Mapping Forest Inventory and Analysis Data Attributes Within the Framework of Double Sampling for Stratification Design

David C. Chojnacky[1], Randolph H. Wynne[2], and Christine E. Blinn[3]

Abstract.—Methodology is lacking to easily map Forest Inventory and Analysis (FIA) inventory statistics for all attribute variables without having to develop separate models and methods for each variable. We developed a mapping method that can directly transfer tabular data to a map on which pixels can be added any way desired to estimate carbon (or any other variable) for a particular area. The method uses a remote-sensing layer coverage to stratify forest from nonforest, a standard double sampling analysis, and a definition of spatial variables for summarizing FIA data within the double sampling framework to link to maps. Further refinements are needed but the new method appears to be accurate, flexible, and straightforward.

Introduction

Forest Inventory and Analysis (FIA) inventories provide population statistics with appropriate sampling errors for many forest attributes according to the underlying sample design. Often, these statistics are presented in tabular format for categories of interest such as geographic areas (States or counties), landownerships, or ecological classifications. FIA's statistical design was developed long before today's explosion in remote-sensing and Geographic Information System (GIS) technologies, however, and contemporary users of FIA data often want more than statistical tables and charts—they want maps.

Generally, FIA data are used in spatial analysis either by mapping selected plots to view plot scatter or to use by using map-based technologies where remote-sensing information is modeled with the aid of FIA plots. A reliable methodology is lacking to more easily map FIA inventory statistics for all attribute variables without having to develop separate models and methods for each variable (such as k-Nearest Neighbor [kNN] classification for forest area, basal area, or volume).

This article presents an overview of a new mapping method that can directly transfer tabular data to a map on which pixels can be added in any way desired to estimate carbon (or any other variable) for a particular area. We describe our approach by way of an example that maps FIA data for the entire State of Nevada.

Have Design, Will Map

FIA uses the double sampling for stratification design, sometimes called two-phase sampling. In this simple approach, a large sample of phase-1 units are selected for an inexpensive measurement and then a smaller phase-2 subsample of the phase-1 units are remeasured in more costly detail. Neyman (1938) devised the method for sampling human populations and Bickford (1952, 1959) first applied it to FIA. Bickford (1952) used aerial photographs to obtain relatively cheap estimates of forest-area strata (phase 1) that were then combined with field plots (phase 2) to obtain forest statistics.

This design has served FIA well over the years (Bechtold and Patterson 2005, Born and Barnard 1983, Chojnacky 1998), but it has not taken advantage of recent advances in remote-sensing and GIS technologies for a spatial display of results. Typically,

[1] Adjunct Faculty, Virginia Polytechnic Institute and State University, Department of Forestry, 144 Rees Place, Falls Church, VA 22046. E-mail: dchoj@cox.net.

[2] Associate Professor, Virginia Polytechnic Institute and State University, Center for Environmental Applications of Remote Sensing, Department of Forestry, Blacksburg, VA 24061. E-mail: wynne@vt.edu.

[3] Postdoctoral Research Associate, Virginia Polytechnic Institute and State University, Center for Environmental Applications of Remote Sensing, Department of Forestry, Blacksburg, VA 24061. E-mail: cblinn@vt.edu.

an FIA table or chart summarizes forest statistics for various forest parameters or categories (fig. 1). Because FIA includes a wealth of species detail for forest cover, it would be desirable, although not necessary, to have map categories to reflect some of this information. Maps are not involved in the estimation of forest cover but they provide an attractive and credible way for a nontechnical audience to see roughly where different species or forest types might occur on a map instead of needing to consult statistical tables.

