Forest Service
US DEPARTMENT OF AGRICULTURE

Northern Research Station | General Technical Report NRS-207 | December 2022

Sampling and Estimation Documentation for the Enhanced Forest Inventory and Analysis Program: 2022

James A. Westfall, John W. Coulston, Gretchen G. Moisen, and Hans-Erik Andersen, compilers

Citation

Westfall, J.A.; Coulston, J.W.; Moisen, G.G.; Andersen, H.-E., compilers. 2022. **Sampling and estimation documentation for the Enhanced Forest Inventory and Analysis Program: 2022**. Gen. Tech. Rep. NRS-GTR-207. Madison, WI: U.S. Department of Agriculture, Forest Service, Northern Research Station. 129 p. https://doi.org/10.2737/NRS-GTR-207.

Abstract

The Forest Inventory and Analysis (FIA) program of the Forest Service, an Agency of the U.S. Department of Agriculture, provides what is arguably the most valuable forest resource dataset in the United States. These data are the basis for numerous inquiries across a wide range of forest-related attributes at various spatial and temporal scales. While user-friendly analytical tools are publicly available to facilitate the use of the data without expert knowledge, there is a need for detailed documentation of the underlying sampling and estimation procedures. The audience for this information entails the entire spectrum of both internal and external FIA data consumers. This document clarifies some aspects of existing documentation, provides the sampling and estimation methods used for key program areas including Urban FIA, National Woodland Owner Survey, Timber Products Output, and Carbon, and provides an examination of burgeoning estimation topics relevant to the FIA program and its users. A broad overview is provided on several advanced estimation approaches of particular interest to the FIA community. While the exposition for each topic is necessarily coarse, links to more detailed research and informational material are provided for readers desiring to further study a specific area of interest.

KEY WORDS: urban forest, timber products, woodland owner, remote sensing, forest carbon

Acknowledgments

The authors are indebted to Aaron Weiskittel (University of Maine) for substantial effort in congregating the expert review committee and providing overall leadership of the review process. We greatly appreciate the careful attention and thoughtful comments of David Affleck (University of Montana), Andrew Finley (Michigan State University), John Paul McTague (Southern Cross Forest Biometrics), Stephen Prisley (National Council for Air and Stream Improvement), and Stephen Fairweather (Fairweather Biometrics) that resulted in considerable improvements to the manuscript. The authors are grateful to the innumerable individuals who have contributed to development of the topic areas over the course of many years (or decades) and to the reviewers who provided valuable insights for improving the document.

Cover Photo

Cover images, from top: Forest in Hardy County, WV, south of Moorefield, WV. USDA Forest Service photo by Karen Kubly; Urban Forest Inventory and Analysis sampling. USDA Forest Service photo by Sjana Schanning; Loading a log truck. USDA Forest Service photo by Consuelo Brandeis; Forest carbon attributes. USDA image by Grant Domke; Postal boxes with National Woodland Owner Surveys. USDA photo by Brett Butler; Remote sensing map. USDA image by Barry T. Wilson.

Manuscript received for publication February 2021

Published by:
USDA Forest Service
One Gifford Pinchot Drive
Madison, WI 53726
December 2022

About the Authors & Compilers

James A. Westfall, Research Forester, USDA Forest Service, Northern Research Station, York, PA

John W. Coulston, Project Leader, USDA Forest Service, Southern Research Station, Blacksburg, VA

Gretchen G. Moisen, Research Forester, USDA Forest Service, Rocky Mountain Research Station, Ogden, UT

Hans-Erik Andersen, Research Forester, USDA Forest Service, Pacific Northwest Research Station, Seattle, WA

Paul L. Patterson, Mathematical Statistician, USDA Forest Service, Rocky Mountain Research Station, Ft. Collins, CO

Charles T. Scott, Program Manager (retired), USDA Forest Service, Northern Research Station, West Chester, PA

Christopher B. Edgar, University of Minnesota, Department of Forest Resources, St. Paul, MN

Brett J. Butler, Research Forester, USDA Forest Service, Northern Research Station, Amherst, MA

Jesse Caputo, Research Forester, USDA Forest Service, Northern Research Station, Amherst, MA

Grant M. Domke, Research Forester, USDA Forest Service, Northern Research Station, St. Paul, MN

Brian F. Walters, Forester, USDA Forest Service, Northern Research Station, St. Paul, MN

James E. Smith, Research Plant Physiologist, USDA Forest Service, Northern Research Station, Durham, NH

Christopher W. Woodall, Research Forester, USDA Forest Service, Northern Research Station, Durham, NH

David M. Bell, Research Forester, USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR

Tracey S. Frescino, Forester, USDA Forest Service, Rocky Mountain Research Station, Ogden, UT

Kelly S. McConville, Reed College, Division of Mathematical and Natural Sciences, Portland, OR

Ronald E. McRoberts, University of Minnesota, Department of Forest Resources, St. Paul, MN

Barry T. Wilson, Research Forester, USDA Forest Service, Northern Research Station, St. Paul, MN

Contents

Chapter 1: Overview 1
James A. Westfall, John W. Coulston, Gretchen G. Moisen, and Hans-Erik Andersen

Chapter 2: Foundational Documentation 4
James A. Westfall, John W. Coulston, Paul L. Patterson, and Charles T. Scott

Chapter 3: Urban Forest Inventory and Analysis 22
Christopher B. Edgar

Chapter 4: National Woodland Owner Survey 35
Brett J. Butler and Jesse Caputo

Chapter 5: Timber Products Output 50
John W. Coulston

Chapter 6: FIA Carbon Attributes 62
Grant M. Domke, Brian F. Walters, James E. Smith, and Christopher W. Woodall

Chapter 7: Emerging Alternative Estimators 76
Gretchen G. Moisen, Hans-Erik Andersen, David M. Bell, John W. Coulston, Tracey S. Frescino, Kelly S. McConville, Ronald E. McRoberts, Paul L. Patterson, James A. Westfall, and Barry T. Wilson

Appendix 1: Forest Area Estimation and Area Control Revisited 102
John W. Coulston and James A. Westfall

Appendix 2: Ratio-to-Size Estimation 123
Charles T. Scott and James A. Westfall

Appendix 3: Metric Equivalents 129

Chapter 1: Overview

James A. Westfall, John W. Coulston, Gretchen G. Moisen, and Hans-Erik Andersen

The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture's Forest Service is responsible for implementation of the national forest inventory of the United States. Initially established under the 1928 McSweeney-McNary Forest Research Act (Public Law 70-466), FIA has been conducting forest inventories for nearly a century. A key change to the program resulted from the 1998 Farm Bill (Public Law 105-185), which prescribed the adoption of the annualized FIA sampling paradigm that many users are familiar with today. Currently, the program is implemented via four regional units corresponding with geographic delineations of the United States (Northern, Southern, Rocky Mountain, and Pacific Northwest) in coordination with the national office (Washington, DC) (Fig. 1). Key programmatic outputs include comprehensive reports for each state published every 5 years—as mandated in the 1998 Farm Bill (Public Law 105-185). The primary focus of these reports is on analyses of statistical estimates of forest resources that arise from the base forest inventory effort; however, within these reports and in other publications are findings from other FIA program surveys and research endeavors designed to address critical information needs or burgeoning topics of interest.

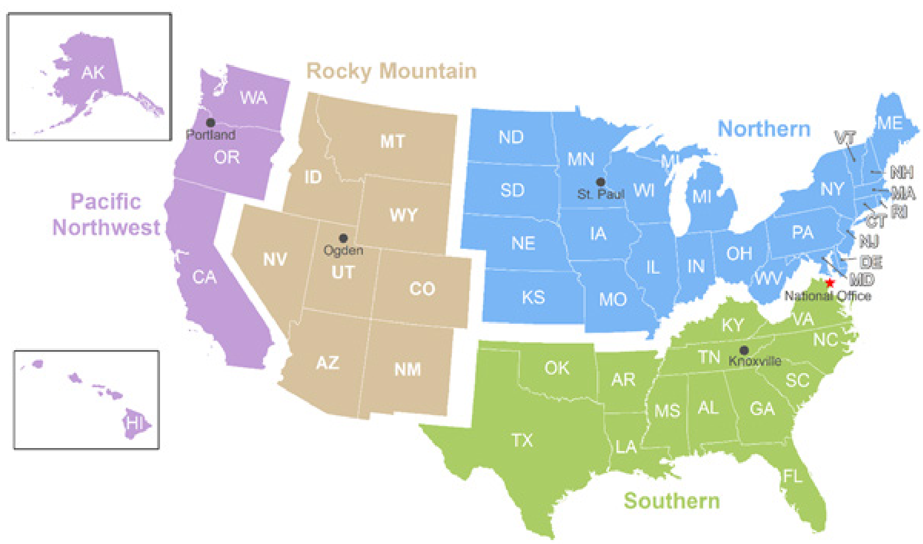

Figure 1.—Map showing geographic extent of the four regional FIA units that comprise the national program. Created by Greg Liknes, USDA Forest Service.

In the two decades subsequent to the 1998 Farm Bill, FIA has continually enhanced not only the forest inventory effort but also its other two core functions of conducting the National Woodland Owner Survey (NWOS) and Timber Products Output (TPO) studies. The 2014 Farm Bill (Public Law 113-79) emphasized further advancement of FIA's ongoing activities and also provided directives for reporting of status and trends of biomass/carbon stocks and land use/land cover, improving statistical precision for sub-state areas, and initiating inventories in urban environments. To address these mandates, the FIA program identified a set of topical areas to focus on further development and refinement. It is important for scientific rigor and transparency to document the history and current status of these central topics, particularly to the extent that similarities and differences with standard FIA inventory procedures (Bechtold and Patterson 2005) are clearly conveyed. Thus, the purpose of this report is to chronicle the fundamental processes and procedures used in several prominent supplementary aspects of the FIA program. Specifically, the relevant subject matter covers the following:

- Urban FIA.
- National Woodland Owner (NWOS).
- Timber Products Output (TPO).
- FIA Carbon Attributes.
- Emerging Alternative Estimators.

The NWOS seeks to better understand population demographics through queries of forest land owners. The TPO program collects data to assess patterns in wood demand and harvesting activities in the context of mill production. Both the NWOS and TPO missions have existed within the FIA program for decades; however, their implementation and statistical methods have evolved with advances in technology and theory. The remaining topics are more recent extensions into broader themes the FIA program is inherently poised to address. The Emerging Alternative Estimators chapter, for example, covers new techniques to address land use and land cover change, small area estimation, and FIA implementation in interior Alaska. As with NWOS and TPO, work in these newer areas is ongoing to establish innovative, scientifically credible procedures as a basis for nationally consistent reporting of key outputs for policy makers, natural resource managers, and the public-at-large. For each topical area, the methods rely on fundamental statistical underpinnings for valid estimation and inference. Documentation of the statistical framework is paramount for establishing transparency and credibility of the outputs. As appropriate to the topic, elements to be addressed may include the following:

- Definition of the population.
- Methods to construct the sampling frame.
- Methods to draw the sample.
- Response design (e.g., four-point cluster plot).
- Design-based estimators.
- Methodology for nonresponse.
- Minimum sample size guidance.

In most cases, relevant information for each subject area has been documented in some form; however, it may not be easily synthesized into a clear understanding of current methods and protocols due to a combination of various outlets, formats, and fragmented dissemination over time. Thus, this publication intends to holistically examine the primary components of key FIA program outputs, as well as describe the current state of research endeavors and their future directions. As noted earlier, many aspects of the FIA program are not static as ongoing research provides improvements in processes and procedures. Recognition of this constant state of continual improvement necessitates a provision for updating the documentation as methods evolve. This is accomplished via establishment of an internet repository (https://www.fia.fs.usda.gov/library/sampling/index.php) where supplementary information can be accessed regarding the current implementation methods for the subject areas covered in this report.

Chapter 2: Foundational Documentation

James A. Westfall, John W. Coulston, Paul L. Patterson, and Charles T. Scott

Within this document, there is considerable reliance on the sampling or estimation procedures, or both, used for the base FIA inventory program that pertain to Phase 1 (stratification), Phase 2 (base sample of ground plots), and Phase 3 (subset of Phase 2; forest health measurements). Detailed documentation on these procedures can be found in Bechtold and Patterson (2005); however, for completeness, ease of reference, and establishment of notational conventions, relevant sampling and estimation techniques described in that publication are summarized here.

Sample Design

Broadly, the target population for the FIA program is defined as all land and water within the official boundaries of the United States. When the annual FIA program was implemented in 1999, spatial balance of the sample was obtained via a grid composed of 5,937 acre (2,403 ha) hexagons superimposed over the area by using a random starting location (Fig. 2). When one or more existing inventory plots were found within a hexagon, one of those points was chosen based on a predefined set of rules (Reams et al. 2005). Sample plot locations were chosen via random selection of a point for hexagons having no pre-existing plots. To accommodate a panelized design in which a portion of the plots are sampled each year and the entire sample would be completed over a specified number of years, each hexagon (with associated sample plot) was assigned to a panel by using techniques that produced approximately uniform spatial coverage of the population within each panel (Fig. 3).

Plot Design

Each plot consists of a four-point cluster where the central point corresponds with the location chosen within each hexagon and the remaining three peripheral points are dispersed at a distance of 120 feet (36.6 m) on azimuths of 120, 240, and 360 degrees (Bechtold and Scott 2005). Centered at each point are subplots having a 24 feet (7.3 m) radius (Fig. 4a). Subplot-based measurements include trees having diameter at breast height (d.b.h.) ≥ 5.0 inches (12.7 cm) and various other site attributes such as forest type and stand size. Each subplot contains a 6.8 feet (2.1 m) radius microplot with center offset 12 feet (3.7 m) at 90 degrees azimuth.

Trees having 1.0 inch (2.5 cm) ≤ d.b.h. ≤ 4.9 inches (12.4 cm) are recorded within the microplot. An optional macroplot based at subplot center with a radius of 58.9 feet (18.0 m) may also be used. The purpose of the macroplot is to measure large trees exceeding a specified d.b.h. threshold.

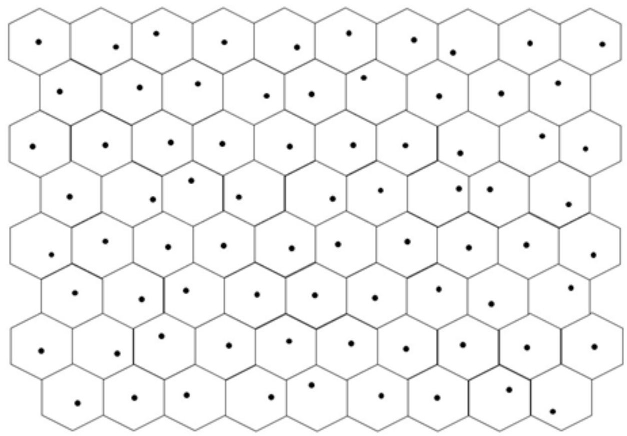

Figure 2.—Hexagonal grid with randomized plot location within each hexagon.

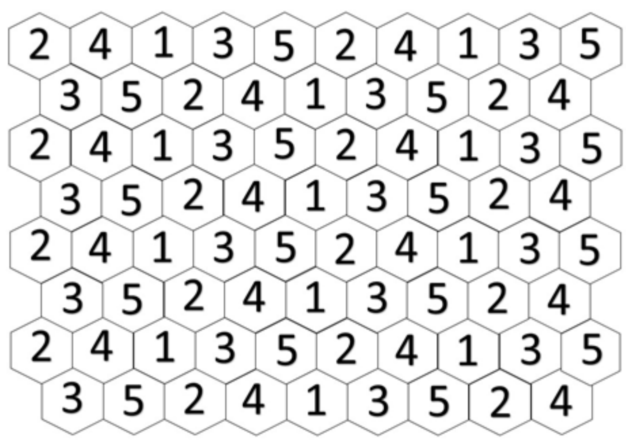

Figure 3.—Assignment of a five-panel system across the hexagonal grid.

A salient detail related to the plot design entails mapping of differing conditions occurring within the plot area. For example, the plot may straddle a forest/nonforest boundary, which would be delineated in the plot measurement process (Fig. 4b). Differences in other attributes such as ownership, forest type, stand size, and tree density may also be used to define condition boundaries. Because the population being sampled includes all land and water within the United States, areas with no forest land are often encountered. On all accessible conditions, the land use (forest, nonforest, or water) and land cover (e.g., tree cover, shrub cover, and barren) are recorded. Conditions meeting the definition of forest land trigger the collection of numerous site- and tree-level variables (USDA Forest Service 2018). Areas that are not forest are assigned a value of zero for all forest-related area and tree attributes. The boundary data are used to calculate the proportion of the plot area in each condition, which facilitates estimation for specified domains (e.g., area of oak-hickory forest type).

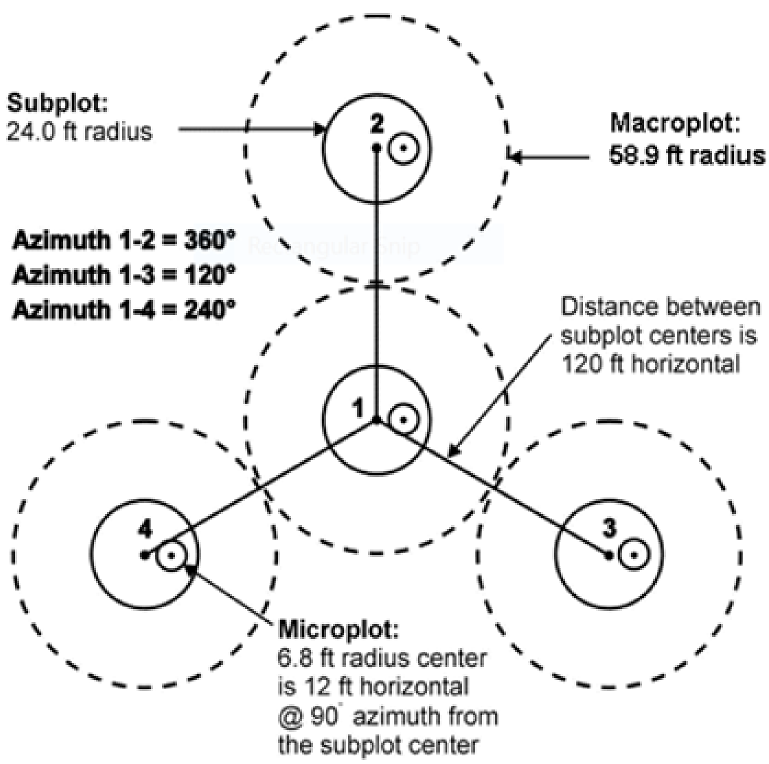

Figure 4a.—The FIA plot design.

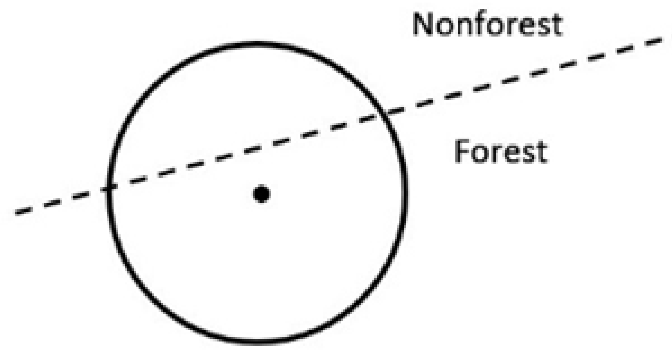

Figure 4b.—Example of forest/nonforest condition boundary intersecting an FIA subplot.

Post-Stratified Estimation

It is common practice in forest inventory to use post-stratification as a means of reducing the variance of estimates (Tomppo et al. 2010). The process entails division of the population into relatively homogenous strata via the use of remotely sensed imagery. Most contemporary applications use digital map products for ease and consistency afforded by automated processing algorithms. This approach will be the method described immediately below; however, readers interested in sampling approaches to post-stratification are directed to Westfall et al. (2019) and references therein.

Generally, the post-stratification proceeds by developing H strata based on classification of pixel values from a selected source of classified digital imagery. For example, post-strata may be defined as forest and nonforest land cover, or a set of canopy cover proportion classes. For each of the $h = 1$ to H strata, the stratum weights (W_h) are calculated based on the proportion of all pixels in the population that are assigned to stratum h. The W_h are considered as known values without error due to the complete pixel coverage of the population. Subsequently, the sample plots are assigned to strata by using the geographic location of the plot center. The ground sample plot is assigned to the same stratum as the pixel that contains the plot center point. This exercise is referred to as post-stratification because the stratification was conducted after the selection of the plot locations.

Upon completion of the post-stratification process, the key pieces of information are the stratum weights (W_h), membership of each plot *i* to a stratum *h*, and the within-stratum sample sizes (n_h). The post-stratification process is repeated each year for the entire set of plots when a new panel of plot measurements are added to the database.

Because the sample covers all lands, the FIA program is primarily interested in domain estimates relevant to forest type, stand size class, and ownership. Examples of domain estimates include the amount of forest land in the population, and the amount of tree biomass growth in longleaf pine forest types in the population. Domain estimation is described separately for area and tree-level attributes in Scott et al. (2005), the primary difference being in how the plot-level values are obtained. Thus, a generalized presentation of the estimation procedures from Scott et al. (2005) is given below with subsequent attention to appropriate calculation of the plot values. The estimate of the population mean for a domain (*d*) of interest for attribute *y* is

$$\bar{Y}_d = \sum_{h=1}^{H} W_h \bar{Y}_{hd} \qquad (1)$$

Where

$$\bar{Y}_{hd} = n_h^{-1} \sum_{i=1}^{n_h} y_{hid} \text{ , and} \qquad (2)$$

y_{hid} = the i^{th} plot observation of attribute *y* in stratum *h* for the domain of interest *d*.

The estimated variance is

$$v(\bar{Y}_d) = n^{-1}[\sum_{h=1}^{H} W_h n_h v(\bar{Y}_{hd}) + \sum_{h=1}^{H} (1 - W_h) n_h n^{-1} v(\bar{Y}_{hd}) \qquad (3)$$

Where

$$v(\bar{Y}_{hd}) = \frac{\sum_{i=1}^{n_h} (y_{hid} - \bar{Y}_{hd})^2}{n_h(n_h - 1)}. \qquad (4)$$

The first term in Equation 3 represents the typical stratified sampling variance (under proportional allocation), whereas the second term accounts for the additional variability arising from random within-stratum sample sizes.

The estimate of the total is then

$$\hat{Y}_d = A_T \bar{Y}_d \tag{5}$$

With estimated variance

$$v(\hat{Y}_d) = A_T^2 v(\bar{Y}_d) \tag{6}$$

Where

$\boldsymbol{A_T}$ = the known total area of the population.

Presentation of estimates and their uncertainty is typically approached through calculation of the standard error of the estimate as the square root of the variance:

$$s_{\hat{Y}_d} = \sqrt{v\left(\hat{Y}_d\right)} \tag{7}$$

Standard errors are then commonly used to estimate percent sampling error and construct confidence intervals, respectively:

$$SE\% = 100\frac{s_{\hat{Y}_d}}{|\hat{Y}_d|}, \text{ and} \tag{8}$$

$$\hat{Y}_d \pm t_{(n-1,1-\propto/2)}\, s_{\hat{Y}_d} \tag{9}$$

Where

t ≈ 2 for a 95 percent (α = 0.05) confidence interval.

It should be noted that large percent SE may be obtained when estimates of $\hat{Y}_d$ are small, e.g., estimates of change over time.

The preceding estimation formulae cover many common applications of FIA data. A numeric example showing calculations to estimate forestland area is provided in Appendix 1. Other types of estimates may be constructed, such as those arising from information needs on a per-tree or per-condition basis. Interested readers are referred to Scott et al. (2005) for details pertaining to these other estimation procedures.

Post-Stratified Estimation of Ratios

In addition to estimating population totals as described, it is often desired to have the estimates expressed on a per-unit-area basis, where the area is also to be estimated (e.g., biomass per acre of forest land, number of trees per acre of ponderosa pine type). These estimates are constructed via a ratio of estimated totals as described above. Typically, a tree-level attribute total serves as the numerator and the total forest land area serves as the denominator; however, other estimations may also be formulated depending on the information need. For ease of presentation, the ratio estimate will be shown using the notation already established. The form of the estimator is

$$\hat{R}_{dd'} = \frac{\hat{Y}_d}{\hat{Y}_{d'}} \quad (10)$$

Where

d, *d'* = the specified domains for the numerator and denominator, respectively.

The estimated variance is

$$v\left(\hat{R}_{dd'}\right) = \frac{1}{\hat{Y}_{d'}^2}\left[v\left(\hat{Y}_d\right) + \hat{R}_{dd'}^2 v\left(\hat{Y}_{d'}\right) - 2\hat{R}_{dd'} cov\left(\hat{Y}_d, \hat{Y}_{d'}\right)\right] \quad (11)$$

The only previously undefined term is

$$cov\left(\hat{Y}_d, \hat{Y}_{d'}\right) = A_T^2 n^{-1}\left[\sum_{h=1}^{H} W_h\, n_h cov(\bar{Y}_{hd}, \bar{Y}_{hd'}) + \sum_{h=1}^{H}(1 - W_h)\, n_h n^{-1} cov(\bar{Y}_{hd}, \bar{Y}_{hd'})\right] \quad (12)$$

Where

$$cov(\bar{Y}_{hd}, \bar{Y}_{hd'}) = \frac{\sum_{i=1}^{n_h} y_{hid} y_{hid'} - n_h \bar{Y}_{hd} \bar{Y}_{hd'}}{n_h(n_h - 1)}\,. \quad (13)$$

Summarizing Data to Plot-Level Observations

The stratified estimator and the ratio estimator are the primary approaches used in online analytical tools such as EVALIDator and DATIM (https://www.fia.fs.usda.gov/tools-data/). However, to apply the domain estimators, the data must first be appropriately summarized to the plot-level . In practice, the plot values are either a proportion (e.g., proportion of the i^{th} plot that was d = forest land) or a per-unit-area value (e.g., aggregate tree volume cubic feet per acre) for d = red maple on forest land for the i^{th} plot). For a given domain, condition- and tree-level observations are summarized to the plot level in order to construct estimates.

The initial plot-level summarization (y_{id}^*) is calculated by summing the attribute of interest across conditions occurring within the plot type (macroplots, subplots, or microplots) that the attribute of interest is measured on for the domain of interest and dividing by the fixed total area of the plot (Eq. 14):

$$y_{id}^* = \frac{\sum_{k=1}^{K} y_{i_j k} \delta_{ikd}}{a_{i_j}} \tag{14}$$

Where

δ_{ikd} = an indicator (0,1) variable which equals 1 when the attribute is in the domain of interest; 0 otherwise,

$y_{i_j k}$ = the observation for attribute y measured on j = macroplots (1 acre), subplots (1/6 acre), or microplots (1/75 acre) for plot i in condition k, and

a_{i_j} = the corresponding total plot area depending on whether the attribute is measured on j = macroplots (1 acre), subplots (1/6 acre), or microplots (1/75 acre).

