IMPLEMENTING A LAND COVER STRATIFICATION ON-THE-FLY

Ronald E. McRoberts and Daniel G. Wendt[1]

ABSTRACT.—Stratified estimation is used by the Forest Inventory and Analysis program of the USDA Forest Service to increase the precision of county-level inventory estimates. Stratified estimation requires that plots be assigned to strata and that proportions of land area in each strata be determined. Classified satellite imagery has been found to be an efficient and effective means of accomplishing both tasks. This study investigated methods for making the county stratifications available for use with user-defined areas without storing the entire pixel-level stratification and processing it separately for each application. Summaries of stratifications at spatial scales of 6,000 acres and less produced stratified estimates at county level and for user-defined areas ranging in size from approximately 300 square miles to approximately 8,000 square miles that were very similar to stratified estimates obtained using the underlying pixel-level stratifications.

The Forest Inventory and Analysis (FIA) program of the USDA Forest Service reports county-level estimates of the means and standard errors for forest land area and live tree volume per unit area for each state. Budgetary constraints and the natural variability among plots prohibit sample sizes sufficient to satisfy national precision standards (USDA-FS 1970) unless the estimation process is enhanced using ancillary data. The FIA program enhances the process using stratified estimation with counties stratified using remotely sensed data. Stratified estimation requires that two tasks be accomplished: each observational unit must be assigned to a stratum, and strata weights must be calculated as the proportions of the area of interest in strata. When satellite imagery is used as the means of stratification, observational units are assigned to strata on the basis of the strata assignments of their associated satellite image pixels, and strata weights are calculated as the proportions of image pixels assigned to strata.

Both FIA and non-FIA users frequently require estimates of inventory variables for areas other than counties, and if the FIA county-level stratifications can be used, the precision of estimates for these user-defined areas will be increased. If storage space and processing time were not constraints, the FIA database (FIADB) (Miles and others 2001) would provide access to FIA's pixel-level strata assignments so that users could stratify their areas of interest in the same manner as FIA stratifies counties. However, because storage space and processing time are constraints, this approach is not feasible at the current time. Therefore, the FIA program seeks methods for summarizing county stratifications and making them available to users without the intense storage and processing requirements.

The objective of this study is to compare stratified estimates obtained using pixel-level stratifications with estimates obtained using summaries of pixel-level stratifications. The comparisons focus on estimates of plot-level means and standard errors for proportion forest land area and volume per unit area using 1998 FIA data for the state of Indiana. Among the observations field crews obtain when measuring plots are individual tree diameters and heights and the proportions of plot areas that satisfy specific ground land use conditions. The tree diameter and height observations are used as predictor variables to obtain model predictions of individual tree volumes. Plot-level volume is obtained by adding the volume predictions over all trees and scaling the sum to a per unit area basis. Plot-level forest land proportions are obtained by aggregating ground land use conditions consistent with the FIA definition of forest land.

[1] Mathematical Statistician, North Central Research Station, U.S. Department of Agriculture, Forest Service, St. Paul, MN 55108. Phone (651) 649-5174; fax: (651) 649-5285; e-mail: rmcroberts@fs.fed.us; and Computer Specialist, R9, U.S. Department of Agriculture, Milwaukee, WI.

STRATIFIED ESTIMATION

The National Land Cover Data (NLCD) is a satellite image-based classification that has been used by the FIA program as a basis for stratifications. 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 is based on nominal 1992 Landsat 5 Thematic Mapper (TM) satellite imagery and a variety of ancillary data. Vogelmann and others (2001) provide an excellent overview and discussion of the NLCD.