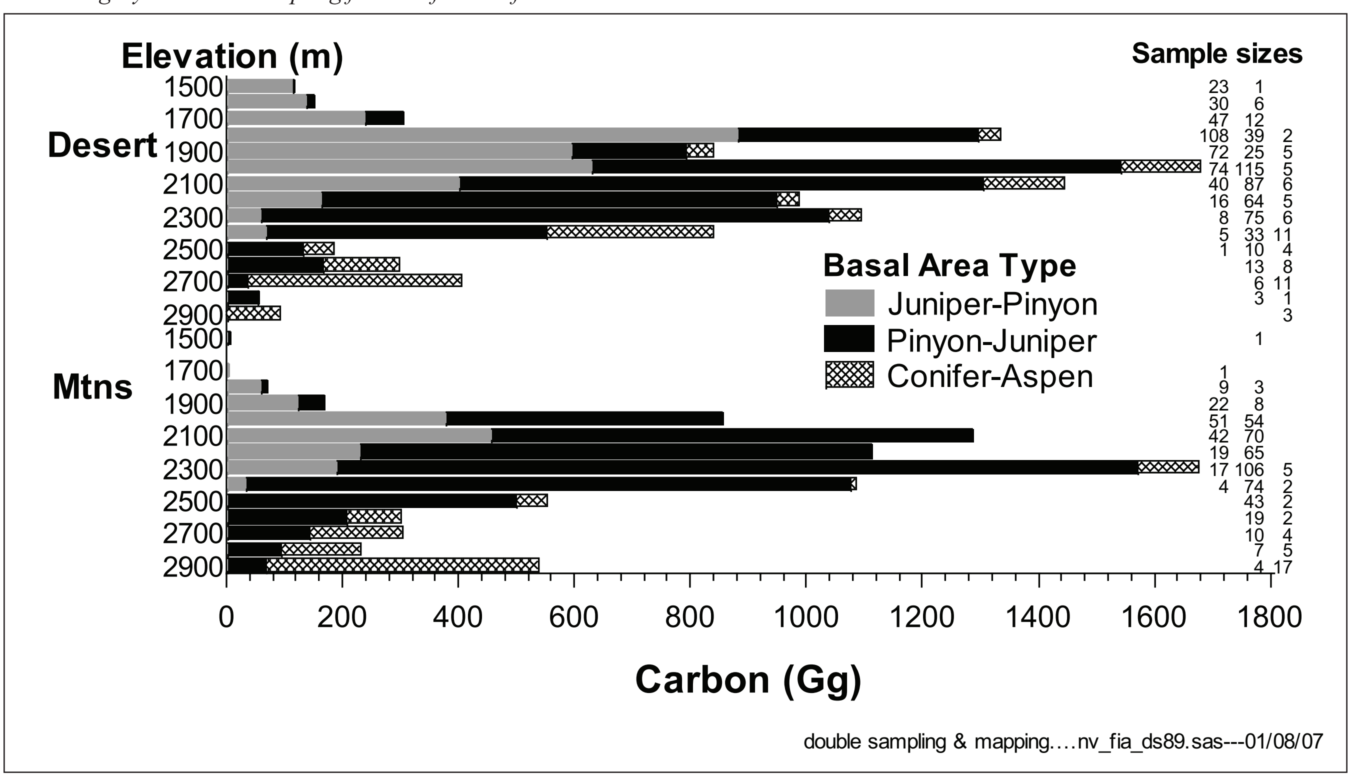
To overcome this shortcoming, researchers are using methods like kNN,, generalized linear models, and other techniques that seek to “fill in” or “expand” the nonsampled area between field plots (Franco-Lopez *et al.* 2001, Haapanen *et al.* 2004, McRoberts *et al.* 2005, Ohmann and Gregory 2002). These methods, however, ignore the double sampling design by using only the phase-2 field plots for developing models that estimate forest attributes directly from remote-sensing measurements.

We think it makes good sense to start with the full-blown double sampling design and look for ways to expand it to mapping applications. We developed a mapping concept that can directly transfer tabular data to a map on which pixels can be added in various ways, by first defining a set of variables called “category variables” that are common to both phase 1 and phase 2. The chart data can be mapped if the variables can be spatially depicted.

Several key concepts underlie our approach to mapping FIA data. First, for phase 1 of double sampling, a remote-sensing layer coverage or aerial photographs are needed for stratifying forest from nonforest.