This results in the plot-level summaries taking on two forms: a proportion or a density. Note the use of the domain indicator variable (δ_{ikd}) results in plot observations of zero for plots where the attribute of interest is not observed in the domain or the domain does not occur within the plot. The plot-level summary is a proportion when $y_{i_j k}$ is area-based (e.g., mapped area of forest land) because both the numerator and denominator of Equation 14 are areas. Proportion-based plot-level summaries are constructed from subplots in all FIA regions except the Pacific Northwest (PNW; California, Oregon, Washington) where the macroplots are used. The plot-level summary is a density when $y_{i_j k}$ is, for example, number of trees because the numerator in Equation 14 is number of trees and the denominator is an area. For some plot-level summaries, such as the number of live trees ≥ 1 inch d.b.h., summaries must be constructed at the microplot scale and the subplot scale because trees 1.0 inch ≤ d.b.h. < 5.0 inches are only recorded on the microplot. In this case

$$y_{id}^{*} = \frac{\sum_{k=1}^{K} y_{i_{subplot}k}\delta_{ikd}}{a_{i_{subplot}}} + \frac{\sum_{k=1}^{K} y_{i_{microplot}k}\delta_{ikd}}{a_{i_{microplot}}} = y_{id_{subplot}}^{*} + y_{id_{microplot}}^{*} \tag{15}$$

A generalized formula that can be used across any combination of plot sizes is

$$y_{id}^{*} = \sum_{j}^{J} \frac{\sum_{k=1}^{K} y_{i_jk}\delta_{ikd}}{a_{i_j}} \tag{16}$$

Adjustment for Partial Plots Outside of the Population and Stratum Assignment

Two additional steps are needed to use the plot-level summaries to construct estimates. The first step is to assign each plot to a stratum *h* (discussed previously). Once the stratum assignment has been made via spatial intersection, $y_{hid}^{*} = y_{id}^{*}$. The remaining step is to adjust y_{hid}^{*} for stratum *h* plots partially outside of the population. Portions of plots may be inaccessible due to denied access or hazardous conditions and these portions are separately mapped unobservable conditions. This means that, for example, the total area measured across the four subplots is not always 1/6 acre. Rather than treating measured plot area as a random variable, stratum-level adjustments are calculated to account for partial plots. A different adjustment is made for the subplot-level, microplot-level, and macroplot-level (PNW only) measurement areas. In general terms, the adjustments are the proportion of measured plot area to total plot area within stratum:

$$\bar{p}_{h_j} = \frac{\sum_{i=1}^{n_h} \sum_{k=1}^{K} a_{hi_jk}^{m}}{n_h a_{i_j}} \tag{17}$$

Where

$a_{hi_jk}^{m}$ = the measured area of condition *k* for plot size *j* within plot *i* assigned to stratum *h* (excluding portions that are inaccessible or out-of-population).

When only one plot-scale is used (e.g., estimated forest area based on the subplot), the final plot-level summary (y_{hid}) is

$$y_{hid} = \frac{y^*_{hi_{subplot}}}{\bar{p}_{h_{subplot}}} = \frac{\sum_{k=1}^{K} y_{i_{subplot}k}\delta_{ikd}}{\bar{p}_{h_{subplot}} a_{i_{subplot}}} \tag{18}$$

And when multiple plot-scales are used (e.g., subplot and microplot to summarize the number of trees with d.b.h. ≥ 1.0 inch), then

$$y_{hid} = \frac{y^*_{hi_{subplot}}}{\bar{p}_{h_{subplot}}} + \frac{y^*_{hi_{microplot}}}{\bar{p}_{h_{microplot}}} = \frac{\sum_{k=1}^{K} y_{i_{subplot}k}\delta_{ikd}}{\bar{p}_{h_{subplot}} a_{i_{subplot}}} + \frac{\sum_{k=1}^{K} y_{i_{microplot}k}\delta_{ikd}}{\bar{p}_{h_{microplot}} a_{i_{microplot}}}. \tag{19}$$

The plot summaries denoted by y_{hid} are the values used when constructing estimates of totals and ratios via Equations 1–13. The only exception to the above method occurs when there is a desire to estimate the area of d = nonresponse, in which case $\bar{p}_{h_j}$ =1 for all strata. Readers interested in computational examples are directed to https://www.fia.fs.usda.gov/library/sampling/index.php.

In addition to calculating estimates of current forest resource attributes, these same formulae can be applied to observed changes from plots measured at two points in time. To obtain valid change estimates, it is necessary that y_{id} is reformulated as a difference observation, i.e., a difference between proportions (e.g., difference in proportion of the i^{th} plot that was d = forest land between an initial and subsequent measurement) or a difference expressed on a per-unit-area basis (e.g., difference in tree volume (cubic feet per acre) on d = forest land for the i^{th} plot between an initial and subsequent measurement). As before, differences between observations at either the condition- or tree-level are summarized to the plot level. Note that only plots (or portions thereof) that were measured at both times can be used to estimate change; thus, plot areas intended to be sampled that remained unobserved due to factors such as denied access and hazardous conditions are excluded from change calculations (except for the specific case of area change estimation for d = nonresponse).

Implementation of Estimators

The estimators described above are applied to populations created from delineated geographic areas having a known area. It is at this level the stratification is applied, i.e., the sum of the stratum weights equals one. In FIA terminology, these areas are known as estimation units and they are usually formed from administrative boundaries such as a county, a group of counties, or in some cases large ownerships such as national forests. Estimation unit boundaries are also constrained within state boundaries, i.e., an estimation unit cannot include more than one state. Due to the independence among estimation units, state-level (or other areas composed of > 1 estimation unit) estimates can be obtained through sums of estimates arising from the estimation units therein. Estimates at the multi-state, regional, and national scales are generally obtained by summing state-level estimates. Thus, estimation units form the essential building blocks to obtain estimates for any geographic area of interest. The estimate in domain d across G estimation units is obtained from

$$\hat{Y}_d = \sum_{g=1}^{G} \hat{Y}_{gd} = \sum_{g=1}^{G} A_{Tg} \bar{Y}_{gd} \tag{20}$$

Where

A_{Tg} = total area in estimation unit g, and

G = total number of estimation units in the area of interest.

The variance of the total is the sum of the variances, since the estimation units are independent of one another.

$$v\left(\hat{Y}_d\right) = \sum_{g=1}^{G} v(\hat{Y}_{gd}) = \sum_{g=1}^{G} A_{Tg}^2 v(\bar{Y}_{gd}) \tag{21}$$

Often, interest is in the mean per unit area (acre) across the estimation units. If the mean is expressed on the basis of the total number of acres in the estimation unit (as in equation 1), the population mean is simply the population total divided by the total area in the population (a weighted mean):

$$\bar{Y}_d = \frac{\hat{Y}}{A_T} = \sum_{g=1}^{G} \frac{A_{Tg}}{A_T} \bar{Y}_{gd} \tag{22}$$

The variance of the mean is easily computed from the variance of the total via division by the square of the total area in the population.

$$v(\bar{Y}_d) = \frac{v(\hat{Y}_d)}{A_T^2} \tag{23}$$

However, often the mean is expressed on a per forested-acre basis. This forms a ratio estimate because the denominator (estimated area of forest land) is also a random variate. As with summing across strata, summing across subpopulations for ratio estimation suggests that the ratios are formed as the last step, i.e., at the population level. The estimated population ratio is

$$\hat{R}_{dd'} = \frac{\hat{Y}_d}{\hat{Y}_{d'}} = \frac{\sum_{g=1}^{G} A_{Tg}\bar{Y}_{gd}}{\sum_{g=1}^{G} A_{Tg}\bar{Y}_{gd'}} \tag{24}$$

The general form of the variance is shown in Equation 11. The variance of the numerator is given in Equation 21 and similarly the variance of the denominator is:

$$v\left(\hat{Y}_{d'}\right) = \sum_{g=1}^{G} v(\hat{Y}_{gd'}) = \sum_{g=1}^{G} A_{Tg}^2 v(\bar{Y}_{gd'}) \tag{25}$$

The above is straightforward; however, the covariance term in Equation 8 remains to be addressed. The following assumes that estimation units are independent; therefore, the covariance between the numerator and denominator is the sum over the estimation unit covariances—analogous to the variances

$$cov\left(\hat{Y}_d, \hat{Y}_{d'}\right) = \sum_{g=1}^{G} A_{Tg}^2 cov(\bar{Y}_{gd}, \bar{Y}_{gd'}) \tag{26}$$

This completes the necessary formulae to calculate the variance of the ratio estimate across multiple estimation units.

Analysts should be aware that FIA analytical tools such as EVALIDator and DATIM (https://www.fia.fs.usda.gov/tools-data/) take a standardized approach of providing estimates at the state level or for other user-defined populations (e.g., a GIS polygon). Hence, estimates from these tools almost always arise via specification of a domain *d* that is applied across all estimation units in the state. The summation across estimation units is accomplished internally and is usually not apparent to the user. Note that estimation units not having any plots containing domain *d* will contribute zero to the both the estimate and its variance.

Equations 1–26 shown above are consistent with those in Scott et al. (2005) and represent the current methods by which FIA generates forest resource estimates. An alternative approach would be to use ratio-to-size estimators (Cochran 1977, section 11.8). The key difference in using ratio-to-size estimators within strata is how partial nonresponse (inaccessible portions of plots) is addressed. The current approach uses $\bar{p}_{h_j}$ (Eq. 17) to compensate for missing portions of plots; however, this source of variation is not directly accounted for when estimating variance. The ratio-to-size estimator yields exactly the same estimates of forest attributes; however, the variance of $\bar{p}_{h_j}$ is incorporated into the variance estimator, yielding a more accurate variance estimate. When there are no partial plots, the variances from both methods are identical. Another feature of the ratio-to-size estimator is that it can be used for estimating ratios of means, such as volume per acre of forest land. The variance estimator is much simpler than what is currently used but yields the same results only in the case of simple random sampling—but not when stratification is used such as in FIA applications. A detailed explanation of ratio-to-size estimation in the context of the FIA inventory is provided in Appendix 1.

Nonresponse

The FIA program gains considerable efficiency via a priori determination of whether a sample plot may contain forest land or is entirely nonforest. Data for entirely nonforest plots is obtained from high-resolution imagery (independently of the poststratification process), whereas plots containing forest land are designated for field measurement. Occasionally, a field plot (or portion thereof) may not be measured for some reason. The most common causes are denied access by the landowner and hazardous conditions such as wildfire or other dangerous circumstances. These unmeasured plot areas are referred to as nonresponse, which is consistent with the survey sampling literature. The treatment of nonresponse conditions in estimation is described by Scott et al. (2005), i.e., plots that are entirely nonresponse are dropped from the estimation and partially nonresponse plots are included in an adjustment factor (Eq. 17). These protocols are implemented at the stratum level; thus, some effort is made to construct strata having similar properties that may influence characteristics of nonresponse plots. An example is using ownership information maps in the creation of the post-strata to separate public and private ownerships, for the primary purpose of accounting for denied access plots mostly occurring on privately owned land (Patterson et al. 2012). This model is similar to the Response Homogeneity Group (RHG) model (Särndal et al. 1992, p. 578), where groups are identified on which the response probabilities are constant. The standard “stratified” estimator is used with the RHGs treated as strata. FIA estimation processes assume the nonresponse areas are missing at random within strata; however, it is acknowledged this assumption may not always be tenable (Bechtold and Scott 2005). Additionally, the FIA approach for estimating change over time only includes plots that were sampled at both the previous and current measurement. Thus, the sample size for change estimation differs from that used to develop estimates of current status and the magnitude of the difference is largely driven by nonresponse rates.

The evolution of nonresponse awareness and treatment varies slightly between the four FIA units. In the early 2010s, the Rocky Mountain (RM) region of FIA decided to investigate the impact of nonresponse. Previously RM-FIA did not include a RHG layer as part of the post-stratification. In most states of the RM region, the majority of nonresponse is due to denied access on privately owned lands; it thus it makes sense to use a public/private ownership map layer. At the same time, New Mexico was being inventoried by using an accelerated schedule and comparisons between ignoring nonresponse bias and the use of a stratification layer to reduce nonresponse bias had a substantial effect on estimates of forest land area. This outcome suggested the use of RHGs considerably reduced nonresponse bias (Goeking and Patterson 2013). RM-FIA is in the process of reviewing the other seven states in its region for the applicability of forming RHGs, the first step being identification of proper mechanisms for treatment of nonresponse. Although, as stated above, the majority of the nonresponse is from denied access, an analysis of the nonresponse for each state will be conducted to see if there are any other nonresponse drivers such as hazardous conditions that need to be accounted for.

In the Pacific Northwest (PNW) FIA unit, the issue of nonresponse in estimation has been mainly addressed through the addition of more informative stratification layers, such as the National Land Cover Database (NLCD) tree canopy cover, historical forest type maps (California, Oregon, and Washington), mean annual temperature (Alaska), elevation (Alaska), ownership, and wilderness layers. However, the possibility of bias remains a concern, especially given the relatively large differences in the nonresponse rate within particular ownership classes (such as corporate vs. noncorporate private landowners) that are not separated in the spatial layers currently used in stratification. Other spatial layers are being considered to reduce possible nonresponse bias attributed to other sources, such as topography (steepness, cliffs) to account for hazardous plots on public land, mapped disturbance information (harvest, cultivation) to explain denial of access on private lands, and additional ownership layers to separate corporate from noncorporate private. There is also an ongoing effort to reduce the number of access-denied plots through more focused outreach and education.

While the nonresponse rate in the Southern FIA region is low (~1.0 percent), there are specific locations where nonresponse is an issue. When there is a nonresponse issue, the typical approach is to construct additional strata related to the nonresponse mechanism so that the assumption of missing at random within strata is more appropriate. For example, in west Texas there is significant nonresponse because of denied access on private forest land. In this case, separate public forest land and private forest land strata are created. An example arising from hazardous conditions is the Okefenokee National Wildlife Refuge in Georgia and northern Florida. While some of the forested plots in this swamp are measured, many are too hazardous to access. The Okefenokee is treated as a separate stratum for estimation purposes.

The stratified estimation process used by Northern FIA comprises estimation units defined by ownership categories (e.g., national forest, other public, and private) and strata defined by 5 percent tree canopy cover classes (0–5, 6–50, 51–65, 66–80, and 81–100) to reduce the possible effects of bias caused by nonresponse (Gormanson et al. 2018).

An analysis of plots designated for field measurement during the period 2007–2017 indicated the rate of nonresponse due to denied access (6.5 percent) and hazardous conditions (1.1 percent) at the national scale has been about 7.6 percent. However, there was considerable spatial variability as shown by states in the northeastern United States tending to have the highest rates of denied access (Delaware, Maryland, and Rhode Island > 20 percent); whereas the southeastern United States had the lowest rates (~ 1.0

percent regionally). Similarly, hazardous conditions were more prevalent in the western United States where phenomena such as wildfire, dangerous wildlife, and extreme terrain are more commonly encountered. Western states had hazardous-condition nonresponse rates of 1.5 percent to 4.5 percent.

Various initiatives to decrease the amount of nonresponse have met with limited success, such that the issue remains a concern to the FIA program. Initial research into the problem and potential solutions were reported by Patterson et al. (2012), which led to increased emphasis on improving post-stratification efforts to better contend with potential bias issues in estimation and recitation of the policy to maintain the existing plot location regardless of nonresponse frequency. Current data suggests that about 5 percent of field plots change status from denied access to accessible (or vice versa) at time of remeasurement. Further, fewer than 1 percent of plots were consistently denied access for three successive measurements. These outcomes suggest a plot replacement effort would be largely ineffective and FIA should continue to preserve the original sample plot selection. For the TPO and NWOS surveys, the nonresponse issue is addressed within the respective chapters.

Literature Cited

Bechtold, W.A.; Patterson, P.L., 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. https://doi.org/10.2737/SRS-GTR-80.

Bechtold, W.A.; Scott, C.T. 2005. **The Forest Inventory and Analysis plot design.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 27–42.

Cochran, W.G. 1977. **Sampling techniques.** New York: John Wiley & Sons. 428 p.

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. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 22 p. https://doi.org/10.2737/RMRS-RN-59.

Gormanson, D.D.; Pugh, S.A.; Barnett, C.J.; [et al.]. 2018. **Statistics and quality assurance for the Northern Research Station Forest Inventory and Analysis Program.** Gen. Tech. Rep. NRS-178. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 25 p. https://doi.org/10.2737/NRS-GTR-178.

Patterson, P.L.; Coulston, J.W.; Roesch, F.A.; [et al.]. 2012. **A primer for nonresponse in the U.S. Forest Inventory and Analysis Program**. Environmental Monitoring and Assessment. 184(3): 1423–1433. https://doi.org/10.1007/s10661-011-2051-5.

Reams, G.A.; Smith, W.D.; Hansen, M.H.; [et al.]. 2005. **The Forest Inventory and Analysis sampling frame.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 11–26.

Särndal, C.E.; Swensson, B.; Wretman, J.H. 1992. **Model assisted survey sampling**. New York: Springer-Verlag. 694 p.

Scott, C.T.; Bechtold, W.A.; Reams, G.A.; [et al]. 2005. **Sample-based estimators used by the Forest Inventory and Analysis national information management system.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 53–77.

Tomppo, E.; Gschwantner, T.; Lawrence, M.; [et al.], eds. 2010. **National forest inventories: pathways for common reporting. Dordrecht: Springer. 612 p.** http://dx.doi.org/10.1007/978-90-481-3233-1.

USDA Forest Service. 2018. **Forest Inventory and Analysis national core field guide. Vol. 1: Field data collection procedures for Phase 2 plots, version 8.0.** https://www.fia.fs.usda.gov/library/field-guides-methods-proc/docs/2018/core_ver8-0_10_2018_final.pdf (accessed December 6, 2021).

Westfall, J.A.; Lister, A.J.; Scott, C.T.; [et al.]. 2019. **Double sampling for post-stratification in forest inventory.** European Journal of Forest Research. 138(3): 375–382. https://dx.doi.org/10.1007/s10342-019-01171-9.

Chapter 3: Urban Forest Inventory and Analysis

Christopher B. Edgar

Urban forests provide numerous environmental benefits. In urban and community areas of the United States, trees reduce residential energy cost by $7.8 billion per year (Nowak et al. 2017), store 643 million tonnes of carbon, and annually sequester 25.6 million tonnes of carbon (Nowak et al. 2013). Further, trees in urban areas of the conterminous United States remove 711 thousand tonnes of air pollution, providing an annual value of $3.8 billion (Nowak et al. 2006). The value of trees as a structural asset in urban forests of the conterminous United States totals $2.4 trillion (Nowak et al. 2002). These and numerous other studies point to the wide range of benefits that urban forests provide and illustrate the important role of urban inventory to quantify status and change in the urban resource.

Urban forests include all trees in urban areas. In the enhanced Forest Inventory and Analysis (FIA) program, tree data are collected on forest land uses only (Bechtold and Scott 2005). Absent extending measurement to all trees in urban areas, only a portion of urban forests are monitored in the national forest inventory (Nowak et al. 2013). In the mid-2000s, several pilot studies were conducted by FIA in partnership with states to test various procedures for urban forest resource monitoring (Cumming et al. 2007, Nowak et al. 2007, Nowak et al. 2011). Although limited to a few states, these pilot studies provided valuable lessons regarding the feasibility of adapting the FIA design to urban environments (Nowak et al. 2016).

The lack of comprehensive coverage of urban forests in the national forest inventory created an information gap that challenged stewardship of the broader forest landscape (USDA Forest Service 2016). From 2000 to 2010, the nation's urban population increased 12.1 percent through a combination of internal growth and outward expansion of urban areas (U.S. Census Bureau 2018). The National Association of State Foresters, in a position statement regarding Farm Bill direction, noted the need to "[s]trengthen forestry outreach, education, research, and inventory programs that enhance the ability of State Foresters to assist private landowners and deliver federal and state programs serving all lands across the rural to urban spectrum" (National Association of State Foresters 2017). Seamless monitoring of forest land and urban forests would fill the information gap and provide critical information needed to help partners and stewards address important issues such as the loss of peri-urban forest, wildland fire, and invasive species.

In recognition of the lack of strategic coverage and the important benefits of urban forests, and with direction from the 2014 U.S. Farm Bill, FIA expanded its focus beyond forest land to include all trees in urban areas through deployment of an urban inventory program. The urban inventory extends the enhanced FIA sampling frame and estimation procedures to urban areas with some modification. Adapting the enhanced FIA program to urban areas leverages the valuable lessons learned implementing annual FIA over the past two decades and lays the foundation for seamless inventory of forest and urban lands.

The purpose of this chapter is to document the sample design and estimation procedures used in the urban inventory. Many of the approaches and techniques employed by the enhanced FIA program that are documented in Bechtold and Patterson (2005) are used. The approach taken here is to identify circumstances where the urban inventory follows enhanced FIA and for the reader to consult the Bechtold and Patterson (2005) publication if more detailed information is desired. Circumstances where the urban inventory deviates from enhanced FIA are noted and technical details provided. Sections are devoted to the sampling frame, plot configuration and measurement, sample-based estimators, combining panels, and continuing investigations. The urban program was initiated in 2014 and is still evolving with investigations into methods and techniques remaining to be conducted. This chapter is accompanied by an online resource (https://www.fia.fs.usda.gov/library/sampling/index.php) where new developments in urban inventory design and estimation are made available.

Sampling Frame

In the urban inventory, populations and subpopulations are formed from urban areas by using one of two models. For the first program goal, that of providing strategic-level coverage of urban forests across the nation, U.S. Census Bureau-defined urbanized areas (UA) and urban clusters (UC) are used. Urbanized areas are densely developed territory containing 50,000 or more people. Urban clusters are densely developed territory with at least 2,500 people but fewer than 50,000 people. As of the 2010 Census, there were 486 urbanized areas and 3,087 urban clusters across the nation covering approximately 68 million acres (U.S. Census Bureau 2018). A second program goal is to provide strategic urban forest inventories for the largest cities in the United States. In this model, a population is formed from a city boundary. FIA is targeting cities with populations of 200,000 or more people as of the 2010 Census. A small number of cities with populations below 200,000 are included to ensure every state has at least one target city in the program. Altogether, approximately 100 cities would be included in the program when partnerships, resources, and funding are fully in place.

The hexagonal sampling frame (Bechtold and Patterson 2005, White et al. 1991) applied to the FIA program is used in the urban inventory. For populations and subpopulations formed from UA/UC, a base spatial sampling intensity of approximately one plot per 5,937 acres is used. For populations formed from city boundaries, the base spatial intensity produces plot numbers insufficient for reliable city-level inference. Sample size decisions are influenced by i-Tree (https://www.itreetools.org/), a set of software tools developed by the USDA Forest Service and partners that provides assessment of urban forest structure, services, and benefits. In an analysis of sample inventory data from 14 cities, Nowak et al. (2008b) reported that 200 0.1-acre circular plots typically produce about a 12 percent error for the estimate of the total number of trees and that more plots provide only marginal gains. Where city boundaries form a population, the sample frame is enhanced by a spatial intensification sufficient to produce around 200 plots. In the spatial intensification process, base hexagons are divided into smaller units with one plot randomly located in each unit.

The sample frame is divided into panels. Each panel provides a spatially complete sample of the population. In standard implementation, one panel is measured at a time and not until that panel is completed is a subsequent panel measured, and no panel is remeasured until all other panels have been measured. Details on the origin and use of the panel system by FIA are documented in Reams et al. (2005).

Post-stratification is used to increase the precision of estimates without increasing sample sizes (Scott et al. 2005). The population is divided into strata of known size by using one or more remote sensing or geospatial layers that cover the population. In the FIA program, considerable research has been conducted into image classification, spatial resolution, and selection of strata boundaries (McRoberts et al. 2006). Recent work has focused on post-stratification for both variance reduction and as a means of minimizing the bias due to differential response rates among major owner groups. Using post-stratification to make the missing-at-random assumption tenable has been recommended as a strategy to mitigate the effects of nonresponse (Patterson et al. 2012). Domke et al. (2014) examined six techniques to compensate for missing observations in carbon estimation, two of which involved replacing missing observations with stratum means. Westfall et al. (2011) examined within-strata sample sizes and presented recommendations on minimum sample size to achieve acceptable levels of bias and variability. Previous FIA work into post-stratification for proportion forest land and cubic net volume (McRoberts et al. 2006, Westfall et al. 2011) and coarse woody debris (Hatfield 2010) estimation inform the post-stratification being applied in the urban inventory.

Using the FIA sampling frame in the urban inventory provides program efficiencies and promotes the goal of seamless monitoring of forest land and urban areas. The FIA sampling frame was formally described as occurring in three phases (Reams et al. 2005). The urban inventory sample design matches Phase 1 and Phase 2 of the enhanced FIA program (Table 1) when satellite pixels rather than photo points are used in Phase 1. The FIA program has a Phase 3; however, there is no analog in the urban inventory. By using the same sample frame, plots from both inventories (i.e., the base-intensity plots) can occur at the same location (i.e., colocated). When that is the case, both plots are installed and measured by using their respective protocols. Measurement of both plots occurs in the same inventory year according to the panel timing of the enhanced FIA program. The number of panels used in the urban inventory is set equal to that of the FIA program in the state or area.

Table 1.—Summary of general attributes associated with the urban inventory

Attribute	Phase 1	Phase 2
Sample type	Satellite pixel	Ground plot
Sample configuration	Pixel	Cluster of four 1/300-acre microplots, one 1/6-acre subplot
Purpose	Stratification of the landscape for the purpose of variance reduction	Samples FIA and i-Tree Eco attributes of interest
Tessellation method	Wall-to-wall	Systematic national hexagonal cell grid
Base-grid intensity	Wall-to-wall	One plot per every 6,000-acre hexagonal cell in urbanized areas and urban clusters (UA/UC); spatial intensification in target cities sufficient to produce at least 200 plots

Plot Configuration and Measurement

Urban inventory plots conform to a national standard configuration (Fig. 5). Each plot consists of one 48-ft radius circular plot. Nested in each plot are four 6.8-ft radius circular microplots offset 12 feet in each of the cardinal directions from plot center. The urban plot configuration differs from the enhanced FIA plot configuration, which consists of four 24-ft radius circular subplots with one subplot centered on plot center and the other three centered 120 feet from plot center at azimuths of 0, 120, and 240 degrees (see Fig. 4a in the Foundational Documentation chapter). The change in configuration was made to reduce the number of owners that would need to be contacted to access urban plots. Although plot configurations are different between the two programs, sampled areas are the same. The four subplots in the enhanced FIA plot configuration total 0.166 acres, which is the same area as the urban plot. For both programs, the microplot area totals 0.013 acres. There is no macroplot in the urban plot design.

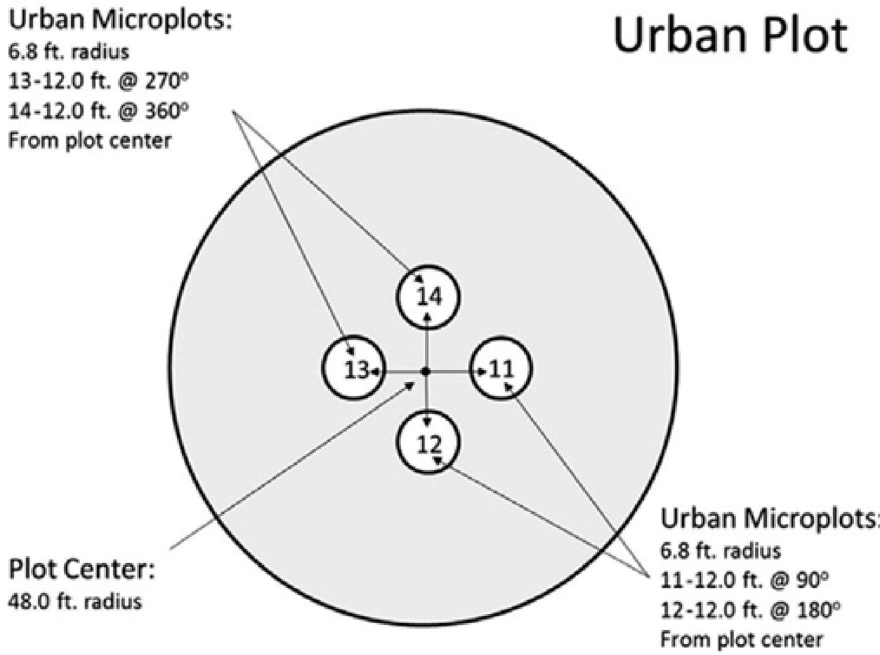

Figure 5.—Plot design used in the urban inventory.

On the urban plot, trees 5.0 inches and larger in diameter at breast height (d.b.h.) or diameter at root collar (d.r.c.) are sampled. Trees 1.0–4.9 in. d.b.h. or d.r.c. are sampled on the microplots. The urban inventory applies many of the same measurements that are used in the FIA program. The urban inventory collects additional measurements related to urban-specific issues and to provide data needed for processing of urban benefits in the i-Tree Eco system (Nowak and Crane 2000, Nowak et al. 2008a). Land use, urban specific damage, and tree location relative to residential buildings are examples of areas where additional measurements are taken. Interested readers can consult the urban inventory field guide for a complete description of measurements taken at sample locations (https://www.fia.fs.usda.gov/library/field-guides-methods-proc/index.php).

