A COMPARISON OF STRATIFICATION EFFECTIVENESS BETWEEN THE NATIONAL LAND COVER DATA SET AND PHOTOINTERPRETATION IN WESTERN OREGON

Paul Dunham, Dale Weyermann, and Dave Azuma [1]

ABSTRACT.—Stratifications developed from National Land Cover Data (NLCD) and from photointerpretation (PI) were tested for effectiveness in reducing sampling error associated with estimates of timberland area and volume from FIA plots in western Oregon. Strata were created from NLCD through the aggregation of cover classes and the creation of 'edge' strata by reclassifying pixels at class boundaries. Strata were created from aerial photography by interpretation of a sample grid for land use and cover attributes. NLCD-based stratifications are less costly than PI and sacrifice little precision on inventory estimates. Neither PI nor NLCD stratifications achieved the FIA target of 10 percent sampling error/billion cubic feet of volume, but both are near the national standard of 3 percent sampling error/million acres of timberland.

Timberland[2] area and growing-stock volume[3] on timberland are two of the key forest resource estimates that the Forest Inventory and Analysis (FIA) program of the USDA Forest Service is mandated to produce. National accuracy standards are in place for sampling error of timberland area and growing-stock volume. In the Pacific Northwest (PNW), the maximum allowable error for timberland area is 3 percent sampling error per million acres of timberland. For growing-stock volume on timberland in PNW a target of 10 percent sampling error per billion cubic feet "…to be achieved as closely as practicable" is specified (Forest Service Handbook 1967). Due to constraints of time, resources, and budget, it has never been possible to meet these standards in PNW through the use of plot data alone. FIA, both nationally and at PNW, has long relied on two-phase sampling (Cochran 1977) to increase precision of inventory estimates. A relatively inexpensive data collection effort based on remotely sensed data (phase 1) serves as a stratification for expensive-to-collect plot data (phase 2). Stratified sampling is a more cost-efficient approach to reducing the sampling error associated with inventory estimates than increasing sample size.

The following equation converts observed sampling error (in percent) to a percent sampling error per specified area or volume (Hansen 2000).

$$E_S = \frac{(E_O)\sqrt{Estimated\ total\ volume\ or\ area}}{\sqrt{(Specified\ volume\ or\ area)}}, \qquad (1)$$

where,

$$E_O = \frac{(\text{Standard Error})}{\text{Estimate}} \qquad (2)$$

E_S = sampling error in percent for the specified area or volume

When evaluating timberland area, the specified area is 1 million acres and an E_S of < 0.03 is mandated; for growing-stock volume, the specified volume is 1 billion cubic feet and an E_S of < 0.10 is desired.

For this study we evaluated stratifications based on two data sources for phase 1. Evaluation was based on cost and effectiveness in reducing the sampling error of estimates of timberland area and of growing-stock volume on timberland.

[1] Forester, Computer Specialist, and Research Forester, respectively. USDA Forest Service, Pacific Northwest Research Station, P.O. Box 3890, Portland, OR 97208. Phone: 503-808-2024; fax: 503-808-2020; e-mail: pdunham@fs.fed.us

[2] Timberland = Forest land capable of growing 20 cubic feet or more per acre per year at culmination in fully stocked, natural stands of continuous crops of trees to industrial roundwood size and quality.

[3] Growing-stock volume = Net volume in cubic feet of live sawtimber and poletimber trees of commercial species and less than 75 percent cull, from a 12-inch stump to a 4-inch top.

The first phase 1 data source was the traditional FIA approach of using a photointerpreted grid; the second data source was the use of the National Land Cover Data set (NLCD) GIS layer for phase 1. The inventory estimates used for testing were timberland area and growing-stock volume on timberland.

STUDY AREA

The study area is approximately 11 million acres in size and consists of the Oregon counties west of the crest of the Cascade Range plus Hood River County, but excludes lands administered by the Forest Service and Bureau of Land Management. Approximately two-thirds is forested; over 90 percent of the forest is classified as timberland and over 50 percent of the timberland is dominated by Douglas-fir (Pseudotsuga menziesii).