FIA obtains stratifications from the NLCD using a three-step process. First, NLCD classes 33 (transitional)[2], 41 (deciduous forest), 42 (evergreen forest), 43 (mixed forest), 51 (shrubland), and 91 (woody wetland) are aggregated into a forest stratum, and the remaining classes are aggregated into a nonforest stratum. Second, a clumping and sieving algorithm (ERDAS 1997) is used to reassign isolated groups of small numbers of contiguous forest pixels to the nonforest stratum and isolated groups of small numbers of contiguous nonforest pixels to the forest stratum. Groups of less than 4 pixels are reassigned because of the approximate correspondence of their aggregated area of 38,750 ft^2 to 1 acre, the minimum area necessary to be designated FIA forest land. Third, two additional strata are created by subdividing the forest stratum into forest and forest edge strata and by subdividing the nonforest stratum into nonforest and nonforest edge strata (Hansen and Wendt 2000, McRoberts and others 2001). These edge strata are created by assigning pixels in the original forest stratum within 2 pixels of the forest/nonforest boundary to a forest edge stratum and by assigning pixels in the original nonforest stratum within 2 pixels of the forest/nonforest boundary to a nonforest edge stratum.

The rationale for creating the edge strata is based on the knowledge that stratification contributes to increasing precision under two conditions: when within-stratum variances are smaller than the overall variance, and/or when strata with large variances represent relatively small proportions of the total population. Plots located in the interior of the forest stratum and away from forest/nonforest boundaries are expected to be predominantly forest, while plots located in the nonforest stratum away from the forest/nonforest boundaries are expected to be predominantly nonforest. In both cases, within-stratum variances are expected to be relatively small, thus satisfying the first condition for which stratification increases precision. Plots located in an edge stratum, i.e., near forest/nonforest boundaries, are expected to exhibit greater variances due to a mix of forest and nonforest conditions and due to the greater probability of errors in assigning plots to strata as a result of plot location errors. Nevertheless, the stratification is expected to concentrate these mixed forest/nonforest and erroneously stratified plots into strata that represent relatively small proportions of the total area, thus satisfying the second condition for which stratification produces increases in precision.

The FIA program assigns plots rather than subplots to strata to avoid the mathematical complexities necessary to accommodate the correlations among subplot observations. Nevertheless, assigning FIA plots to strata may not be trivial, because each plot is covered by multiple TM pixels. For the sake of simplicity, each plot is assigned to the stratum of the pixel corresponding to the plot center for this study. Strata weights are then calculated as the proportions by strata of pixels with centers in the county. For future reference, this stratification method is designated the pixel method and is the standard of comparison for other methods.

Stratified estimates of means and variances means are obtained using formulae from Cochran (1977)

$$\bar{x}_{str} = \sum_{h=1}^{H} w_h \bar{x}_h, \qquad (1)$$

and

$$Var(\bar{x}_{str}) = \sum_{h=1}^{H} w_h^2 \frac{\hat{\sigma}_h^2}{n_h}, \qquad (2)$$

where

$$\bar{x}_h = \sum_{k=1}^{n_h} x_{hk}, \qquad (3)$$

$$\hat{\sigma}_h^2 = \frac{\sum_{k=1}^{n_h} \left(x_{hk} - \bar{x}_h \right)^2}{n_h - 1}, \qquad (4)$$

[2] The correct numerical designation for the transitional class is 33; its designation as 31 in Vogelmann and others (2001) is attributed to a manuscript error (Vogelmann, EROS Data Center, U.S. Geological Survey, personal communication, 10 October 2001).

$\overline{x}_{str}$ denotes the stratified mean, h=1,...,H denote strata, w_h denotes the h^{th} stratum weight, n_h denotes the number of plots assigned to the h^{th} stratum, k=1,...,n_h indexes observations within the h^{th} stratum, $\overline{x}_h$ denotes the mean for the h^{th} stratum, and $\hat{\sigma}_h^2$ denotes the variance for the h^{th} stratum.