The urban inventory uses mapped condition classes similar to the enhanced FIA program. One significant area of deviation is treatment of nonforest conditions. In the enhanced FIA program, measurement of trees occurs on forest conditions only. Urban forests include all trees in urban areas and thus urban inventory measurements are made on all conditions, forest and nonforest. The urban inventory is collecting more detailed information on the occurrence, type, and condition of nonforest land use classes than is done in the FIA program. This reflects the fact that trees in urban areas occur both inside and outside forest land and information from all plots and conditions is needed for estimation of the full range of attributes produced by urban forests.

Design-Based Estimation

The estimators used in the FIA program have been adopted in the urban inventory for attributes collected under a probabilistic design. Populations may contain both enhanced FIA plots and urban inventory plots; however, for the purposes of estimation in the urban inventory, the plot types are not pooled or combined. The estimation of urban area attributes is based on data from urban inventory plots only. The estimation approach follows the methodology described in the Foundational Documentation chapter, which uses Equations 1–26 as appropriate for either population totals or ratios.

There is a need and expectation that growth, removals, and mortality estimates will be produced once sufficient remeasurement data become available. Like the FIA program, the urban inventory is designed to produce components of change. Two sequential measurements of a plot provide the essential data needed to compute change components. Scott et al. (2005) provide detailed information on the individual components recognized in FIA and how those components are combined to produce various estimates of change. The urban inventory will likely adopt the estimation methods for growth, removals, and mortality used in the enhanced FIA, although some modification may be needed. For example, tree mortality may be accompanied by removal, especially in areas where standing dead trees are a hazard to people and property. The extent to which this occurs and subsequent impact on component of change calculation is an issue for investigation.

Combining Panels

The FIA program does not prescribe a single procedure for combining panels for the reason that a single technique may not work for all situations encountered in a national forest inventory (McRoberts et al. 2005). Patterson and Reams (2005) briefly describe several approaches to combining panels and important considerations. Estimates that are produced in standard or routine reporting in the urban inventory program best match the temporally indifferent method in the sense that all panels are pooled and the same stratification is applied. Note that the temporally indifferent method is equivalent to the moving average method when weights proportional to the number of plots in each panel are used (Patterson and Reams 2005).

Continuing Investigations

The urban inventory started in 2014 with measurements in Austin, Texas, and Baltimore, Maryland. As of 2020, the urban inventory is operational in 40 cities covering all the major regions of the United States. The latest information on implementation, as well as other program information, is available at the National Urban Forest Inventory and Analysis Program website hosted by the USDA Forest Service's Northern Research Station (https://www.fia.fs.usda.gov/program-features/urban/).

The rapid expansion of the urban inventory can at least in part be attributed to the decision to adapt the enhanced FIA program. Leveraging the FIA sample frame and the accumulated expertise in data acquisition and information management enabled quick deployment of the inventory in urban areas. As an integration of FIA and i-Tree, urban inventory is benefitting from the lessons learned from each of those successful systems. The urban inventory is expanding and there remain issues to be resolved. Several areas for potential future investigation include efficient sample size in target cities; auxiliary data for post-stratification; change component computation; allocation of population model output to domains; and methods of combining panels. Seamless monitoring of rural and urban forests is a goal of the FIA program, suggesting the need for investigation of methods for producing estimates that combine the data from both FIA and urban inventory (e.g., Westfall et al. 2018). Urban areas are expanding and that presents a challenge to the urban inventory – the boundaries of the population can change. FIA is currently using U.S. Census spatial layers from 2010 to define urban populations with plans to adopt new layers in 2020. Considerable focus has been placed on development of information models and databases with the flexibility to store versions of the population boundaries. Investigation into estimation methods for both inventory and change component estimation in preparation for changing boundaries will be needed. As advances are made, newly established techniques and methods will be documented and made available on the website (https://www.fia.fs.usda.gov/library/sampling/index.php) that accompanies this report.

Acknowledgments

Mark Majewsky and Mark Hatfield, both with the Northern Research Station, Forest Inventory and Analysis unit, answered questions about the overall program and technical details of the sample design and estimation. Each reviewed excerpts of the chapter and provided valuable feedback. Funding was provided by the USDA Forest Service Project 15 FIA—Forest Biometrics Program Support (RJVA 15-JV-11242305-100) and Minnesota Agricultural Experiment Station Project MIN-42-078.

Literature Cited

Bechtold, W.A.; Patterson, P.L., 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. https://doi.org/10.2737/SRS-GTR-80.

Bechtold, W.A.; Scott, C.T. 2005. **The Forest Inventory and Analysis plot design.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 27–42.

Cumming, A.B.; Nowak, D.J.; Twardus, D.B.; [et al.]. 2007. **Urban forests of Wisconsin: pilot monitoring project 2002.** NA-FR-0507. Newtown Square, PA: U.S. Department of Agriculture Forest Service, Northeastern Area State and Private Forestry Report. 33 p.

Domke, G.M.; Woodall, C.W.; Walters, B.F.; [et al.]. 2014. **Strategies to compensate for the effects of nonresponse on forest carbon baseline estimates from the national forest inventory of the United States.** Forest Ecology and Management. 315: 112–120. https://doi.org/10.1016/j.foreco.2013.12.031.

Hatfield, M.H. 2010. **Post-stratified estimation of coarse woody debris volume using the down woody materials sample of Forest Inventory and Analysis.** Minneapolis, MN: University of Minnesota. 169 p. M.S. thesis.

McRoberts, R.E.; Bechtold, W.A.; Patterson, P.L.; [et al.]. 2005. **The enhanced Forest Inventory and Analysis Program of the USDA Forest Service: historical perspective and announcement of statistical documentation.** Journal of Forestry. 3(6): 304–308.

McRoberts, R.E.; Holden, G.R.; Nelson, M.D.; [et al.]. 2006. **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. 36: 2968–2980. https://doi.org/10.1139/x05-222.

National Association of State Foresters. 2017. **Farm Bill | National Association of State Foresters.** https://www.stateforesters.org/where-we-stand/farm-bill/ (accessed April 20, 2017).

Nowak, D.J.; Appleton, N.; Ellis, A.; [et al.]. 2017. **Residential building energy conservation and avoided power plant emissions by urban and community trees in the United States.** Urban Forestry & Urban Greening. 21: 158–165. https://doi.org/10.1016/j.ufug.2016.12.004.

Nowak, D.J.; Bodine, A.R.; Hoehn, R.E.; [et al.]. 2016. **Austin's urban forest, 2014.** Resource Bulletin NRS-100. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 55 p. http://dx.doi.org/10.2737/NRS-RB-100.

Nowak, D.J.; Crane, D.E. 2000. **The urban forest effects (UFORE) model: quantifying urban forest structure and functions.** In: Hansen, M; Burk, T., eds. Integrated tools for natural resources inventories in the 21st century. Proceedings of IUFRO conference. Gen. Tech. Rep. NC-212. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 714–720.

Nowak, D.J.; Crane, D.E.; Dwyer, J.F. 2002. **Compensatory value of urban trees in the United States.** Journal of Arboriculture. 28(4): 194–199. https://doi.org/10.48044/jauf.2002.028.

Nowak, D.J.; Crane, D.E.; Stevens, J.C. 2006. **Air pollution removal by urban trees and shrubs in the United States.** Urban Forestry & Urban Greening. 4: 115–123. https://doi.org/10.1016/j.ufug.2006.01.007.

Nowak, D.J.; Crane, D.E.; Stevens, J.C.; [et al.]. 2008a. **A ground-based method of assessing urban forest structure and ecosystem services.** Arboriculture & Urban Forestry. 34(6): 347–358. https://doi.org/10.48044/jauf.2008.048.

Nowak, D.J.; Cumming, A.B.; Twardus, D.B.; [et al.]. 2007. **Monitoring urban forests in Indiana: pilot study 2002, part 2: statewide estimates using the UFORE model.** NA-FR-01-07. Newtown Square, PA: U.S. Department of Agriculture Forest Service, Northeastern Area State and Private Forestry Report. 13 p.

Nowak, D.J.; Cumming, A.B.; Twardus, D.B.; [et al.]. 2011. **Urban forests of Tennessee, 2009.** Gen. Tech. Rep. SRS-149. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 52 p. https://doi.org/10.2737/SRS-GTR-149.

Nowak, D.J.; Greenfield E.J.; Hoehn, R.E.; [et al.]. 2013. **Carbon storage and sequestration by trees in urban and community areas of the United States.** Environmental Pollution. 178: 229–236. https://doi.org/10.1016/j.envpol.2013.03.019.

Nowak, D.J.; Walton, J.T.; Stevens, J.C.; [et al.]. 2008b. **Effect of plot and sample size on timing and precision of urban forest assessments.** Arboriculture & Urban Forestry. 34(6): 386–390. https://doi.org/10.48044/jauf.2008.052.

Patterson, P.L.; Coulston, J.W.; Roesch, F.A; [et al.]. 2012. **A primer for nonresponse in the U.S. Forest Inventory and Analysis Program.** Environmental Monitoring and Assessment. 184: 1423–1433. https://doi.org/10.1007/s10661-011-2051-5.

Patterson, P.L.; Reams, G.A. 2005. **Combining panels for Forest Inventory and Analysis estimation.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 69–73.

Reams, G.A.; Smith, W.D.; Hansen, M.H.; [et al.]. 2005. **The Forest Inventory and Analysis sampling frame.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 11–26.

Scott, C.T.; Bechtold, W.A.; Reams, G.A.; [et al.]. 2005. **Sampled-based estimators used by the Forest Inventory and Analysis national information management system.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 53-77.

U.S Census Bureau. 2018. **Growth in urban population outpaces rest of nation, Census Bureau reports.** https://www.census.gov/newsroom/releases/archives/2010_census/cb12-50.html (accessed April 24, 2019).

USDA Forest Service. 2016. **Forest Inventory and Analysis strategic plan.** FS-1079. Washington, D.C.: U.S. Department of Agriculture, Forest Service. 48 p. https://www.fia.fs.usda.gov/library/bus-org-documents/docs/strategic-plan-docs/FIA%20Strategic%20Plan%20FS-1079.pdf (accessed December 6, 2021).

Westfall, J.A.; Patterson, P.L.; Coulston, J.W. 2011. **Post-stratified estimation: within-strata and total sample size recommendations.** Canadian Journal of Forest Research. 41: 1130–1139. https://doi.org/10.1139/x11-031.

Westfall, J.A.; Patterson, P.L.; Edgar, C.B. 2018. **Integrating urban and national forest inventory data in support of rural-urban assessments.** Forestry: An International Journal of Forest Research. 91(5): 641-649. https://doi.org/10.1093/forestry/cpy023.

White, D.; Kimerling, A.J.; Overton, W.S. 1991. **Cartographic and geometric components of a global sampling design for environmental monitoring.** Cartography and Geographic Information Systems. 19(1): 5–22.

Chapter 4: National Woodland Owner Survey

Brett J. Butler and Jesse Caputo

The U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis (FIA) program conducts the National Woodland Owner Survey (NWOS). The NWOS is the social complement to the plot-based, biophysical-focused, forest inventory component of the FIA program. It has been implemented by FIA since 1993 (Birch 1996) with subsequent iterations completed in 2006 (Butler 2008), 2013 (Butler et al. 2016), and 2018 (Butler et al. 2021). Beginning in 2019, the NWOS switched to an annual system, as is done with the FIA forest inventory. The goals of the NWOS are to generate information related to the attitudes, behaviors, and general characteristics of private forest ownerships in the United States. This information is used to help design policies, implement programs, and provide services aimed at private landowners.

This chapter summarizes the sampling and estimation procedures for the base/rural NWOS. After defining the populations of interest, the sample design is described, the estimators are presented, issues related to nonresponse are discussed, and recommended sample sizes are presented. Additional information about the sampling and estimation procedures is available in Butler and Caputo (2021) and Butler et al. (2021).

Definition of the Population

The NWOS has multiple populations of interest: family forest ownerships, large corporate forest ownerships, other corporate forest ownerships, and other private forest ownerships (Table 2). An ownership is defined as a legal entity that has specified rights associated with the land, such as the ability to sell and manage it. In the case of family forest ownerships, the entity can be composed of one or more individual owners. For the base NWOS, the primary population of interest is family forest ownerships. The corporate category is further divided into large (≥ 45,000 acres) and smaller ownerships (Caputo et al. 2017). NWOS results are typically reported in terms of ownerships and acreage. The sampling procedures for the Urban National Landowner Survey, which are not covered in this chapter, are very similar to the approach used for the base NWOS, but the populations of interest are different.

Table 2.—Descriptions of populations of interest for the USDA Forest Service, National Woodland Owner Survey.

Population	Description
Family forest ownerships	Individuals and families, including trusts, estates, and family partnerships with 1+ acres of forest land
Large corporate forest ownerships	Corporate, including forest industry companies, timber investment management organizations (TIMOs), real estate investment trusts (REITSs), Native Corporations in Alaska, and private universities with 45,000+ acres of forest land
Other corporate forest ownerships	Corporate, including forest industry companies, TIMOs, REITs, Native Corporations in Alaska, and private universities with 1–44,999 acres of forest land
Other private forest ownerships	Nongovernmental conservation / natural resources organization and unincorporated partnerships / associations / clubs with 1+ acres of forest land

Sampling Frame

The sampling procedure used by the NWOS (Fig. 6) is built upon the same framework used for the FIA forest inventory described in the Foundational Documentation chapter of this report. An area-based sampling frame, defined as the land area in a state, forms the basis for sampling. Land area includes forest land, nonforest land, and non-Census water; it does not include Census water.

Sample Selection

The NWOS sample begins with the FIA sample points associated with 5 years' worth of forest inventory plots. There is a 2-year time lag between plot inventory and survey implementation to ensure the field crews can inventory the plots prior to survey implementation and provide some separation in terms of burden on the respondents. Where possible, the ownership information is updated prior to survey implementation. The time lag should have little impact on the survey results, especially given the long tenure of most ownerships and because the ownership information is being updated.

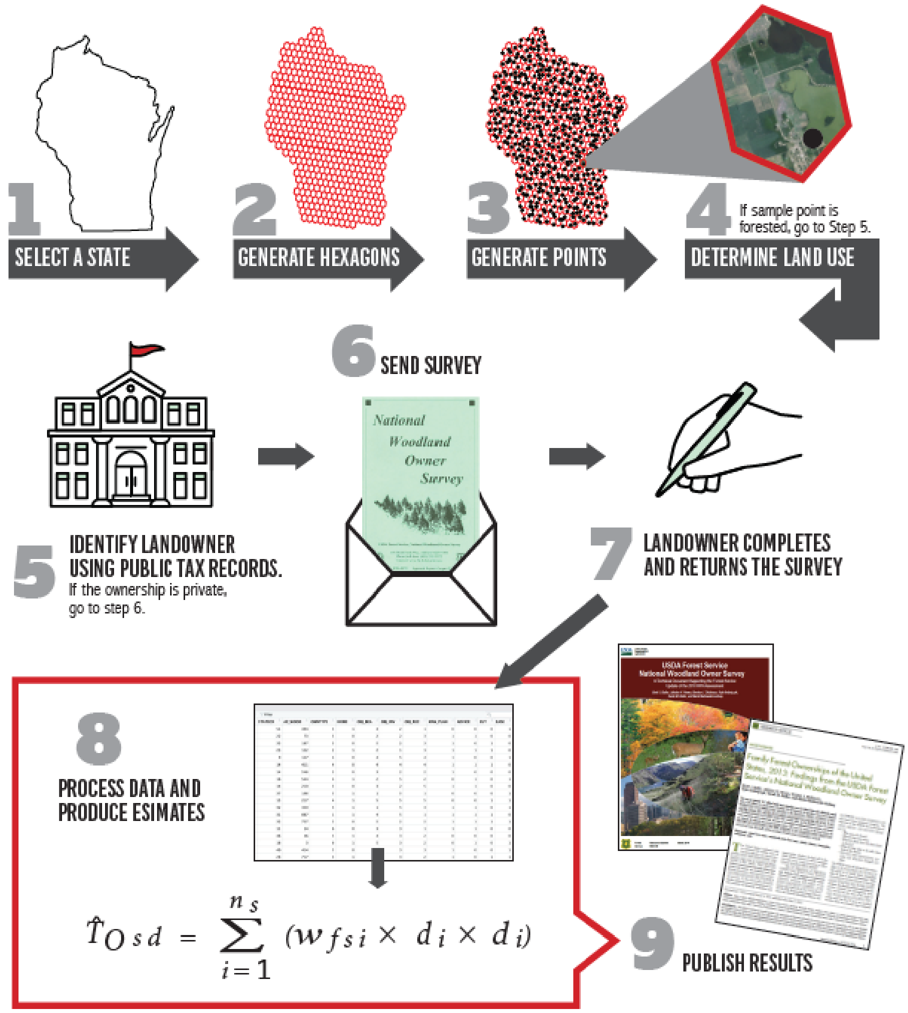

Figure 6.—USDA Forest Service, National Woodland Owner Survey sampling procedure. Only private ownerships with forested sample points on their land are asked to participate. Created by Penny MichAalak (Penny Michalak Design).

The FIA inventory sample design consists of hexagons that cover a state, sample points randomly located in each hexagon, and inventory plots associated with each sample point (see Foundational Documentation chapter). The base hexagons have an area of 5,937 acres. The sampling is implemented on an annual basis with 10 to 20 percent of the points sampled each year. The sample is spatially distributed across the entire state each year.

Where necessary, the NWOS sample is "augmented" to increase the number of sample points to reach the target of 250 respondents per state. Augmentation occurs by repeating the basic sampling procedure with a decreased hexagon size, with these smaller hexagons nested within the base hexagons in a fully tessellated manner. The hexagon size is based on the target sample size (see below), expected response rates, and expected area in the stratum of interest. One point is located randomly in each of the empty (i.e., not containing a point associated with a forest inventory plot) smaller hexagons. These additional points are iteratively added to the sample (by using the panel designation of the parent hexagon) until the desired intensity is achieved.

The NWOS, like the FIA inventory, is implemented on an annual basis. For the NWOS, 20 percent of the sample points, spatially distributed across a state, are used each year with a full cycle being completed once every 5 years. This 5-year remeasurement cycle is implemented across all states, regardless of the remeasurement cycle for the FIA plots in a given state. At the end of the 5-year cycle, remeasurement commences using the same base FIA and augmented sample points; owners who still own the sample point are asked to complete another survey and new owners are invited to complete a survey.

The sample points are post-stratified based on land use and ownership; stratum areas are not known a priori. For the FIA inventory sample points, the land use (forest/nonforest) data come directly from the variables collected by the FIA prefield staff and field crews (i.e., are the same data used in the estimates derived from the FIA forest inventory). The initial land use is classified with aerial photography, and ground-validation occurs for plots that are likely to be forested or are potentially forested. For the augmented points, aerial photography is again used to determine land use, but no ground-truthing occurs. For the forested sample points, the ownership of record is determined based on publicly available property ownership records. If the ownership is private, it becomes part of the list of ownerships surveyed as part of the NWOS. Some FIA inventory sample points are not sampled due to denied access, hazardous conditions, or other reasons, and are treated similarly to the augmented points: the land use is determined by using image analysis and, as with all points, the ownership is derived from property records. Although included in the sample, no surveys are sent to ownerships associated with these points.

Response Design

The basic unit of response is an ownership. For family, other corporate, and other private forest ownerships, the NWOS asks them to respond for all the forest land they own in a specified state. In the case that a family, other corporate, or other private forest ownership has forest land in multiple states that is part of the sample, they would be asked to respond separately for each state. Large corporate forest ownerships are asked to respond for all the forest land they own in the United States, along with how many acres they own in each state.

Design-Based Estimators

The NWOS sampling procedure uses an area-based sampling frame that results in a sample design with inclusion probabilities proportional to size of forest holdings—the greater the forest acreage owned, the more likely an ownership will be surveyed. In order to generate accurate estimates, the post-stratification, sampling weights, and other attributes of the sample design need to be incorporated into the estimators. Estimates of totals, proportions, means, and quantiles and their associated variances are made in terms of ownerships and acreages. Estimates are generated separately by state and stratum and can then be combined to get estimates for broader groupings.

A weighting approach (Valliant et al. 2013) is used for the estimates of totals, means, proportions, and quantiles with a bootstrapping approach (Efron and Tibshirani 1986) used for the associated variance estimates. An R package for implementing these procedures is available on GitHub (Butler and Caputo 2020).

There is not a one-to-one relationship between sample points and ownerships—i.e., there is the possibility of having more than one sample point on an ownership's land. For example, an ownership may have many acres of forest land across the state and sample points fall on more than one parcel. It is also possible that a parcel straddles hexagon boundaries and has more than one sample point fall on that parcel. The probability of multiple points increases as the acreage of the holding increases. This too is accounted for in the estimators.

Weights

Weights incorporate the sample design, adjustments for response rates, alternative stratum area estimates, and unit nonresponse biases. These are calculated separately for each stratum or population of interest by state (e.g., family forest ownerships in Wisconsin). The base weights are the inverse of the stratification-adjusted design weights multiplied by the number of forested sample points that fall on an ownership's land (Eq. 27). The design weights (Eq. 28) are a function of the acreage owned within a stratum, the estimated acreage of the stratum, and the total number of sample points in the stratum:

$$\omega_{hi} = \frac{1}{\pi_{hi}} p_{hi} \tag{27}$$

$$\pi_{hi} = \frac{a_{hi}}{\hat{A}_h / n_h} = \frac{a_{hi} n_h}{\hat{A}_h} \tag{28}$$

Where

ω_{hi} = base weight for ownership *i* in stratum *h*,

π_{hi} = design weight for ownership *i* in stratum h,

p_{hi} = number of sample points owned by ownership *i* in stratum h,

$\hat{A}_h = A\frac{n_h}{n}$ = estimated area of land in stratum h,

A = total land area,

n_h = number of sample points in stratum *h*, and

n = total number of sample points.

All of the calculations occur at the state level; to simplify notation, a subscript denoting state is suppressed.

To get the total area of forest land to equal values from the FIA forest inventory estimates, an alternative value, $\hat{A}'_h$, is substituted for $\hat{A}_h$ in Equation 28. This allows the reported values to be the same across multiple reports. Both sets of estimates, i.e., based on $\hat{A}_h$ and $\hat{A}'_h$, are unbiased estimates for area in a stratum, but the differences, which are likely to be small, can complicate interpretation. Although Equation 28 can be simplified by substituting $A\frac{n_h}{n}$ for $\hat{A}'_h$, this would negate the ability to adjust with $\hat{A}'_h$. For the variance estimation described below, it is important to exclude $\hat{A}'_h$, otherwise the variability associated with estimating $\hat{A}_h$ is lost.

To account for response rates, the base weight is multiplied by the inverse of the response rate:

$$\omega'_{hi} = \frac{\omega_{hi}}{RR_h} \quad (29)$$

Where

ω'_{hi} = response rate adjusted base weight for ownership *i* in stratum *h*,

$RR_h = \frac{R_h}{R_h + NR_h}$ = response rate in stratum *h*,

R_h = number of sample points owned by respondents in stratum *h*, and

NR_h = number of sample points owned by nonrespondents in stratum *h*.

This approach for response rate adjustments is similar to the partial plot adjustments described in the Foundational Documentation chapter.

Estimates of ownership totals are the summation of the products of the adjusted base weights, domain variables, and the variables of interest (Eq. 30a). As defined, the domain variable is a binary variable indicating an ownership's inclusion in the domain of interest. Likewise, the attribute of interest is defined by a binary variable, y_{hi}, where 1 indicates the presence of the attribute of interest and is 0 otherwise. The estimates of acreage totals are similar to ownership totals with the inclusion of area owned, a_{hi}, in the product chain (Eq. 30b):

$$\hat{T}_{Ohd} = \sum_h \sum_i \omega'_{hi} d_{hi} y_{hi} \quad (30a)$$

$$\hat{T}_{Ahd} = \sum_h \sum_i \omega'_{hi} a_{hi} d_{hi} y_{hi} \quad (30b)$$

Where

$\hat{T}_{Ohd}$ = estimated total number of ownerships in domain d in stratum h,
$\hat{T}_{Ahd}$ = estimated total acreage in domain d in stratum h,

d_{hi} = a binary variable indicating inclusion of ownership i in domain d in stratum h,

y_{hi} = a binary variable indicating the presence of a specific attribute for ownership i in stratum h, and

a_{hi} = area owned by ownership i in stratum h.

Proportions are derived from the total estimators above. The total number (or acreage) of ownerships with a given attribute is divided by the total number (or acreage) of ownerships in the domain. For example, we may want to know what proportion of family forest ownerships in a state have a written forest management plan (using Eq. 31a) or what proportion of family forest acres in a state are owned by someone who has a written forest management plan (using Eq. 31b):

$$\hat{p}_{Ohd} = \frac{\sum_h \sum_i \omega'_{hi} d_{hi} y_{hi}}{\sum_h \sum_i \omega'_{hi} d_{hi}} \tag{31a}$$

$$\hat{p}_{Ahd} = \frac{\sum_h \sum_i \omega'_{hi} a_{hi} d_{hi} y_{hi}}{\sum_h \sum_i \omega'_{hi} a_{hi} d_{hi}} \tag{31b}$$

Where

$\hat{p}_{Ohd}$ = estimated proportion of ownerships in domain d in stratum h, and

$\hat{p}_{Ahd}$ = estimated proportion of acreage in domain d in stratum h.

Means are calculated the same way as proportions except that the variable representing the attribute of interest, y_{hi}, is a numeric value, as opposed to simply binary as above. For example, to calculate the mean size of forest holdings, a_{hi} is substituted for y_{hi} in Equation 31a.

Quantiles, including median values, are estimated with an iterative approach. Successive values for the variable of interest are tested until the probability of interest, e.g., 0.5 for medians, is reached in terms of the proportion of ownerships, or acreage.

Variance Estimation

A resampling, bootstrap approach (Efron and Tibshirani 1986) is used to estimate variances. This approach was selected due to the complex sample design of the NWOS and the resulting complications in developing a closed-form variance estimator. With bootstrapping, a specified number of replicates are created by randomly selecting sample points, with replacement, from the original full sample list until the original sample size is reached; this is done at the state level. Based on attenuation of coefficients of variation[1], 1,000 replicates were deemed adequate to estimate variances (Butler and Caputo 2021). The value for the estimate of interest is calculated for each replicate and the variance is calculated based on the replicate estimates.

Post-stratification for variance reduction is used for the FIA forest inventory (see Foundational Documentation chapter). This technique is something that could be investigated for future iterations of the NWOS.

Nonresponse

Unit and item nonresponse are issues that need to be addressed for the NWOS, as they do for virtually all surveys. Unit nonresponse is when an ownership included in the sample does not respond. Ownerships that fail to meet the minimum 75 percent threshold for completeness of their questionnaire are treated as nonrespondents (Butler et al. 2021). Item nonresponse occurs when an ownership returns a questionnaire but fails to respond to all of the questions asked of them.

Unit Nonresponse

The overall cooperation rate for the 2018 NWOS was 40 percent (Butler et al. 2021). The state-level cooperation rates ranged from less than 30 percent in Connecticut, Hawaii, and Nebraska to over 60 percent in coastal Alaska, Nevada, North Dakota, and South Dakota (Fig. 7). Cooperation was calculated as the number of completed responses divided by the sum of the numbers of completed responses, partial responses, and nonresponses; this corresponds to the American Association for Public Opinion Research's COOP3 equation (AAPOR 2016, p. 63).

[1] Coefficient of variation is defined here as the estimated standard error of the total divided by the estimate of the total.

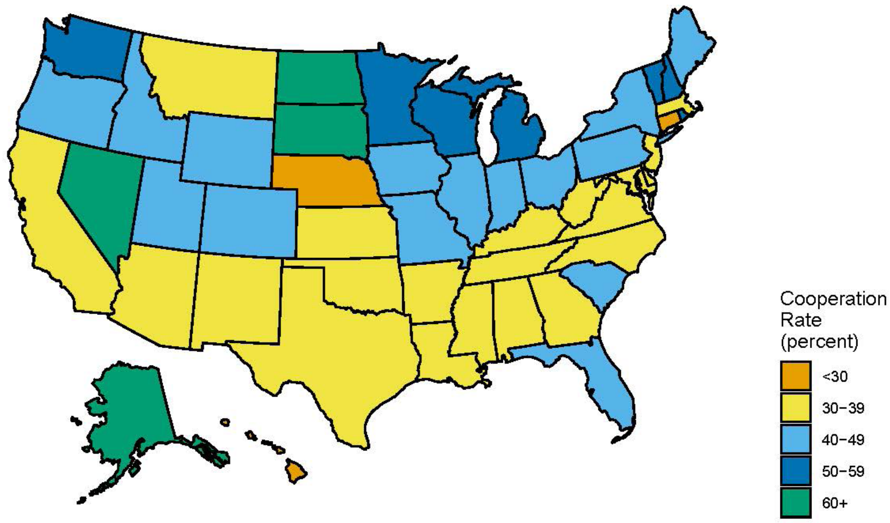

Figure 7.—Cooperation rates for the 2018 USDA Forest Service, National Woodland Owner Survey (Butler et al. 2021). Created by Brett Butler, USDA Forest Service.