Second, all calculations are done within the double sampling framework, which includes all the variables that FIA estimates. For our example, we estimate carbon; however, results could apply to volume, stand structure indexes, understory, deadwood,

Figure 1.—*Example of Forest Inventory and Analysis (FIA) tabular data for Nevada. Carbon amounts calculated for categories of Nevada FIA data are grouped into basal-area types, elevation classes, and two ecoregion classes (“Mtns” includes the Great Basin Mountains Section in the center of Nevada and “Desert” includes the rest of the State). Sample sizes are numbers of field plots for each category. A double sampling for stratification formula was used to summarize data.*



or any other variables that FIA estimates. Also, the double sampling design provides a framework to calculate variance estimates for each result. In other words, we have the power of a sample design backing our mapping method and do not have to start from scratch to develop statistical estimates for our maps.

Our third step is a new addition to double sampling. For this initial illustration, we link the double sampling design to maps by defining spatial variables for summarizing FIA data within the double sampling framework. We call these variables “category variables” for bridging the link between sampling design and map; the variables must be “categorical” to include groups of FIA data based on sufficient numbers of plots. For Nevada, we selected three category variables: forest cover type, Bailey’s (1995) ecoregion, and 100-m elevation classes.

The Nevada Example

First, a Standard Analysis

A standard double sampling analysis was done for combining the phase 1 and 2 data. In phase 1, Moderate Resolution Imaging Spectroradiometer (MODIS) imagery (FIA program 2006b) was used for distinguishing forest from nonforest; for Nevada, 4.5 million pixels, 250 m^2 in size, covered the State. For phase 2, 14,138 FIA plots were available online (FIA program 2006a) for Nevada. Although the official survey date for Nevada was 1989, actual measurements ranged from 1979 to 1997, with 85 percent measured in 1982. Ownership of the plots included 69 percent Bureau of land Management, 13 percent private, and the rest national forest or other public ownership. The sampling design was roughly a grid that included much nonforest desert or rangeland; 88 percent of the plots were nonforest and the remainder were classified into three forest types according to species mix and dominate tree basal area.

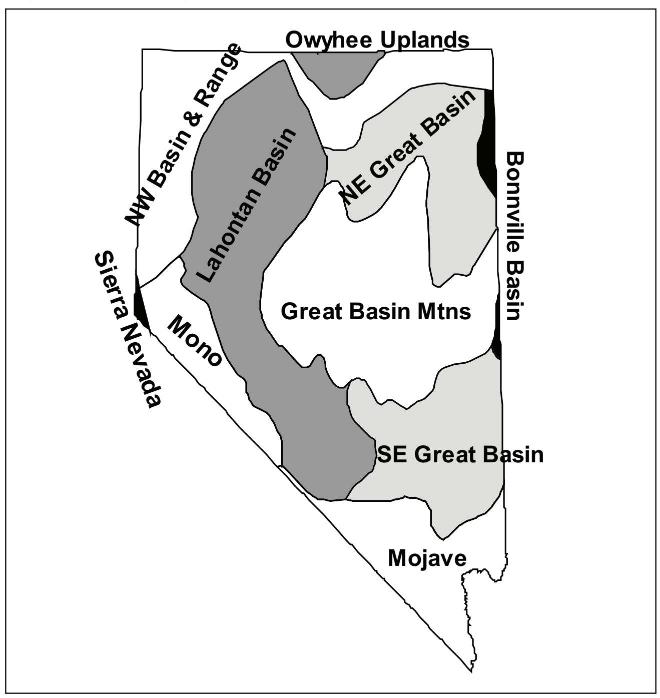
Basal area was calculated at diameter at breast height, with a conversion (Chojnacky and Rogers 1999) even though most of the woodland tree species were measured at ground line. Forest type or basal-area type then was defined in the following manner:

- Pinyon-juniper (P-J) plots: 953 included less than 67 percent juniper basal area and/or pinyon and/or mountain mahogany.
- Juniper-pinyon (J-P) plots: 589 were those in which more than 67 percent of the basal area on the plot included juniper species, could be pinyon, but was not mountain mahogany.
- Conifer-aspen (C-A) plots: 104 had basal area that was predominately ponderosa pine, fir, or other conifer species and/or aspen.
- Other plots: 2 included desert acacia species but were ignored for this study.

This classification roughly followed elevation. Predominately juniper (and perhaps some pinyon or J-P) was at lower elevations but decreased higher up, where pinyon-juniper mix (or P-J) became predominate. Mountain mahogany tended to mix with pinyon on the upper end of P-J, but we did not separate with another class. The other conifers and aspen (or C-A) occupied the highest elevations.