STRATIFICATION SUMMARIES

The traditional FIA approach to summarizing county-level stratifications is via area expansion factors. For the h^{th} stratum in the c^{th} county, the area expansion factor, F_{ch}, is calculated as,

$$F_{ch} = \frac{A w_{ch}}{n_{ch}} \qquad (5)$$

where A is the total county area and w_{ch} and n_{ch} are the stratum weight and the number of plots, respectively, for the h_{th} stratum in the c^{th} county. F_{ch} expresses the land area in the c^{th} county represented by each plot in the h^{th} stratum and is assigned to all plots in that county-stratum combination. The area expansion factors provide a crude summary of county stratifications. However, it is not necessarily the case that $F_{c'h}=F_{c''h}$ where c' and c" designate different counties. To calculate stratified estimates using area expansion factors, strata assignments of plots from the county stratification are used, and strata weights are estimated as ratios of the sum of area expansion factors for all plots assigned to strata and the sum of area expansion factors for all plots over all strata. Within-strata means and variances are calculated using (4) and (5) except that each plot observation is weighted by the appropriate F_{ch}. This method for summarizing a stratification is designated the area expansion factor method.

Stratifications may also be summarized by counting and storing the numbers of pixels by strata for summary units at a variety of spatial scales. For example, beginning with an arbitrarily selected starting point, the State of Indiana was covered by a set of non-overlapping square kilometers, the numbers of pixels with centers in each square kilometer were counted by strata, and the strata pixel counts for each square kilometer were stored for future use. This method for summarizing a stratification is designated the square kilometer method.

The FIA sampling design is based on an array of regular hexagons covering the conterminous United States. Each hexagon includes 5,937.2 ac and contains at least one permanent field (fig. 1). This network of plots is designated the Federal base sample and provides complete, consistent coverage of all lands in the United States. To facilitate intensification of the sample for some states, the hexagons have been subdivided into three parallelograms designated subhexagons (fig. 2). Stratifications are summarized at the subhexagon and hexagon level in the same manner as for square kilometers and are designated the subhexagon and hexagon methods, respectively.

Figure1.–*Indiana FIA hexagons.*

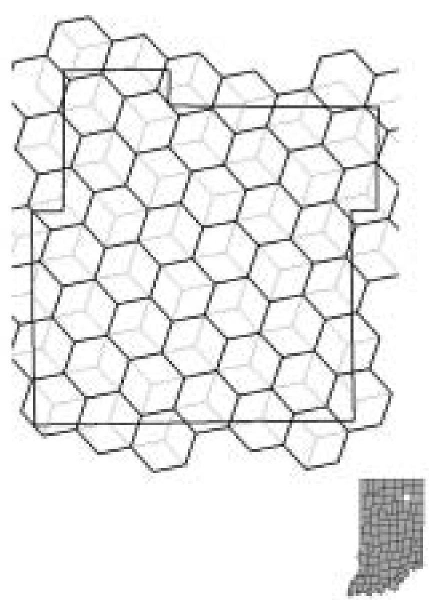

Figure 2.–*FIA hexagons and sub-hexagons for Whitley County, Indiana.*

STRATIFIED ESTIMATION USING STRATIFICATION SUMMARIES

To obtain stratified estimates for user-defined areas with stratification summaries, the same two stratification tasks must be accomplished: plots must be assigned to strata, and strata weights must be calculated. The first task is easily accomplished by using the same strata assignments as for the county stratifications. Several approaches may be used to accomplish the second task. One approach is to select the summary units with centers in the user-defined area, add the pixel counts by strata over these units, and calculate strata weights as the proportions of pixels assigned to strata. This approach, based on including or excluding a summary unit's pixel counts in the overall total on the basis of whether the center of the unit is inside the user-defined area, is designated the center approach and is used in combination with the square kilometer, subhexagon, and hexagon summary units.

The pixel counts by strata using the center approach will not be the same as those obtained using the pixel method, because the exterior boundaries of the spatially aggregated summary units will not coincide exactly with the boundaries of the user-defined area. To simplify the discussion that follows, two categories of summary units are distinguished: interior summary units are wholly within the user-defined area, while boundary summary units are only partially within the user-defined area. Boundary units cause two kinds of errors in pixel counts for a user-defined area. Boundary units with centers inside the user-defined area cause pixel counts by strata over all units to include counts for some pixels outside the user-defined area, and boundary units with centers outside the user-defined area cause pixel counts by strata over all units to exclude counts for some pixels inside the user-defined area. However, stratified estimation does not require pixel counts by strata, because strata weights are calculated as proportions of pixel counts. Thus, even though counts for some pixels are erroneously included and some are erroneously excluded in the pixel counts by strata over summary units, the strata weights may still be approximately correct.