DATA

Phase 2 Data (Field Plots)

For this study 1,449 plots visited by FIA crews between 1995 and 1997 were used. FIA field plots were on a nominal 3.4-mile grid; each plot consists of five subplots dispersed over a 5.2-acre area. At each plot, the proportion of plot area by land use (timberland, non-productive forest, or nonforest) was assessed. Where plots were at least partially forested, attributes such as percent forest, percent timberland, and area by forest type (predominant overstory species) were assessed. Trees occurring on forested subplots were measured.

Phase 1 Data (Photointerpretation)

A 0.84-mile grid of phase 1 photoplots was established on 1994 vintage 1:40,000 scale black and white aerial photo transparency stereo pairs. To ensure that photoplots were established at locations corresponding to their true coordinates, digitizing photogrammetry software was used to compensate for scale and parallax effects (Warner and Carson 1992). Within the study area there were 23,480 photoplots. At each photoplot, attributes such as land use, crown cover, broad forest type (conifer vs. hardwood), stand stage of development, and conifer height classes were assessed for a 5.2-acre circle[4]. Approximately 6 percent of these photoplots were also measured as phase 2 field plots on the ground.

Phase 1 Data (National Land Cover Data Set)

The National Land Cover Data set is a land cover GIS layer developed by the United States Geological Survey (USGS) from leaf-on and leaf-off LANDSAT 5 TM imagery circa 1992. This layer was derived primarily through the unsupervised classification of the imagery and through the use of ancillary spatially referenced data such as elevation, population density, soil attribute maps, and National Wetlands Inventory data. The resulting data set is a land cover map for the conterminous U.S. with 30-m resolution and 21 cover classes (Vogelmann and others 2001). Accuracy assessment was conducted by the USGS based on interpretation of 1990 National Aerial Photography Program photographs. Overall accuracy at Anderson Level 1 classification was 81 percent for the United States east of the Mississippi and Ohio Rivers, but it dropped to 60 percent for the 21 individual classes (Vogelmann and others 2001). Accuracy assessment has not yet been completed for the western U.S.

METHODS

Stratification by Photointerpretation

In the testing of photointerpreted data as the phase 1 sample, photoplots and their associated field plots were grouped into strata based on similarity of photointerpreted attributes. Stratum areas were calculated by assigning fractions of the total sample area in proportion to the number of photoplots within each stratum. Population totals and variances were calculated for timberland area and growing-stock volume using equations for double sampling for stratification (Cochran 1977). It should be noted that this stratification technique does not produce a map of the strata; the size of the strata can be calculated but their boundaries cannot.

Two stratifications were developed and tested from the photointerpreted phase 1 data. The first stratification (PI-Area) was designed to provide the most precise estimate of timberland area. Each photoplot was assigned to a stratum based on the proportion of the photoplot that was interpreted to be timberland. Twelve strata were created in this manner, ranging from 0 percent to 100 percent timberland. The

[4] A procedures manual entitled "Instructions for Western Oregon Primary Sample Data Collection" is on file with the USDA Forest Service, Pacific Northwest Research Station, Portland, OR.

second stratification (PI-Volume) was designed to provide the most precise estimate of growing-stock volume. Photoplots were assigned to strata based on interpretation of land use (timberland, non-productive forest, or nonforest), broad forest type (conifer, hardwood, or nonstocked), crown closure, conifer heights, and stage of stand development (nonstocked, seedling-sapling, poletimber, or sawtimber).

