Forest Ecology and Management 315 (2014) 112–120

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Strategies to compensate for the effects of nonresponse on forest carbon baseline estimates from the national forest inventory of the United States

Grant M. Domke *, Christopher W. Woodall, Brian F. Walters, Ronald E. McRoberts, Mark A. Hatfield

USDA Forest Service, Northern Research Station, 1992 Folwell Ave., St. Paul, MN 55108, USA

ARTICLE INFO

Article history:
Received 6 September 2013
Received in revised form 19 December 2013
Accepted 23 December 2013

Keywords:
Imputation
Missing data
Greenhouse gas reporting
Forest Inventory and Analysis (FIA)

ABSTRACT

Forest ecosystem carbon (C) stocks and stock change in the United States (US) have been documented using Intergovernmental Panel on Climate Change (IPCC) procedures and guidance with 1990 as a baseline reference for all United Nations Framework Convention on Climate Change reports. In the US, estimates of forest C stocks and stock change are obtained from data collected and maintained by the Forest Inventory and Analysis (FIA) program of the US Forest Service. Over the course of the IPCC monitoring period, the FIA program made a transition from state-by-state multiyear periodic inventories selected on a rotating basis – with reporting standards largely tailored to regional requirements – to nationally consistent, annual inventories (where a proportion of plots is measured in each state each year) designed for large-scale strategic requirements. Lack of measurements on all forest land during the periodic inventories, along with plot access difficulties and misidentification of forest plots as nonforest due to poor aerial imagery, have resulted in missing data (i.e., nonresponse) throughout the FIA database. Nonresponse, which in some US states is greater than 20%, may lead to differences in estimates of forest C stock change due to the procedural transition from periodic to annual inventories. As an initial step towards rectifying the differences in estimates, we examined several strategies to compensate for missing observations using the most recent annual inventory data from the Lake States region of the US. Results varied by state in the study but given the annual reporting cycle and requirements to compile national estimates of forest C, it was deemed that techniques, where non-observed samples are removed from estimation procedures, provided the optimal combination of statistical performance and efficiency. While the initial analysis focused on the Lake States region, several compensation strategies described may be useful in bridging the gap between national C flux estimates from periodic and annual forest inventories.

Published by Elsevier B.V.

1. Introduction

As signatories to the United Nations Framework Convention on Climate Change (UNFCCC), the United States (US) has been providing annual estimates of forest ecosystem carbon (C) stocks and stock change (USEPA, 2013) to the UNFCCC in accordance with IPCC Good Practice Guidance (IPCC, 2006). Carbon stocks and stock change are estimated by pool each year across a defined reporting period that runs from 1990 to the present (IPCC, 2006). In the US, C stocks and stock change are estimated from data collected and maintained by the Forest Inventory and Analysis (FIA) program of the US Forest Service, which conducts the national forest inventory (NFI) of the US (USDA Forest Service, 2012). 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 (Gillespie, 1999) – to nationally consistent, annual inventories designed for large-scale strategic requirements (McRoberts et al., 2010). Lack of measurements on all forest land during the periodic inventories, along with plot access difficulties and misidentification of forest plots as nonforest due to poor aerial imagery, have resulted in missing data throughout the FIA database (McRoberts, 2003; Birdsey, 2004; Patterson et al., 2012; Goeking and Patterson, 2013). These data gaps may contribute to structural differences in estimates of forest C stock change between periodic and annual inventories that are procedural artifacts as opposed to changes in natural resources. Because the US's forest carbon baseline (USEPA, 2013) serves to inform policy as well as carbon science, improving the accuracy and scientific rigor of the baseline is paramount (Woodall, 2012).

* Corresponding author. Tel.: +1 651 649 5138; fax: +1 651 649 5140.
E-mail addresses: gmdomke@fs.fed.us (G.M. Domke), cwoodall@fs.fed.us (C.W. Woodall), bfwalters@fs.fed.us (B.F. Walters), rmcroberts@fs.fed.us (R.E. McRoberts), mahatfield@fs.fed.us (M.A. Hatfield).

0378-1127/$ - see front matter Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.foreco.2013.12.031

Fundamentally, the lack of appropriate data to estimate an annual forest C baseline can be viewed as a missing observation research problem (Van Deusen, 1997; Little and Rubin, 2002). The issue of missing observations (i.e., nonresponse) has been persistent through the course of forest inventories around the world (McRoberts, 2003; Eskelson et al., 2009; Tomppo et al., 2010; Beets et al., 2011; Barrett and Maltamo, 2012; Patterson et al., 2012; Goeking and Patterson, 2013) whereas forest C baselines suffer from the same dilemma: how should missing observations be accommodated when using large scale forest inventories to estimate forest resources whether C or sawtimber volume? The extent to which nonresponse is a serious problem depends to some degree on the underlying reasons for nonresponse. In the US, the most common reason for nonresponse is that private forest land owners refuse field crews access to their land (Patterson et al., 2012). However, for most countries, individual property rights are not nearly as strong as in the US. Thus, contrary to the situation in the US, NFI field crews in only very few countries must obtain permission of landowners to access plots on private forestland. For these countries, two results follow; first, the reasons for nonresponse are limited to natural causes such as floods and hazardous terrain, and second, the proportion of nonresponses is so small that the problem is generally ignored. In the southern hemisphere, for example, NFIs are sufficiently rare and of such recent initiation that the degree to which nonresponse is a problem has not been rigorously evaluated.

As a surrogate for examining techniques to mitigate nonresponse by year for C baselines, data that exist from forest inventories collected over multiple years may be used to inform annual C baseline strategies. Within annual forest inventories, missing observations are optimally estimated with data of interest for all sample units that were observed. Unfortunately, nonresponse is inevitable in most large inventories which has led to a well-defined order of operations for determining strategies to compensate for missing observations (Sande, 1982; Lemeshow, 1985; Rubin, 1987; Särndal et al., 1992; Lesser, 2001; Little and Rubin, 2002; McRoberts, 2003; Eskelson et al., 2009; Patterson et al., 2012). In the annual FIA inventory, nonresponse is largely the result of denied access on private forest lands and, to a lesser extent, hazardous areas (McRoberts, 2003; Patterson et al., 2012; Goeking and Patterson, 2013). On private forest lands, FIA field crews typically make a single attempt to obtain permission to gain access to the plot via letter, phone call, or location visit. Landowners may deny the field crews access to measure plots on their private forest land and in those cases plots are listed as denied access. Plot locations on forest land deemed hazardous (e.g., flooded) may be revisited later in the field season when conditions allow access but this is not always possible. Once an attempt has been made and the plot or portion of the plot (i.e., condition – area classification mapped on each plot using discrete variables such as land use, forest type, and/or ownership group to enable division of forest into various domains of interest, USDA Forest Service, 2013) – hereafter collectively referred to as plot – cannot be observed and measured, it is recorded as nonsampled and given a reason code in the FIA database (USDA Forest Service, 2013). Identifying plausible strategies to compensate for nonresponse first requires quantification of nonresponse, followed by an assessment of the properties of the nonresponse elements and an understanding of the nonresponse mechanisms (Little and Rubin, 2002). Because missing data is an issue for both annual forest inventories and UNFCCC forest C baselines, examining potential strategies for nonresponse compensation is warranted.