Unit nonresponse is addressed in the weights through an adjustment based on the response rates (Eq. 29). This naively assumes no nonresponse bias. A common approach to assessing nonresponse bias is to contact a portion of the nonrespondents via an alternative mode and make comparisons between the responses received for the different modes (Dillman et al. 2014). For the base NWOS, telephone interviews are conducted with a subset of the sampled ownerships that responded via mail. For the telephone nonresponse assessment for the 2018 NWOS, most variables did not significantly vary between the mail and phone respondents and none of the differences had a large effect size (Butler et al. 2021). Overall, the NWOS appears to capture a representative sample of forest landowners in most regards, with nonrespondents being slightly less active on their land and less sure of future plans. Comparison across different modes provides an important qualitative assessment, but cannot be used to make adjustments.

To make unit nonresponse adjustments, a response propensity modeling approach is adopted. To do so, data are required for all potential respondents in the sample (Lohr 1999). The auxiliary variables used for the NWOS unit nonresponse adjustments are: sample origin (FIA forest inventory or augmentation), population density, ecoregion, and size of the parcel at plot center. A response model, using random forests (Breiman 2001), with these auxiliary variables is created and the adjustment factors are calculated as the inverse of the probability of response as predicted by the model. These adjustment factors are combined with the weights (Eq. 32a) and the weights are recalibrated to ensure the total areas in the strata do not change (Eq. 32b). Additional details are available in Butler and Caputo (2021) and Butler et al. (2021).

$$\omega''_{hi} = \omega'_{hi} \times \text{nr\_adj}_{hi} \tag{32a}$$

$$\omega'''_{hi} = \omega''_{hi} \times \frac{\sum_h \sum_i (\omega'_{hi} \times a_{hi})}{\sum_h \sum_i (\omega''_{hi} \times a_{hi})} \tag{32b}$$

Where

ω''_{hi} = unit nonresponse, response rate adjusted base weight for ownership i in stratum h, and nr_adj_{hi}= unit nonresponse adjustment for ownership i in stratum h.

Item Nonresponse

Question-level item nonresponse rates average 4 percent for the 2018 NWOS (Butler et al. 2021). This excludes the “other (please specify)” variables, which were never intended for quantitative analysis and were dropped from the analyses. Nineteen percent of the questions (n = 18) have item nonresponse rates of over 5 percent, and 3 percent of the questions (n = 3) have item nonresponse rates in excess of 10 percent (Fig. 8). Due to excessive missingness, the results from the “Nontimber forest products—reason for collecting,” “Management plan implementation,” and “Management plan writer” questions should be viewed cautiously.

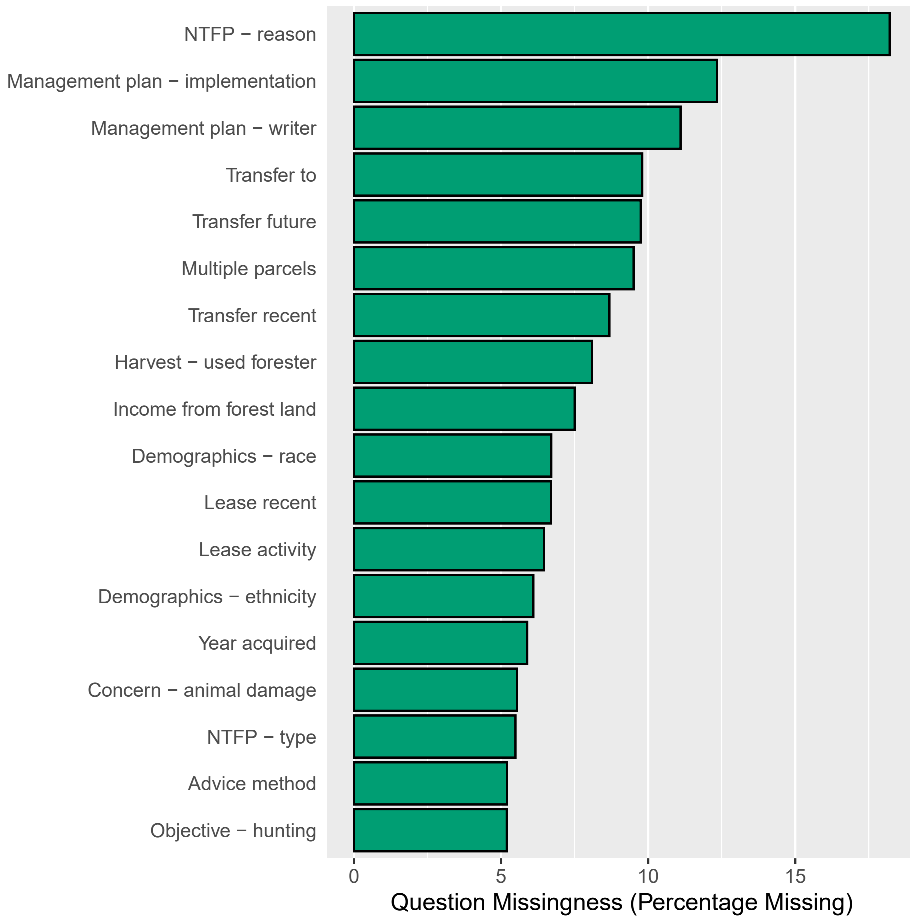

Figure 8.—Item nonresponse rates for questions with missingness of over 5 percent from the 2018 USDA Forest Service, National Woodland Owner Survey (Butler et al. 2021). Created by Brett Butler, USDA Forest Service.

There are different approaches available for addressing item nonresponse. Multiple imputation with chained equations (mice) is a robust and increasingly common approach (Azur et al. 2011, van Buuren 2018) and is the one used for the NWOS. This technique uses a series of models that predict the missing values based on the observed values. Random forests (Breiman 2001) is used for the underlying models because it is a robust technique that can easily handle discrete and continuous variables. The whole process is implemented through the mice package in R (van Buuren and Groothuis-Oudshoorn 2011). A total of five imputed data sets are created; this is the number of imputations deemed sufficient by Rubin (1987) for datasets with low to moderate amounts of item nonresponse and good convergence of imputations, of which the NWOS qualifies. Following imputation, values are checked to ensure internal consistency. Additional details of the NWOS imputation implementation are available in Butler et al. (2021).

Minimum Sample Size Guidance

The target sample sizes for the NWOS are based on desired error rates and logistical/financial constraints. The coefficients of variation for estimates of number and acreage of family forest ownerships are used for this assessment. At a sample size of 100 respondents in a state, the major reduction in the coefficient of variation, defined as the estimated standard error of the total divided by the estimate of the total, is obtained. After a sample size of 250 respondents, the rate of reduction of the coefficient of variation is greatly reduced. At a sample size of 500 respondents, the estimates will obtain a minimum coefficient of variation of 0.05 for acreage of family forest ownerships with 1+ and 10+ acres. In terms of ownerships, estimates for 1+ acre ownerships exceed the 0.05 threshold, but estimates for ownerships with 10+ acres are within this threshold. States with at least 10 responses are included in regional and national totals, but a minimum of 100 respondents is the threshold to publish state-level results. Similar thresholds can be applied to all custom data retrievals and analyses.

Literature Cited

American Association for Public Opinion Research (AAPOR). 2016. **Standard definitions: final dispositions of case codes and outcome rates for surveys. 9th edition.** Oakbrook Terrace, IL: AAPOR. 81 p. https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf (accessed December 6, 2021).

Azur, M.J.; Stuart, E.A.; Frangakis, C.; Leaf, P.J. 2011. **Multiple imputation by chained equations: What is it and how does it work?** International Journal of Methods in Psychiatric Research. 20(1): 40–49. https://dx.doi.org/10.1002%2Fmpr.329.

Birch, T.W. 1996. **Private forest-land owners of the United States, 1994.** Resource Bulletin NE-134. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 183 p. https://doi.org/10.2737/NE-RB-134.

Breiman, L. 2001. **Random forests.** Machine Learning. 45(1): 5–32. https://doi.org/10.1023/A:1010933404324.

Butler, B.J. 2008. **Family forest owners of the United States, 2006.** Gen. Tech. Rep. NRS-27. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 73 p. https://doi.org/10.2737/NRS-GTR-27.

Butler, B.J.; Butler, S.M.; Caputo, J.; [et al.]. 2021. **Family forest ownerships of the United States, 2018: results from the USDA Forest Service, National Woodland Owner Survey.** Gen. Tech. Rep. NRS-199. Madison, WI: U.S. Department of Agriculture, Forest Service, Northern Research Station. 52 p. [plus 4 appendixes]. https://doi.org/10.2737/NRS-GTR-199.

Butler, B.J.; Caputo, J. 2020. **NWOS: An R package for working with USDA Forest Service, National Woodland Owner Survey data.** https://github.com/familyforestresearchcenter/nwos (accessed August 13, 2020).

Butler, B.J.; Caputo, J. 2021. **Weighting for the USDA Forest Service, National Woodland Owner Survey.** Gen. Tech. Rep. NRS-198. Madison, WI: U.S. Department of Agricutlure, Forest Service, Northern Research Station. 24 p. https://doi.org/10.2737/NRS-GTR-198.

Butler, B.J.; Hewes, J.H.; Dickinson, B.J.; [et al.]. 2016. **USDA Forest Service National Woodland Owner Survey: national, regional, and state statistics for family forest and woodland ownerships with 10+ acres, 2011–2013.** Resource Bulletin NRS-99. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 39 p. https://doi.org/10.2737/NRS-RB-99.

Caputo, J.; Butler, B.J.; Hartsell, A.J. 2017. **How large is large? Identifying large corporate ownerships in FIA datasets.** Res. Pap. NRS-29. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 6 p. https://doi.org/10.2737/NRS-RP-29.

Dillman, D.A.; Smyth, J.D.; Christian, L.M. 2014. **Internet, phone, mail, and mixed-mode surveys: The tailored design method.** 4th ed. Hoboken, NJ: Wiley & Sons. 528 p.

Efron, B.; Tibshirani, R. 1986. **Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy.** Statistical Science. 1(1): 54–75. https://doi.org/10.1214/ss/1177013815.

Lohr, S.L. 1999. **Sampling: design and analysis.** Pacific Grove: Duxbury Press. 494 p.

Rubin, D.B. 1987. **Multiple imputation for nonresponse in surveys.** New York: Wiley. 320 p. https://doi.org/10.1002/9780470316696.

Valliant, R.; Dever, J.A.; Kreuter, F. 2013. **Practical tools for designing and weighting survey samples.** New York: Springer. 670 p. https://doi.org/10.1007/978-1-4614-6449-5.

van Buuren, S. 2018. **Flexible imputation of missing data.** Boca Raton, FL: CRC Press. 415 p. https://doi.org/10.1201/9780429492259.

van Buuren, S.; Groothuis-Oudshoorn, K. 2011. **mice: multivariate imputation by chained equations in R.** Journal of Statistical Software. 45(3): 1–67. https://doi.org/10.18637/jss.v045.i03.

Chapter 5: Timber Products Output

John W. Coulston

Wood product markets affect forest sector jobs (Hodges et al. 2012, Sorenson et al. 2016, Woodall et al. 2012), shape the composition and structure of future forests (Wear et al. 2016), and are strong drivers of investments in forest management (FAO 2009). Monitoring timber products output (TPO) is key to understanding the current utilization of raw material to support these markets. In the United States, TPO monitoring has been a constituent program within the Forest Inventory and Analysis program (FIA) since 1948. Estimates from the TPO program have provided the essential foundation for U.S. timber market analyses and projections (e.g., Abt et al. 2009, Adams and Haynes 1996, Buongiorno 1996, Ince et al. 2011, McCarl et al. 2000), sustainability analyses (e.g., Shifley and Moser 2016, USDA Forest Service 2012, Wear and Greis 2002, Wear and Greis 2013), policy analysis (Boyd and Hyde 1989, Haynes 2003), and local wood basket analysis of potential market expansion. The goal of the TPO effort is to estimate volumes by timber product (Table 3), species group (e.g., hardwood, softwood), and county of origin.

Population and Sampling Frame

The population of interest for timber products output estimation is all facilities receiving roundwood from the United States. The roundwood received can be in log form, bolt form, or chipped roundwood (Bentley and Johnson 2011). The population includes primary mills and other entities such as power companies, traders, and direct log exporters that also receive roundwood originating from the United States. Primary facilities include sawmills, pulp mills, veneer and plywood mills, composite panel mills, biomass and energy plants, pole, post, and piling mills, and other miscellaneous mills that accept roundwood (Table 3).

Table 3.—Terminology used in timber products output monitoring

Term	Definition
Bioenergy/Fuelwood	Roundwood products and mill residue byproducts used to produce some form of energy (heat, steam, etc.) in residential, industrial, or institutional settings.
Byproducts	Primary wood products, e.g., pulp chips, animal bedding, and fuelwood, recycled material from mill residues.
Composite panels	Roundwood products manufactured into chips, wafers, strands, flakes, shavings, or sawdust and then reconstituted into a variety of panel and engineered lumber products.
Industrial roundwood products	Any primary use of the main stem of a tree, such as saw logs, pulpwood, veneer logs, intended to be processed into primary wood products such as lumber, wood pulp, or sheathing, at primary wood using mills.
Post, poles, pilings	Roundwood products milled (cut or peeled) into standard sizes (lengths and circumferences) to be put in the ground to provide vertical and lateral support in buildings, foundations, utility lines, and fences. May also include nonindustrial (unmilled) products.
Pulpwood	A roundwood product that will be reduced to individual wood fibers by chemical or mechanical means. The fibers are used to make a broad generic group of pulp products that includes paper products, as well as fiberboard, insulating board, and paperboard.
Sawlog	A roundwood product, usually 8 feet in length or longer, processed into a variety of sawn products such as lumber, cants, pallets, railroad ties, and timbers.
Veneer log	A roundwood product either rotary cut, sliced, stamped, or sawn into a variety of veneer products such as plywood, finished panels, veneer sheets, or sheathing.

A list of roundwood receiving facilities serves as the sampling frame and the frame is maintained by the FIA program in cooperation with partner organizations (e.g., state forestry agencies). Maintaining the sampling frame is important for producing quality timber product statistics. The frame is updated annually by the FIA program and partners by contacting each facility to ensure they are still in business with the same mill capacity consumption, number of employees, and other relevant information. Establishments no longer in business or that are inactive are coded appropriately and not considered part of the sample frame for that year. New facilities can be identified in several ways. First, when contacting existing facilities, the practitioner can also ask about new facilities in the area. The practitioner may also rely on professional organizations, trade journals, and the internet to identify planned mill construction. New facilities that are identified are contacted and added to the frame.

Sample Design and Selection

The TPO program uses a stratified random sampling approach (Coulston et al. 2018) with a sampling fraction of 0.4 (40 percent sample). The population of roundwood receiving facilities is divided into subpopulations where subpopulations are defined by the state the facility resides in (or country when considering Canadian mills) and the primary roundwood product the facility receives. Each subpopulation is then stratified based on facility measure of size (e.g., amount of roundwood received or capacity). The goal of the stratification is to arrange the facilities into strata where the total within stratum measure of size (MOS) is similar. Facilities with an MOS greater than 10 million cubic feet of roundwood are placed in their own strata; however, this threshold may be adjusted by subpopulation. Strata containing a single facility are considered sampled with certainty ($N_h = n_h = 1$, where N_h and n_h are the number of facilities and the number of sampled facilities, respectively, for stratum h). The remaining facilities in the subpopulation are then arranged into strata and a fixed sample size of $n_h = 2$ is randomly selected.

For example, suppose the subpopulation of interest is active facilities in South Carolina that receive saw logs and that there are $N = 17$ of these facilities (Table 4). The target sample size is $n = N*0.4 = 6.8$ facilities. To identify the strata boundaries, the facility list is placed in descending MOS order. Note that facility 1 has an MOS of 12 million cubic feet. This facility is sampled with certainty, placed in its own stratum (stratum A), and then the stratum boundary is drawn (Table 4). The sample size subject to randomization is then $n_{randomized} = n\text{-}n_{\mathbf{certain}} = 7\text{-}1 = 6$. The cumulative MOS for each stratum h subject to randomization is approximated by

$$\sum MOS_h = 2\sum_{i=1}^{n_{randomized}} MOS_i/n_{randomized}. \tag{33}$$

Table 4.—Hypothetical example of the stratification and sample selection process for a subpopulation

Facility number	Subpopulation	Facility MOS (mmcf)	Stratum	Sample	Inclusion probability
1	South Carolina Saw logs	12	A	X	1.00
2	South Carolina Saw logs	9	B	X	0.67
3	South Carolina Saw logs	9	B	X	0.67
4	South Carolina Saw logs	9	B		0.67
5	South Carolina Saw logs	8	C		0.40
6	South Carolina Saw logs	6	C	X	0.40
7	South Carolina Saw logs	5	C		0.40
8	South Carolina Saw logs	4	C		0.40
9	South Carolina Saw logs	4	C	X	0.40
10	South Carolina Saw logs	4	D	X	0.25
11	South Carolina Saw logs	4	D		0.25
12	South Carolina Saw logs	4	D		0.25
13	South Carolina Saw logs	3	D		0.25
14	South Carolina Saw logs	3	D		0.25
15	South Carolina Saw logs	3	D		0.25
16	South Carolina Saw logs	3	D		0.25
17	South Carolina Saw logs	3	D	X	0.25

In the example (Table 4), the target cumulative MOS is 27 million cubic feet. Starting with facility 2, the MOS is accumulated until the total is approximately 27 million cubic feet and then a stratum line is drawn. This results in facilities 2 through 4 being assigned to stratum B. Starting with facility 5, the procedure is repeated, and the stratum C boundary is drawn once the cumulative MOS is at least 27 million cubic feet. Stratum C contains facilities 5 through 9. Stratum D contains the remaining facilities (10–17). For each stratum subject to randomization a sample size of n_h = 2 is randomly selected (denoted by X in Table 4). The overall effect of this approach to stratification is that facilities that have a larger MOS also have a greater inclusion probability (Table 4).

The base sampling fraction for each subpopulation is 0.4. However, because of the sampled with certainty strata, all pulpmills that receive pulpwood (primary product) will be sampled (i.e., all roundwood receiving pulpmills exceed 10 million cubic feet of receipts). The minimum sample size for any state is 20, the minimum sample size for any product is 5, and the minimum sample size within strata is 2 except for sampled with certainty strata.

TPO monitoring is intended to be flexible so that market shifts and emerging markets can be captured. To this end, the sampling fraction of the TPO sample can be adjusted for a specific subpopulation or across subpopulations to meet information needs. For example, a state may choose to implement a 100 percent sample (complete enumeration) or a state may choose to increase the sampling fraction for a specific product. Additional subpopulations can also be added to capture emerging markets. TPO practitioners work directly with partners (e.g., state forestry organizations) to implement intensifications.

Data Collection Process and Timing

Sampling is conducted annually based on an up-to-date sampling frame resulting in a new set of sampled facilities each year (i.e., the samples are independent). Typically, the practitioner will update the sampling frame in the fall of each year and the annual sample is identified once the frame is updated. In January, the practitioner will send out postcards to inform selected facilities of the upcoming survey and survey questionnaires two weeks after initial postcard mailings. The questionnaire is the primary survey instrument. The TPO survey form is designed to determine the location, size, and composition of the primary wood-using facilities in each state, the volume of roundwood harvested by product, species, and geographic location, and the volume and disposition of wood residues generated during primary processing annually. All facility-level information is confidential and used only to construct aggregate estimates.

Estimation

There are five main population parameters that are estimated: receipts, production, imports, exports, and retained. These parameters may be estimated from either a volume (e.g., cubic feet) or weight perspective (e.g., tons). The retained roundwood volume is the amount of material processed in the state where the roundwood originated from. Exports are the roundwood volume from a state that was sent to facilities in other states or out of the country (e.g., Canada or direct log exports). Imports are the roundwood volume from other states that was received by facilities in the state of interest. Receipts are the total roundwood volume received by facilities in the state of interest (retained + domestic imports). Production is the roundwood volume from the state of interest used for timber products (retained + exports). In addition, these parameters may in some cases be estimated by product, species, and in the case of production, by county.

The standard direct estimators for stratified random sampling from Cochran (1977) are used to construct estimates for each subpopulation. Under the stratified random sample design each facility (*i*) belongs to a single subpopulation (*s*) and strata (*h*). In the notation below we rely on the *s* subscript to recognize different subpopulation and we do not carry the *s* subscript through all variables as *i* and *h* are defined uniquely for each subpopulation. The estimated total $\hat{Y}_s$ is then

$$\hat{Y}_s = \sum_h^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} \tag{34}$$

with estimated variance

$$v(\hat{Y}_s) = \sum_h^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) v(\bar{Y}_h) \tag{35}$$

Where

$$v(\bar{Y}_h) = \frac{1}{n_h(n_h - 1)} \sum_{i=1}^{n_h} (y_{hi} - \bar{Y}_h)^2 \text{ , and} \tag{36}$$

$$\bar{Y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} \text{ is the mean for stratum } h. \tag{37}$$

Estimates of the total and the variance of the total for a subpopulation can be further aggregated to construct estimates of total roundwood receipts for a state across products, for example. In this case both the estimates and variances are additive across subpopulations of interest.

Equations 34–37 are modified to construct estimates for domains (*d*) within the subpopulation. For example, the volume of hardwood saw logs imported to facilities in Virginia from other states is a domain of study. The volume of softwood saw logs from Wake County, North Carolina received by saw mills in North Carolina is another example of a domain of study. To construct domain estimates, we introduce a domain indicator to be used in the construction of domain (*d*) totals and variances of totals:

$$\hat{Y}_{sd} = \sum_h^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}\, \delta_{hid} \qquad (38)$$

With estimated variance

$$v\left(\hat{Y}_{sd}\right) = \sum_h^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) v(\bar{Y}_{hd}) \qquad (39)$$

Where

δ_{hid} = domain indicator which takes the value of 1 when facility *i* in stratum *h* is in domain *d* and zero. Otherwise,

$$v(\bar{Y}_{hd}) = \frac{1}{n_h(n_h - 1)} \sum_{i=1}^{n_h} (y_{hi}\delta_{hid} - \bar{Y}_{hd})^2 \text{, and} \qquad (40)$$

$$\bar{Y}_{hd} = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}\, \delta_{hid}. \qquad (41)$$

is the mean for stratum *h*. Estimated totals and their variances are additive across subpopulations for a domain of interest.

Each facility selected with certainty forms its own stratum with $N_h = n_h = 1$ and when the distribution of the MOS is highly skewed $N_h = n_h = 2$ can also occur. When $N_h = n_h$ the stratum is completely enumerated and hence the contribution to the variance of the estimate is assumed to be zero.

Change estimates are constructed by using the independent samples collected at two points in time. Because sampled with certainty strata are complete enumerations and strata subject to randomization are constructed each year there is no covariance term in the estimation of change. Change estimates are

$$\Delta\hat{Y}_{t_2,t_1} = \hat{Y}_{t_2} - \hat{Y}_{t_1} \tag{42}$$

With estimated variance

$$v\left(\Delta\hat{Y}_{t_2,t_1}\right) = v\left(\hat{Y}_{t_1}\right) + v\left(\hat{Y}_{t_2}\right). \tag{43}$$

Most of the parameter estimates used for TPO reporting are constructed with domain estimation (Table 5). However, estimates of state roundwood receipts and state roundwood receipts by product will generally rely on Equations 34–37.

Table 5.—Appropriate design-based estimators for constructing typical timber products estimates

Parameter	State	Product	Species	County
	equation number (total, variance)			
Receipts	34, 35	34, 35[a]	38, 39	--
Retained	38, 39	38, 39	38, 39	--
Exports	38, 39	38, 39	38, 39	--
Imports	38, 39	38, 39	38, 39	--
Production	38, 39	38, 39	38, 39	38, 39

[a]If facilities process multiple products then equations 38 and 39 should be used.

Nonresponse

Nonresponse is a concern in any survey whether sample-based or census-based. There are two types of nonresponse: item nonresponse and unit nonresponse. Unit nonresponse refers to a complete missing value for a selected sample unit from the frame. Item nonresponse refers to missing values for specific variables collected for each sample unit (Cochran 1977, Kish 1995).

The primary goal of any survey is to minimize nonresponse, yet a nonresponse plan should be developed. The TPO nonresponse plan contains the following elements:

- Evaluation of the survey questionnaire to ensure the questions are understandable and follow a logical format.
- Evaluation of respondent's burden.
- A communication plan that informs respondents of the importance of the survey.
- Follow-up schedule for cases of both unit and item nonresponse. This includes reminders, follow-up phone calls, and in person visits.

The core approach for unit nonresponse is to collapse strata such that $n_h \geq 2$. This approach assumes that responses are missing at random within strata. The core optional approach is to model missing values based on historical information. In cases of item nonresponse, the missing values are modeled. As of the writing of this chapter, research is underway to evaluate key mechanisms behind nonresponse, the reasonableness of the missing-at-random assumption, and the efficacy of imputation methods for nonresponse (Rubin 1986).

Literature Cited

Abt, R.C.; Cubbage, F.W.; Abt, K.L. 2009. **Projecting southern timber supply for multiple products by subregion.** Forest Products Journal. 59(7-8): 7–16.

Adams, D.; Haynes. R. 1996. **The 1993 timber assessment market model: structure, projections, and policy simulations.** Gen. Tech. Rep. PNW-GTR-368. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 58 p. https://doi.org/10.2737/PNW-GTR-368.

Bentley, J.W.; Johnson, T.G. 2011. **Mississippi's timber industry—an assessment of timber product output and use, 2009.** Resource Bulletin SRS-181. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 31 p. https://doi.org/10.2737/SRS-RB-181.

Boyd, R.G.; Hyde, W.F. 1989. **Forestry sector intervention: the impacts of public regulation on social welfare.** Ames, IA: Iowa State University Press. 295 p.

Buongiorno, J. 1996. **Forest sector modeling: a synthesis of econometrics, mathematical programming, and system dynamics methods.** International Journal of Forecasting. 12: 329–343. https://doi.org/10.1016/0169-2070(96)00668-1.

Cochran, W.G. 1977. **Sampling techniques.** New York: John Wiley & Sons. 428 p.

Coulston, J.W.; Westfall, J.A.; Wear, D.N.; [et al.]. 2018. **Annual monitoring of U.S. timber production: rationale and design.** Forest Science. 64(5): 533–543. https://doi.org/10.1093/forsci/fxy010.

Food and Agriculture Organization [FAO]. 2009. **State of the world's forests: 2009.** Rome: Food and Agriculture Organization of the United Nations. 152 p. https://www.fao.org/3/i0350e/i0350e00.htm.

Haynes, R.W. 2003. **An analysis of the timber situation in the United States: 1952–2050.** Gen. Tech. Rep. PNW-GTR-560. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 254 p. https://doi.org/10.2737/PNW-GTR-560.

Hodges, D.G.; Hartsell, A.J.; Brandeis, C.; [et al.]. 2012. **Recession effects on the forest and forest products industries of the South.** Forest Products Journal. 61(8): 614–624. https://doi.org/10.13073/0015-7473-61.8.614.

Ince, P.J.; Kramp, A.D.; Skog, K.E.; [et al.]. 2011. **U.S. Forest products module: a technical document supporting the Forest Service 2010 RPA Assessment.** Res. Pap. FPL-RP-662. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 61 p. https://doi.org/10.2737/FPL-RP-662.

Kish, L. 1995. **Survey sampling.** New York: John Wiley & Sons. 643 p.