Stratification by NLCD

When testing NLCD as the phase 1 sample, we made several changes to the GIS layer to match FIA land use definitions and to eliminate unnecessary classes. Of the 21 NLCD classes, 6 corresponded generally to what FIA would define as forest in western Oregon: deciduous forest, evergreen forest, mixed forest, woody wetlands, shrubland, and transitional. The other NLCD classes were aggregated into a single nonforest class. The shrubland and transitional classes were included as possible forest because these areas are often recent clearcuts where forest is expected to regenerate. Three groupings of these classes were tested as the basis for stratification. The first grouping (forest/nonforest) was a simple forest versus nonforest two-class scheme where all six potential forest classes were combined into the forest class. The second grouping (forest/other-forest/nonforest) combined the classes most likely to be timberland (deciduous forest, evergreen forest, and mixed forest) into the forest class; the transitional, shrubland, and woody wetland classes were combined into an other forest class (three classes). The third grouping (deciduous/evergreen/mixed) used deciduous forest, evergreen forest, mixed forest, other forest, and nonforest as separate classes (five classes).

Areas smaller than FIA's minimum mapping requirements (less than an acre, equivalent to four 30-m pixels of the same class) were re-assigned to adjacent classes. This was accomplished using the clump, sieve, and neighborhood functions of ERDAS Imagine (ERDAS 1997).

Following Hansen (2000), we investigated the effectiveness of using edge strata for each of the three class groupings. Edge strata were created at the margins of each class using a search function (ERDAS 1997). In cases where registration or coordinate errors cause a field plot to be assigned to the wrong stratum or where a field plot overlaps class boundaries, the anomalous value contributed by that plot can substantially increase the variance of the stratum associated with that plot. Edge strata allow these field plots to be isolated within smaller strata where they will have less impact on the overall variance. The width of edge strata tested ranged from 2 to 4 pixels (60 to 120 m).

Stratum sizes were calculated by summing pixel counts within each stratum with each pixel counting as 0.22239 acres. Population totals and variances were calculated for timberland area and growing-stock volume using equations for stratified estimation (Cochran 1977).

No Stratification

As a basis for comparison, timberland area and growing-stock volume were also summarized using no stratification. Population totals and variances were calculated using equations for a simple random sample (Cochran 1977).

RESULTS

Timberland Area

Both photointerpretation stratifications met the FIA standard for 3 percent sampling error per 1 million acres of timberland. Performance of the PI-Volume and PI-Area stratifications was nearly equal: each reduced the sampling error by just over 40 percent compared to a random sample (no stratification). In this situation, stratification designed for volume on timberland (PI-Volume) was also effective as a stratification for timberland area (table 1).

None of the stratifications based on NLCD met the 3 percent standard for sampling error; however, the best performers came quite close to the target at 3.2 percent, a sampling error 34 percent lower than the random sample (fig. 1). When edge strata were not used, all three groupings of the forest classes performed about equally with sampling errors of 3.5 percent per million acres. The addition of edge strata improved the performance of the forest/nonforest and the forest/other forest/nonforest groupings, with 4-pixel edges having more benefit than 2-pixel edges. Edge strata had little effect on the deciduous/evergreen/mixed grouping (table 1).

Growing-Stock Volume

Neither photointerpretation based stratification met FIA's 10 percent sampling error per billion cubic feet target, but the stratification intended for volume (PI-Volume) was close at

Table 1.—*Timberland area in western Oregon by stratification technique (outside national forests and Bureau of Land Management lands)*

Stratification method	Area in millions of acres	Standard error in millions of acres	Sampling error per million acres
No stratification	6,981	128	4.9%
Forest/Nonforest no edge	6,887	93	3.5%
Forest/Nonforest 4-pixel edge	6,982	85	3.2%
Forest/Other forest/Nonforest no edge	6,891	92	3.5%
Forest/Other forest/Nonforest 4-pixel edge	6,991	85	3.2%
Deciduous/Evergreen/Mixed no edge	6,895	91	3.5%
Deciduous/Evergreen/Mixed 4-pixel edge	6,955	89	3.4%
PI-Area	7,034	72	2.7%
PI-Volume	7,013	73	2.8%

12.9 percent sampling error per billion cubic feet. The PI-Volume stratification had a sampling error 22 percent lower than that of a random sample (fig. 1).