We examined several approaches that compensate for missing observations with respect to the accuracy and precision of estimates of C stocks per unit area using data from the FIA annual inventory in the Lake States region (Michigan, Minnesota, and Wisconsin) of the US (Fig. 1). The specific objectives of the study were to: (1) quantify nonresponse in the annual FIA data, (2) describe the properties of the missing observations, (3) identify strategies to compensate for nonresponse, (4) describe the process of incorporating compensation strategies into the FIA sampling framework, (5) assess each compensation strategy under increasing levels of simulated nonresponse, and (6) describe how the selected strategies may be employed to compensate for missing observations in the periodic inventories dating back to the 1990 C baseline used in greenhouse gas reporting (e.g., UNFCCC).

2. Methods

Because this study is an initial step toward rectifying the differences in forest C stock change due to the procedural transition from periodic to annual inventories, we examined several well-established strategies to compensate for missing observations using the most recent annual inventory data. We chose to restrict our initial assessment to the annual inventory because it includes a nationally consistent sampling frame and plot design so the methodologies established for compensating for missing observations could be applied nationally without substantial modification. Furthermore, causes of nonresponse were thought to be more consistent across regions of the US in the annual inventory, thus providing a parsimonious approach to identifying potential nonresponse patterns and selection of strategies to compensate for nonresponse in the data.

2.1. Data

The FIA program employs a three phase inventory, with each phase contributing to the subsequent phase. Phase 1 is a variance reduction step where satellite imagery is used to assign Phase 2 plots to strata (Bechtold and Patterson, 2005). A stratum is a defined geographic area (e.g., state or estimation unit) that includes plots with similar attributes; in the Lake States region, strata are defined by predicted percent canopy cover. Data in this study came from Phase 2 plots measured in each of the two most recent annual inventory cycles (2002–2006 and 2007–2011) in the Lake States region of the US. Phase 2 plots are distributed approximately every 2428 hectares across the 48 conterminous states of the US (Fig. 1). Each Phase 2 permanent ground plot comprises a series of smaller plots (i.e., subplots) where tree- and site-level attributes – such as diameter at breast height (dbh) and tree height – are measured at regular temporal intervals (USDA Forest Service, 2013).

In the Lake States region, gross tree volume is estimated using a model with tree dbh, site index (as a proxy for tree height), and basal area as explanatory variables (Woodall et al., 2011). Gross volume estimates are adjusted to account for volume loss due to rotten and missing cull defect and the sound volume estimates are converted to oven-dry biomass using the component ratio method (Heath et al., 2009; Woodall et al., 2011). Tree biomass is then multiplied by 0.5 to convert to C and live tree-level C estimates are summed within the plot and then converted to a per unit area basis. The plot-level estimate obtained as the sum of tree-level estimates is assumed to be an observation without error (McRoberts and Westfall, 2014).

2.2. Stratification

Because the precision standards (USDA Forest Service, 1970) established by the FIA program may not be satisfied for estimates of some parameters, the estimation process is enhanced through stratification. Stratification is used to reduce the variance of estimates of parameters such as C stocks, by partitioning the popula-

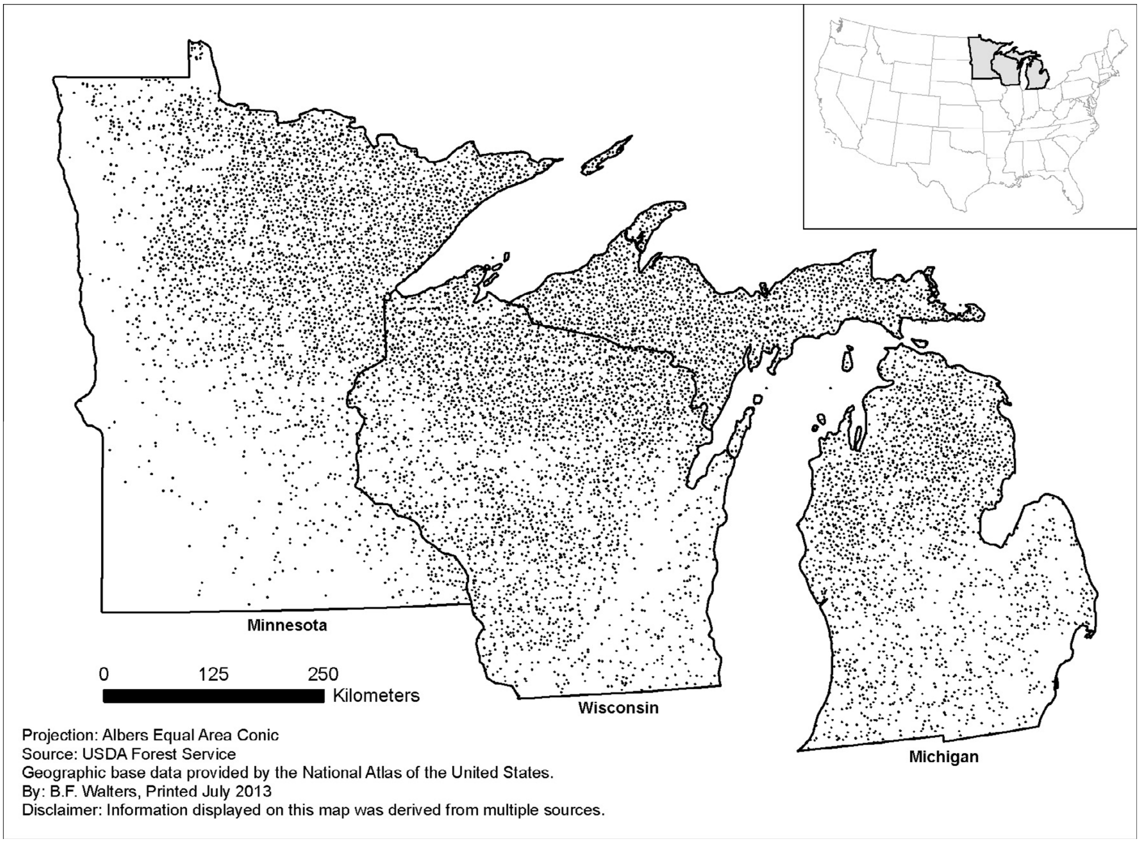

Fig. 1. Approximate plot locations with at least one forested condition in the Lake States (Michigan, Minnesota, Wisconsin) study region.

tion into strata based on auxiliary information such as estimated percent tree canopy cover (Bechtold and Patterson, 2005). FIA plots are assigned to strata based on the estimates of percent canopy cover of the population unit (i.e., pixel) containing the plot center 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 estimated percent canopy cover (i.e., 0–5%, 6–50%, 51–65%, 66–80%, and 81–100%; Table 1). Additional stratification factors include sampling intensity (i.e., area represented by each plot) and political boundaries (USDA Forest Service, 2013).

