ASSESSING ESTIMATION TECHNIQUES FOR MISSING PLOT OBSERVATIONS IN THE U.S. FOREST INVENTORY

Grant M. Domke, Christopher W. Woodall, Ronald E. McRoberts, James E. Smith, and Mark A. Hatfield[1]

Abstract.—The U.S. Forest Service, Forest Inventory and Analysis Program made a transition from state-by-state periodic forest inventories—with reporting standards largely tailored to regional requirements—to a nationally consistent, annual inventory tailored to large-scale strategic requirements. Lack of measurements on all forest land during the periodic inventory, along with access issues and misidentification of forest plots as nonforest, have resulted in plot-level data gaps spread in the FIA database. In this study, we examined several approaches that compensate for missing observations with respect to the deviation and precision of stratified estimates of carbon stocks per unit area using data from the FIA database. Preliminary estimates of live tree carbon stocks per unit area calculated using all missing data approaches were well within one standard error of the baseline estimates for the Lake States study region.

INTRODUCTION

Forest ecosystem carbon (C) stocks and stock change have been documented by the Intergovernmental Panel on Climate Change (IPCC) using 1990 as a baseline reference for all IPCC reports. In the United States, estimates of forest C stocks and stock change are obtained from data collected and maintained by the U.S. Forest Service, Forest Inventory and Analysis (FIA) program. Over the course of the IPCC monitoring period, the FIA program made a transition from state-by-state periodic inventories—with reporting standards largely tailored to regional requirements—to nationally consistent, annual inventories tailored to large-scale strategic requirements (Bechtold and Patterson 2005). Lack of measurements on all forest land during the periodic inventory, along with access issues and misidentification of forest plots as nonforest due to poor aerial imagery, have resulted in plot-level data gaps throughout the FIA database. These data gaps contribute to large differences in estimates of carbon stock change between periodic and annual inventories. In this study, we examined several approaches that compensate for missing observations with respect to the accuracy and precision of stratified estimates of carbon stocks per unit area using data from the FIA database. The objectives of the study were to: 1) identify patterns of missingness in the FIA data; 2) examine approaches for replacement; 3) assess approaches under increasing levels of missingness; and 4) document strategies for replacement in periodic and annual forest inventory data within the context of National Greenhouse Gas Inventory.

METHODS

Data

Data came from base intensity FIA plots measured in each of the two most recent annual inventory cycles (2002-2006 and 2007-2011) in the Lake States region (Michigan, Minnesota, and Wisconsin). These plots are quasi-systematically distributed approximately every

[1] Research Forester (GMD and CWW), Mathematical Statistician (REM), and Forester (MAH), U.S. Forest Service, Northern Research Station, 1992 Folwell Ave., St. Paul, MN 55108; Plant Physiology/Modeling (JES), Northern Research Station, Durham, NH. GMD is corresponding author: to contact, call 651-649-5138 or email at gmdomke@fs.fed.us.

2,428 hectares across the 48 conterminous states of the U.S. Each plot comprises a series of smaller plots (i.e., subplots) where tree- and site-level attributes—such as diameter at breast height (d.b.h.) and tree height—are measured at regular temporal intervals (Bechtold and Patterson 2005).

Because the precision standards established by the FIA program are rarely satisfied with the base intensity plot sample size, the estimation process is enhanced through stratification. Stratification is used to reduce the variance of attributes, such as C stocks, by portioning the population into strata (Bechtold and Patterson 2005). Each FIA plot is assigned to a stratum using the National Land Cover Database (Homer et al. 2004) or other Forest Service databases (Ruefenacht et al. 2008). In the Lake States region, strata are assigned based on percent canopy cover (i.e., 0-5, 6-50, 51-65, 66-80, and 81-100 percent). Strata are typically grouped into estimation units which are determined by a combination of sampling intensity (i.e., number of plots) and geographical boundaries (Woudenberg et al. 2010).