Of the 14,138 FIA plots, 1,645 forested plots were grouped into 34 strata defined as forest and nonforest; the 9 ecoregion sections (fig. 2); and a lower, mid-, and upper elevation zone corresponding to percentiles of Nevada elevation divided into thirds. Plots per stratum ranged from 2 to 481; the median was 13 plots.

Figure 2.—*Ecoregion sections for Nevada.*



The double sampling analysis was categorized into three basal-area types and 100-m elevation classes; a third variable corresponding to the ecoregion sections was considered but grouped into two classes to include some plots for each 100-m elevation class. The center Great Basin Mountain section (666 plots) was separated from the rest of the State in what we called Great Basin Desert (980 plots).

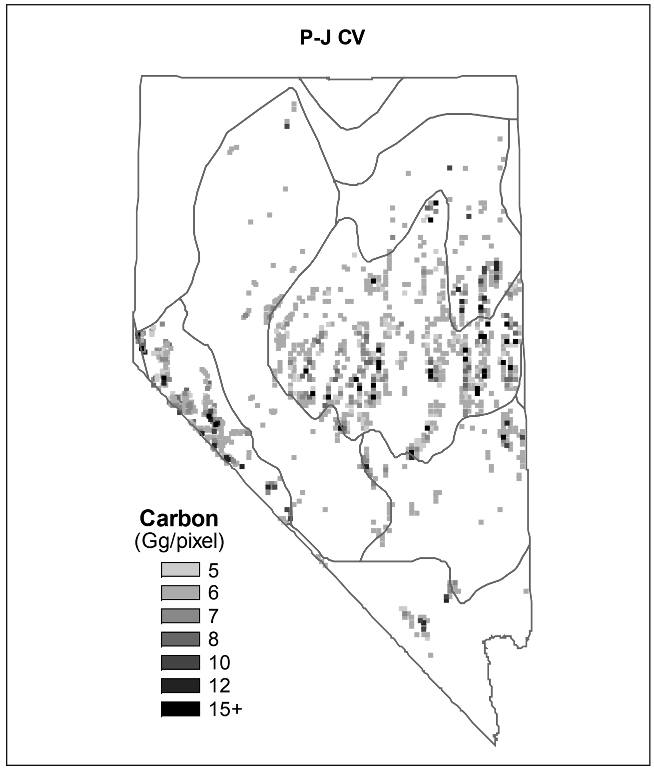
Next, Mapping the Results for Pinyon-Juniper

The double sampling results for carbon estimation (in Gg, or 10^9 g) show trends for the categories chosen (fig. 1, earlier). Although the chart provides useful information showing elevation distribution among J-P, P-J, and C-A forests, we next wanted to map the results.

The MODIS data used for phase 1 contained elevation and ecoregion information; however, they did not distinguish the forest types shown in figure 1. Therefore, we added a classification of forest into the three basal-area types by using logistic regression from work on another ongoing study. The model prediction corresponding to basal-area type with the largest probability of occurrence was assigned to each MODIS pixel for mapping. Then figure 1 data were assigned to appropriate pixels corresponding to elevation, ecoregion, and basal-area type.

Figure 3 shows carbon mapped for P-J in classes as carbon per pixel, which allows one to add the pixels in any manner to estimate carbon from the map. This feature is possible because the category means in figure 3 are simply divided by the number of pixels for that category on the map. The figure also includes coefficients of variation for the categories shown in figure 1.

Figure 3.—*Carbon for pinyon-juniper basal-area type (in fig. 1) mapped for Nevada Forest Inventory and Analysis data on left; coefficient of variation (CV) on right. Although it is difficult to illustrate map data in gray scale, the legends correspond to the range of data.*



Discussion

We have shown that one can (1) in advance of analysis, define categories common to phase 1 and phase 2; (2) estimate means and variances for the categories from double sampling; and (3) map results for category variables having spatial significance.