Stratified estimates of mean aboveground live tree C per unit area, $\overline{C}$, and variance, $V(\overline{C})$, were calculated following Cochran (1977):

$$\overline{C} = \sum_{j=1}^{J} w_j \overline{C}_j, \tag{1}$$

and

$$V(\overline{C}) = \sum_{j=1}^{J} w_j^2 \frac{\hat{\sigma}_j^2}{n_j}, \tag{2}$$

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, $\overline{C}_j$ was the estimate of the mean C per unit area for plots assigned to the jth stratum, n_j was the number of plots in the jth stratum, and $\hat{\sigma}_j^2$ was the within-stratum variance estimate for the jth stratum:

$$\hat{\sigma}_j^2 = (n_j - 1)^{-1} \sum_{i=1}^{n_j} (C_{ij} - \overline{C}_j)^2, \tag{3}$$

Table 1
Summary statistics of the mean carbon per unit area (Mg ha^{-1}) and standard error of the mean for each of the Lake States (Michigan, Minnesota, Wisconsin) by stratum.

State and stratum	Canopy cover (%)	Stratum proportion	Number of plots	Live tree C per unit area (Mg ha^{-1})		
				2006	2011	
				Mean	Mean	Standard error
Michigan		1.00	4386	16.98	13.91	1.14
1	0–5	0.64		8.23	7.16	0.61
2	6–50	0.07		10.39	8.21	0.44
3	51–65	0.04		18.08	16.81	0.84
4	66–80	0.10		27.72	23.83	0.58
5	80–100	0.15		41.78	37.90	0.70
Minnesota		1.00	3926	11.35	9.34	0.93
1	0–5	0.63		7.90	5.81	0.51
2	6–50	0.03		9.83	6.99	0.94
3	51–65	0.03		11.50	9.55	0.73
4	66–80	0.12		16.06	14.49	0.48
5	80–100	0.19		20.33	18.03	0.41
Wisconsin		1.00	3834	15.60	13.68	0.76
1	0–5	0.60		7.79	6.89	0.34
2	6–50	0.02		10.41	8.52	0.88
3	51–65	0.03		11.46	11.87	0.94
4	66–80	0.08		18.56	17.70	0.71
5	80–100	0.27		31.25	28.30	0.50

where C_{ij} was the live tree C per unit area for the ith plot within the jth stratum.

In most cases, missing observations are ignored and an adjustment factor is incorporated into the $\overline{C}$ term to compensate for nonresponse or situations where all or part of a plot falls outside the population (e.g., plots straddling an international boundary or national forest boundary where the population of interest is the national forest) (Bechtold and Patterson, 2005). The nonresponse adjustment factor is given by:

$$\overline{P}_j = (an_{jh})^{-1}\sum_{i}^{n_{jh}} a_{ij}, \tag{4}$$

where $\bar{P}_j$ is the nonresponse adjustment factor, a is the plot area, n_{jh} is the number of plots in the jth stratum excluding the hth missing plots, and a_{ij} is the area sampled for the ith plot in the jth stratum. The incorporation of the nonresponse adjustment factor into the stratum level estimates is given by:

$$\overline{C}_j = (an_{jh})^{-1}\sum_{i=1}^{n_{jh}} \frac{C_{ij}}{\overline{P}_j}, \tag{5}$$

In the FIA program the variability in $\overline{P}_j$ is ignored when calculating sampling variance essentially making $\overline{P}_j$ a constant. A mean plot area ($\overline{P}_j$) of 1 indicates there are no partial missing observations and all plots are within the population. The nonresponse adjustment factor was developed as a mechanism to compensate for potential bias in estimators introduced by ignoring portions of plots spanning population boundaries. Incorporating $\overline{P}_j$ into the stratum level estimates, $\overline{C}$ ensures that plots that are inside the population of interest are included in the estimation (Bechtold and Patterson, 2005).

2.3. Nonresponse assumptions

The way in which missing observations are treated in inventory data depends on why the observations are missing (Little and Rubin, 2002). In general, nonresponse may be described as: (1) missing at random (MAR), or (2) missing not at random (MNAR). Data MAR are dependent on known values in the inventory. In this case, accounting for the values which may be causing the missing data will not induce bias into estimators. Data MNAR are dependent on attributes which were not measured, thus making it impossible to adjust for bias in estimators associated with compensation strategies. The fact that the FIA program ignores missing observations in the estimation process suggests that data are assumed to be MAR. This is a fairly common assumption in stratified estimation (Särndal et al., 1992); however, there are components within the FIA sampling design that may challenge this assumption. In particular, Patterson et al. (2012) point out potential differences in response probabilities between sample locations that were sent to the field for data collection because they met the FIA definition of forest and those not sent to the field because the sample location did not meet the FIA definition of forest. Those differences have the potential to manifest themselves when all plots – sent to the field (forest and non-forest) and not sent to the field (non-forest) – are combined within strata because the sample may be larger for plots not sent to the field (due to errors in stratification maps) resulting in underestimates of forest area. A similar problem could occur for estimates of $\overline{C}$ if a percentage of plots sent to the field are not sampled (MAR or MNAR) and represent a unique forest condition. Either way, the nonresponse would result in a reduced sample size and may bias estimates of $\overline{C}$ for the unique forest condition. If the plots sent to the field were MNAR this bias would need to be accommodated in the estimates of $\overline{C}$ for the unique forest condition. Nevertheless, accounting for as much of the missing data mechanism as possible typically produces sound results (Little, 1995; Rubin, 1996). Based on the strategic decision to ignore missing observations in the stratification process – which assumes MAR – and the support for choosing a mechanism that best matches the plurality of the missing data in the Lake States region, we assume MAR in this study.

2.4. Missing data strategies

Strategies used to compensate for missing observations in forest inventories generally fall into two categories – ignoring missing observations or replacing them. In this study, we examined six techniques: (1) ignore missing observations, (2) replace with previous estimates, (3) replace with imputed estimates at random, (4) replace with imputed estimates from the selected group mean, (5) replace with stratum estimates at random, and (6) replace with stratum estimates group mean. Estimates obtained from each compensation technique were calculated with and without the nonresponse adjustment factor (designated: *adj* and *unadj*, respectively) and further divided by broad ownership (i.e., public and private forest land) domain (i.e., area classifications for partitioning forest land into categories; designated: *domain*) for a total of 24 unique estimates for each state's C population estimate (i.e., evaluation) (Table S.1). The division by ownership domain was included primarily to account for potential bias due to denied access only on private forest lands – which is the most common reason for missing observations in the annual FIA data (Patterson et al., 2012) – and also because private forest lands may be managed differently than public lands in some areas (Heath et al., 2011a,b), which may influence $\overline{C}$ with respect to ownership.

Ownership determinations were made using ownership and nonsampled reason codes in the FIA database (USDA Forest Service, 2013). Because of lack of an accurate ownership layer, we assumed that all missing observations without ownership and nonsampled reason codes were on private forest land. This delineation was only relevant for the stratum subdivision by ownership domain.

2.4.1. Baseline estimates for comparison

The $\overline{C}$ estimates generated by each missing data approach were compared to baseline estimates by stratum (*baseline*). Stratified *baseline* estimates, $\overline{C}$ and $V(\overline{C})$ were calculated using observations for all plots across the five canopy cover strata with and without nonresponse adjustments. These *baseline* estimates served as standards for comparison for estimates obtained with the techniques used to compensate for missing observations (Table 1).