Two additional approaches are investigated. The first additional approach includes pixel counts for boundary summary units on the basis of whether any part of the unit is inside the user-defined area. This approach is designated the exterior approach, because the boundaries of the spatially aggregated summary units coincide with or are exterior to the boundaries of the user-defined area. The exterior approach counts all pixels in boundary units, regardless of whether the unit center is inside or outside the user-defined area. The exterior approach is used only in combination with hexagon summary units. The corresponding interior approach is not investigated because of the risk of excluding so many pixels that the counts by strata will not accurately represent the user-defined area. This risk is particularly problematic for user-defined areas with narrow components for which the ratio of interior boundary units to boundary summary units is small.

The second additional approach attempts to compensate for pixel inclusion and exclusion errors. With this approach, the pixel counts by strata for boundary summary units are adjusted by multiplying them by the proportion of the unit in the user-defined area. The additional computation necessary to determine the proportion for each boundary unit may produce more accurate pixel counts and hence more accurate strata weights. This approach is designated the proportional approach. Care should be taken to distinguish the proportional approach from an approach that obtains pixel counts by strata for boundary units by first determining the portion of the unit in the user-defined area and then counting only pixels in the selected portion. Although this is exactly the approach that would be used under ideal conditions, it is also exactly the storage- and processing-intensive approach for which this study seeks alternatives. The proportional approach is used only in combination with hexagon summary units.

ANALYSES

Two sets of analyses are used to compare stratified estimates obtained using the pixel stratification method with estimates obtained using stratification summary methods. For both sets of analyses, the comparisons focus on estimates of plot-level means and standard errors for proportion forest land area and volume per unit area in Indiana. The purpose of the first set of analyses is to evaluate the bias in the stratified mean and standard error estimates that result from summarizing stratifications using the square kilometer, subhexagon, and three hexagon methods. Bias was evaluated using the distributions over the 92 Indiana counties of three statistics: (1) the relative bias, calculated as the ratio of the difference in means for the pixel method and each of the summarization methods and the mean for the pixel method; (2) the t-statistic, calculated as the ratio of the difference in means for the pixel method and each of the summary methods and the standard error of the mean for the pixel method; and (3) the ratio of relative efficiencies for the pixel method and each of the summary methods where relative efficiency is the ratio of the variance of the mean assuming simple random sampling and the variance obtained using stratified estimation. The area expansion factor method was not used for these analyses, because it produces the same estimates for counties as does the pixel method.

The second set of analyses compares methods for areas enclosed in two sets of concentric circles of radii 10, 20, 30, 40, and 50 mi. One set is centered in a sparsely forested area in northern Indiana, while the other is centered in a more heavily forested area in southern Indiana that includes the Hoosier National Forest (fig. 3). The purpose of this set of analyses is to compare estimates of the means and standard errors for areas that mimic user-defined areas. Circular areas of radius 10 miles include 314.2 mi^2, and with a sampling intensity of one plot for

