hat{p}_{iv}\right)^2} \quad (4)$$

where a_i is the area of the ith county, $A = \sum a_i$ is the total Wisconsin area, and $\hat{p}_{ir}$ and $\hat{p}_{iv}$ denote the estimated proportion of forest land area in the ith county obtained from FIA, CDL, or NLCD (r) and FIA, CDL, or NLCD (v), respectively, for each of three pair-wise comparisons of estimates.

Stratified Estimation

The purpose of stratified estimation is to reduce the variance of sample-based estimates without increasing sample size. For this study, each FIA plot was assigned to one of four strata for both the NLCD and CDL corresponding to the pixel in which the center of the plot was located. Stratum proportions were estimated by counting the number of pixels in each stratum and dividing by the total number of pixels in the study area; these proportions are used as stratum weights when calculating stratified estimates of means and variances. In addition, relative efficiency (RE) was calculated by determining the variance of an estimate under the assumption of simple random sampling and dividing by the variance using stratified estimation. All estimation followed the standard FIA estimation procedures presented by Scott *et al.* (2005). Estimates, variances, standard errors, and RE for forest area, number of trees, timber volume, and tree biomass were calculated for the entire State as well as for FIA survey units (fig. 1).

Figure 1.—*FIA survey units for the State of Wisconsin.*

Results

Pixel-Based Area Estimation

Estimates of Wisconsin forest land area were 16.0, 13.0, and 16.9 million acres for FIA, CDL, and NLCD, respectively. On a per-county basis, estimates differed by -63 to 74 percent for FIA and CDL, -48 to 289 percent for FIA and NLCD, and -75 to 54 percent for CDL and NLCD (fig. 2). Wisconsin county comparisons of the three forest land area estimates resulted in area-weighted RMSD values of 8.2 percent for FIA compared with CDL, 4.1 percent for FIA compared with NLCD, and 9.6 percent for CDL compared with NLCD.

Stratified Estimation

Comparing FIA's plot-based estimate using NLCD and CDL for poststratification, the results for number of trees, forest

area, biomass, and volume are all comparable and within one standard error of one another (fig. 3). Comparing the RE produced by the two different stratification layers, results are consistent across all four variables and all five FIA survey units (table 1). Specifically, CDL resulted in higher RE for number of trees, forest area, biomass, and volume in all units, with one exception. The RE for the forest area estimate was substantially higher when using NLCD compared with CDL in unit 2 (fig. 4).

Figure 2.—*Percent difference in estimates of forest land area between Forest Inventory and Analysis and Cropland Data Layer, Wisconsin. Each dot represents a city with population greater than 50,000.*

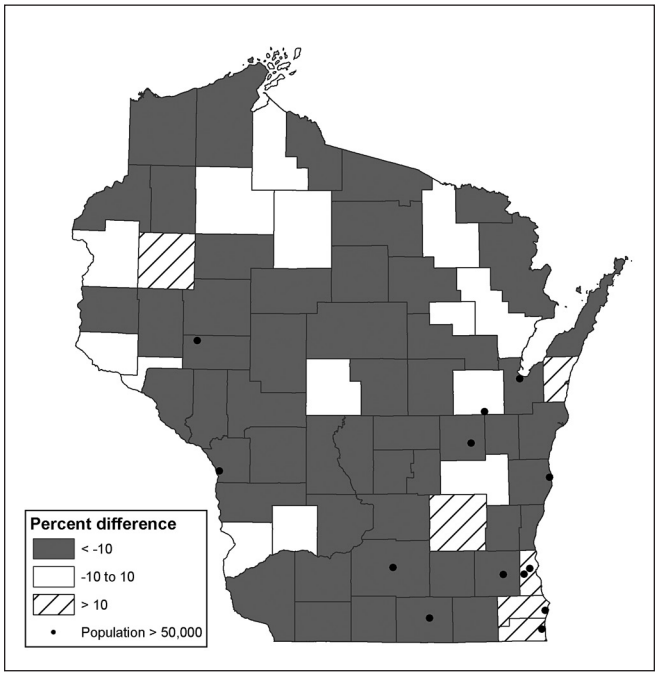


Table 1.—*The relative efficiency of stratified estimates of number of trees, forest area, biomass, and volume in Wisconsin using the 2005 CDL and the 1992 NLCD. The first number in each pair (**bold**) corresponds to the estimate generated using the CDL and the second number corresponds to the estimate generated using the NLCD.*