2.4.2. Ignore missing observations

A common approach to missing observations in inventory data is to ignore them in the analysis (Little and Rubin, 2002). This approach treats missing observations as if they had not been selected for the sample, thus reducing the sample size. The first technique (*ignore*) removed plots not observed or measured, and strata estimates were calculated with the reduced sample (Table S.1). The reduced sample size can lead to $\overline{C}$ estimates with larger standard errors and estimator bias if the excluded plots differ systematically from the completely observed cases. To account for potential bias due to ownership differences, strata estimates were also calculated by ownership domain (*ignore_domain*).

2.4.3. Replace with previous estimates

Replacing missing observations with estimates from the previous sample period may be an option in cases where data are collected at the same location over the course of the annual inventory (i.e., permanent sample plots). The FIA sampling frame is divided into five panels with one panel being measured each year, in sequence, until all five panels have been completed and the process is repeated (Bechtold and Patterson, 2005). In the Eastern US, the panel system translates into a 5-year cycle length. In the Western US, each panel is divided into two 5-year subpanels, resulting in a 10-year cycle. In the Lake States, most plots are on their third annual remeasurement so using the previous $\overline{C}$ estimate to replace the missing observation on plots in the current inventory is an option, provided that data were collected during the last cycle. The second approach (*last* and *last_domain*) used $\overline{C}$ estimates from the most recent 5-year cycle, when available, to replace missing observations on plots in the current inventory (Table S.1).

2.4.4. Replace with imputed estimates

A group of replacement approaches, broadly categorized as hot deck imputation (Sande, 1982; Andridge and Little, 2010), were developed using attributes from observations in the current FIA inventory period. Imputation techniques have been explored by Van Deusen (1997) as a means for replacing missing observations in the FIA database and Reams and McCollum (2000) specifically evaluated a hot deck approach using FIA data. The replacement strategy (*match*) in this study utilized known attributes available for the missing observations in the FIA database (i.e., stratum, inventory year, and proportion of the forest condition on the plot). A group of five observations with attributes most similar to the know attributes from the missing observation was selected, one of the five observations was chosen at random, and the $\overline{C}$ estimate from the known observation was imputed to the missing observation (Table S.1). The same approach was expanded to include ownership domain (*match_domain*) to account for potential bias due to ownership differences. Missing observations were replaced independently five times for the *match* and *match_domain* techniques and five separate stratified estimates from the mean and variance were compiled following Rubin (1987):

$$\overline{C} = k^{-1}\sum_{k=1}^{t}\overline{C}^{k}, \tag{6}$$

and

$$V(\overline{C}) = k^{-1}\sum_{k=1}^{t}V(\overline{V}^{k}) + t^{-1}(t+1)\sigma^{2}_{\overline{V}}, \tag{7}$$

where $\overline{C}^k$ and $V(\overline{C}^k)$ were the stratified estimates of the mean and variance for the kth completion for the data set and $\sigma^2_{\overline{V}}$ was the variance among the five stratified estimates of the mean.

Building on the *match* and *match_domain* replacement techniques, a k-nearest neighbor method was employed where the mean of the five estimates ($k = 5$) was imputed to the missing observation for each stratum (*match_mean*) and stratum + domain (*match_domain_mean*).

2.4.5. Replace with stratum estimates

Another compensation strategy relied on mean estimates calculated for observed plots within each stratum (*stratum*) to replace missing observations (Table S.1). This approach has the potential to contribute to bias in the estimator if missing observations come from a particular ownership, so the method was further divided by ownership domain (*stratum_domain*). Assuming the nonresponse patterns identified in Patterson et al. (2012) hold and the majority of missing observations are due to denied access on private forest land, the *stratum_domain* approach should yield a nearly unbiased estimator of the stratum mean by ownership domain. That said, the variance estimator of the *stratum_domain* will be biased downward because all missing observations are replaced with the same stratum estimate. To assess that bias, an additional set of estimates were calculated with an independent and randomly generated number from a normal distribution with σ representing uncertainty from the *stratum* and *stratum_domain* means as well as the observed variability around the means.

Expanding on the *stratum* and *stratum_domain* approaches, the same imputation strategy was employed – however, rather than randomly selecting a $\overline{C}$ estimate from the pool of five similar observations, the mean of the five estimates was imputed to the missing observation for each stratum (*stratum_mu*) and stratum + domain (*stratum_domain_mu*) following models (6) and (7).

2.5. Simulation and analysis

The most common reason for nonresponse in the current inventory in the Lake States region was denied access on private forest land. Nonresponse patterns in the Lake States are consistent with Patterson et al. (2012), who recently documented denied access rates for plots on private forest lands from 6% to 21% across most of the states in the US. To emulate nonresponse patterns in the Lake States region, nonresponse was simulated on private forest land across the range of nonresponse (i.e., 0–25%) documented by Patterson et al. (2012) and the performance of each compensation strategy was compared to the *baseline* and other compensation approaches.

2.5.1. Simulation process

Nonresponse was simulated using all observations on forest land from the current FIA inventory for the Lake States region. Given nonresponse patterns in the region, the proportion of missing observations was calculated as the ratio of the number of observations classified as missing on private forest land to the total number of observations on all forest land – no missing observations were simulated on public land. The simulation of nonresponse (i.e., denied access) on private forest land was assumed to be random. The simulation process included: (1) calculating weights for each strata and strata + domain, (2) calculating the nonresponse adjustment factors, (3) calculating the *baseline* stratified estimates, $\overline{C}$ and $V(\overline{C})$ for all observations, (4) randomly selecting observations as missing on private forest land, (5) calculating $\overline{C}$ and $V(\overline{C})$ for each compensation approach, (6) repeating steps 4 and 5 1000 times, and (7) retaining all estimates across the range of simulated nonresponse for each compensation strategy.

2.5.2. Statistical analyses

The means of the distributions of the 1000 simulated stratified estimates, $\overline{C}$ and $V(\overline{C})$ were calculated for each compensation strategy. The standard error of $\overline{C}$ was estimated for each approach as the square root of mean $V(\overline{C})$ for the 1000 simulations. The precision and bias of the estimators associated with each compensation strategy were compared with the *baseline* mean estimates using this metric, which is equivalent to a conservative t-test for the mean difference of two groups.

3. Results

Data analyses were restricted to 12146 Phase 2 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. Approximately 39.7% (4818 plots) of the plots in the study were on public forest land while 60.3% (7328 plots) were on private land. The proportion of missing observations on private forest land for the 2007–2011 inventory period was 3.4% (90 observations) in Wisconsin, 6.1% (114 observations) in Minnesota, and 10.4% (293 observations) in Michigan. Nearly all missing observations (91.3%) were due to private landowners denying field crews access to lands, with the remaining sample locations deemed hazardous (5.4%), skipped because measurement of the plots was not completed prior to the time an annual sampling panel was closed at the end of the field season and submitted for processing (0.4%), or not sampled for a reason other than one of the specific reasons listed in the FIA database (2.8%).