McCarl, B.A.; Adams, D.M.; Alig, R.J.; [et al.]. 2000. **Effects of global climate change on the U.S. forest sector: response functions derived from a dynamic resource and market simulator.** Climate Research. 15: 195–205. https://doi.org/10.3354/cr015195.

Rubin, D.B. 1986. **Basic ideas of multiple imputation for nonresponse.** Survey Methodology. 12(1): 37–47.

Shifley, S.R.; Moser, W.K. 2016. **Future forests of the northern United States.** Gen. Tech. Rep. NRS-151. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 388 p. https://doi.org/10.2737/nrs-gtr-151.

Sorenson, C.B.; Keegan, C.E.; Morgan, T.A.; [et al.]. 2016. **Employment and wage impacts of timber harvesting and processing in the United States.** Journal of Forestry. 114(4): 474–482. https://doi.org/10.5849/jof.14-082.

USDA Forest Service. 2012. **Future of America's forest and rangelands: Forest Service 2010 Resources Planning Act Assessment.** Gen. Tech. Rep. WO-87. Washington, DC. 198 p. https://doi.org/10.2737/WO-GTR-87.

Wear, D.N.; Greis, J.G., eds. 2002. **Southern forest resource assessment—technical report.** Gen. Tech. Rep. SRS-53. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 635 p. https://doi.org/10.2737/SRS-GTR-53.

Wear, D.N.; Greis, J.G., eds. 2013. **The southern forest futures project: technical report.** Gen. Tech. Rep. SRS-GTR-178. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 542 p. https://doi.org/10.2737/SRS-GTR-178.

Wear, D.N.; Prestemon, J.P.; Foster, M.O. 2016. **U.S. forest products in the global economy.** Journal of Forestry. 114(4): 483–493. https://doi.org/10.5849/jof.15-091.

Woodall, C.W.; Luppold, W.G.; Ince, P.J.; [et al.]. 2012. **An assessment of the downturn in the forest products sector in the northern region of the United States.** Forest Products Journal. 61(8): 604–613. https://doi.org/10.13073/0015-7473-61.8.604.

Chapter 6: FIA Carbon Attributes

Grant M. Domke, Brian F. Walters, James E. Smith, and Christopher W. Woodall

Given the broad interest in forest carbon estimation as well as international reporting requirements from treaties such as the United Nations Framework Convention on Climate Change, the FIA program is continually evaluating existing methods and models used to estimate carbon attributes in the FIADB and testing new methods and modeling approaches to characterize forest carbon attributes of interest. Tree- and condition-level carbon attributes in the FIADB Forest Inventory and Analysis (FIA) database (FIADB) are estimated by using models and tree-, condition-, and plot-level core variables that have been measured or estimated along with auxiliary site-level variables (e.g., mean temperature; soil order) which have been spatially joined to actual plot locations. The carbon attributes reported in the FIA program are used to establish baselines in a regulatory carbon market (Marland et al. 2017), in national and international reporting (FAO 2016, USDA OCE 2016, U.S. EPA 2020), state and local reporting (Kurtz et al. 2015, Morin et al. 2017), as well as many other research and inventory and monitoring activities.

This chapter briefly describes the carbon attributes currently in the FIADB, how they are compiled, and how those values are used to obtain population estimates and their variances. This chapter also describes ongoing work to develop new approaches for quantifying carbon attributes.

Design-Based Estimators

Population estimates for carbon attributes are obtained by using the same sample-based estimators employed by the FIA program (as shown in the Foundational Documentation chapter). Carbon attributes are compiled with models at either the tree- (e.g., $C_{\mathbf{agb}}$) or condition level (e.g., $C_{\mathbf{litter}}$). One of the key differences is that tree-level observations usually need to be aggregated and put into a per-unit-area basis, whereas the model prediction at the condition-level is already in the appropriate units, such as tons per acre of forest land. The model prediction can serve as the plot-level summary, y_{id}^* (see Foundational Documentation chapter, Equation 14) in cases where the plot is entirely forested and has only one condition. However, for plots that are only partially forested with a single forest condition, the model prediction must be multiplied by the forest condition proportion to obtain y_{id}^* When more than one forested condition is present, the weighted average value based on the condition proportions must be calculated to obtain y_{id}^{*} Analysts should be aware of this distinction when estimating population totals or ratios as described in the Foundational Documentation chapter, Equations 1 through 26.

Currently, the FIA program provides information on: (1) carbon in aboveground and belowground live tree biomass, (2) carbon in above and belowground understory biomass, (3) carbon in aboveground and belowground standing dead tree biomass, (4) carbon in downed dead woody materials, (5) carbon in litter, and (6) carbon in organic soil (Table 6).

Carbon in Aboveground and Belowground Live Tree Biomass

Carbon in the aboveground and belowground biomass of live trees (C_{agb} and C_{bgb} pounds per tree) is provided for trees with a diameter ≥ 1.0 inch. C_{agb} includes the aboveground portion, excluding foliage; while C_{bgb} consists of coarse roots ≥ 0.1 inch in root diameter. The component ratio method (CRM) is currently used to estimate aboveground and belowground carbon in live tree components in the FIADB (Heath et al. 2009, Woodall et al. 2011—see for detailed examples). The CRM involves measuring tree attributes (e.g., diameter at breast height [d.b.h.], tree height) in the field, using those attributes to estimate gross volume (cubic feet), converting gross volume to sound volume of wood by deducting rotten and missing cull volume, and then using specific gravity (Miles and Smith 2009) of wood and bark and the weight of water to convert sound volume to oven-dry bole biomass. Biomass in the stump (D_s) and tops and limbs (D_t) of trees ≥ 5.0 inches d.b.h. is estimated as a proportion of bole biomass that uses component ratios from Jenkins et al. (2003) and Raile (1982). It is important to note that foliage is not currently included as a component of C_{agb} in the CRM in the FIA program, but it is included in national greenhouse gas estimation and reporting efforts (U.S. EPA 2020). Biomass in the foliage (D_f) can be estimated as a proportion of bole biomass that uses component ratios from Jenkins et al. (2003). All tree biomass components obtained from Jenkins et al. (2003) and Raile (1982) must be multiplied by an

Table 6.—Summary of carbon attributes currently included in the Forest Inventory and Analysis Database (FIADB) and notation used to refer to the carbon attributes. See the Notes column for details on alternative methods and models and plans for the future.

Carbon attribute	Units	FIADB name	Notation used in this chapter	Notes
Carbon in aboveground live tree biomass	lbs per tree	CARBON_AG	C_{agb}	New models and methods are currently being explored and will be adopted by the FIA program in the future.
Carbon in belowground live tree biomass	lbs per tree	CARBON_BG	C_{bgb}	New models and methods are currently being explored and will be adopted by the FIA program in the future.
Carbon in standing dead trees	tons per acre of forest land	CARBON_STANDING_DEAD	C_{sd}	Carbon in standing dead tree biomass can be calculated directly for individual trees (see Domke et al. 2011, Domke et al. 2012).
Carbon in aboveground live understory biomass	tons per acre of forest land	CARBON_UNDERSTORY_AG	C_{uagb}	New models and methods are currently being explored and will be adopted by the FIA program in the future.
Carbon in belowground live understory biomass	tons per acre of forest land	CARBON_UNDERSTORY_BG	C_{ubgb}	New models and methods are currently being explored and will be adopted by the FIA program in the future.
Carbon in downed dead woody materials	tons per acre of forest land	CARBON_DOWN_DEAD	C_{dd}	Carbon in downed dead woody materials can be calculated directly for individual pieces (see Woodall et al. 2008, 2013, 2019).
Carbon in litter	tons per acre of forest land	CARBON_LITTER	C_{litter}	Carbon in litter can be calculated directly for FIA plots and a new model based on FIA field data has been developed for condition-level estimates and will be adopted by the FIA program in the future (see Domke et al. 2016).
Carbon in mineral soil	tons per acre of forest land	CARBON_SOIL_ORG	Csoil	Carbon in mineral soil can be calculated directly for FIA plots and a new model based on FIA field data has been developed for condition-level estimates and will be adopted by the FIA program in the future (see Domke et al. 2017).

adjustment factor to estimate CRM biomass. Biomass of saplings (D_{sap}; trees < 5.0 inches d.b.h.) is based on models from Jenkins et al. (2004) by using the observed d.b.h. The model prediction is then multiplied by a sapling adjustment factor (Heath et al. 2009, Woodall et al. 2011). Woodland species biomass (D_w) for stems < 5.0 inches d.b.h. is estimated in a similar fashion as sapling biomass. These species have a volume estimate from ground to tip, which is used to estimate total aboveground biomass similar to the bole biomass (D_b) calculation. C_{agb} is estimated by multiplying by 0.5 (Woodall et al. 2011) for each tree type as follows:

$$C_{agb} = 0.5 \times (D_b + D_s + D_t) \tag{44}$$

$$C_s = 0.5 \times D_{sap} \tag{45}$$

$$C_w = 0.5 \times D_w \tag{46}$$

The oven-dry belowground (coarse root) biomass (D_b) is based on a proportion of bole biomass (using component ratios from Jenkins et al. 2003) and an adjustment factor. Carbon in belowground biomass is then estimated as:

$$C_{bgb} = 0.5 \times D_b \tag{47}$$

Carbon in Aboveground and Belowground Standing Dead Tree Biomass

Carbon in aboveground (C_{agb}) and belowground (C_{bgb}) standing dead tree biomass (pounds per tree) is compiled following the same methods as carbon in live tree biomass. In addition, density reduction factors and structural loss adjustments are applied to account for decay and structural loss in standing dead trees by decay class in the FIADB (Domke et al. 2011).

Carbon in standing dead trees (C_{sd}; tons per acre of forest land) is also compiled for all forested conditions in the FIADB using a model based on geographic area, forest type, and (except for nonstocked conditions) growing stock volume (Smith et al. 2006). Note that the condition-level C_{sd} attribute is not calculated from direct field measurements of standing dead trees on FIA plots (Domke et al. 2011, Domke et al. 2012, Woodall et al. 2012b), so it is advised that the individual-tree approach that uses Equations 44–47 described above be used with density reduction factors and structural loss adjustments (Domke et al. 2011) to compile carbon in aboveground and belowground standing dead tree biomass at the tree-level.

Carbon in Aboveground and Belowground Live Understory Biomass

Carbon in live understory aboveground and belowground biomass (C_{uagb} and C_{ubgb}; tons per acre of forest land) includes the aboveground and belowground portions, respectively, of seedlings and woody shrubs. These carbon density values (per unit area of forest land) are obtained from a model based on geographic area, forest type, and (except for nonstocked and pinyon-juniper stands) live tree carbon density (Smith et al. 2006). These values are components of the U.S. National Greenhouse Gas Inventory (U.S. EPA 2020) and are not based on direct field measurements of seedlings and woody shrubs.

Carbon in Downed Dead Woody Materials

Carbon in downed dead wood (C_{dd}; tons per acre of forest land) includes woody material ≥ 3.0 inches in diameter with a lean angle greater than 45 degrees from vertical, stumps and their roots ≥ 3.0 inches in diameter, and slash piles. This is a carbon density value (per-unit-area of forest land) from a model based on geographic area, forest type, and live tree carbon density (Smith et al. 2006). This attribute is a component of the U.S. National Greenhouse Gas Inventory (U.S. EPA 2019) and is not calculated from direct field measurements of downed woody materials, stumps and their roots, or slash piles.

Downed woody material (including fine/coarse woody materials and slash piles) have been measured on a subset of forested plots in the FIA program since 2001 (Woodall and Monleon 2008, Woodall et al. 2013, 2019). Downed woody material C density (i.e., per-unit-area) values obtained from field measurements of downed dead wood pieces have been found to be statistically significantly different from C_{dd} described above (Domke et al. 2013). The C_{dd} values generally overestimate downed woody material carbon density on sites with small amounts of material and underestimate downed woody material on sites with large amounts of material. The per-unit-area divergence is also evident at the state level but, collectively, the difference between methods for all states is less than

9 percent. The relatively small absolute difference between model- and field-based estimates at the per-unit-area and population scales is not a result of good model fits; rather, it reflects the models overestimating the contribution of C from coarse woody material pieces (diameter ≥ 3 inches) and underestimating the contribution of C downed woody material in slash piles. To compensate for these differences between the values compiled from individual downed dead wood piece measurements and C_{dd} estimates, state-specific adjustment factors have been developed (see Table 4 in Domke et al. 2013) and are used to compile downed dead wood estimates in the U.S. National Greenhouse Gas Inventory (U.S. EPA 2019).

Carbon in Litter

Carbon in litter (C_{litter}; tons per acre of forest land) includes the organic material on the floor of the forest, including fine woody debris, humus, and fine roots in the organic forest floor layer above mineral soil. Although the FIA program has been measuring litter attributes, including carbon content and bulk density, on a subset of forested FIA plots since 2001 (O'Neill et al. 2005, Woodall et al. 2012a), a model using literature values and forest type that was used previously in greenhouse gas reporting is still made available. The C_{litter} attribute is a carbon density (per-unit-area of forest land) value from a model based on geographic area, forest type, and (except for nonstocked and pinyon-juniper stands) stand age (Smith and Heath 2002). This attribute is not calculated from direct field measurements of organic material on the floor of the forest above mineral soil.

Improved methods to estimate carbon based on direct measurements of litter ($C_{field\ litter}$) on forested plots in the FIA program are described in Domke et al. (2016). In general, the C_{litter} values described above show a substantial overestimate relative to estimates obtained directly from FIA plot measurements of litter variables resulting in statistically significant differences between the C_{litter} estimates and those obtained from FIA plot measurements (Domke et al. 2016). As a result of these differences, a new approach to estimate litter carbon density (Domke et al. 2016) has been developed based directly on litter measurements from forested plots in the FIA program along with auxiliary climate variables. This approach is used in the U.S. National Greenhouse Gas Inventory (U.S. EPA 2019) and the values obtained from this method will replace the C_{litter} values in the FIADB in the future (Domke et al. 2016).

Carbon in Soil Organic Matter

Carbon in soil (C_{soil}; tons per acre of forest land) includes the fine organic material below the soil surface to a depth of 39 in. This is a per-unit-area estimate from a model based on soil inventory data, geographic area, and forest type (Amichev and Galbraith 2004, Smith et al. 2006) and was previously used in national greenhouse gas reporting. The C_{soil} attribute is not calculated from direct field measurements of fine organic material below the soil surface on FIA plots.

The FIA program has been consistently measuring soil attributes since 2001 and has amassed an extensive inventory of soil observations in forest land in the conterminous United States and southeast and southcentral coastal Alaska (Domke et al. 2017, O'Neill et al. 2005, USDA Forest Service 2011). Soil samples are collected on a subset of forested FIA plots, and soil cores are taken to a depth of 8 inches adjacent to subplot 2 (see Fig. 4a in the Foundational Documentation chapter) on each of these plots. Methods to estimate carbon in soil on forested plots in the FIA program are described in Domke et al. (2017). The C_{soil} estimates described above are to a depth of 39 inches and the values obtained from the soil cores on FIA plots are to a depth of 8 inches; it is not possible to make direct comparisons of the two values. That said, it is clear when comparing the values by soil order that the model used to obtain C_{soil} values substantially underestimates the contribution of carbon in soil on forested FIA plots (Domke et al. 2017). This downward bias can be attributed to several factors. First, the C_{soil} values described above were developed by using STATSGO data (Schwarz et al. 1995), which has a wide distribution, but much of the data is from nonforest land. These values of soil organic carbon are means over large map units intended for broad planning and management uses covering state, regional, and multi-state areas and are not expected to provide accurate estimates of soil organic carbon for specific locations (Domke et al. 2017, Homann et al. 1998). Second, soil organic carbon estimates were organized by broad forest type in the C_{soil} model whereas C content and bulk density measurements were used to obtain estimates of carbon in soil from the FIA plots (Domke et al. 2017). Finally, given the variability observed in soil carbon estimates, it is likely that the model currently used to estimate soil carbon in the FIADB does not include important interactions between variables (e.g., temperature, precipitation) that directly and indirectly influence soil carbon dynamics (Domke et al. 2017, Jobbágy and Jackson 2000, Parton et al. 2007). As a result of these differences, a new approach to estimate carbon in soil organic matter per-unit-area has been developed based directly on soil measurements from forested plots in the FIA program (USDA Forest Service 2011). This approach is used in the U.S. National Greenhouse Gas Inventory (U.S. EPA 2019) and the values obtained from this method will replace the C_{soil} values in the FIADB in the future (Domke et al. 2017).

Further Research

The carbon attributes described in this chapter can be used in numerous applications such as national greenhouse gas monitoring and reporting, forest carbon cycle research, or project-level carbon monitoring and reporting efforts. As new data become available within the FIA program, the methods and models currently used to estimate carbon attributes in the FIADB are evaluated and refined or replaced to better represent the conditions observed on FIA plots (Domke et al. 2011, 2013, 2016, 2017). To that end, there are ongoing efforts within the FIA program to evaluate and improve methods and models described in this chapter.

The methods and models currently used to estimate aboveground and belowground live and dead tree volume, biomass, and carbon have been evaluated and alternative methods are being explored (Frank et al. 2018, Radtke et al. 2017, Martin et al. 2021, Weiskittel et al. 2015, Westfall et al. 2016, Zhao et al. 2018a, Zhao et al. 2018b). Nationally consistent methods and models are expected to replace existing methods and models in the near future and these will include estimators of foliage biomass and carbon. Research is also underway to evaluate and improve characterization of carbon in aboveground and belowground understory biomass by using FIA plot data and auxiliary information (Johnson et al. 2017, Russell et al. 2014). This includes the use of regionally specific understory vegetation data from the FIA program as well as destructively sampled data from understory studies and digital photo-series data. Results suggest that the current models used to estimate carbon in the understory biomass are overestimating the contribution of carbon in understory biomass (Johnson et al. 2017). The sampling methods for downed woody material have recently changed (Woodall et al. 2019) with efforts underway to revise models to characterize carbon in downed woody material (Smith et al. 2022) and use direct measurements of downed woody materials to characterize the sources of uncertainty associated with new methods and models (Campbell et al. 2019). Finally, while the methods and models used to characterize carbon in litter and soil in the FIA program have recently been revised (Domke et al. 2016, 2017) and will be adopted in the FIADB in the future, there is ongoing work to expand the new methods (Cao et al. 2019) and to use remeasurements from the FIA program to evaluate change in these important carbon pools.

Literature Cited

Amichev, B.Y.; Galbraith, J.M. 2004. **A revised methodology for estimation of forest soil carbon from spatial soils and forest inventory data sets.** Environmental Management. 33(1): S74-S86. http://dx.doi.org/10.1007/s00267-003-9119-0.

Campbell, J.L.; Green, M.B.; Yanai, R.D.; [et al.]. 2019. **Estimating uncertainty in the volume and carbon storage of downed coarse woody debris.** Ecological Applications. 29(2): e01844. https://doi.org/10.1002/eap.1844.

Cao, B.; Domke, G.M.; Russell, M.B.; [et al.]. 2019. **Spatial modeling of litter and soil carbon stocks on forest land in the conterminous United States.** Science of the Total Environment. 654: 94–106. http://dx.doi.org/10.1016/j.scitotenv.2018.10.359.

Domke, G.M.; Perry, C.H.; Walters, B.F.; [et al.]. 2017. **Toward inventory-based estimates of soil organic carbon in forests of the United States.** Ecological Applications. 27(4): 1223–1235. https://doi.org/10.1002/eap.1516.

Domke, G.M.; Walters, B.F.; Perry, C.H.; [et al.]. 2016. **Estimating litter carbon stocks on forest land in the United States.** Science of the Total Environment. 557–558: 469–478. https://doi.org/10.1016/j.scitotenv.2016.03.090.

Domke, G.M.; Woodall, C.W.; Smith, J.E. 2011. **Accounting for density reduction and structural loss in standing dead trees: implications for forest biomass and carbon stock estimates in the United States.** Carbon Balance and Management. 6: 14. https://doi.org/10.1186/1750-0680-6-14.

Domke, G.M.; Woodall, C.W.; Smith, J.S. 2012. **Recent changes in the estimation of standing dead tree biomass and carbon stocks in the U.S. forest inventory.** In: Morin, R.; Likens, G., eds. Moving from status to trends: 2012 Forest Inventory and Analysis (FIA) symposium. Gen. Tech. Rep. NRS-P-105. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station: 164-169.

Domke, G.M.; Woodall, C.W.; Walters, B.F.; [et al.]. 2013. **From models to measurements: comparing down dead wood carbon stock estimates in the U.S. forest inventory.** PLoS ONE. 8(3): e59949. https://doi.org/10.1371/journal.pone.0059949.

Food and Agriculture Organization [FAO]. 2016. **Global forest resources assessment 2015: How are the world's forests changing?** 2nd ed. Rome: U.N. Food and Agriculture Organization. http://www.fao.org/3/a-i4793e.pdf (accessed May 10, 2019).

Frank, J.; Castle, M.E.; Westfall, J.A.; [et al.]. 2018. **Variation in occurrence and extent of internal stem decay in standing trees across the eastern U.S. and Canada: evaluation of alternative modelling approaches and influential factors.** Forestry: An International Journal of Forest Research. 91(3): 382–399. https://doi.org/10.1093/forestry/cpx054.

Heath L.S.; Hansen, M.H.; Smith, J.E.; [et al.]. 2009. **Investigation into calculating tree biomass and C in the FIADB using a biomass expansion factor approach.** In: McWilliams, W.; Moisen, G.; Czaplewski, R., comps. Forest Inventory and Analysis (FIA) symposium 2008. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 24.

Homann, P.S.; Sollins, P.; Fiorella, M.; [et al.]. 1998. **Regional soil organic carbon storage estimates for western Oregon by multiple approaches.** Soil Science Society of America Journal. 62(3): 789–796. https://andrewsforest.oregonstate.edu/publications/2543.

Jenkins, J.C.; Chojnacky, D.C.; Heath, L.S.; [et al.]. 2003. **National scale biomass estimators for United States tree species.** Forest Science. 49(1): 12–35.

Jenkins, J.C.; Chojnacky, D.C.; Heath, L.S.; [et al.]. 2004. **Comprehensive database of diameter-based biomass regressions for North American tree species.** Gen. Tech. Rep. NE-319. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 45 p. https://doi.org/10.2737/NE-GTR-319.

Jobbágy, E.G.; Jackson, R.B. 2000. **The vertical distribution of soil organic carbon and its relation to climate and vegetation.** Ecological Applications. 10(2): 423–436. https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2.

Johnson, K.; Domke, G.M.; Russell, M.B.; [et al.]. 2017. **Estimating aboveground live understory vegetation carbon in the United States.** Environmental Research Letters. 12(12): 125010. https://doi.org/10.1088/1748-9326/aa8fdb.

Kurtz, C.; Moser, W.K.; Hansen, M.H.; [et al.]. 2015. **Forest resources within the Lake States Ceded Territories 1980–2013.** Resource Bulletin NRS-96. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 89 p. https://doi.org/10.2737/NRS-RB-96.

Marland, E.; Domke, G.M.; Hoyle, J.; [et al.]. 2017. **Understanding and analysis: the California Air Resources Board forest offset protocol.** Cham, Switzerland: Springer. 72 p. https://doi.org/10.1007/978-3-319-52434-4.

Martin, A.R.; Domke, G.M.; Doraisami, M.; [et al.]. 2021. **Carbon fractions in the world's dead wood.** Nature Communications. 12(1): 1-9. https://doi.org/10.1038/s41467-021-21149-9.

Miles, P.D.; Smith, W.B. 2009. **Specific gravity and other properties of wood and bark for 156 tree species found in North America.** Res. Note NRS-38. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 35 p. https://doi.org/10.2737/NRS-RN-38.

Morin, R.S.; Domke, G.M.; Walters, B.F. 2017. **Forests of Vermont, 2016.** Resource Bulletin Update FS-119. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 4 p. https://doi.org/10.2737/FS-RU-119.

O'Neill K.P.; Amacher M.C.; Perry, C.H. 2005. **Soils as an indicator of forest health: a guide to the collection, analysis, and interpretation of soil indicator data in the Forest Inventory and Analysis Program.** Gen. Tech. Rep. NC-258. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 53 p. https://doi.org/10.2737/NC-GTR-258.

Parton W.; Silver W.L.; Burke I.C.; [et al.]. B. 2007. **Global-scale similarities in nitrogen release patterns during long-term decomposition.** Science. 315: 361–364. https://doi.org/10.1126/science.1134853.

Radtke, P.; Walker, D.; Frank, J.; [et al.]. 2017. **Improved accuracy of aboveground biomass and carbon estimates for live trees in forests of the eastern United States.** Forestry: An International Journal of Forest Research. 90(1): 32–46. https://doi.org/10.1093/forestry/cpw047.

Raile, G.K. 1982. **Estimating stump volume.** Res. Pap. NC-224. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 7 p. https://doi.org/10.2737/NC-RP-224.

Russell, M.B.; D'Amato, A.W.; Schulz, B.K.; [et al.]. 2014. **Quantifying understory vegetation in the U.S. Lake States: a proposed framework to inform regional forest carbon stocks.** Forestry: An International Journal of Forest Research. 87(5): 629–638. https://doi.org/10.1093/forestry/cpu023.

Schwarz, G.E.; Alexander, R.B. 1995. **State Soil Geographic (STATSGO) database for the conterminous United States.** No. 95-449. https://doi.org/10.3133/ofr95449.

Smith, J.E.; Domke, G.M.; Woodall, C.W. 2022. **Predicting downed woody material carbon stocks in forests of the conterminous United States.** Science of The Total Environment. 803(7): 150061. https://doi.org/10.1016/j.scitotenv.2021.150061.

Smith, J.E.; Heath, L.S. 2002. **A model of forest floor carbon mass for United States forest types.** Res. Paper NE-722. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 37 p. https://doi.org/10.2737/NE-RP-722.

Smith, J.E.; Heath, L.S.; Skog, K.E.; [et al.]. 2006. **Methods for calculating forest ecosystem and harvested carbon with standard estimates for forest types of the United States.** Gen. Tech. Rep. NE-343. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 216 p. https://doi.org/10.2737/NE-GTR-343.

U.S. Environmental Protection Agency (U.S. EPA). 2020. **Inventory of U.S. greenhouse gas emissions and sinks: 1990–2018.** EPA 430-R-20-002. https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2018.

USDA Forest Service. 2011. **Phase 3 field guide—soil measurements and sampling.** V5.1. https://www.fia.fs.usda.gov/library/field-guides-methods-proc/docs/2012/field_guide_p3_5-1_sec22_10_2011.pdf (accessed April 24, 2019).

USDA Office of the Chief Economist (OCE). 2016. **U.S. agriculture and forestry greenhouse gas inventory: 1990–2013.** USDA, Office of the Chief Economist, Climate Change Program Office. Technical Bulletin No. 1943. 137 p.

Weiskittel, A.R.; MacFarlane, D.W.; Radtke, P.J.; [et al.]. 2015. **A call to improve methods for estimating tree biomass for regional and national assessments.** Journal of Forestry. 113(4): 414–424. https://doi.org/10.5849/jof.14-091.

Westfall, J.A.; McRoberts, R.E.; Radtke, P.J.; [et al.]. 2016. **Effects of uncertainty in upper-stem diameter information on tree volume estimates.** European Journal of Forest Research. 135(5): 937–947. https://doi.org/10.1007/s10342-016-0985-4.

Woodall, C.W.; Domke, G.M.; MacFarlane, D.W.; [et al.]. 2012b. **Comparing field- and model-based standing dead tree carbon stock estimates across forests of the United States.** Forestry. 85: 125–133. https://doi.org/10.1093/forestry/cpr065.

Woodall, C.W.; Heath, L.S.; Domke, G.M.; [et al.]. 2011. **Methods and equations for estimating aboveground volume, biomass, and carbon for forest trees in the U.S.'s national inventory, 2010.** Gen. Tech. Rep. NRS-88. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 30 p. https://doi.org/10.2737/NRS-GTR-88.

Woodall, C.W.; Monleon V.J. 2008. **Sampling protocol, estimation, and analysis procedures for the down woody materials indicator.** Gen. Tech. Rep. NRS-22. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 68 p. https://doi.org/10.2737/NRS-GTR-22.

Woodall, C.W.; Monleon, V.J.; Fraver, S.; [et al.]. 2019. **The downed and dead wood inventory of forests in the United States.** Nature Scientific Data. 6: 180303. https://doi.org/10.1038/sdata.2018.303.