Stratified estimates of aboveground live tree (≥12.7 cm d.b.h.) C per unit area, $\bar{C}$, and variance, $Var(\bar{C})$, were calculated following Cochran (1977):

$$\bar{C} = \sum_{j=1}^{J} w_j \bar{C}_j$$

and

$$Var(\bar{C}) = \sum_{j=1}^{J} w_j^2 \frac{\hat{\sigma}_j^2}{n_j}$$

where $j = 1, \ldots, J$ denoted stratum, w_j was the weight for the jth stratum, calculated as the proportion of pixels assigned to the stratum, $\bar{C}_j$ was the mean carbon per unit area for plots assigned to the jth stratum, and $\hat{\sigma}_j^2$ was the within-stratum variance for the jth stratum.

Missing Data Strategies

Strategies used to compensate for missing plot observations in forest inventory estimation generally fell into two categories, ignoring plots with missing observations or replacing missing observations. In this study, we examined five approaches: 1) treat plots with missing observations as if they had not been selected for the sample (IGNORE); 2) replace missing plot observations with the observation for the same plot from the previous inventory (PREVIOUS); 3) replace missing observations with the stratum mean (STRATUM); 4) randomly draw from a pool (nearest neighbors) of observed plots most similar to the plot with the missing observation (NEAREST); and 5) compute the expected values for missing plot observations by repeatedly updating maximum-likelihood parameter estimates and imputing expected values until convergence is achieved (EM). Each approach was further divided (beyond strata) by ownership domain to account for differences in forest land management which may result in different C estimates. This subdivision also accounts for bias in instances when all missing plot observations fall on a particular ownership (e.g., denied access on private forest land).

Analysis and Comparisons

The C estimates generated by each missing data approach were compared to the base estimates (BASE) by stratum and stratum+ownership (i.e., public and private). Stratified base estimates of $\bar{C}$ and $Var(\bar{C})$ were calculated using observations for all base intensity plots across the five canopy-cover strata. This BASE estimate served as the standard for comparison for estimates obtained with the techniques that compensate for missing plot observations. Estimates were first compared visually by generating a graph of the distribution for the BASE estimates and the distributions of the different missing data approaches. Estimates for the missing data approaches were then compared with the BASE estimates and each other for proportions of missing plot observations ranging from 0 to 25 percent, which encompassed the range of nonresponse in the FIA program reported by Patterson et al. (2012). Specifically, the proportion was calculated as the ratio of the number of plots classified as missing (i.e., hazardous, denied access, skipped) to the total number of plots selected for measurement. A Monte Carlo procedure (McRoberts 2003) was used to simulate random denied access on private forest lands to compare the deviation and precision of the mean values produced by each missing data approach at each level of missingness (0-25 percent) against the BASE estimates of $\bar{C}$. In the initial analysis, standard errors of the mean estimates produced by the missing data approaches at each missingness level were used to compare with the BASE estimates. Mean estimates greater than one standard error from the BASE mean were consider significantly different from the BASE estimate. All analyses were conducted using R statistical software (R Development Core Team 2012)

RESULTS

Preliminary analyses were restricted to 12,323 base intensity plots where at least one accessible forest land condition (i.e., area classification on each plot such as forest type or ownership group used for analytical purposes) was present during the annual inventory period. The proportion of missing base intensity plots for the most recent inventory was 5 percent in Minnesota, 6 percent in Wisconsin, and more than 11 percent in Michigan. Nearly all missing observations (94 percent) were due to private landowners denying field crews access to lands with the remaining plots deemed hazardous by field crews (3 percent) or skipped due to seasonal access (3 percent). The distribution of missing plot observations by county suggests that denied access areas are not uniformly distributed throughout the study region (Fig. 1).