The distribution of missing plot observations by county suggests that denied access areas may not be uniformly distributed throughout the study region (Fig. S.1). A Pearson's product-moment correlation analysis was conducted to assess the linear relationship between county-level nonresponse and human population using 2010 census data (USDC, 2013). There was no statistically significant relationship between the two variables in Minnesota or Wisconsin but in Michigan, where there was a larger number of forested plots relative to Minnesota or Wisconsin, there was a positive correlation between county-level nonresponse and human population by county, $r = 0.35$, $n = 83$, $p < 0.001$, suggesting that nonresponse may increase with increasing population.

The means of the distributions of stratified estimates ($\overline{C}$) across the range of simulated nonresponse stabilized well before the 1000 simulations for each state and missing data approach in the study with respect to the estimates (Figs. 2, S.2 and S.5), suggesting the 1000 simulations was more than adequate for the analysis. The means of the distributions of simulated stratified estimates ($\overline{C}$) responded differently with increasing nonresponse for each state in the analysis (Figs. 3, S.3 and S.6). In all this study's states, replacing missing observations with previous estimates ($\overline{C}$) resulted in substantial divergence from the *baseline* and the standard errors (Table 1) indicated statistically significant differences between $\overline{C}$ estimates and the *baseline* for all but three estimates (i.e., *last_domain_adj* and *last_domain_unadj* estimates were within one standard error of the *baseline* for the 0.15, 0.20, and 0.25 simulated nonresponse in Minnesota) so the *last* approaches were omitted from the rest of the analysis. In Michigan, as the proportion of nonresponse increased the means of the distributions of simulated stratified estimates ($\overline{C}$) increased with respect to the *baseline* (Fig. 3). In Minnesota, the opposite occurred, with most of the means of the distributions of simulated stratified estimates ($\overline{C}$) decreasing as the proportion of nonresponse increased with respect to the *baseline* (Fig. S.3). In Wisconsin, means of the distributions of simulated stratified estimates ($\overline{C}$) from the *stratum_adj*, *stratum_domain_adj*, *stratum_mu_adj*, *and stratum_mu_domain_adj* approaches generally increased with increasing simulated nonresponse while the remaining approaches generally decreased with respect to the *baseline* (Fig. S.6).

While relatively large deviations with respect to the *baseline* were observed for the means of the distributions of simulated stratified estimates () with increasing nonresponse, the estimates from all of the approaches, with the exception of the *last* approaches, were well within one standard error of the *baseline* mean in all three states (Figs. 4, S.4 and S.7). This result indicated there were no statistically significant differences between the missing data approaches and the *baseline*. There were, however, several apparent trends among the remaining estimates for the different missing data approaches. The *match* techniques, particularly those that did not account for ownership domain, performed the best among the missing data approaches across the range of nonresponses (Figs. 4, S.4 and S.7). The *ignore*, *stratum*, *and stratum_mu* approaches performed similarly across the range of simulated nonresponse with no clear improvement when accounting for ownership. The unadjusted means of the distributions of simulated stratified estimates ($\overline{C}$) were slightly smaller in all cases than the adjusted estimates which was expected given that $\bar{P}_{oj}$ is always greater than 1 when missing plots are accounted for in the estimates.

4. Discussion

All of the compensation strategies, with the exception of the *last* approaches, resulted in means of the distributions of stratified estimates of $\overline{C}$ within one standard error of the *baseline* estimates across the range of simulated nonresponse in the three states included in the study. While performance of the missing data approaches across the range of simulated nonresponse was more than adequate with respect to the *baseline* estimates in all but a few cases, there was not a single approach that performed optimally in all three states. This result highlights both potential differences in nonresponse patterns across states as well as differences in the forest inventory data within each state. Given the nationally consistent sampling protocols in the FIA program and the nonsampling codes listed in the inventory data for each plot and condition across the nation, it is reasonable to assume that factors beyond sampling error led to differences between states. The number of sampled plots and observed nonresponse in Minnesota and Wisconsin were similar, however the stratum weights by percent canopy cover and estimates of $\overline{C}$ differed, which may have contributed to differences in the statistical performance of the compensation strategies between the two states. These differences were driven, in large part, by differences in forest type composition and live tree volume (Miles and VanderSchaaf, 2012; Perry, 2013) and perhaps, to a lesser extent, by differences in forest management practices in

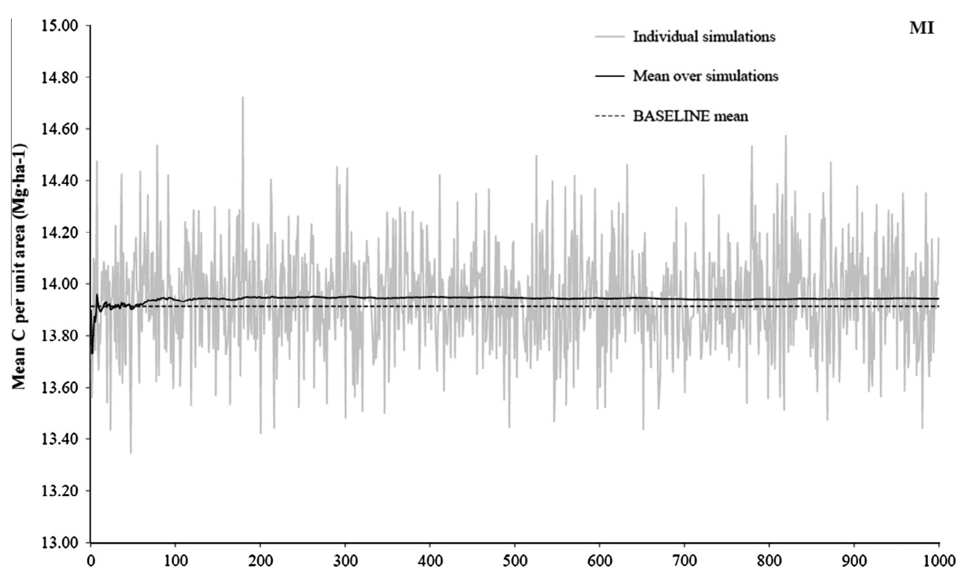

Fig. 2. Means of distributions of *match_random_adj* stratified estimates of mean carbon density (Mg ha^{-1}) by simulation – 20% simulated nonresponse for Michigan.