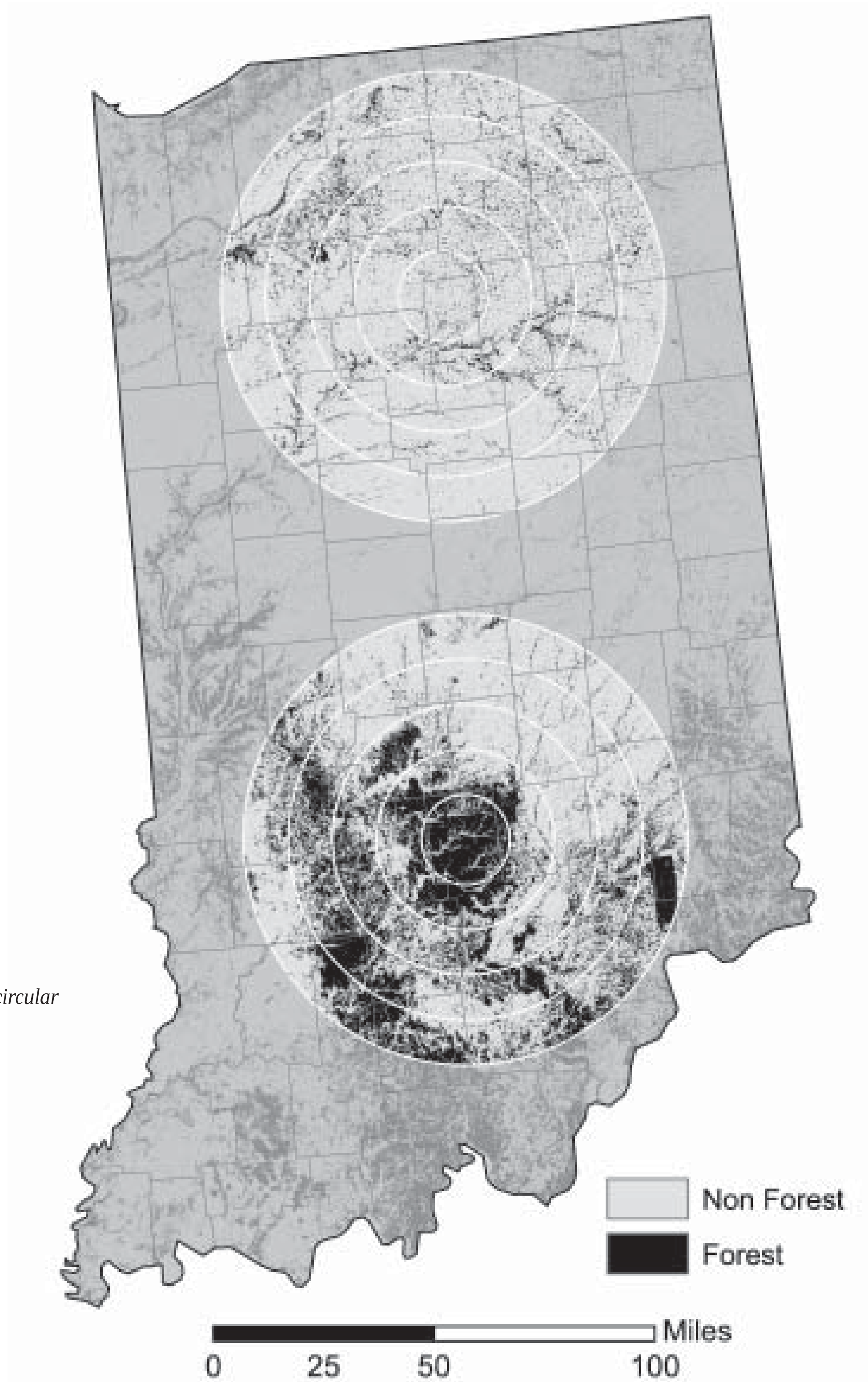

Figure 3.–*User-defined circular areas for Indiana.*

approximately every 5,937.2 ac, are expected to include approximately 34 plots. Assuming moderate variability in the number of plots per stratum, circular areas of radius 10 mi are about the smallest areas that ensure five plots per stratum, the FIA minimum.

RESULTS

The first set of analyses indicate that bias for the square kilometer, subhexagon, and three hexagon methods is minimal (tables 1a and 1b). As expected, bias is less for methods that

Table 1a.—*Volume per unit area results over the 92 Indiana counties*

Statistic	km^2	Subhexagon	Method center	Hexagon proportional	Exterior
Relative bias					
Maximum	0.017	0.028	0.058	0.084	0.291
Mean	0.000	-0.002	-0.005	-0.002	0.016
Minimum	-0.019	-0.0036	-0.083	-0.081	-0.126
t-statistic					
Maximum	0.049	0.150	0.227	0.239	1.258
Mean	-0.002	-0.010	-0.027	-0.015	0.037
Minimum	-0.088	-0.141	-0.367	-0.386	-1.071
Relative efficiency					
Maximum	1.039	1.079	1.170	1.179	1.314
Mean	0.990	0.993	0.997	0.995	0.980
Minimum	0.960	0.945	0.886	0.844	0.606