None of the stratifications based on NLCD met the 10 percent standard, with the best performing NLCD stratification, at 14.1 percent, a little worse than PI-Volume. Unlike NLCD stratification for timberland area, less aggregation of the forest classes led to better results when stratifying for volume (table 2). The deciduous/evergreen/mixed grouping worked well for volume stratification, most likely because of the large difference in growing-stock volume between conifer and hardwood stands represented by the evergreen and deciduous forest classes, and the inclusion of recently clearcut timberland in the other forest class. The addition of edge strata improved the performance of all three of the forest groupings with the most benefit coming from the widest edge strata. The best performing NLCD stratification for volume was the

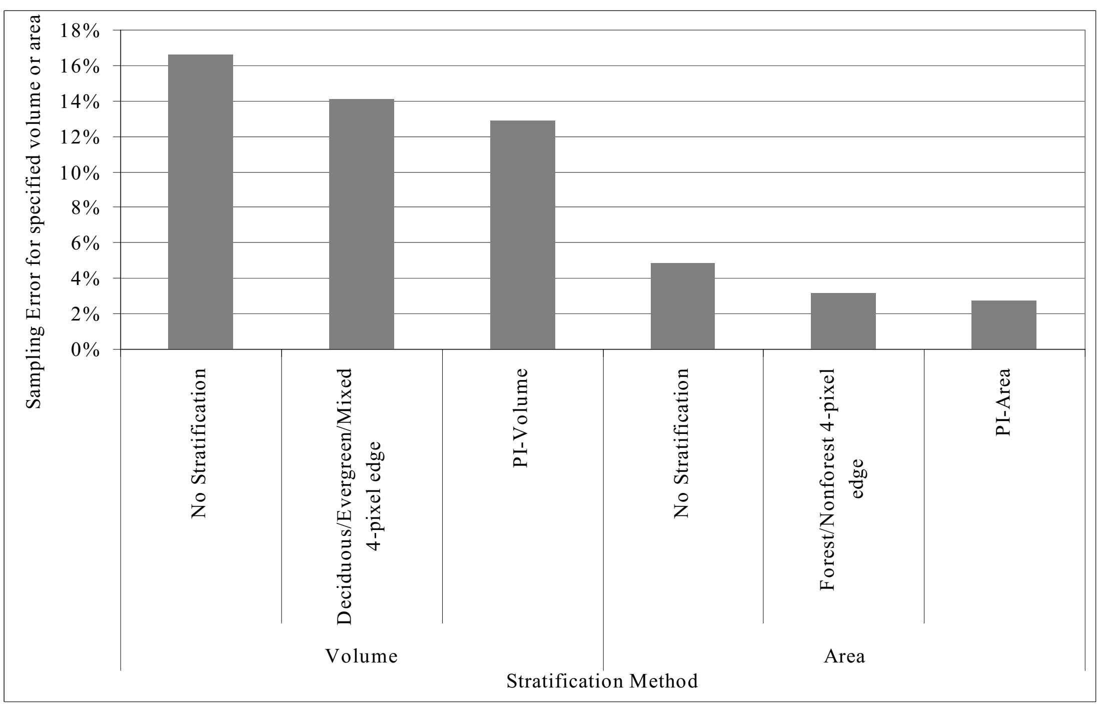

Figure 1.—*Sampling error per million acres of timberland or billion cubic feet of growing-stock volume for no stratification, best NLCD stratifications, and best photointerpreted stratifications.*

Table 2.—*Growing-stock volume in western Oregon by stratification technique (outside national forests and Bureau of Land Management lands)*

Stratification method	Volume in million cubic feet	Standard error in million cubic feet	Sampling error per billion cubic feet
No stratification	20,046	742	16.6%
Forest/Nonforest no edge	19,729	668	15.0%
Forest/Nonforest 4-pixel edge	20,055	648	14.5%
Forest/Other forest/Nonforest no edge	19,743	658	14.8%
Forest/Other forest/Nonforest 4-pixel edge	20,207	637	14.2%
Deciduous/Evergreen/Mixed no edge	19,874	657	14.7%
Deciduous/Evergreen/Mixed 4-pixel edge	20,046	632	14.1%
PI-Area	20,150	628	14.0%
PI-Volume	20,440	582	12.9%

deciduous/evergreen/mixed grouping with 4-pixel edges. This stratification achieved a 15-percent reduction in sampling error compared to a random sample (fig.1).