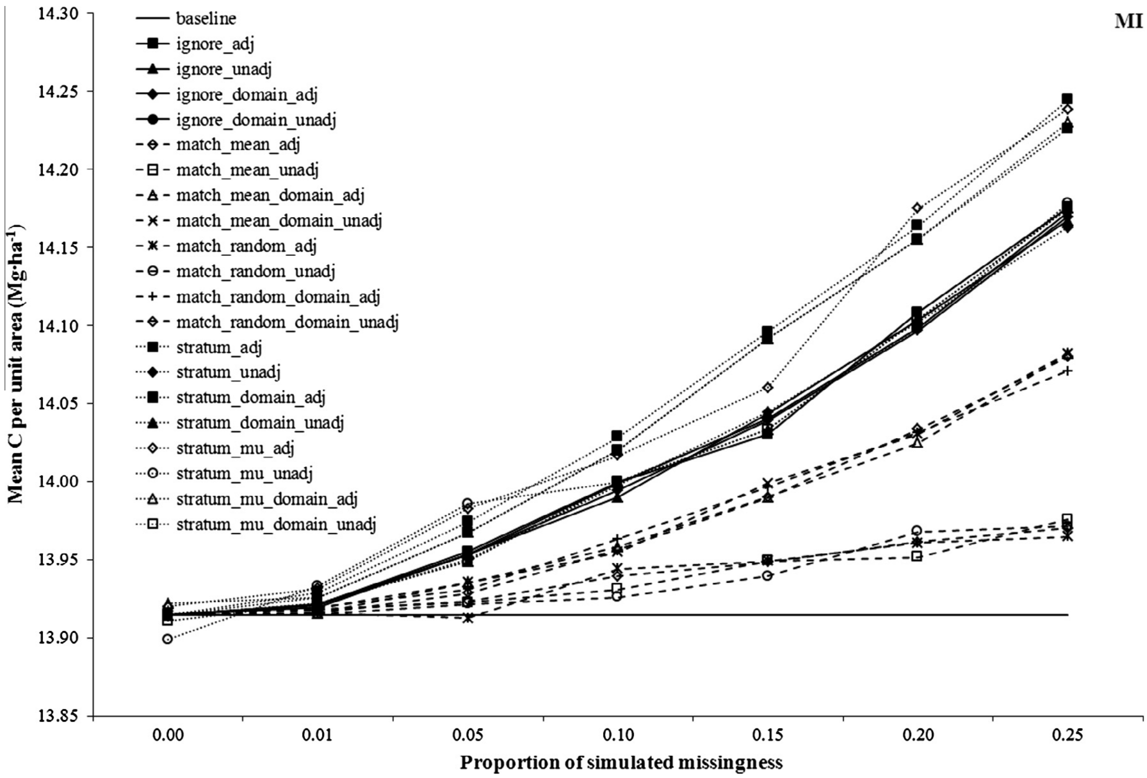

Fig. 3. Means of distributions of stratified estimates of mean carbon density (Mg ha^{-1}) for each missing data approach by proportion of simulated nonresponse (0.00–0.25) for Michigan. Note that estimates compiled for each of the compensation strategies may overlap.

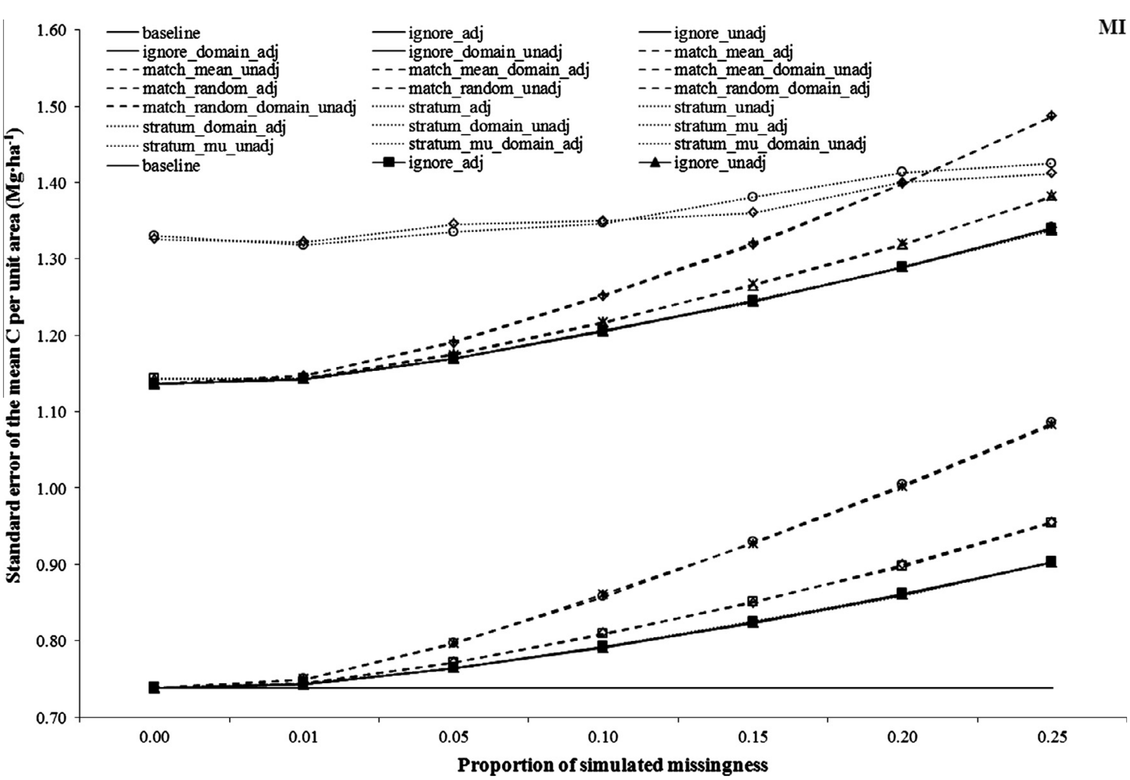

Fig. 4. Means of distributions of stratified estimates of the standard error of mean carbon density (Mg ha^{-1}) by proportion of simulated nonresponse (0.00–0.25) for Michigan. Note that estimates compiled for each of the compensation strategies may overlap.

the 2 states. In Michigan, where the observed nonresponse was 10.4%, there was a greater number of plots when compared to Minnesota and Wisconsin and stratified estimates of $\overline{C}$ were also larger. The latter was not surprising given the differences in forest type composition and greater live tree volume in the state (Pugh, 2013) relative to Minnesota and Wisconsin, but interestingly, there was also a weak, though statistically significant positive linear relationship between observed nonresponse and human population by county in Michigan. This positive correlation, which suggests that nonresponse (i.e., denied access) increases with increasing population, may be coincidental although there is evidence that as resources (e.g., forest land) become limited, in this case due to increasing population, landowners may be less willing to allow access to those resources (Anderson and Hill, 1975).

The majority of missing observations in the study were the result of denied access on what was assumed to be privately owned forest land – no ownership information exists in the FIA database for missing observations. While the large percentage of nonresponse may support the decision to restrict simulated nonresponse to private forest land, the compensation strategies that incorporated the ownership domain performed no better and, in many cases, performed worse than the same method calculated at the stratum level (i.e., estimates grouped according to percent canopy cover). In a similar study using FIA data from Indiana, McRoberts (2003) found that accounting for ownership improved stratum estimates relative to the study *baseline*. There were no statistically significant differences in estimates of live tree C density by ownership in this study (Fig. S.8), which suggests that accounting for ownership in the estimation process did not improve the statistical performance of estimates with respect to the *baseline* or those estimates that did not account for ownership. That said, the differences between studies may suggest that the ownership assumptions in this study were not appropriate or require more careful evaluation. Furthermore, the original imagery used to assign plots to strata may need to be analyzed to determine ownership for sampled forest land plots that are missing data, including ownership information. This would require a substantial investment in FIA program resources to accurately assign each sampled plot with missing data to the correct ownership. Such an investment in time and money must be weighed against the limited potential for improvement over estimates that do not incorporate ownership. The differences between studies may also be due to differences in forest conditions, forest management practices, ownership patterns, or some combination of these factors between regions.