Table 1b.—*Proportion forest land area results over the 92 Indiana counties*

Statistic	km^2	Subhexagon	Method center	Hexagon proportional	Exterior
Relative bias					
Maximum	0.012	0.030	0.058	0.059	0.276
Mean	0.000	-0.002	-0.005	-0.002	0.016
Minimum	-0.019	-0.0035	-0.074	-0.077	-0.121
t-statistic					
Maximum	0.152	0.191	0.250	0.361	2.152
Mean	-0.002	-0.021	-0.048	-0.026	0.074
Minimum	-0.134	-0.255	-0.590	-0.641	-1.122
Relative efficiency					
Maximum	1.039	1.070	1.143	1.159	1.261
Mean	0.990	0.994	0.995	0.995	0.985
Minimum	0.978	0.944	0.894	0.917	0.681

summarize at smaller spatial scales. Among the three approaches used with hexagon summary units, the hexagon-exterior results are clearly inferior, while results obtained with the hexagon-center and hexagon-proportional methods are similar. The distributions of t-statistics indicate that differences between means obtained using the pixel method and means obtained using the summary methods are never statistically significant ($\alpha = 0.05$). It is noted that the denominator used in calculating this t-statistic includes the standard error of the mean obtained with the pixel method only. No account is made of the uncertainty associated with the standard error obtained with the stratification summary method. If the latter were included, the t-statistics would be even less statistically significant. The ratios of relative efficiencies indicate that, on average, virtually no precision is lost with the stratification summaries. Even with the inferior hexagon-exterior method, the relative efficiencies are always ±18 percent of the relative efficiency obtained with the pixel method. Because of its inferiority relative to the other two hexagon methods, the hexagon-exterior method is not discussed further. The hexagon-center and the hexagon-proportional methods produce similar results, but the hexagon-center method is preferable because of its less intense processing requirements. Therefore, only the square kilometer, sub-hexagon, and hexagon-center methods are further evaluated.

The second set of analyses simulated user-defined areas and yielded two important findings: first, results for the square kilometer, subhexagon, and hexagon-center summary methods are similar and are similar to results obtained with the pixel method, which is the standard of comparison; and second, results for the area expansion factor method are inferior to results obtained with all the other methods (tables 2a and 2b). Results obtained with the spatial summary methods are slightly more similar to the results with the pixel method for smaller scale summary units. For practical purposes, however, there is little to distinguish among results for these three summary methods. These similarities hold for estimates of both means and standard errors.

CONCLUSIONS

Three conclusions may be drawn from this study. First, the traditional FIA area expansion factor method should be abandoned because it is inferior to the other methods. Second, results obtained with the square kilometer, subhexagon, and hexagon-center methods were all acceptable approximations to the results obtained with the pixel stratifications. Third, the hexagon-center is the overall preferred method: it produces highly satisfactory results relative to the standard of comparison, the pixel method; it has direct linkages to the FIA national sampling design; and it requires the least storage and processing time of all the spatial summary methods considered. The quality of results for the hexagon-center method suggests that it should be investigated in other areas with different topographies, tree species, and forest management practices.

The necessity of summarizing stratifications is a temporary phenomenon that will be alleviated as the costs of computer storage and processing decrease. Nevertheless, it is worth noting the difference in storage requirements for the underlying stratification and a summary of the stratification at the hexagon level. The storage requirement for summarizing a stratification at the hexagon level is four cells, one for each of the four stratum pixel counts. For the underlying stratification, one cell is required for each of the 26,696 30 m x 30 m TM pixels contained in a hexagon of 5,937.2 ac. The magnitude of this ratio, 26,696:4, which is also an approximation of the factor by which computer processing requirements may be reduced, cannot be ignored.