Woodall, C.W.; Perry C.H.; Westfall J.A. 2012a. **An empirical assessment of forest floor carbon stock components across the United States.** Forest Ecology and Management. 269: 1–9. https://doi.org/10.1016/j.foreco.2011.12.041.

Woodall, C.W.; Walters, B.F.; Oswalt, S.N.; [et al.]. 2013. **Biomass and carbon attributes of downed woody materials in forests of the United States.** Forest Ecology and Management. 305: 48–59. https://doi.org/10.1016/j.foreco.2013.05.030.

Zhao, D.; Lynch, T.B.; Westfall, J.; [et al.]. 2018a. **Compatibility, development, and estimation of taper and volume equation systems.** Forest Science. 65(1): 1–13. https://doi.org/10.1093/forsci/fxy036.

Zhao, D.; Westfall, J.A.; Coulston, J.W.; [et al.]. 2018b. **Additive biomass equations for slash pine trees: comparing three modeling approaches.** Canadian Journal of Forest Research. 49(1): 27–40. https://doi.org/10.1139/cjfr-2018-0246.

Chapter 7: Emerging Alternative Estimators

Gretchen G. Moisen, Hans-Erik Andersen, David M. Bell, R, John W. Coulston, Tracey S. Frescino, Kelly S. McConville, Ronald E. McRoberts, Paul L. Patterson, James A. Westfall, and Barry T. Wilson

New estimation strategies are evolving in the FIA program. This evolution is motivated by new information needs about change (e.g., transitions out of forest land use), remote areas (e.g., wilderness), smaller geographic areas (e.g., counties and disturbance boundaries) and time scales (e.g., those shorter than the inventory cycle length). The 2014 Farm Bill (Public Law 113-79) contains three provisions that specifically articulate these needs:

- Complete the transition to a fully annualized forest inventory program and include inventory and analysis of interior Alaska
- Understand and report on changes in land cover and use
- Implement procedures to improve precision in substate estimates

In all cases, FIA is investigating new statistical techniques, increasing computing capacity, and improving auxiliary remotely sensed data to help FIA meet these needs. Because the classes of estimators covered here are just emerging in the FIA program and are not yet operational, this chapter takes a very different form than the others in this collection. The purpose of this chapter is to provide a brief overview of topics critical to these three Farm Bill provisions, review FIA's progress to date on estimation strategies that are not yet implemented operationally, and set up a dynamic mechanism to track progress through a new online repository for material referenced here, as well as for rapidly-evolving tutorials and application details that enables people to track FIA's advances under these numerous topics.

Model-Assisted Estimation

As the quantity and quality of relevant remotely sensed data increases, it is important for FIA to explore how these data can be combined with ground plot data to improve the precision of estimates of forest parameters. Model-assisted inference provides one possible framework where an assisting model links these two data sources. For FIA's production processes, it currently uses a simple post-stratified estimator (Bechtold and Patterson 2005) to estimate population means and totals. The post-stratified estimator is a model-assisted estimator where the assisting model is the group mean model (Särndal et al. 1992, pages 264–269), which relies on a single, categorical auxiliary variable—such as the classes obtained from a forest/nonforest or tree canopy cover map.

Inference under the model-assisted framework is design-based where the randomness in the estimators is based solely on the random selection of population units for the sample. No stochastic structure is assumed for variables and instead the values of the response variables and the auxiliary variables in the population are treated as fixed quantities. Therefore, a key feature of model-assisted inference is that a probability-based sample is collected. To construct model-assisted estimators, the response and auxiliary variables must be available for every sample unit and the auxiliary variables collected as either a large probability sample (e.g., double sampling) and/or on every unit in the population. Depending on the form of the assisting model, the auxiliary variable may only be needed in summary form, such as population totals or means.

The assisting model is built on the sampled data and then predictions of the response variable are generated for every population unit. The model-assisted estimator of the population mean is composed of two pieces: the mean of the predicted values over the population and the survey-weighted mean of the residuals. The second component ensures that the estimator is approximately unbiased for a wide range of assisting models and sampling designs, regardless of how well the assisting model captures the true relationship between the response variable and the auxiliary variables.

The variance of the model-assisted estimator is impacted by the predictive accuracy of the assisting model, warranting the need for good model building practices when determining which assisting model to employ. To this end, many assisting models have been explored for forest inventory, such as linear regression (Andersen et al. 2009, Gobakken et al. 2012, Gregoire et al. 2011, Saarela et al. 2015), logistic regression (McRoberts 2010, McRoberts and Walters 2012, McRoberts et al. 2016c), nonlinear regression (McRoberts et al. 2013, Moser et al. 2017), penalized linear regression (McConville et al. 2017), K-nearest neighbors (Baffetta et al. 2009, McRoberts et al. 2015b, McRoberts et al. 2016b, McRoberts et al. 2017), modified generalized regression estimator (Wojcik et al. 2022), and generalized semiparametric additive models (Breidt et al. 2007, Kangas et al. 2016, Opsomer et al. 2007).

To help make model-assisted estimators more accessible, McConville et al. (2020) provide a tutorial on parametric model-assisted estimators with guidance on their use in forest inventory applications. Seven estimators are covered, including: Horvitz-Thompson, ratio, post-stratification, regression, lasso, ridge, and elastic net. An R package called *mase* (Model Assisted Survey Estimation, McConville et al. 2018) is available on the Comprehensive R Archival Network (CRAN), enabling easy computation of these estimators, along with closed form and bootstrap variances. The *mase* package has also been embedded in the R package FIESTA for FIA applications (described in the Computing Resources section of this chapter).

Examples of the use of model-assisted estimators in forest inventory applications can be found in the dynamic content, here: https://www.fia.fs.usda.gov/library/sampling/index.php.

Model-Based Estimation

Assumptions underlying model-based inference, also characterized as model-dependent inference, differ considerably from the more familiar design-based or probability-based inference used for model-assisted estimation. First, the basis for the validity of design-based inference is a probability sample, but the basis for the validity of model-based inference is correct specification of the model. Second, design-based inference assumes a single possible value for each population unit, but model-based inference assumes an entire distribution of possible values for each unit. Third, randomization for design-based inference is accomplished via selection of population units into the sample, while with model-based inference the sample is considered fixed with randomization occurring via realization of observations from the distribution characterizing the population units selected for the sample.

An important consequence of the first assumption is that model-based inference does not require probability samples. Although probability samples may be used and, in fact, may be preferable and purposive, other nonprobability samples may also produce entirely valid model-based inferences provided the model is appropriately specified. The absence of a requirement for a probability sample means that model-based inference can be used for applications for which design-based inference is not possible. First, model-based inference can be used when the sample is sufficient for constructing a model but was not acquired with a probability sampling design. Examples are remote and/or inaccessible populations such as interior Alaska, Siberia, or tropical regions for which probability sampling is not feasible because of logistical and/or cost issues (Andersen et al. 2013, McRoberts et al. 2014). Second, model-based inference can be used when the sample size is insufficient for design-based inference, as discuss in more detail in the following section; examples typically entail estimation for small areas such as individual forest stands (Breidenbach et al. 2016, McRoberts 2006).

An important consequence of the second assumption is how population parameters are interpreted. Although model-based inference assumes a finite population, the fact that each population unit has an entire distribution of possible observations means that an infinite number of finite populations could be realized. Thus, with model-based inference, population parameters are random, whereas with design-based inference they are constants. This conceptual framework is characterized as a superpopulation, and model-based inference has occasionally been characterized as superpopulation inference.

With model-based inference, a model is used to predict the response variable for all population units. The point estimator for the population mean is then simply the mean over the predictions for all population units. An important aspect of model-based inference is that when the model is correctly specified, population estimators are unbiased, but when the model is misspecified, the adverse effects on inference may be substantial (Valliant 2009). Because the model-based estimator of the population mean cannot be assured to be unbiased, mean square error (MSE) rather than variance is used to characterize the uncertainty of model-based estimators.

An emerging use of model-based inference is as a component of hybrid inference (Corona et al. 2014, Fattorini 2012, Ståhl et al. 2016). This form of inference combines design-based and model-based methods and has four key features: (1) a probability sample of population units for which only auxiliary information is available, i.e., observations of the response variable are not available; (2) a prediction technique that uses the auxiliary information to predict the response variable for the sample units; (3) a design-based estimator of the population parameter that uses the predictions for the sample units; and (4) a design-based estimator of the

MSE to estimate the effects of sampling variability together with a model-based estimator to estimate the effects of the uncertainty of the sample unit predictions (McRoberts et al. 2016a).

Although the term hybrid inference was not always used, Breidenbach et al. (2014), McRoberts and Westfall (2014), McRoberts et al. (2015a, 2016a), and Ståhl et al. (2014) all documented a common inventory application for which hybrid inference perhaps should be, but generally has not been, used. They used allometric models to predict volume or biomass for a probability sample consisting of inventory plots, used design-based estimators with the allometric model predictions to estimate mean volume or biomass per unit area, and then used both design-based and model-based estimators to estimate the standard error of the estimate of the mean. Hence, both sampling and measurement error were accommodated.

Examples of the use of model-based estimators in forest inventory applications can be found in the dynamic content, here: https://www.fia.fs.usda.gov/library/sampling/index.php.

Small Area Estimation

Increasingly, FIA is being asked to answer resource questions for small areas or domains, i.e., subsets of the population for which there are too few sample plots from which to construct estimates with adequate precision using a direct estimator. Interest in this stems from needs to quantify ecological conditions, disturbances, and timber supplies for increasingly smaller areas or over specific time intervals. Statistical techniques tailored to produce these small area estimates are needed for rapid assessments such as assessing damage caused by fire and wind events, for quantifying insect damage and the associated rate of spread, and for ecological assessments such as estimating the extent of old growth forests and wildlife habitat for indicator species. In addition, forest industry has expressed substantial interest in quantifying wood supplies surrounding current and potential mill sites, and both the Southern Group of State Foresters and the Northeastern States Research Cooperative have expressed the need for county-level harvest estimates where too few sample plots are collected to warrant FIA's standard direct estimators, described below. User perspectives on the need for small area estimates in the private forest sector and National Forest Systems were published recently in Prisley et al. 2021 and Wiener et al. 2021, respectively. Thus, requirements for small area estimates by FIA's traditional and future clients are expected to increase.

Rao (2003, pages 1–141) gives a comprehensive account of small area models. Small area estimation can be conducted under design-based (i.e., Särndal et al. 1992, pages 20–22) or model-based (i.e., Valliant 2009) modes of inference. Even when sample sizes are small, FIA currently relies primarily on model-assisted direct estimators constructed under a design-based framework, bringing in ancillary data through simple post-stratification. Direct estimators use values of the response variable only from the sample units in the domain. Indirect estimators, on the other hand, borrow strength by using values of the response variable from sample units outside the domain by means of a linking superpopulation model, thus increasing the "effective" sample size.

Some indirect estimators rely upon an implicit superpopulation model, such as a synthetic estimator that uses a reliable direct estimator for a large area to derive an indirect estimator for a small area assuming the small area has similar characteristics. Other indirect estimators rely on an explicit superpopulation model that accounts for the variability in the relationship between auxiliary and response variables among small areas. These models can be specified as either unit-level or area-level models, depending on whether the auxiliary data are available for the individual population units (e.g., plots), or only at the aggregate level for each small area (e.g., counties). Such models permit the estimation of area-specific MSE values, unlike in the case of purely synthetic estimators where the estimated MSE is averaged over all small areas. However, distinct models must be identified and calibrated for each unique delineation of small areas within a population.

A common framework for many explicit superpopulation models is the general linear mixed model, which includes both fixed effects of auxiliary variables across small areas as well as random effects associated with each small area. Empirical Best Linear Unbiased Predictor (EBLUP) estimators can be used to simultaneously estimate the parameters associated with the fixed and random effects (Henderson 1975). These models can be fitted either from a frequentist or Bayesian perspective.

Under the frequentist interpretation, probability refers to the long-term frequency of an observation (the evidence) given repeated outcomes from an experiment, or repeated samples from a population (the hypothesis). The underlying parameters that describe this repeated process are assumed to be unknown but fixed. An early example of using satellite imagery to improve precision in survey estimates of crop acres was provided by Battese et al. (1988). EBLUPs are receiving increasing attention as a tool to produce small area estimates in forest inventory applications (Breidenbach and Astrup 2012, Goerndt et al. 2013, Mauro et al. 2017). EBLUPs also have some intuitive appeal in that they can be algebraically expressed as a weighted average of a direct and synthetic estimator, with more weight going to the synthetic estimator as number of plots decreases.

When fitting an explicit superpopulation model under the Bayesian interpretation, probability refers to the plausibility of a set of underlying parameters describing the random process (the hypothesis) given the fixed set of observations available (the evidence). In this view, the parameters are not fixed but come from a distribution of possible values. Empirical Bayes is equivalent to the EBLUP approach via general linear mixed models. This is not a fully Bayesian approach because it does not incorporate a prior probability distribution of model parameters, but instead estimates values for model parameters based on the sample units by using maximum likelihood, which is fundamentally frequentist. In the Hierarchical Bayes approach, a prior probability distribution of model parameters is specified. This, in conjunction with the values of the sample units, induces a posterior probability distribution of the small area parameter of interest via Bayes theorem.

Small area estimation through explicit superpopulation models and estimators like the EBLUP has been made more accessible through statistical packages such as sae (Molina and Marhuenda 2015) and JoSae (Breidenbach 2018). Both packages have been embedded in the R package FIESTA for FIA applications, described in the Computing Resources section. Synthetic estimators are being implemented through an Amazon Web Services cloud computing environment using Esri's Raster Analytics platform as described in the BIGMAP project in the Computing Resources section.

Guldin (2021) and Dettmann et al. (2022) provide recent reviews of small area estimation in forest inventory applications, and investigations into improving precision in FIA estimates over small domains continues to rise. For example, in the Pacific Northwest, Bell et al. (2022) compare Horvitz Thompson, generalized regression, and k-nearest neighbor synthetic estimates of aboveground live carbon, while Temesgen et al. (2021) use Fay-Herriot models of aboveground biomass and volume specific to stand-level inventories. In the Interior Western United States, estimates for multiple forest attributes were explored using a modified generalized regression estimator over counties (Wojcik et al. 2021), and area-level Hierarchical Bayesian and EBLUPs were compared to post-stratification over ecological subsections (White et al. 2021). In the northern United States, Harris et al. (2021) compare design- and model-based estimates in support of the National Woodland Owner Survey. In the Southern United States, Cao et al. (2022) improve precision in volume estimates for counties using spatial area-level small area estimators. For multiple regions or nationally, Gaines and Affleck (2021) estimate postfire tree density through temporal borrowing strategies, Stanke et al. (2022) develop spatial Fay-Herriot models of forest carbon stocks for counties, and Frescino et al. (2022) deliver a large collection of model-assisted and model-based estimates of forest attributes over ecosubsections, counties, and HUC 10 watersheds through dashboards.

These and other examples of the use of small area estimation in forest inventory applications can be found in the dynamic content, here: https://www.fia.fs.usda.gov/library/sampling/index.php.

Photo-Based Sampling and Estimation

Photo-based estimation involves both an alternative sample design and estimator. The sample design relies on techniques to (1) obtain a sample of individual locations within a region and (2) establish at each location a photo plot with multiple points (or dots) within the photo-plot boundary. At each point within the photo plot (illustrated in Fig. 9), a photo interpreter assesses a set of characteristics, including: (1) condition attributes or land use, e.g., forest type or privately owned; and (2) an object type or land cover, e.g., vegetation type, species of tree, or bare ground. If photos are available at two time periods, then change-based characteristics, such as agent of change, can be assessed by the interpreter and assigned to each point within the photo plot. The estimators for a single point in time are the same as for two points in time.

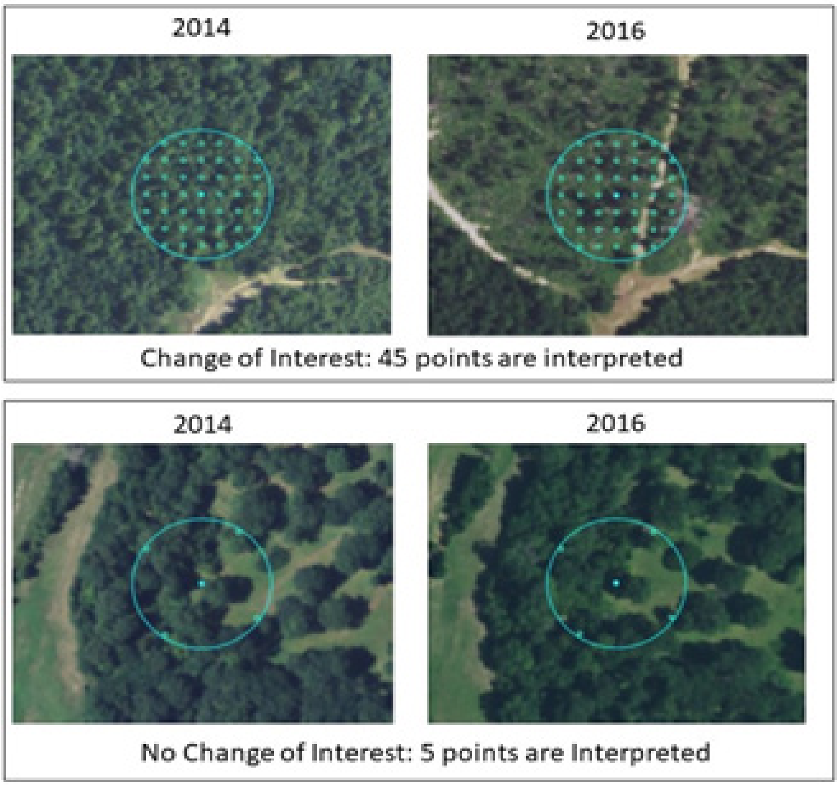

Figure 9.—Example ICE (Image-Based Change Estimation) plots on images acquired from the National Agriculture Imagery Program (NAIP). Forty-five dots are observed on plots where some change of interest has occurred, while only five dots are observed where no change of interest has occurred between the two NAIP image dates. Created by Tracey Frescino, USDA Forest Service.

The first application of photo-based estimation in FIA was the Nevada Photo-based Inventory Pilot (NPIP). An outgrowth of NPIP was the Image-Based Change Estimation (ICE) pilot program. ICE uses imagery from the National Agriculture Imagery Program (NAIP) to assess changes in land use and land cover and agent of change between two different years of NAIP. The sample of individual locations within a region for both NPIP and ICE are the FIA plot locations in the region (or possibly a subset of the FIA plot locations). At each photo-plot location, a sample of points is established within the photo plot. For details of the sample procedures used in NPIP, see Frescino et al. (2009). For details of the sample procedure for ICE, see USDA Forest Service (2017a). The derivation of the statistical estimators for NPIP and ICE is based on the Cordy (1993) estimators for continuous populations and Stevens and Urquhart (2000) results for support regions (Patterson 2012).

For an example of estimates produced for NPIP, see Frescino et al. (2016). Recently, the USDA Forest Service Geospatial Technology and Applications Center and IW-FIA developed summary reports that provide snapshots of basic analysis from the ICE data; see USDA Forest Service (2017b).

A recent pilot study in North Central Georgia (Moisen et al. 2020) compared FIA field and photo measures to data collected by using both the ICE protocol as well as historical Landsat-based observations collected through the image interpretation tool, TimeSync (Cohen et al. 2010), to evaluate how these three data sources could be used to best estimate land use and land cover (LULC) change. The study revealed that in order to report LULC trends in North Central Georgia with adequate precision and temporal coherence, data were needed on all the FIA plots each year over a long time series and broadly collapsed LULC classes. Discussions continue in the FIA program on the role of image interpretation for LULC change analyses. Further work is also underway to incorporate model-assisted methods and nonresponse (e.g., clouds and uninterpretable points) in image-based estimators, and substantiate theoretical results through simulation.

Photo-based estimators have been incorporated in the R package FIESTA (described in the Computing Resources section).

Examples of the use of photo-based (and other image-based) estimators in forest inventory applications can be found in the dynamic content, here: https://www.fia.fs.usda.gov/library/sampling/index.php.

Temporally Specific Estimation

Patterson and Reams (2005) discuss various methods to combine panels in order to construct estimates. They do not prescribe a core methodology but rather provide three alternatives: the moving average method, the temporally indifferent method, and modeling. To date, most of the estimates produced by the FIA program were produced with the temporally indifferent (TI) method. The TI method pools all panels into essentially one large periodic inventory measured over all the years in the periodic cycle. The post-stratified estimator is then used on the pooled inventory. This method is indifferent to time in that the cycle length depends on whether the population is in the eastern United States (where cycle length is 5 or 7 years) or in the western United States (where cycle length is 10 years) and all observations spanning those years are combined without considering how current those panels are relative to the date of interest. The TI method does not produce an estimate for any specific year and emerging trends are lagged when estimates are produced with the TI method (Van Deusen 2002).

In many cases there is a need to increase the temporal precision of estimates to be specific for a given year, and Johnson et al. (2003), Smith and Conkling (2004), and Van Deusen (1999) provide examples. The most obvious approach to constructing an estimate for a specific year is to use a single panel of data collected in the year of interest. However, these estimates can have large year to year variability given the small sample size. A weighted moving average may be used to improve the temporal specificity, where more weight is given to more recent data and less weight to the older data. The assignment of weights can be done by using a number of functions, including a linear or exponential, and options for weighted moving averages are currently available in the rFIA package described in the Computing Resources section.

The mixed estimator (Van Deusen 1999) offers another alternative. The mixed estimator requires two models: an observation model and a transition model. The time series of annual data arising from the panel design is used to construct the simple linear observation model, which defines the relationship between the forest parameter of interest and time. The transition model describes how the parameter in the observation model changes over time. Van Deusen (1999) constrained the parameters in this set of equations with first, second, and third derivative constraints, which yield a flat trend, a linear increasing or decreasing trend, or a quadratic trend respectively. Annual estimates and their variances are simultaneously estimated via maximum likelihood. Van Deusen's (1999) approach generally smooths individual panel estimates and is informed by the patterns revealed in the sample data.

Examples of constructing temporally specific estimates in forest inventory applications can be found in the dynamic content, here: https://www.fia.fs.usda.gov/library/sampling/index.php.

Application of Alternative Estimators for the FIA Inventory of Interior Alaska

The FIA program is mandated by Congress to assess current status and trends for all forest lands of the United States. However, implementing an inventory with the standard sampling intensity is cost-prohibitive in more remote regions of the nation, such as interior Alaska (estimated to have 110 million acres of forest), due to challenging logistics and high transportation costs (the lack of roads requires that virtually every plot is accessed via helicopter). For this reason, in interior Alaska, FIA has implemented a modified sampling design that uses a reduced sampling intensity for field plots (1 plot per 30,000 acres), supplemented with high-resolution airborne imagery (lidar, hyperspectral, thermal) acquired through a research collaboration with scientists from NASA-Goddard Space Flight Center in a strip sampling mode (Cook et al. 2013, Pattison et al. 2018). Although the field plots are established on a regular hexagonal grid and therefore standardized FIA estimation approaches (Bechtold and Patterson 2005) can be used for estimates over large areas, sample sizes within smaller management units (national parks, wildlife refuges, state forests, etc.) will be relatively small, leading to estimates with large standard errors, especially for smaller domains such as forest type. In addition, the periodic nature of the inventory (plot remeasurement interval likely greater than 10–12 years) has increased the interest in techniques that can provide more frequent information on forest change over this vast region.

For these reasons, several alternative approaches to estimation—with the aim of improving the timeliness and reliability of the FIA estimates through the use of sampled, high-resolution remote sensing measurements—are currently being developed and evaluated to support the interior Alaska FIA inventory, including two-stage, model-assisted (Ringvall et al. 2016, Cahoon and Baer 2022), hybrid estimation (Ene et al. 2018), and Bayesian hierarchical model-based approaches (Babcock et al. 2018). Recent advances in the use of Gaussian Nearest Neighbor processes have significantly increased the computational efficiency, and feasibility, of implementing Bayesian hierarchical modeling approaches over large regions (Finley et al. 2019). In addition, recent studies have indicated that lidar sampling may have significant value as a monitoring tool in interior Alaska, especially in capturing forest change due to wildfire (Alonzo et al. 2017).

Computing Resources

FIESTA

Forest Inventory ESTimation for Analysis (FIESTA) (Frescino et al. 2015, Frescino et al. 2020) is an R package that was originally developed to support the production of estimates consistent with tools available from the FIA National Program, such as Forest Inventory Data Online (FIDO) and EVALIDator. FIESTA provides an alternative data retrieval and reporting tool that is functional within the R environment, allowing customized applications and compatibility with other R-based analyses. Over the last few years, the tool has expanded to include new modules that accommodate many of the topics covered in this chapter on alternative estimators. FIESTA is available for download with vignettes for model-assisted estimation, photo-based estimation, small area estimation, and nonresponse. FIESTA was recently used to produce a large collection of model-assisted and model-based estimates of key forest attributes for counties, ecosubsections and watersheds across the conterminous, which were then delivered through user-friendly dashboards (Frescino et al 2022). The source code for the back-end estimation done in FIESTA is publicly available via the FIESTAutils R package on the CRAN (Frescino et al 2022). The FIESTA package will be distributed in its entirety on GitHub (https://github.com/USDAForestService/FIESTA) and CRAN (https://cran.r-project.org, a stand-alone desktop application of FIESTA is currently rolling out, and FIESTA is also being integrated into ArcGIS Pro.

Links to FIESTA documentation and distributed products can be found in the dynamic content here: https://www.fia.fs.usda.gov/library/sampling/index.php.

rFIA

A recently developed, publicly available R package, rFIA (https://doi.org/10.1016/j.envsoft.2020.104664, https://rfia.netlify.com/), is gaining popularity among FIA users for estimation, plot-level summaries, and visualization. The package can be downloaded by anyone from CRAN or Github, and its use is not restricted to Forest Service employees and partners. In addition to the temporally indifferent method, rFIA can produce estimates for individual annual panels and implements multiple forms of a moving average estimator, providing flexibility in the temporal specificity of estimates. In Stanke et al. (2022), rFIA was recently used to develop spatial Fay-Herriot models of forest carbon stocks for counties across the conterminous United States.

rFIA-related material can be found in the dynamic content, here: https://www.fia.fs.usda.gov/library/sampling/index.php.

GEE

Google Earth Engine's (GEE) cloud-based archive and programming interface eliminate image acquisition and processing tasks that previously took up most of the budget for large land cover mapping projects. At a time when FIA is being asked to do more with less, GEE represents an opportunity to achieve estimation goals for LULC change and other issues in a budget-conscious way. Best-practice protocols are currently being developed for FIA to ensure safe use of coordinate data with GEE. This protocol development work will provide the foundation of requirements written into the security plans of both internal and external users of FIA data in cloud-based platforms like GEE.

A link to GEE examples in forest inventory applications can be found in the dynamic content, here: https://www.fia.fs.usda.gov/library/sampling/index.php.

BIGMAP

The Big Data Mapping and Analytics Platform project (BIGMAP), a partnership between Forest Service and Esri, is leveraging cloud computing to accelerate processing of large geospatial datasets to facilitate the development, validation, and distribution of new data products and estimation techniques. Among other applications, Forest Service researchers are implementing nearest neighbor imputation methods for mapping forest carbon pools based on FIA plot data and time series of Landsat imagery (as defined by Wilson et al. 2013 and Wilson et al. 2018). These methods include model-based inference for small area estimation (McRoberts et al. 2007). Computational challenges associated with model-based inference arise from both (1) the production of forest attribute maps, and the associated manipulation of massive spatio-temporal datasets (e.g., 30-m Landsat satellite imagery), and (2) spatially explicit calculations needed for estimating variances, such as the calculation of covariances between pairs of pixels. While such challenges have limited the applicability of some methods to regional scales, the Amazon Web Services cloud computing environment leveraged by this project facilitates national-scale efforts.

A link to BIGMAP-related materials can be found in the dynamic content, here: https://www.fia.fs.usda.gov/library/sampling/index.php.