Currently, FIA field crews typically make a single attempt to obtain permission to gain access to plots on private forest lands via letter, phone call, or location visit. Landowners may deny the field crews access to measure plots on their private forest land and in those cases plots are listed as denied access. Once an attempt has been made and the plot cannot be observed and measured, it is recorded as nonsampled and given a reason code in the FIA database. At this time there is no formal process in the FIA program for contacting private forest landowners to gain access to their land. Denied access may be reduced if multiple attempts are made, perhaps initially by mail, and subsequently with a phone call or follow-up letter. In a wetlands study conducted by the EPA in North Dakota, Lesser (2001) found that on the sites where permission to access was granted in 1995 and 1996, 68% and 82%, respectively, of consents were obtained by the initial mail contact. An additional 32% (1995) and 18% (1996) of landowners granted access after follow-up telephone calls. In this study, there was a formal process established to reduce nonresponse well before site visits to achieve an adequate sample and accurate inference to wetlands in North Dakota. There are several well-established strategies commonly used in surveys to reduce nonresponse (Groves, 1989; Lessler and Kalsbeek, 1992; Dillman, 2000). These include using priority mailings that ensure the request was received by the landowner, making multiple attempts to convey the importance of the survey, describing the information needed and how it will be used in the mailing, personalization of the correspondence, and/or material or financial incentives. These techniques have all been shown to reduce nonresponse in sample surveys (Dillman, 2000) however the time and costs required to gain access to private forest land must be weighed against the efforts to compensate for nonresponse in the annual inventory.

As this work is expanded to the periodic inventories to improve baseline estimates of C in the US national greenhouse gas inventory (NGHGI) as signatories to UNFCCC, assessing the distribution of missing plot observations and the range of nonresponse will be necessary. There are likely patterns of nonresponse within the NGHGI of the US, albeit for a variety of different reasons, which may require subdivision beyond basic stratification to account for bias. Federal forest lands across the US, for example, tend to have higher C density estimates than private forest lands (Heath et al., 2011a,b). In extreme cases, where entire National Forests may be missing from the periodic inventories, using an aggregate estimate of C density that does not account for ownership (e.g., intensified harvesting due to differences in management practices), may result in substantially different estimates.

The US currently uses a Tier 3 approach to forest C estimation following the *IPCC Good Practice Guidance for LULUCF* (IPCC, 2006). This approach requires the use of a combination of empirical inventory-based information and models to estimate forest C stocks and stock changes in the NGHGI. Inevitably, nonresponse occurs in large-scale strategic inventories such as the one implemented by the US national forest inventory program. The results from this study can be used to compensate for missing observations throughout the periodic inventories in the US with the goal of enhancing the consistency between periodic and annual estimates of forest C stocks and stock changes dating back to the 1990 NGHGI baseline year. Better alignment between periodic and annual inventory data is expected to facilitate, and be facilitated by, efforts to improve Tier 3 reporting by incorporating disturbance history and forest land area changes via auxiliary data sources (e.g., Landsat satellite imagery, Schroeder et al., 2012) and associated products (e.g., LandTrendr, Cohen et al., 2010) over the NGHGI reporting period. In contrast to the national forest inventory situation in the US and a relatively small number of other countries (e.g., New Zealand), the nations that utilize Tier 1 or 2 estimation methods may not be subjected to the issue of nonresponse, although see Beets et al. (2011). In many European nations, national forest inventory crews have the legal right to measure plots on private land without landowner permission. This eliminates a major factor potentially contributing to nonresponse in forest inventories. Nonresponse due to denied access is also likely to be low in nations where the majority of forest land is held by the government (e.g., Canada). In both cases, hazardous conditions due to terrain, season, or both are likely to be the primary factors contributing to nonresponse in national forest inventories. In those cases, techniques to compensate for nonresponse must account for the conditions (e.g., high elevation, steep grade) restricting access to the sample locations.

5. Conclusions

This study represents an initial attempt to document the performance of several well-established approaches to compensate for missing inventory data within the context of NGHGI development. As nations endeavor to empirically monitor forest C stocks through national forest inventories, they will increasingly face the reality of nonresponse in forest inventories. Nonresponse varies across the landscape, hence it can affect which techniques are selected to compensate for nonresponse in forest C inventories. Although study results varied by state, due to the requirements of compiling a NGHGI annually it was deemed that the *ignore* techniques (*ignore_unadj*, *ignore_adj*, *ignore_domain_unadj*, and *ignore_domain_adj*), where non-observed samples are removed from estimation procedures, provided the optimal combination of statistical performance and simplicity. Future work evaluating the utility of imputation strategies that incorporate auxiliary data sources such as Landsat products into forest C estimation over the entire NGHGI reporting period should be explored to reduce uncertainties associated with NGHGI baselines since 1990.

Acknowledgements

The authors thank Dale Gormanson for assistance with map and dataset development. We also thank John Stanovick, Paul Van Deusen, and Andrew Hill and three anonymous reviewers for comments that improved the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foreco.2013.12.031.