The model we used is preliminary but the method appears to be unbiased. It results in maps that are similar in concept to mapping FIA data by county, congressional district, or watershed but are more accurate because the phase 1 categories are used to separate forest from nonforest and further divide forest land into similar categories. Since the categorical variables are spatially defined in phase 1, the resulting maps cover a landscape with polygons defined by means, totals, or standard errors for attributes compiled from the sample design.

Defining good, robust categories is challenging, however. For example, instead of using ecoregion sections, it probably would have been more effective to use ecological provinces (the next scale up in Bailey's [1995] hierarchy) and to include neighboring States for full province analyses. Further study is needed on the definition of category variables for splitting the data because these definitions greatly affect sample sizes in splitting the data, which affects map precision.

A useful feature of the map is the ability to add "carbon per pixel" for any area of the map to estimate population totals. A variance for this estimate is not so simple; a variance estimate should exist but we have not yet studied this.

Finally, mapping basal-area type for Nevada using MODIS data was challenging because we needed the additional logistic regression model to classify "MODIS forest" into basal-area type. This classification worked reasonably well except for a few vegetation categories with small sample sizes at upper and lower elevations where missing values resulted for mapping.

Conclusion

We have initiated a new method to combine map-based technology into the double sampling design that enables us to more easily map FIA inventory statistics for all attribute variables without having to develop separate models and methods for each variable. Because the approach builds on FIA's double sampling design, it offers all the power of the FIA database for map-based applications, which, we think, is a big advantage over some other modeling methods for expanding field plot grids.

Acknowledgments

The authors thank Ron Tymcio of the Forest Service, U.S. Department of Agriculture's Interior West Forest Inventory and Analysis Unit in Ogden, UT, who provided the MODIS data for Nevada; the Nevada Bureau of Land Management for funding studies in Nevada that allowed us to acquire data for this study; and Mary Carr of the Forest Service's Content Analysis Team Ecosystem Management Coordination Publishing Arts group for editing the manuscript.

Literature Cited

Bailey, R.G. 1995. Descriptions of the ecoregions of the United States. 2nd ed. Misc. Pub. 1391. Washington, DC: U.S. Department of Agriculture, Forest Service. 108 p.

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

Bickford, C.A. 1952. The sampling design used in the forest survey of the northeast. Journal of Forestry. 50(4): 290-293.

Bickford, C.A. 1959. A test of continuous inventory for national forest management based upon aerial photographs, double sampling, and remeasured plots. In: Proceedings, Society of American Forestry. Bethesda, MD: Society of American Forestry: 143-148.

Born, J.D.; Barnard, J.E. 1983. FINSYS-2: subsystem TABLE-2 and OUTPUT-2. Gen. Tech. Rep. NE-84. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 133 p.

Chojnacky, D.C. 1998. Double sampling for stratification: a forest inventory application in the interior west. Res. Pap. RP-RMRS-7. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 15 p.

Chojnacky, D.C.; Rogers, P. 1999. Converting tree diameter measured at root collar to diameter at breast height. Western Journal of Applied Forestry. 14(1): 14-16.

Forest Inventory and Analysis (FIA) program. 2006a. Forest Inventory and Analysis national program. http://fia.fs.fed.us. (17 January 2007).

FIA program. 2006b. Interior west forest inventory and analysis, map products: interior west region. http://www.fs.fed.us/rm/ogden/map-products/intwest/iwmaps.shtml. (17 January 2007).

Franco-Lopez, H.; Ek, A.R.; Bauer, M.E. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of Environment. 77(3): 251-274.

Haapanen, R.; Ek, A.R.; Bauer, M.E.; Finley, A.O. 2004. Delineation of forest/nonforest land use classes using nearest neighbor methods. Remote Sensing of Environment. 89: 265-271.

McRoberts, R.E.; Holden, G.R.; Nelson, M.D.; Liknes, G.C.; Gormanson, D.D. 2005. Using satellite imagery as ancillary data for increasing the precision of estimates for the Forest Inventory and Analysis program of the USDA Forest Service. Canadian Journal of Forest Research. 35(12): 2968-2980.

Neyman, J. 1938. Contribution to the theory of sampling human populations. Journal of American Statistical Association. 33: 101-116.

Ohmann, J.L.; Gregory, M.J. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, U.S.A. Canadian Journal of Forest Research. 32(4): 725-741.