Summary

FIA's response to the Farm Bill's call for improved LULC change estimates, full implementation of the FIA inventory in interior Alaska, as well as improved precision over small geographic areas, have all been on an accelerated research schedule. This chapter introduced several of the rapidly evolving topics. Whether a model-assisted, model-based, or hybrid approach should be taken depends on the research question and available data. Ståhl et al. (2016) provide a thorough comparison of these different frameworks and their applicability to different inventory scenarios. Research updates will continually be posted to FIA's website to track the organizations progress on these important fronts.

Literature Cited

Alonzo, M.; Morton, D.C.; Cook, B.D.; [et al.]. 2017. **Patterns of canopy and surface layer consumption in a boreal forest fire from repeat airborne lidar.** Environmental Research Letters. 12(6). http://dx.doi.org/10.1088/1748-9326/aa6ade.

Andersen, H.-E.; Barrett, T.; Winterberger, K.; [et al.]. 2009. **Estimating forest biomass on the Western Lowlands of the Kenai Peninsula of Alaska using airborne lidar and field plot data in a model-assisted sampling design.** In: Proceedings of the IUFRO Division 4 Conference: Extending forest inventory and monitoring over space and time: 19–22.

Andersen, H.-E.; Reutebuch, S.E.; McGaughey, R.J.; [et al.]. 2013. **Monitoring selective logging in western Amazonia with repeat lidar flights.** Remote Sensing of Environment. 151: 157–165. https://doi.org/10.1016/j.rse.2013.08.049.

Babcock, C.; Finley, A.; Andersen, H.-E.; [et al.]. 2018. **Geostatistical estimation of forest biomass in interior Alaska combining Landsat-derived tree cover, sampled airborne lidar and field observations.** Remote Sensing of Environment. 212: 212–230. https://doi.org/10.1016/j.rse.2018.04.044.

Baffetta, F.; Fattorini, L.; Franceschi, S.; [et al.]. 2009. **Design-based approach to k-nearest neighbours technique for coupling field and remotely sensed data in forest surveys.** Remote Sensing of Environment. 113(3): 463–475. https://doi.org/10.1016/j.rse.2008.06.014.

Battese, G.E.; Harter, R.M.; Fuller, W.A. 1988. **An error-components model for prediction of county crop areas using survey and satellite data.** Journal of the American Statistical Association. 83(401): 28–36. https://doi.org/10.2307/2288915.

Bechtold, W.A.; Patterson, P.L., 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. https://doi.org/10.2737/SRS-GTR-80.

Bell, D.M.; Wilson B.T.; Werstak, C.E.; [et al.]. 2022. **Examining k-nearest neighbor small area estimation across scales using national forest inventory data.** Frontiers in Forests and Global Change. 5: 763422. https://doi.org/10.3389/ffgc.2022.763422.

Breidenbach, J. 2018. **JoSAE: Unit-level and area-level small area estimation.** R package version 0.3.0. https://CRAN.R-project.org/package=JoSAE (accessed May 2019).

Breidenbach, J.; Anton-Fernandez, C.; Petersson, H.; [et al.]. 2014. **Quantifying the model-related variability of biomass stock and change estimates in the Norwegian National Forest Inventory.** Forest Science. 60(1): 25–33. https://doi.org/10.5849/forsci.12-137.

Breidenbach, J.; Astrup, R. 2012. **Small area estimation of forest attributes in the Norwegian National Forest Inventory.** European Journal of Forest Research. 131(4): 1255–1267. http://dx.doi.org/10.1007/s10342-012-0596-7.

Breidenbach, J.; McRoberts, R.E.; Astrup, R. 2016. **Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume.** Remote Sensing of Environment. 173: 274–281. https://doi.org/10.1016/j.rse.2015.07.026.

Breidt, F.J.; Opsomer, J.D.; Johnson, A.A.; [et al.]. 2007. **Semiparametric model-assisted estimation for natural resource surveys.** Survey Methodology. 33(1): 35–44. https://www150.statcan.gc.ca/n1/pub/12-001-x/2007001/article/9850-eng.pdf (accessed December 6, 2010).

Cahoon, S.; Baer, K. 2022. **Forest resources of the Tanana unit, Alaska: 2018.** Gen. Tech. Rep. PNW-1005. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 92 p. https://doi.org/10.2737/PNW-GTR-1005.

Cao, Q.; Dettmann, G.T.; Radtke, P.J.; [et al.]. 2022. **Increased precision in county-level volume estimates in the U.S. National Forest Inventory with area-level SAE.** Frontiers in Forests and Global Change. https://doi.org/10.3389/ffgc.2022.769917.

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

Cohen, W.B.; Yang, Z.Q.; Kennedy, R. 2010. **Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync—tools for calibration and validation.** Remote Sensing of Environment. 114: 2911–2924. https://doi.org/10.1016/j.rse.2010.07.010.

Cook, B.D., Corp, L.W.; Nelson, R.F.; [et al.]. 2013. **NASA Goddard's Lidar, Hyperspectral and Thermal (G-LiHT) airborne imager.** Remote Sensing. 5: 4045–4066. http://dx.doi.org/10.3390/rs5084045.

Cordy, C.B. 1993. **An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe.** Statistics & Probability Letters. 18: 353–362. https://doi.org/10.1016/0167-7152(93)90028-H.

Corona, P.; Fattorini, L.; Franceschi, S.; [et al.]. 2014. **Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives.** Canadian Journal of Forest Research. 44: 1303–1311. http://dx.doi.org/10.1139/cjfr-2014-0203.

Dettmann, G.T.; Radtke, P.J.; Coulston, J.W.; [et al.]. 2022. **Review and synthesis of estimation strategies to meet small area needs in forest inventory.** Frontiers in Forests and Global Change. 5: 813569. https://doi.org/10.3389/ffgc.2022.813569.

Efron, B.; Tibshirani, R. 1994. **An introduction to the bootstrap.** Boca Raton, FL: Chapman and Hall/CRC. 456 p. https://doi.org/10.1201/9780429246593.

Ene, L.T.; Gobakken, T.; Andersen, H.-E.; [et al.]. 2018. **Large-area hybrid estimation of aboveground biomass in interior Alaska using airborne laser scanning data.** Remote Sensing of Environment. 204: 741–755. http://dx.doi.org/10.1016/j.rse.2017.09.027.

Fattorini, L. 2012. **Design-based or model-based inference? The role of hybrid approaches in environmental surveys.** In: Fattorini, L., ed. Studies in honor of Claudio Scala. Siena, Italy: Department of Economics and Statistics, University of Siena: 173–214.

Finley, A.O.; Datta, A.; Cook, B.C.; [et al.]. 2019. **Efficient algorithms for Bayesian nearest neighbor Gaussian processes.** Journal of Computational and Graphical Statistics. 28: 401–414. https://doi.org/10.1080/10618600.2018.1537924.

Frescino, T.S.; McConville, K.S.; White, G.W.; [et al.]. 2022. **Small area estimates for national applications: A database to dashboard strategy for FIA using FIESTA.** Frontiers in Forests and Global Change. 5: 779446. https://doi.org/10.3389/ffgc.2022.779446.

Frescino, T.S.; Moisen, G.G.; Megown, K.A.; [et al.]. 2009. **Nevada photo-based inventory pilot (NPIP) photo sampling procedures.** Gen. Tech. Rep. RMRS-GTR-222. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 30 p. https://doi.org/10.2737/RMRS-GTR-222.

Frescino, T.S.; Moisen, G.G.; Patterson, P.L.; [et al.]. 2016. **Nevada photo-based inventory pilot (NPIP) resource estimates (2004–2005).** Gen. Tech. Rep. RMRS-GTR-344. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 59 p. https://doi.org/10.2737/RMRS-GTR-344.

Frescino, T.S.; Moisen, G.G.; Patterson, P.A.; [et al.]. 2020. **Demonstrating a progressive FIA through FIESTA: a bridge between science and production.** In: Brandeis, T.J., comp. 2020. Celebrating progress, possibilities, and partnerships: Proceedings of the 2019 Forest Inventory and Analysis (FIA) science stakeholder meeting. E-Gen. Tech. Rep. SRS-256. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station: 199–200.

Frescino, T.S.; Patterson, P.L.; Moisen, G.G.; [et al.]. 2015. **FIESTA—an R estimation tool for FIA analysts.** In: Stanton, S.M.; Christensen, G.A., comps. FIA symposium 2015. Gen. Tech. Rep. PNW-GTR-931. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. P. 72. https://www.fs.usda.gov/pnw/pubs/pnw_gtr931/pnw_gtr931_023.pdf (accessed August 10, 2022).

Frescino, T.S.; Toney C.; White, G.W. 2022. **FIESTAutils: Utility Functions for Forest Inventory Estimation and Analysis.** R package version 1.0.0. https://CRAN.R-project.org/package=FIESTAutils.

Gaines, G.C.; Affleck, D.L.R. 2021. S**mall area estimation of postfire tree density using continuous forest inventory data.** Frontiers in Forests and Global Change. 4: 761509. https://doi.org/10.3389/ffgc.2021.761509.

Gobakken, T.; Næsset, E.; Nelson, R.; [et al.]. 2012. **Estimating biomass in Hedmark County, Norway, using national forest inventory field plots and airborne laser scanning.** Remote Sensing of Environment. 123: 443–456. https://doi.org/10.1016/j.rse.2012.01.025.

Goerndt, M.; Monleon, V.; Hailemariam, T. 2013. **Small-area estimation of county-level forest attributes using ground data and remote sensed auxiliary information.** Forest Science. 59: 536–548. https://doi.org/10.5849/forsci.12-073.

Gregoire, T.G.; Ståhl, G.; Næsset, E.; [et al.]. 2011. **Model-assisted estimation of biomass in a lidar sample survey in Hedmark County, Norway.** Canadian Journal of Forest Research. 41(1): 83–95. https://doi.org/10.1139/X10-195.

Guldin, R.W. 2021. **A systematic review of small domain estimation research in forestry during the twenty-first century from outside the United States.** Frontiers in Forests and Global Change. 4: 695929. https://doi.org/10.3389/ffgc.2021.695929.

Harris, V.; Caputo, J.; Finley, A.; [et al.]. 2021. **Small-area estimation for the USDA Forest Service, National Woodland Owner Survey: Creating a fine-scale land cover and ownership layer to support county-level population estimates.** Frontiers in Forests and Global Change. 4: 745840. https://doi.org/10.3389/ffgc.2021.745840.

Henderson, C R. 1975. **Best linear unbiased estimation and prediction under a selection model.** Biometrics. 31(2): 423–447. https://doi.org/10.2307/2529430.

Johnson, D.S.; Williams, M.S.; Czaplewski, R.L. 2003. **Comparison of estimator for rolling samples using forest inventory and analysis data.** Forest Science. 49(1): 50–63.

Kangas, A.; Myllymäki, M.; Gobakken, T.; [et al.]. 2016. **Model-assisted forest inventory with parametric, semiparametric, and nonparametric models.** Canadian Journal of Forest Research. 46(6): 855–868. http://dx.doi.org/10.1139/cjfr-2015-0504.

Mauro, F.; Monleon, V.J.; Temesgen, H.; [et al.]. 2017. **Analysis of area level and unit level models for small area estimation in forest inventories assisted with lidar auxiliary information.** PloS ONE. 12(12): 1–14. https://doi.org/10.1371/journal.pone.0189401.

McConville, K.; Moisen, G.G.; Frescino, T.S. 2020. **A tutorial in model-assisted estimation with application to forest inventory.** Forests. 11: 244. https://doi.org/10.3390/f11020244.

McConville, K.S.; Breidt, F.J.; Lee, T.C.M.; [et al.]. 2017. **Model-assisted survey regression estimation with the lasso.** Journal of Survey Statistics and Methodology. 5: 131–158. https://doi.org/10.1093/jssam/smw041.

McConville, K.S.; Tang, B.; Zhu, G.; [et al.]. 2018. **MASE: Model-assisted survey estimators.** https://cran.r-project.org/package=mase (accessed May 2019).

McRoberts, R.E. 2006. **A model-based approach to estimating forest area.** Remote Sensing of Environment. 103: 56–66. https://doi.org/10.1016/j.rse.2006.03.005.

McRoberts, R.E. 2010. **Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data.** Remote Sensing of Environment. 114(5): 1017–1025. https://doi.org/10.1016/j.rse.2009.12.013.

McRoberts, R.E.; Chen, Q.; Domke, G.M.; [et al.]. 2016a. **Hybrid estimators for mean aboveground carbon per unit area.** Forest Ecology and Management. 378: 44–56. https://doi.org/10.1016/j.foreco.2016.07.007.

McRoberts, R.E.; Chen, Q.; Walters, B.F. 2017. **Multivariate inference for forest inventories using auxiliary airborne laser scanning data.** Forest Ecology and Management. 401: 295–303. https://doi.org/10.1016/j.foreco.2017.07.017.

McRoberts, R.E.; Domke, G.M.; Chen, Q.; [et al.]. 2016b. **Using genetic algorithms to optimize k-nearest neighbors configurations for use with airborne laser scanning data.** Remote Sensing of Environment. 184: 387–395. https://doi.org/10.1016/j.rse.2016.07.007.

McRoberts, R.E.; Moser, P.; Zimermann Oliveira, L. 2015a. **A general method for assessing the effects of uncertainty in individual-tree volume model predictions on large-area volume estimates with a subtropical forest illustration.** Canadian Journal of Forest Research. 45: 44–51. https://doi.org/10.1139/cjfr-2014-0266.

McRoberts, R.E.; Næsset, E.; Gobakken, T. 2013. **Inference for lidar-assisted estimation of forest growing stock volume.** Remote Sensing of Environment. 128: 268–275. https://doi.org/10.1016/j.rse.2012.10.007.

McRoberts, R.E.; Næsset, E.; Gobakken, T. 2014. **Estimation for inaccessible and nonsampled forest areas using model-based inference and remotely sensed auxiliary information.** Remote Sensing of Environment. 154: 226–23. https://doi.org/10.1016/j.rse.2014.08.028.

McRoberts, R.E.; Næsset, E.; Gobakken, T. 2015b. **Optimizing the k-nearest neighbors technique for estimating forest aboveground biomass using airborne laser scanning data.** Remote Sensing of Environment. 163: 13–22. https://doi.org/10.1016/j.rse.2015.02.026.

McRoberts, R.E.; Tomppo, E.O.; Finley, A.O.; [et al.]. 2007. **Estimating areal means and variances of forest attributes using the k-nearest neighbor technique and satellite imagery.** Remote Sensing of Environment. 111: 466–480. https://doi.org/10.1016/j.rse.2007.04.002.

McRoberts, R.E.; Vibrans, A.C.; Sannier, C.; [et al.]. 2016c. **Methods for evaluating the utilities of local and global maps for increasing the precision of estimates of subtropical forest area.** Canadian Journal of Forest Research. 46(7): 924–932. https://doi.org/10.1139/cjfr-2016-0064.

McRoberts, R.E.; Walters, B.F. 2012. **Statistical inference for remote sensing-based estimates of net deforestation.** Remote Sensing of Environment. 124: 394–401. https://doi.org/10.1016/j.rse.2012.05.011.

McRoberts, R.E.; Westfall, J.A. 2014. **Effects of uncertainty in model predictions of individual tree volume on large area volume estimates.** Forest Science. 60(1): 34–42. https://doi.org/10.5849/forsci.12-141.

Moisen, G.G.; McConville, K.S.; Schroeder, T.A.; [et al.]. 2020. **Estimating land use and land cover change in north central Georgia: Can remote sensing observations augment traditional forest inventory data?** Forests. 11(8): 856. https://doi.org/10.3390/f11080856.

Molina, I.; Marhuenda, Y. 2015. **Sae: An R package for small area estimation.** The R Journal. 7:81–98. http://dx.doi.org/10.32614/RJ-2015-007.

Moser, P.; Vibrans, A.C.; McRoberts, R.E.; [et al.]. 2017. **Methods for variable selection in LiDAR-assisted forest inventories.** Forestry. 90: 112–124. https://doi.org/10.1093/forestry/cpw041.

Opsomer, J.D.; Breidt, F.J.; Moisen, G.G.; [et al.]. 2007. **Model-assisted estimation of forest resources with generalized additive models (with discussion).** Journal of the American Statistical Association. 102: 400–416. https://doi.org/10.1198/016214506000001491.

Patterson, P.L. 2012. **Photo-based estimators for the Nevada photo-based inventory.** Res. Pap. RMRS-RP-92. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 14 p. https://doi.org/10.2737/RMRS-RP-92.

Patterson, P.L; Reams, G.A. 2005. **Combining panels for forest inventory and analysis estimation.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 69–74.

Pattison, R.; Andersen, H.-E.; Gray, A.; Schulz, B.; Smith, R.; Jovan, S., tech. coords. 2018. **Forests of the Tanana Valley State Forest and Tetlin National Wildlife Refuge Alaska: results of the 2014 pilot inventory.** Gen. Tech. Rep. PNW-GTR-967. Portland, OR: USDA Forest Service, Pacific Northwest Research Station. 80 p. https://doi.org/10.2737/PNW-GTR-967.

Prisley, S.; Bradley, J.; Clutter, M.; [et al.]. 2021. **Needs for small area estimation: Perspectives from the US private forest sector.** Frontiers in Forests and Global Change. 4: 746439. https://doi.org/10.3389/ffgc.2021.746439.

Rao, J.N.K.; Molina, I. 2015. **Small area estimation.** 2nd Edition. Hoboken, New Jersey: John Wiley & Sons, Inc. 441 p. https://doi.org/10.1002/9781118735855.

Ringvall, A.; Ståhl,, G.; Ene, L; [et al.]. 2016. **A post-stratified ratio estimator for model-assisted biomass estimation in sample-based airborne laser scanning surveys.** Canadian Journal of Forest Research. 46: 1386–1395. https://doi.org/10.1139/cjfr-2016-0158.

Saarela, S.; Grafström, A.; Ståhl, G.; [et al.]. 2015. **Model-assisted estimation of growing stock volume using different combinations of lidar and landsat data as auxiliary information.** Remote Sensing of Environment. 158: 431–440. https://doi.org/10.1016/j.rse.2014.11.020.

Särdnal, C.E.; Swensson, B.; Wretman, J.H. 1992. **Model assisted survey sampling.** New York: Springer-Verlag. 694 p.

Smith, W.D.; Conkling, B.L. 2004. **Analyzing forest health data.** Gen. Tech. Rep SRS-077. Asheville, NC: USDA Forest Service, Southern Research Station. 33 p. https://doi.org/10.2737/SRS-GTR-77.

Ståhl, G.; Heikkinen, J.; Petersson, H.; [et al.]. 2014. **Sample-based estimation of greenhouse gas emissions from forests—a new approach to account for both sampling and model errors.** Forest Science. 60: 3–13. http://dx.doi.org/10.5849/forsci.13-005.

Ståhl, G.; Saarela, S.; Schnell, S.; [et al.]. 2016. **Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation.** Forest Ecosystems. 3: 5. https://doi.org/10.1186/s40663-016-0064-9.

Stanke, H.; Finley, A.O.; Domke, G.M. 2022. **Simplifying small area estimation with rFIA: a demonstration of tools and techniques.** Frontiers in Forests and Global Change. 5: 745874. https://doi.org/10.3389/ffgc.2022.745874.

Stevens, D.L.; Urquhart, S. 2000. **Response designs and support regions in sampling continuous domains.** Environmetrics. 11: 13–41. https://doi.org/10.1002/(SICI)1099-095X(200001/02)11:1%3C13::AID-ENV379%3E3.0.CO;2-8.

Temesgen, H.; Mauro, F.; Hudak, A.T.; [et al.]. 2021. **Using Fay–Herriot models and variable radius plot data to develop a stand-level inventory and update a prior inventory in the Western Cascades, OR, United States.** Frontiers in Forests and Global Change. 4: 745916. https://doi.org/10.3389/ffgc.2021.745916.

USDA Forest Service. 2017a. **Image-based Change Estimation (ICE) protocol guide.** Version 2. Unpublished guide on file at: Geospatial Technology and Applications Center, Salt Lake City, UT. 19 p.

USDA Forest Service. 2017b. **Image-based Change Estimation (ICE) summary for New Hampshire, 2012 to 2014.** Version 1. Unpublished report on file at: Geospatial Technology and Applications Center, Salt Lake City, UT. 4 p.

Valliant, R. 2009. **Model-based prediction of finite population totals.** In: Rao, C.R., ed. Handbook of statistics: sample surveys: inference and analysis. Volume 29, Part B: 11–31. https://doi.org/10.1016/S0169-7161(09)00223-5.

Van Deusen, P.C. 1999. **Modeling trends with annual survey data.** Canadian Journal of Forest Research. 29: 1824–1828. https://doi.org/10.1139/x99-142.

Van Deusen, P.C. 2002. **Comparison of some annual forest inventory estimators.** Canadian Journal of Forest Research. 32: 1992–1995. https://doi.org/10.1139/x02-115.

White, G.W.; McConville, K.S.; Moisen, G.G.; [et al.] 2021. **Hierarchical Bayesian small area estimation using weakly informative priors in ecologically homogeneous areas of the Interior Western forests.** Frontiers in Forests and Global Change. 4: 752911. https://doi.org/10.3389/ffgc.2021.752911.

Wiener S.W.; Bush, R.; Nathanson, A.; [et al.]. 2021. **United States Forest Service use of forest inventory data: Examples and needs for small area estimation.** Frontiers in Forests and Global Change. 4: 763487. https://doi.org/10.3389/ffgc.2021.763487.

Wilson, B.T.; Knight, J.F.; McRoberts, R.E. 2018. **Harmonic regression of Landsat time series for modeling attributes from national forest inventory data.** ISPRS Journal of Photogrammetry and Remote Sensing. 137: 29–46. https://doi.org/10.1016/j.isprsjprs.2018.01.006.

Wilson, B.T.; Woodall, C.W.; Griffith, D.M. 2013. **Imputing forest carbon stock estimates from inventory plots to a nationally continuous coverage.** Carbon Balance and Management. 8: 1. http://dx.doi.org/10.1186/1750-0680-8-1.

Wojcik, O.C.; Olson, S.D.; Nguyen, P.V.; [et al.]. 2022. **GREGORY: A modified generalized regression estimator approach to estimating forest attributes in the Interior Western US.** Frontiers in Forests and Global Change. 4: 763414. https://doi.org/10.3389/ffgc.2021.763414.

Appendix 1: Forest Area Estimation and Area Control Revisited

John W. Coulston and James A. Westfall

Forest area is a key population parameter estimated by the FIA program. Both sampling and estimation methodologies have changed over time and there remains some misperception around how the methodologies presented by Scott et al. (2005) differ from previous approaches for estimating forest area. Frayer and Furnival (1999) and USDA Forest Service (1992) provide an overview of historical statistical designs. Frayer and Furnival (1999), as a general statement, suggested that most states were inventoried with a double sampling design where aerial photo plots were used for stratification and field plots were used for collecting basic forest mensuration variables. This practice led to the idea of a “Phase 1” forest area estimate that arises from a stratification process based on a large number of photo plots. Reams (2000) provides an example of “Phase 1” forest area estimation. With the shift to the rotating panel design and the adoption of the post-stratified estimator, the concept of a “Phase 1” area estimate has been eliminated. The goal of this Appendix is to revisit area estimation procedures from Scott et al. (2005) with an emphasis on forest area estimation and area control. The practitioner is the target audience of this Appendix and as such we take guidance from Freese (1962), provide example calculations for forest area estimates, and re-present equations that are also given in the Foundational Documentation chapter.

Simple Random Sampling Case

To increase understanding of area estimates, area estimation under simple random sampling (SRS) is initially presented and then extended to constructing estimates with the post-stratified estimator. This approach was taken because when only one stratum is used with the post-stratified estimator, an SRS estimate is obtained. Further, the rotating panel design is assumed to produce an equal probability sample (McRoberts et al. 2006) and estimators for equal probability samples generally reduce to the same form as SRS estimators. The general steps for constructing an SRS estimate for a population total that uses FIA data include the following:

1. Calculate the plot (i) area proportion for the domain (d) of interest (y_{id}).

2. Estimate the mean of the plot-level values $(\bar{Y}_d)$.
3. Estimate the variance of the mean $(v(\bar{Y}_d))$.
4. Expand the mean and the variance of the mean to a total and variance of the total.

In step 1, the domain of interest is identified. Here we focus on forest as the domain of interest, but the approach is valid for any area-based estimate. For simplicity, we assume all plots were sampled and adjustments for nonresponse are unnecessary. When forest is the domain of interest the measured forest proportion of each plot is summarized:

$$y_{id} = \sum_{k=1}^{K} y_{ik}\,\delta_{ikd} \tag{48}$$

Where

y_{ik} = the proportion of each plot *i* in condition *k*, and

δ_{ikd} = an indicator (0, 1) variable that takes the value of 1 when the condition is forest.

In simple terms, y_{id} is the sum of the forested condition proportions for each sample plot in the population, including entirely nonforest plots. In Scott et al. (2005), y_{id} has an additional summation to first sum the conditions for each *j* subplot (or macroplot) on plot *i* but the approaches yield equivalent results.

$$y_{id} = \sum_{k=1}^{K} y_{ik}\,\delta_{ikd} = \sum_{j=1}^{4}\sum_{k=1}^{K} y_{i_jk}\delta_{i_jkd}.$$

The mean proportion forest $\bar{Y}_d$ is estimated by

$$\bar{Y}_d = n^{-1}\sum_{i=1}^{n} y_{id} \tag{49}$$

Where

n = the total number of plots (forest and nonforest plots) in the population.

The variance of the mean proportion forest $(v(\bar{Y}_d))$ is estimated by

$$v(\bar{Y}_d) = \frac{\sum_{i=1}^{n}(y_{id}-\bar{Y}_d)^2}{n(n-1)}\ . \tag{50}$$

In the divisor of Equation 50, n is used to estimate the variance of the mean from the variance of the sample $\frac{\sum_{i=1}^{n}(y_{id}-\bar{Y}_d)^2}{n-1}$. In this manner the standard error of the estimate is then $v(\hat{Y}_d)^{0.5}$.

To expand the mean proportion forest and the variance of the mean to population totals, the area of the population (A_T) is needed. For estimation purposes, A_T is considered known and is typically taken from official U.S. Census Bureau statistics. Given a known A_T, the total forest area is

$$\hat{Y}_d = A_T \bar{Y}_d \tag{51}$$

With estimated variance

$$v(\hat{Y}_d) = A_T^2 v(\bar{Y}_d) \tag{52}$$

and the standard error of the estimate is $v(\hat{Y})^{0.5}$.

Example Forest Area Estimate Under SRS

Consider a hypothetical population made up of two counties with an inventory of 30 sampled plots (Table 7). County 1 is 87,420 acres, county 2 is 98,580 acres, and the population is A_T = 186,000 acres.

Based on Equation 49, the mean proportion forest is

$$\bar{Y}_d = 30^{-1}(1 + 1 + 0 + 0.75 + \cdots + 0) = 0.513333$$

Using Equation 50 the estimated variance of the mean proportion forest is

$$v(\bar{Y}_d) = \frac{(1-.513)^2 + (1-.513)^2 + (0-.513)^2 + (1-.513)^2 + \cdots + (0-.513)^2}{30(30-1)}$$

= 703.985 X 10^-5

The total estimated forest area is 186,000 acres × 0.513333 = 95,480 acres with the estimated variance being 186,0002 acres^2 × 703.985 × 10^{-5} = 243,550,651 acres^2. The standard error of the estimate is (243,550,651 acres^2)$^{0.5}$ = 15,606 acres.

Table 7.—Dataset used for example forest area estimates using simple random sampling estimation and post-stratified estimation. Note that the condition data have already been summarized to plot-level proportion forest using Equation 48.

County	Plot	Forest proportion
1	1	1
1	2	1
1	3	0
1	4	0.75
1	5	0.25
1	6	0.3
1	7	1
1	8	0
1	9	0
1	10	0
1	11	0.1
1	12	0.8
1	13	1
2	14	1
2	15	0
2	16	0
2	17	1
2	18	1
2	19	1
2	20	0
2	21	1
2	22	0
2	23	0.25
2	24	1
2	25	0
2	26	0.75
2	27	1
2	28	1
2	29	0.2
2	30	0

Post-stratified Estimation

The post-stratified estimators for area parameters are presented in Scott et al. (2005) and are presented earlier in the Foundational Documentation chapter equations 1–6. These estimators are relevant to single panel estimates or estimates based on a full set of panels assuming temporal indifference (Patterson and Reams 2005). There are two components of stratified estimation as presented in Scott et al. (2005) that are not covered under the SRS example above. These components are the stratification process and the adjustment for partially nonsampled plots.