References

Anderson, T.L., Hill, P.J., 1975. The evolution of property rights: a study of the American West. J. Law Econ. 18 (1), 163–179.
Andridge, R.R., Little, R.J., 2010. A review of hot deck imputation for survey nonresponse. Int. Stat. Rev. 78 (1), 40–64.
Barrett, T., Maltamo, M., 2012. Missing data in forest ecology and management: advances in quantitative methods. For. Ecol. Manage. 272, 1–2.
Bechtold, W.A., Patterson, P.L., 2005. The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC, p. 85.
Beets, P.N., Brandon, A.M., Goulding, C.J., Kimberley, M.O., Paul, T.S.H., Searles, N., 2011. The inventory of carbon stock in New Zealand's post-1989 planted forest for reporting under the Kyoto protocol. For. Ecol. Manage. 262 (6), 1119–1130.
Birdsey, R., 2004. Data gaps for monitoring forest carbon in the United States: an inventory perspective. Environ. Manage. 33 (1), S1–S8.
Cochran, W.G., 1977. Sampling Techniques, third ed. Wiley, New York.
Cohen, W.B., Yang, Z., Kennedy, R., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSnyc–Tools for calibration and validation. Remote Sensing Environ. 114, 2911–2924.
Dillman, D.A., 2000. Mail and Internet Surveys, second ed. John Wiley & Sons, New York, NY, USA.
Eskelson, B.N.I., Temesgen, H., Lemay, V., Barrett, T.M., Crookston, N.L., Hudak, A.T., 2009. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. Scan. J. For. Res. 24 (3), 235–246.
Gillespie, A.J.R., 1999. Overview of the annual inventory system established by FIA. J. For. 97, 16–20.
Goeking, S.A., Patterson, P.L., 2013. Stratifying to reduce bias caused by high nonresponse rates: a case study from New Mexico's forest inventory. Res. Note RMRS-RN-59. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO. 22 p.
Groves, R.M., 1989. Survey Errors and Survey Costs. John Wiley & Sons, New York, NY, USA.
Heath, L.S., Hanson, M.H., Smith, J.E., Smith, W.B., Miles, P.D., 2009. Investigation into calculating tree biomass and C in the FIADB using a biomass expansion factor approach. In: McWilliams, W., Moisen, G., Czaplewski, R. (Eds.), Forest Inventory and Analysis (FIA) Symposium 2008. RMRS-P-56CD U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, p. 839.
Heath, L.S., Smith, J.E., Skog, K.E., Nowak, D.J., Woodall, C.W., 2011a. Managed forest carbon estimates for the US greenhouse gas inventory, 1990–2008. J. For. 109 (3), 167–173.
Heath, L.S., Smith, J.E., Woodall, C.W., Azuma, D.L., Waddell, K.L., 2011b. Carbon stocks on forestland of the United States, with emphasis on USDA Forest Service ownership. Ecosphere 2 (1), art6.
Homer, C., Huang, C., Yang, L., Wylie, B., Coan, M., 2004. Development of a 2001 national landcover database for the United States. Photogram. Eng. Remote Sens. 44, 3999–4005.
Intergovernmental Panel on Climate Change (IPCC), 2006. Good Practice Guidance for Land Use, Land-Use Change and Forestry. <http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html> (accessed 26.07.13).
Lemeshow, S., 1985. Nonresponse (in sample surveys). In: Kotz, S., Johnson, N.L. (Eds.), Encyclopedia of Statistics, vol. 6. John Wiley & Sons, New York, pp. 333–336.
Lesser, V.M., 2001. Applying survey research methods to account for denied access to research sites on private property. Wetlands 21 (4), 639–647.
Lessler, J.T., Kalsbeek, W.D., 1992. Nonsampling Error in Surveys. John Wiley & Sons, New York, NY, USA.
Little, R.J., 1995. Modeling the drop-out mechanism in repeated-measures studies. J. Am. Stat. Assoc. 90 (431), 1112–1121.
Little, R.J., Rubin, D.B., 2002. Statistical Analysis with Missing Data, vol. 4. Wiley, New York.
McRoberts, R.E., 2003. Compensating for missing plot observations in forest inventory estimation. Can. J. For. Res. 33, 1990–1997.
McRoberts, R.E., Westfall, J.A., 2014. The effects of uncertainty in model predictions of individual tree volume on large area volume estimates. For. Sci. <http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html>.
McRoberts, R.E., Hansen, M.H., Smith, W.B., 2010. United States of America. In: Tomppo, E., Gschwanter, T., Lawrence, M., McRoberts, R.E. (Eds.), National Forest Inventories: Pathways for Common Reporting. Chapter 37. Springer, Heidelberg, p. 607.
Miles, P.D., VanderSchaaf, C.L., 2012. Minnesota's forest resources, 2012. Res. Note NRS-175. U.S. Department of Agriculture, Forest Service, Northern Research Station. 4 p.
Patterson, P.L., Coulston, J.W., Roesch, F.A., Westfall, J.A., Hill, A.D., 2012. A primer for nonresponse in the US forest inventory and analysis program. Environ. Monit. Assess. 184, 1423–1433.
Perry, C.H., 2013. Wisconsin's forest resources, 2012. Res. Note NRS-193. U.S. Department of Agriculture, Forest Service, Northern Research Station. 4 p.
Pugh, S.A., 2013. Michigan's forest resources, 2012. Research Note NRS-165. U.S. Department of Agriculture, Forest Service, Northern Research Station. 4 p.
Reams, G.A., McCollum, J.M., 2000. The use of multiple imputation in the southern annual forest inventory system. In Integrated tools for natural resources inventories. In: Hansen, M.H., Burk, T. (Eds.), Proceedings of the IUFRO Conference, 16–20 August 1998, Boise, Idaho. Gen. Tech. Rep. NC-212 USDA Forest Service, North Central Research Station. pp. 228–233.
Rubin, D.B., 1987. Multiple Imputation in Non-response Surveys. Wiley, New York.
Rubin, D.B., 1996. Multiple imputation after 18+ years. J. Am. Stat. Assoc. 91 (434), 473–489.
Ruefenacht, B., Finco, M.V., Nelson, M.D., et al., 2008. Conterminous US and Alaska forest type mapping using forest inventory and analysis data. Photogram. Eng. Remote. Sens. 74, 1379–1388.
Sande, I.G., 1982. Imputation in surveys: coping with reality. Am. J. Stat. 36, 145–152.
Särndal, C.E., Swensson, B., Wretman, J., 1992. Model-assisted Survey Sampling. Springer-Verlag, New York.
Schroeder, T.A., Moisen, G.G., Healey, S.P., Cohen, W.B., 2012. Adding value to the FIA inventory: combining FIA data and satellite observations to estimate forest disturbance. In: Morin, R., Likens, G. (Eds.), Moving From Status to Trends: 2012 Forest Inventory and Analysis (FIA) Symposium. Gen. Tech. Rep. NRS-P-105. U.S. Department of Agriculture, Forest Service, Northern Research Station, 478 p.
Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E., 2010. National Forest Inventories: Pathways for Common Reporting. Springer, Verlag.
USDA Forest Service, 1970. Operational procedures. First report of the Blue Ribbon Panel on forest inventory and analysis. Forest Service Handbook 4809.11. Washington D.C. 11:1–1. (Chapter 10).
USDA Forest Service, 2012. Forest Inventory and Analysis fiscal year 2011 business report. FS-999. Washington D.C.
USDA Forest Service, 2013. FIA Database Description and Users Manual for Phase 2., Ver. 5.1.5. <http://www.fia.fs.fed.us/library/database-documentation/current/ver5.1.5/FIADB_user%20manual_5-1-5_p2_02_2013_FINAL.pdf>. (accessed 02.05.13).
USDC United States Census Bureau, 2013. 2010 census data. <http://www.census.gov/2010census/data/> (accessed 03.05.13).
US Environmental Protection Agency, 2013. Inventory of U.S. greenhouse gas emissions and sinks: 1990–2011. Chapter 7. Land use, landuse change, and forestry. Annex 3.12. Methodology for estimating net carbon stock changes in forest land remaining forest lands. U.S. EPA, Washington, DC <http://www.epa.gov/climatechange/ghgemissions/usinventoryreport.html>. (accessed 24.10.13).
Van Deusen, P.C., 1997. Annual forest inventory statistical concepts with emphasis on multiple imputation. Can. J. For. Res. 27, 379–384.
Woodall, C.W., 2012. Where did the U.S. forest biomass/carbon go? J. For. 110, 113–114.
Woodall, C.W., Heath, L.S., Domke, G.M., Nichols, M.C., 2011. Methods and equations for estimating aboveground volume, biomass, and carbon for trees in the U.S. forest inventory, 2010. Gen. Tech. Rep. NRS-88. U.S. Department of Agriculture, Forest Service, Northern Research Station, p. 30.