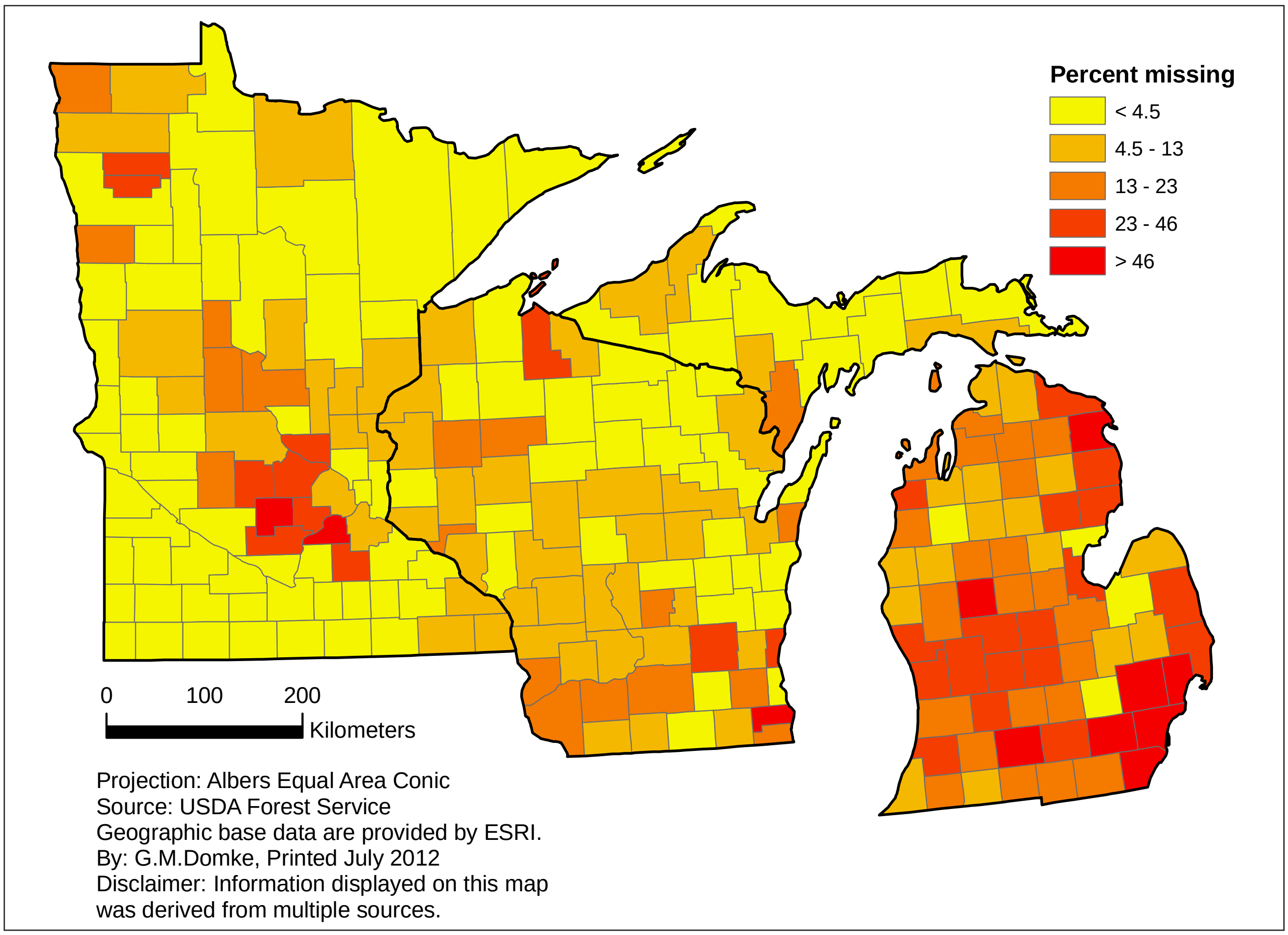

Figure 1.—Proportion of missing plot observations due to denied access by county in the Lake States region of the U.S. for the most recent FIA inventory period, 2007-2011.

Stratified base estimates of $\bar{C}$ increased with increasing canopy cover for the most recent inventory period in each of the three Lake States (Table 1). Preliminary estimates calculated using the IGNORE, STRATUM, NEAREST approaches, at current missingness levels, were within one standard error of the BASE estimate of $\bar{C}$ using observations for all plots. This suggests there were no statistically significant differences among estimates obtained using the missing data approaches initially investigated in the study. That said, the IGNORE approach was computationally more efficient than the STRATUM approach, which was more efficient than the NEAREST approach.