FIA survey unit (WI)	Relative Efficiency (**CDL**/NLCD)			
	Number of trees	Forest area	Biomass	Volume
1	**1.28**/1.22	**2.46**/2.40	**1.45**/1.36	**1.33**/1.28
2	**1.23**/1.21	**1.59**/2.17	**1.42**/1.34	**1.30**/1.27
3	**1.52**/1.36	**2.31**/2.14	**1.62**/1.44	**1.53**/1.38
4	**1.60**/1.53	**2.48**/2.30	**1.77**/1.68	**1.65**/1.55
5	**1.41**/1.30	**1.96**/1.66	**1.67**/1.47	**1.67**/1.47

CDL = Cropland Data Layer. FIA = Forest Inventory and Analysis. NLCD = National Land Cover Dataset. WI = Wisconsin.

Figure 3.—*Stratified estimates of the number of trees, forest area, biomass, and volume for Wisconsin using both the National Land Cover Dataset and Cropland Data Layer for stratum assignment of Forest Inventory and Analysis plots and weighting of strata.*

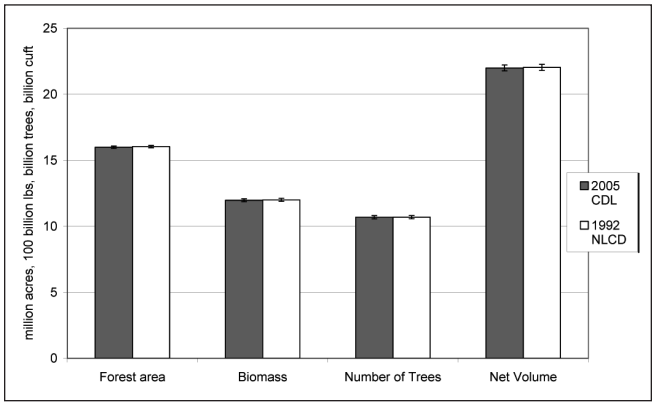
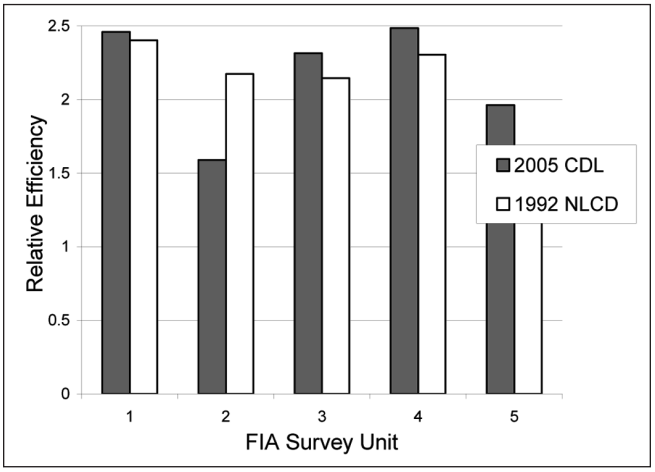


Figure 4.—*The relative efficiency of the stratified estimate of forest area for all Forest Inventory and Analysis survey units in Wisconsin using both the National Land Cover Dataset and Cropland Data Layer stratification layers.*



Conclusions

Compared with the FIA estimate of Wisconsin forest land area, the CDL estimate was 18.7 percent smaller and the NLCD estimate was 5.2 percent larger. Differences in county estimates were much larger, but area-weighted RMSD was only 8.2 percent for FIA compared with CDL and only 4.1 percent for FIA compared with NLCD. The greatest differences appeared in counties with sparse forest cover, especially those counties containing urban tree cover that does not meet FIA's definition of forest land use. Additional geospatial processing steps could separate urban tree cover from other forest land, accounting for these definitional differences.