DISCUSSION

Although stratification using NLCD brings less reduction in sampling error than stratification using photointerpretation, it can be accomplished at a fraction of the cost (table 3). The photointerpretation of western Oregon required a large investment to purchase 2,400 aerial photos, establish the sample points on the photos, develop sampling protocols, and interpret the photography. The NLCD approach achieves nearly as much reduction in sampling error at one-twentieth the cost of stratification by photointerpretation. At least for western Oregon, NLCD appears to be a very attractive stratification alternative to traditional methods.

Other factors besides cost and precision can influence choice of stratification approach. Aerial photography has a much finer resolution than NLCD and can be used for applications other than stratification. For example, PNW-FIA has used the photography purchased for stratification as a data source for analysis of land development over time, including change in structure counts on photography from three decades (Azuma and others 1999). The NLCD approach works reasonably well as a stratification tool in western Oregon where the vast

Table 3.—*Cost comparison of stratification methods*

Stratification method	Estimated cost per million acres	Study area size in acres	Total cost to stratify study area
Photointerpretation			
Photo Acquisition	$ 1,945		
Photo Setup	$ 14,140		
Photointerpretation	$ 2,203		
Total	$ 18,288	11,090,003	$ 202,814
NLCD			
Reproject, mosaic & mask	$ 251		
Classify/post-process filtering/edging	$ 503		
Administration/coordination	$ 101		
Total	$ 854	11,090,003	$ 9,471

majority of forest land has a high degree of tree crown cover. It is unlikely that NLCD alone would work well in areas where forest typically has low crown cover that may go undetected by satellite imaging. For example, of the 2.8 million acres of juniper forest in eastern Oregon nearly half has crown cover less than 20 percent (Gedney and others 1999). The cost advantages of the current NLCD version are due to the high degree of subsidy in its development and the sharing of cost between agencies. As new versions are developed (one is planned based on year 2000 imagery), FIA may be asked to bear more of the cost of its development. The cost of stratification using NLCD may increase with newer versions but is not expected to approach the cost of photo-interpretation. At this time it is unknown how long the NLCD product based on 1992 imagery will remain an effective Phase 1 data source.

LITERATURE CITED

Azuma, D.L.; Birch, K.R.; DelZotto, P.; Herstrom, A.A.; Lettman, G.L. 1999. Land use change on non-federal land in western Oregon, 1973-1994. Salem, OR: Oregon Department of Forestry. 55 p.

Cochran, W.G. 1977. Sampling techniques. 3d ed. New York, NY: John Wiley and Sons. 428 p.

ERDAS, Inc. 1997. ERDAS field guide, 4th ed. Atlanta, GA: ERDAS Inc. 656 p.

USDA Forest Service. 1967. Forest Service Handbook. FSH 4813.1. Washington, DC: U.S. Department of Agriculture, Forest Service. 128 p.

Gedney, D.R.; Azuma, D.L.; Bolsinger, C.L.; McKay, N. 1999. Western juniper in eastern Oregon. Gen. Tech. Rep. PNW-464. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 53 p.

Hansen, M.H. 2001. Remote sensing precision requirements for FIA estimation. In: Reams, G.A.; McRoberts, R.E.; Van Deusen, P.C., eds. Proceedings of the Second annual fourest inventory and analysis symposium. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 43-51.

Vogelmann, J.E.; Howard, S.M.; Yang, L.; Larson, C.R.; Wylie, B.K.; Van Driel, N. 2001. Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering and Remote Sensing. 67(6): 650-661.

Warner, W.S.; Carson, W.W. 1992. MDSD guide. As, Norway: JORDFORSK Center for Soil and Environmental Research. 1432. 104 p.