DISCUSSION

Early results suggest there are a number of strategies for dealing with missing plot observations in annual forest inventory data. The IGNORE approach at current nonresponse levels was computationally more efficient than the other two approaches initially evaluated. Assuming the PREVIOUS and EM approaches perform comparably to the STRATUM and NEAREST techniques, the likely outcome of the initial phase of this study will be that the IGNORE approach is optimal for dealing with missing plot observations (at current nonresponse levels in the study area) due to denied access in annual forest inventory data. This approach has merit assuming the distribution of missing plot observations is random. If not, an alternative approach and/or subdivision of strata or domain may be necessary to account for bias from missing plot observations due to denied access. Furthermore, all missing data approaches must be examined across the range of potential nonresponse in order to evaluate which approach or approaches may be useful at the national level. Assessing the distribution of missing plot observations and the range of nonresponse is important since it is likely there are patterns of missingness in the periodic inventory, albeit for a variety of different reasons, which may

Table 1.—Stratum statistics for the most recent FIA inventory (2007-2011) in the Lake States

			Live tree C per unit area (Mg ha^{-1})		
			2006	2011	
State and Stratum	Weight	Number of plots	Mean	Mean	Standard Error
Michigan	1.00	4,454	28.36	28.80	0.42
Canopy cover 0 - 5	0.64	391	9.77	12.42	0.90
Canopy cover 6 - 50	0.07	600	11.66	12.24	0.65
Canopy cover 51 - 65	0.04	415	20.03	20.50	1.02
Canopy cover 66 - 80	0.10	1,264	26.63	27.46	0.66
Canopy cover 81 - 100	0.15	1,784	41.22	40.85	0.74
Minnesota	1.00	3,966	17.24	18.04	0.32
Canopy cover 0 - 5	0.63	370	8.85	10.59	0.82
Canopy cover 6 - 50	0.03	97	9.45	11.05	1.36
Canopy cover 51 - 65	0.03	202	11.42	13.42	1.01
Canopy cover 66 - 80	0.12	1,231	16.29	17.15	0.56
Canopy cover 81 - 100	0.19	2,082	20.21	20.53	0.47
Wisconsin	1.00	3,903	23.51	25.23	0.39
Canopy cover 0 - 5	0.60	686	10.25	11.86	0.56
Canopy cover 6 - 50	0.02	135	11.53	12.44	1.32
Canopy cover 51 - 65	0.03	182	10.98	15.31	1.22
Canopy cover 66 - 80	0.08	619	19.88	21.31	0.82
Canopy cover 81 - 100	0.27	2,281	30.19	31.86	0.55
Lake States	1.00	12,323	23.24	24.21	0.22

require similar subdivisions to account for bias. While the initial analyses focused on a few approaches using annual inventory data, the full suite of missing data approaches will be evaluated using the annual inventory and then applied to the periodic inventory to assess whether any of the missing data approaches better align estimates for forest C stocks and stock change between periodic and annual inventories in the United States.

ACKNOWLEDGMENTS

The authors would like to thank Dale Gormanson and Brian Walters for assistance with map and dataset development. We also thank Rob Slesak and David Wilson for reviewing the draft manuscript.

LITERATURE CITED

Bechtold, W.A.; Patterson, P.J., 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.

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

Homer, C.; Huang, C.; Yang, L.; Wylie, B.; Coan, M. 2004. **Development of a 2001 national landcover database for the United States.** Photogrammetric Engineering and Remote Sensing. 44: 3999-4005.

McRoberts, R.E. 2003. **Compensating for missing plot observations in forest inventory estimation.** Canadian Journal of Forest Research. 33: 1990-1997.

R Development Core Team. 2012. **R: a language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing.

Ruefenacht, B.; Finco, M.V.; Nelson, M.D.; Czaplewski, R.; Helmer, E.H.; Blackard, J.A.; Holden, G.R.; Lister, A.J.; Salajanu, D.; Weyermann, D.; Winterberger, K. 2008. **Conterminous US and Alaska forest type mapping using forest inventory and analysis data.** Photogrammetric Engineering and Remote Sensing. 74: 1379-1388.

Woudenberg, S.W.; Conkling, B.L.; O'Connell, B.M.; LaPoint, E.B.; Turner, J.A.; Waddell, K.L. 2010. **The Forest Inventory and Analysis database: database description and users manual version 4.0 for Phase 2.** Gen. Tech. Rep. RMRS-GTR-245. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 336 p.