Although, when compared to NLCD, CDL's pixel-based estimate of forest area is substantially different from the FIA estimate, each land cover data set results in statistically equivalent Wisconsin estimates of the four FIA variables when used as a stratifying layer in the sample-based estimation process. This observation lends support to the notion that FIA's estimation procedures result in an unbiased estimate, even if the stratification layer used is biased. Furthermore, because the CDL outperformed NLCD with respect to RE for the four variables, perhaps we can conclude that agreement in pixel-based forest area with sample-based estimates is not a controlling factor in reducing variance.

Why did the CDL outperform NLCD with respect to variance reduction in most cases? Because of its focus on agricultural lands, the CDL represents nonforest land with good accuracy, resulting in an accurate forest/nonforest interface. We speculate the edge classes created along this interface are more likely to contain highly variable FIA plots compared with the NLCD-derived edge strata. Confining highly variable plots to edge strata, which occupy relatively little area, results in a net reduction in the variance of estimates of primary FIA variables.

In summary, CDL appears to be a viable option for FIA to use in stratified estimation, with several cautions. First, comparisons were only made for data in a single State. Additional work should be done to determine the consistency of CDL across States. Also, because CDL is updated each year, clouds would likely be an issue, and a more robust procedure than was used in this study would have to be implemented to fill this data gap. Finally, if the FIA program is aiming for national consistency, CDL presents a problem in that data is not available for every State. An intradepartmental partnership with the goal of expanding CDL to all States would be necessary to conform to an FIA objective of national consistency.

Literature Cited

Bechtold, W.A.; Patterson, P.L., eds. 2005. The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Craig, M.E. 2001. The NASS cropland data layer program. In: Proceedings, third international conference on geospatial information in agriculture and forestry. Ann Arbor, MI: Veridian. [On CD-ROM].

Häme, T.; Stenberg, P.; Andersson, K.; Rauste, Y.; Kennedy, P.; Folving, S.; Sarkeala, J. 2001. AVHRR-based forest proportion map of the Pan-European area. Remote Sensing of Environment. 77: 76-91.

Homer, C.; Huang, C.; Yang, L.; Wylie, B.; Coan, M. 2004. Development of a 2001 national land-cover database for the United States. Photogrammetric Engineering & Remote Sensing. 70: 829-840.

Loveland, T.L.; Shaw, D.M. 1996. Multiresolution land characterization: building collaborative partnerships. In: Scott, J.M.; Tear, T.; Davis, F., eds. Gap analysis: a landscape approach to biodiversity planning. Proceedings, American Society for Photogrammetry and Remote Sensing/Gap Analysis Program symposium. Moscow, ID: National Biological Service: 83-89.

Miles, P.D. 2001. Forest inventory mapmaker users guide. Gen. Tech. Rep. NC-221. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. http://www.ncrs2.fs.fed.us/4801/fiadb/available_data.html. (6 November 2006).

Mueller, R.; Ozga, M. 2002. Creating a cropland data layer for an entire state. Proceedings, 2002 American congress on surveying and mapping/American Society for Photogrammetry and Remote Sensing conference and technology exhibition. Bethesda, MD: American Society for Photogrammetry and Remote Sensing. [On CD-ROM].

Scott, C.T.; Bechtold, W.A.; Reams, G.A.; Smith, W.D.; Westfall, J.A.; Hansen, M.H.; Moisen, G.G. 2005. Sample-based estimators used by the Forest Inventory and Analysis National Information Management System. In: Bechtold, W.A.; Patterson, P.L., eds. The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 43-67.

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