Finally, one caveat must be noted. These results are predicated on the assumption that the proportions of pixels by strata obtained using stratification summaries will be approximately the same as for the underlying pixel stratification. The degree to which this assumption is valid is expected to be directly proportional to the ratio of interior summary units to boundary summary units. Additional investigations are necessary to determine the ratio at which the assumption is no longer sufficiently valid to produce acceptable results.

LITERATURE CITATIONS

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

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

Table 2a.—*Volume per unit area for Indiana circular areas*

Method	Radius					
	10 mi		30 mi		50 mi	
	Mean ±	SE	Mean ±	SE	Mean ±	SE
Northern area						
Area Expansion Factor	154.8 ±	24.7	166.6 ±	17.1	160.1 ±	13.5
Pixel	159.7 ±	26.2	215.2 ±	21.4	203.0 ±	18.4
km^2	159.3 ±	26.2	215.2 ±	21.4	203.0 ±	18.4
Subhexagon	161.6 ±	26.2	214.0 ±	21.3	203.0 ±	18.4
Hexagon-center	163.4 ±	26.4	214.1 ±	21.4	203.6 ±	18.4
Southern area						
Area Expansion Factor	1,258.7 ±	172.8	741.0 ±	39.3	591.5 ±	19.9
Pixel	1,122.5 ±	241.5	767.1 ±	40.2	628.6 ±	20.8
km^2	1,122.9 ±	241.2	766.9 ±	40.2	628.8 ±	20.8
Subhexagon	1,117.9 ±	239.0	764.0 ±	40.1	629.3 ±	20.9
Hexagon-center	1,113.9 ±	243.7	768.1 ±	40.3	628.9 ±	20.8

Table 2b.—*Proportion forest land for Indiana circular areas*

Method	Radius					
	10 mi		30 mi		50 mi	
	Mean ±	SE	Mean ±	SE	Mean ±	SE
Northern area						
Area Expansion Factor	0.010 ±	0.022	0.089 ±	0.007	0.095 ±	0.005
Pixel	0.010 ±	0.024	0.115 ±	0.007	0.115 ±	0.005
km^2	0.010 ±	0.024	0.115 ±	0.007	0.115 ±	0.005
Subhexagon	0.010 ±	0.024	0.114 ±	0.007	0.115 ±	0.005
Hexagon-center	0.010 ±	0.025	0.114 ±	0.007	0.115 ±	0.005
Southern area						
Area Expansion Factor	0.729 ±	0.065	0.389 ±	0.014	0.319 ±	0.007
Pixel	0.708 ±	0.104	0.394 ±	0.015	0.328 ±	0.007
km^2	0.708 ±	0.103	0.394 ±	0.015	0.328 ±	0.007
Subhexagon	0.704 ±	0.102	0.392 ±	0.015	0.328 ±	0.007
Hexagon-center	0.705 ±	0.104	0.394 ±	0.015	0.328 ±	0.007

Hansen, M.H.; Wendt, D.G. 2000. Using classified Landsat Thematic Mapper data for stratification in a statewide forest inventory. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C., eds. Proceedings of the first annual forest inventory and analysis symposium. Gen. Tech. Rep. NC-213. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 20-27.

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 of the ASPRS/GAP Symposium; Charlotte, NC. Moscow, ID: National Biological Service: 83-89.

McRoberts, R.E.; Wendt, D.G.; Nelson, M.D.; Hansen, M.H. 2001. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. Remote Sensing of Environment. 80: 1-9

Miles, P.D.; Brand, G.J.; Lerich, C.L.; Bednar, L.F.; Woudenberg, S.W.; Govler, J.F.; Ezzell, E.N. 2001. The forest inventory and analysis database: description and users' manual version 1.0. Gen. Tech. Rep. NC-218. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 130 p.

USDA Forest Service (USDA-FS). 1970. Operational Procedures, Forest Service Handbook 4809.11, Chapter 10:11-1. Washington, DC: U.S. Department of Agriculture, Forest Service,

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: 650-662.

.