Stratification Process

The goal of stratification is to form homogeneous strata by using auxiliary information to reduce the variance of the estimate. Generally, the precision of forest area estimates is increased 1.4- to 3.0-fold by using post-stratified estimation (Brooks et al. 2016, Coulston 2008, McRoberts et al. 2006). Westfall et al. (2011) recommended that strata be constructed such that there are ≥ 10 plots within each stratum. As of the writing of this report, FIA typically constructs a stratification map based on one or more geospatial datasets. Often, one component of the stratification is based on a map of forest cover (Homer et al. 2015) or tree cover (Coulston et al. 2012) derived from the National Land Cover Database (NLCD). Another common component is an ownership map, such as the Protected Areas Database (DellaSala et al. 2001) where public and private lands are delineated.

When wall-to-wall maps are used for post-stratification, the weight of each stratum is assumed to be known and each inventory plot is assigned to a single stratum. For a simple example, consider a forest cover / nonforest cover raster map that completely covers a population. Each inventory plot is assigned to one stratum h (forest cover stratum or nonforest cover stratum). The weight of each stratum (W_h) is calculated directly from the stratification map. It follows that the weight of the forest cover stratum is the number of forest cover grid cells in the raster map (pixels) in the population divided by the total number of pixels in the population. The weight of the nonforest cover stratum is calculated similarly. Hence, $\sum_h^H W_h = 1$ for the population.

It is important to note that in practice the strata are not completely homogeneous. This occurs for several reasons:

- Classification differences—The FIA program defines forest based on use, which differs from defining forest based on cover (Coulston et al. 2014, Nelson and Reams 2017). For example, under a forest use definition, areas that have been harvested and replanted remain in forest land use whereas these same areas do not meet the definition of forest cover (e.g., trees > 5 m tall and occupying > 20 percent of vegetative cover).
- Error associated with the wall-to-wall maps—For example, Wickham et al. (2017) found that the accuracy of the 2011 NLCD Land Cover map was 88 percent across broad land cover classes. Likewise, ownership boundary layers contain geospatial error arising from incorrect property boundaries.
- Error in plot locations—Each plot is assigned to a single stratum based on the location of the plot center. When there are errors in the plot locations, the potential for assigning plots to an incorrect stratum increases. McRoberts (2010) suggests that the general plot coordinate accuracy is within the range of 8 to 20 m.
- Multiple condition plots—Plots may only be assigned to one stratum, yet multiple conditions can be recorded on a plot. For example, a plot may have both forest and nonforest conditions but only be assigned to a forest or nonforest stratum.
- Temporal discontinuity—The FIA program collects data annually, yet the maps used for stratification are only updated on a periodic basis. For example, the most current NLCD Land Cover maps in 2020 are from the 2016 NLCD. This leads to situations where the plot data reflect more current ground conditions than the map used for stratification.

It is common to have forest land use plots in a nonforest stratum and vice versa. This does not bias results but rather increases the variance of the estimate.

Partially Sampled Plots

There are certain situations where field inventory crews cannot access a sample location (in whole or in part). This can occur because of denied access by a landowner, a hazardous situation, or the location falls outside the population of interest (Scott et al. 2005). These nonsampled conditions are addressed in two ways. Plots that are completely nonsampled are not used when estimating forest area. They are assumed to be missing at random within the stratum (Patterson et al. 2012). For plots that are partially nonsampled, an adjustment is calculated for each stratum within the population. In practical terms, the adjustment is the average proportion of the plots within a stratum that were sampled and is discussed further in the following section.

Estimators

Equation 4.1 from Scott et al. (2005) extends Equation 48 to identify the stratum to which plot *i* belongs and the adjustment $(\bar{p}_h)$ for partial plots outside of the population within the stratum:

$$Y_{hid} = \frac{\sum_{k=1}^{K} y_{hik}\delta_{hikd}}{\bar{p}_h} \tag{53}$$

Where

$$\bar{p}_h = \frac{\sum_{i=1}^{n_h}\sum_{k=1}^{K} y_{hik}\delta_{hikd}}{n_h} \, .$$

When calculating $\bar{p}_h$ it is important to note that the domain indicator (δ) changes. In calculating the adjustment for partial plots, δ takes on a value of 1 if the k^{th} condition was in the population and sampled and zero otherwise. For example, consider a plot that has three conditions. Condition 1 is forest land use with y_{hik} = 0.5, condition 2 is nonforest land use with y_{hik}= 0.25, and condition 3 is nonsampled with y_{hik} = 0.25. When calculating $\bar{p}_h$, δ_{hikd} = 1 for both condition 1 (forest land use) and condition 2 (nonforest land use) irrespective of the domain of interest because those are sampled conditions on the plot. Likewise, for any *d*, condition 3 is nonsampled and hence δ_{hikd} = 0. Overall, this example plot has a sampled proportion of 0.75. The division by $\bar{p}_h$ in Equation 53 is a method to account for partially sampled plots and this approach is used so that plot area is not a random variable.

Equations 4.3 and 4.4 in Scott et al. (2005) are extensions of Equation 49 and Equation 50 and are used to estimate the plot-level mean and the variance of the plot-level mean within strata, respectively:

$$\bar{Y}_{hd} = n_h^{-1} \sum_{i=1}^{n_h} y_{hid} \tag{54}$$

$$v(\bar{Y}_{hd}) = \frac{\sum_{i=1}^{n_h} (y_{hid} - \bar{Y}_{hd})^2}{n_h(n_h - 1)} \tag{55}$$

In effect, the SRS estimators are just used within each stratum. The estimate of the population mean for the domain of interest is then a weighted average based on the strata weights (W_h):

$$\bar{Y}_d = \sum_h^H W_h \bar{Y}_{hd} \tag{56}$$

With estimated variance

$$v(\bar{Y}_d) = n^{-1} \left[\sum_h^H W_h n_h v(\bar{Y}_{hd}) + \sum_h^H (1 - W_h) \frac{n_h}{n} v(\bar{Y}_{hd}) \right] \tag{57}$$

Equation 51 is then used to estimate the total forest area from the mean forest area (Eq. 56), and Equation 52 is used to estimate the variance of the total forest area from Equation 57. Equations 54–57 are also presented in the Foundational Documentation chapter.

Example Forest Area Estimate Under Poststratification

To provide an example of forest area estimation under post-stratification, the same hypothetical population described for SRS estimation is used. Each plot from Table 7 is assigned to a single stratum (Table 8). The stratification is a simple forest / nonforest approach where the stratum weights and sample sizes are $W_{h=Forest}$ = 0.57, $W_{h=Nonforest}$ = 0.43, $n_{h=Forest}$ = 18, and $n_{h=Nonforest}$ = 12.

Table 8.—Stratification assignment for each plot in the example dataset from table 7.

County	Plot	Forest proportion	Stratum
1	1	1	Forest
1	2	1	Forest
1	3	0	Non-forest
1	4	0.75	Forest
1	5	0.25	Non-forest
1	6	0.3	Forest
1	7	1	Non-forest
1	8	0	Forest
1	9	0	Non-forest
1	10	0	Non-forest
1	11	0.1	Non-forest
1	12	0.8	Forest
1	13	1	Forest
2	14	1	Forest
2	15	0	Forest
2	16	0	Non-forest
2	17	1	Forest
2	18	1	Forest
2	19	1	Forest
2	20	0	Non-forest
2	21	1	Forest
2	22	0	Forest
2	23	0.25	Non-forest
2	24	1	Forest
2	25	0	Non-forest
2	26	0.75	Forest
2	27	1	Forest
2	28	1	Forest
2	29	0.2	Non-forest
2	30	0	Non-forest

The first step is to estimate the mean proportion forest for each stratum based on Equation 54:

$$\bar{Y}_{h=Forest,d} = 18^{-1}(1 + 1 + 0.75 + 0.3 + \cdots + 1) = 0.755555$$

$$\bar{Y}_{h=Non-forest,d} = 12^{-1}(0 + 0.25 + 1 + 0 + \cdots + 0) = 0.15$$

At this point note that both strata contain forested plots and contribute to the overall forest area estimate. The estimated variance of the mean proportion forest for each stratum based on Equation 55 is then

$$v(\bar{Y}_{h=Forest,d}) = \frac{(1-0.76)^2+(1-0.76)^2+(0.75-0.76)^2+(0.3-0.76)^2+\cdots+(1-0.76)^2}{18(18-1)} = 842.956 \times 10^{-5}$$

$$v(\bar{Y}_{h=Non-forest,d}) = \frac{(0-0.15)^2+(0.25-0.15)^2+(1-0.15)^2+(0-0.15)^2+\cdots+(0-0.15)^2}{12(12-1)} = 685.606 \times 10^{-5}$$

The mean proportion for the population is estimated by using Equation 56:
$\bar{Y}_d = (0.57 \times 0.76) + (0.43 \times 0.15) = 0.495167$

The estimated variance is obtained by using Equation 57:

$$v(\bar{Y}_d) = 30^{-1}\Bigg[\big((0.57 \times 18 \cdot 842.956 \times 10^{-5}) + (0.43 \times 12 \times 685.606 \times 10^{-5})\big)$$

$$+\left(\left((1-0.57)\frac{18}{30} \times 842.956 \cdot 10^{-5}\right) + \left((1-0.43)\frac{12}{30} \times 685.606 \cdot 10^{-5}\right)\right)\Bigg]$$

$$= 418.675 \times 10^{-5}$$

The total forest area is 186,000 acres × 0.495167 = 92,101 acres with an estimated variance of $186{,}000^2$ $\text{acres}^2 \times 418.675 \times 10^{-5}$ = 144,844,803 acres^2. The standard error of the estimate is $(144{,}844{,}803\ \text{acres}^2)^{0.5}$ = 12,035 acres

Recall that the goal of the stratification is to reduce the variance of the estimate. The variance of the post-stratified estimate of the mean was 0.004 as compared to 0.007 under SRS suggesting a relative efficiency of approximately 0.007/0.004 = 1.75.

Expansion Factors

Expansion factors are commonly used in survey statistics and are the sampling weights (i.e., the inverse of the inclusion probability). An expansion factor approach can easily be developed by combining Equations 51, 54, and 56 to yield:

$$\hat{Y}_d = \sum_h^H A_T\, W_h n_h^{-1} \sum_{i=1}^{n_h} y_{hid}$$

$$= \sum_h^H \frac{A_h}{n_h} \sum_{i=1}^{n_h} y_{hid}$$

$$= \sum_h^H EF_h \sum_{i=1}^{n_h} y_{hid}$$

$$= \sum_h^H \sum_{i=1}^{n_h} y_{hid}\, EF_h$$

Where

$A_h = W_h A_T$ = the area of stratum *h*, and

$EF_h = A_h/n_h$ = the expansion factor for stratum *h*.

The expansion factors in this context are not sampling weights but rather post-stratification weights. As such, post-stratification weights and hence expansion factors can change when different stratification approaches are used, when the map used to construct the stratification changes, and when n_h changes because of different response rates over time.

As described in Burrill et al. (2018), the expansion factors are often joined directly to condition- or plot-level data summaries and further simplify calculations to

$$\hat{Y}_d = \sum_{i \in d} y_{hid} EF_{hid} \tag{58}$$

We demonstrate the typical use of the expansion factor approach that uses the population described in Table 8. As previously noted, A_T =186,000 acres, $W_{h=forest}$ = 0.57, $n_{h=forest}$ = 18, $W_{h=Nonforest}$ = 0.43, and $n_{h=Nonforest}$ = 12.

The expansion factor for the forest stratum

EF_h=Forest = 0.57 × 186,000 acres/18 = 5,890 acres (Result 1)

The expansion factor for the nonforest stratum

EF_h=Nonforest = 0.43 × 186,000 acres/12 = 6,665 acres (Result 2)

The notation $i \in d$ in the first term in Equation 58 denotes that only plots in the domain = forest are summed. That is, only plots where proportion forest > 0. In Table 8, only 20 plots have a proportion forest > 0. The area of forest is then

$$\hat{Y}_d = (1 \times 5890) + (1 \times 5890) + (0.75 \times 5890) + (0.25 \times 6665) + \cdots + (0.2 \times 6665)$$

= 92,101 acres.

In most instances the practitioner uses the expansion factor approach without knowledge of geographic boundary used to define the population. As Scott et al. (2005) note, "… the use of expansion factors prohibits accurate variance estimation." This is an issue when the practitioner selects a geographic domain of study that contains subsets of one or more population boundaries used to calculate the expansion factors. When this happens, A_T and W_h are no longer known but estimated.

Expansion Factor-Based Variance Calculation

Consider the population described in Table 8 and suppose there is a need to estimate the forest area in county 2. As previously defined by Result 1 and Result 2, the expansion factors for the forest and nonforest strata are 5,890 ac and 6,665 acres, respectively. The first step in the expansion factor approach is to subset Table 8 so that only plots from county 2 are used. The total forest area is then estimated by using Equation 58:

$$\hat{Y}_d = (1\times\ 5{,}890) + (1\times\ 5{,}890) + (1\times\ 5{,}890) + (1\times\ 5{,}890) + \cdots + (0.2\times\ 6{,}665)$$

$$= 54{,}537\ acres \qquad \text{(Result 3)}$$

A_T and W_h are needed to estimate $v(\hat{Y}_d)$. Based on the derivation of the expansion factors

$A_h = n_h EF_h$, $A_T = \sum A_h$, and $W_h = A_h / A_T$

$$\hat{A}_{h=forest} = 11\times\ 5{,}890\ acres = 64{,}790\ acres$$

$$\hat{A}_{h=nonforest} = 6\times\ 6{,}665\ acres = 39{,}990\ acres$$

$$\hat{A}_T = 64{,}790\ ac + 39{,}990\ acres = 104{,}780\ acres \qquad \text{(Result 4)}$$

$$\widehat{W}_{h=Forest} = {}^{64{,}790\ acres}\!/_{104{,}780\ acres} = 0.6183 \qquad \text{(Result 5)}$$

$$\widehat{W}_{h=Nonforest} = {}^{39{,}990\ acres}\!/_{104780\ acres} = 0.3817 \qquad \text{(Result 6)}$$

In the above calculations we have added accents to denote that A_T and W_h are now estimated rather than known. Ignoring that A_T and W_h are now estimated, the practitioner may rely on Equations 54, 55, and 57 to construct the variance of the forest area estimate for county 2:

$$\bar{Y}_{h=Forest,d} = 11^{-1}(1 + 0 + 1 + 1 + \cdots + 1) = 0.79545$$

$$\bar{Y}_{h=Non-forest,d} = 6^{-1}(0 + 0 + 0.25 + 0 + \cdots + 0) = 0.075$$

$$v\left(\bar{Y}_{h=Forest,d}\right) = \frac{(1-0.795)^2 + (0-0.795)^2 + (1-0.795)^2 + (1-0.795)^2 + \cdots + (1-0.795)^2}{11(11-1)} = 0.01457$$

$$v\left(\bar{Y}_{h=Non-forest,d}\right) = \frac{(0-0.075)^2 + (0-0.075)^2 + (0.25-0.075)^2 + (0-0.075)^2 + \cdots + (0-0.075)^2}{6(6-1)} = 0.00229$$

$$v(\bar{Y}_d) = 17^{-1}\left[\left((0.6183 \times\ 11 \cdot 0.01457) + (0.3817 \times\ 6 \times\ 0.00229)\right) + \left(\left((1 - 0.6183)\frac{11}{17} \times\ 0.01457\right) + \left((1 - 0.3817)\frac{6}{17} \times\ 0.00229\right)\right)\right] = 0.00638.$$

The estimated variance of the total forest area estimate for county 2 is

$v\left(\hat{Y}_d\right) = 104780^2\ acres^2 \times\ 0.00638 = 70{,}019{,}557\ acres^2.$ (Result 7)

The standard error is 8,367.8 acres. However, error arising from the estimation of W_h and A_T has not been accounted for with the expansion factor approach.

Domain-Based Variance Calculation

Scott et al. (2005) note inaccurate variance estimates when they use the expansion factor approach. The issue arises because it is not obvious how the populations are defined when data are downloaded from FIA's online database. Population information is stored in a series of population tables within the database (see Burrill et al. 2018 for more information). To properly construct the estimated variance, the practitioner should realize that a domain of interest has changed from forest area in the population to forest area in county 2. With this domain of interest, the domain indicator in Equation 48 is changed so that δ_{ikd} takes a value of 1 when the observation on plot i in condition k is in the domain of interest and zero otherwise. The use of Equation 48 sets all plot-level observations to zero when they are not forest in county 2. Once the domain indicator has been adjusted, the variance calculation follows Equations 54, 55, and 57:

$$\bar{Y}_{h=Forest,d} = 18^{-1}(0+0+0+0+\cdots+1+1) = 0.4861,$$

$$\bar{Y}_{h=Non-forest,d} = 12^{-1}(0+0+0+0+\cdots+0.2+0) = 0.0375$$

$$v\left(\bar{Y}_{h=Forest,d}\right) = \frac{(0-0.486)^2+(0-0.486)^2+(0-0.486)^2+(0-0.486)^2+\cdots+(1-0.486)^2+(1-0.486)^2}{18(18-1)} = 0.01408$$

$$v\left(\bar{Y}_{h=Non-forest,d}\right) =$$

$$\frac{(0-0.0375)^2+(0-0.0375)^2+(0-0.0375)^2+(0-0.0375)^2+\cdots+(0.2-0.0375)^2+(0-0.0375)^2}{12(12-1)} = 0.00065$$

$$v(\bar{Y}_d) = 30^{-1}\Bigg[\big((0.57\times 18\cdot 0.01408)+(0.43\times 12\cdot 0.00065)\big) + \left(\left((1-0.57)\frac{18}{30}\times 0.01408\right)+\left((1-0.43)\frac{12}{30}\times 0.00065\right)\right)\Bigg] = 0.00505$$

The estimated variance of the total forest area estimate for county 2 is

$$v\left(\hat{Y}_d\right) = 186{,}000^2\ acres^2 \times 0.00505 = 174{,}833{,}568\ acres^2 \qquad \text{(Result 8)}$$

The standard error of the estimate is 13,223 acres.

Comparing Results

As stated in the Simple Random Sampling Case section above, the total area of county 2 is 98,580 acres and the stratum weights are 0.57 and 0.43 for the forest and nonforest stratum, respectively. However, based on Result 3, the estimated total area of county 2 is 104,780 acres and the estimated strata weights are 0.62 and 0.38 for the forest and nonforest strata, respectively (Results 5 and 6). The estimated variance for the domain-based variance (Result 8) is 2.5 times larger than the expansion factor estimated variance (Result 7). These results support the statement by Scott et al. (2005) that the expansion factor approach can lead to inaccurate variance estimates.

Area Control

The term "area control" refers to the spatial scales at which A_T is considered known.[2] At its most basic level, area control refers to the size and geographic extent of the population for post-stratified estimation. The way in which populations are defined, for estimation purposes, differs by region. For example, in the western United States where counties may be large enough to meet the sample size recommendations (Westfall et al. 2011), a single county may be defined as a population. In the South where counties are often small, populations tend to be defined by survey units (aggregates of counties that roughly follow physiographic boundaries). When the population, and hence A^T, is defined by county aggregates, there is no area control at the county-level.

We demonstrated, in the Comparing Results section above, how estimated values of total area can differ from known values of total area. For example, the estimated total area of county 2 is 104,780 acres and the known value is 98,580 acres. When estimating the total and the variance of the total, it is important to use a known A_T when possible because the estimated total arising from Equation 51 is the product of A_T and the estimate of the mean. The variance of the total is the product of A_T^2 and the variance of the mean (Eq. 52). There are two primary options available to ensure that a known A_T is used when constructing estimates.

[2] Our use of the term area control relates only to constructing estimates of forest parameters. Area control as presented in this Appendix does not have any relationship to terms area control and volume control used in forest regulation.

The first and preferred option is to define the population based on the desired scale of area control and develop the post-stratification weights for that population. Following our previous example of estimating the forest area in county 2, the practitioner would consider county 2 as the population, develop the post-stratification weights specifically for the county 2 population, and then employ Equations 53–58 to construct estimates. However, care should be taken with this approach to ensure adequate sample size.

The second approach is to use the ratio estimator (Foundational Documentation chapter, Equations 10–13) and treat both the forest area and the total area as estimated parameters. This approach, from a forest area perspective, produces mean proportion forest for the domain (and the variance of the mean). The known area of the domain may then be used to construct estimates of the total forest area using Equations 51 and 52. Returning our attention to the estimation of forest area in county 2, the ratio estimator is

$$\hat{R}_{dd'} = \frac{\hat{Y}_d}{\hat{Y}_{d'}}$$

Where

d, d' = the specified domains for the numerator and denominator, respectively.

When the forest area estimate for county 2 (Result 3) is used as the numerator and estimated total area of county 2 (Result 4) is used as the denominator, the ratio estimate (proportion forest in county 2) is

$$\hat{R}_{dd'} = \frac{54{,}537 \text{ ac}}{104{,}780 \text{ ac}} = 0.52049.$$

The estimate of the total forest area is then the product of the known area of the county 2 (98,580 acres) and $\hat{R}_{dd'}$.

$$\hat{Y}_{dr} = 98{,}580\ acres \times\ 0.52049 = 51309.7\ acres$$

Here we have added the *r* subscript to $\hat{Y}_d$ to denote that the estimate arises from the ratio estimator.

As discussed in the Foundational Documentation chapter, the estimated variance is

$$v(\hat{R}_{dd'}) = \frac{1}{\hat{Y}_{d'}^2}\left[v(\hat{Y}_d) + \hat{R}_{dd'}^2 v(\hat{Y}_{d'}) - 2\hat{R}_{dd'} cov(\hat{Y}_d, \hat{Y}_{d'})\right]$$

Where

$$cov(\hat{Y}_d, \hat{Y}_{d'}) = A_T^2 n^{-1}[\sum_{h=1}^H W_h\, n_h cov(\bar{Y}_{hd}, \bar{Y}_{hd'}) + \sum_{h=1}^H (1 - W_h)\, n_h n^{-1} cov(\bar{Y}_{hd}, \bar{Y}_{hd'})],$$

and $cov(\bar{Y}_{hd}, \bar{Y}_{hd'}) = \frac{\sum_{i=1}^{n_h} y_{hid} y_{hid'} - n_h \bar{Y}_{hd} \bar{Y}_{hd'}}{n_h(n_h - 1)}$.

To calculate $v(\hat{R}_{dd'})$ parameter estimates are needed for $v(\hat{Y}_d)$, $v(\hat{Y}_{d'})$ $cov(\bar{Y}_{hd}, \bar{Y}_{hd'})$ and $cov(\hat{Y}_d, \hat{Y}_{d'})$. Result 8 provides $v(\hat{Y}_d)$. To estimate $v(\hat{Y}_{d'})$ Equations 52, 54, 55, and 57 are used. Additionally, the domain indicator must be changed so that all plots in county 2 (not just the forested plots) have a value of 1 and the remaining plots have a value of zero (i.e., y_{hid} = 1 if the plot is in county 2 and zero otherwise). The estimate of $v(\hat{Y}_{d'})$ = 310,778,949 ac^2. The within-stratum covariance is

$$cov(\bar{Y}_{h=forest,d}, \bar{Y}_{h=forest,d'})$$

$$= \frac{[(0\times\ 0) + (0\times\ 0) + \cdots + (1\times\ 0.75) + (1\times\ 1) + (1\times\ 1)] - 18\times\ 0.486\times\ 0.611}{18(18-1)}$$

= 0.0111

$$cov(\bar{Y}_{h=non-forest,d}, \bar{Y}_{h=non-forest,d'})$$

$$= \frac{[(0\times\ 0) + (0\times\ 0) + \cdots + (1\times\ 0) + (1\times\ 0.2) + (1\times\ 0)] - 12\times\ 0.0375\times\ 0.5}{12(12-1)}$$

= 0.0017

The total covariance is then

$$cov(\hat{Y}_d, \hat{Y}_{d'}) = 186{,}000^2 30^{-1}[(0.57\times\ 18\cdot 0.011) + (0.43\times\ 12\times\ 0.0017)$$

$$+ \left((1-0.57)\times\ 18\times 30^{-1}\times\ 0.011\right) + \left((1-0.43)\times\ 12\times 30^{-1}\times\ 0.0017\right)]$$

$$= 145{,}471{,}840\ acres^2$$

And the estimated variance of $\hat{R}_{dd'}$ is

$$v(\hat{R}_{dd'}) = \frac{1}{104{,}780^2}[174{,}833{,}568 + (0.2709 \times\ 310{,}778{,}949) - (2 \times\ 0.52049 \times$$

$$145{,}471{,}840)] = 0.0098.$$

The variance of $\hat{Y}_{dr}$ is then estimated as

$$v(\hat{Y}_{dr}) = 98{,}580^2 \times\ 0.0098 = 95{,}237{,}085 \text{ acres}^2$$

The standard error of the estimate is 9,759 acres. Using the ratio approach, the practitioner can construct domain estimates with area control provided the area of the domain is known. This approach can be more precise than relying on the basic domain estimation approach presented in the Domain-Based Variance Calculation section where the standard error of the forest area estimate for county 2 is 13,223 acres.

Area control always occurs at the population level and aggregates of populations also have area control. To our knowledge, the lowest common denominator of area control is at the state level. In other words, there is a set of populations used for estimation purposes that, when summed, maintain area control at the state level. This is an important characteristic for state-level reporting requirements.

Literature Cited

Brooks, E.B.; Coulston, J.W.; Wynne, R.H.; [et al.]. 2016. **Improving the precision of dynamic forest parameter estimates using LandSat.** Remote Sensing of Environment. 179: 162–169. https://doi.org/10.1016/j.rse.2016.03.017.

Burrill, E.A.; Wilson, A.M.; Turner, J.A.; [et al.]. 2018. **FIA database description and users guide for Phase 2 (version 7.2).** U.S. Department of Agriculture, Forest Service. 950 p. Available at https://www.fia.fs.usda.gov/library/database-documentation/ (accessed May 10, 2019).

Coulston, J.W. 2008. **Forest inventory and stratified estimation: a cautionary note.** Res. Note SRS-16. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 8 p. https://doi.org/10.2737/SRS-RN-16.

Coulston, J.W.; Moisen, G.G.; Wilson, B.T.; [et al.]. 2012. **Modeling percent tree canopy cover: A pilot study.** Photogrammetric Engineering and Remote Sensing. 78(7): 715–727.

Coulston, J.W.; Reams, G.A.; Wear, D.N.; [et al]. 2014. **An analysis of forest land use, forest land cover and change at policy-relevant scales.** Forestry. 87(2): 267–276. https://doi.org/10.1093/forestry/cpt056.

DellaSala, D.A.; Staus, N.L.; Strittholt, J.R.; [et al.]. 2001. **An updated protected areas database for the United States and Canada.** Natural Areas Journal. 21: 124–135.

Frayer, W.E.; Furnival, G.M. 1999. **Forest survey sampling designs: a history.** Journal of Forestry. 97(12): 4–10.

Freese, F. 1962. **Elementary forest sampling.** Agriculture Handbook No. 232. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experimental Station. 91 p.

Homer, C.; Dewitz, J.; Yang, L.; Jin [et al.]. 2015. **Completion of the 2011 National Land Cover Database for the conterminous United States—representing a decade of land cover change information.** Photogrammetric Engineering and Remote Sensing. 81(5): 345–354. http://dx.doi.org/10.14358/PERS.81.5.345.

McRoberts, R.E. 2010. **The effects of rectification and Global Positioning System errors on satellite image-based estimates of forest area.** Remote Sensing of Environment. 114: 1710–1717. https://doi.org/10.1016/j.rse.2010.03.001.

McRoberts, R.E.; Holden, G.R.; Nelson, M.D.; [et al.]. 2006. **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. 36: 2968–2980. https://doi.org/10.1139/x05-222.

Nelson, M.D.; Reams, G.A. 2017. **Is the area of U.S. forests increasing or decreasing?** Forestry Source. 22: 16–17.

Patterson, P.L.; Coulston, J.W.; Roesch, F.A.; [et al.]. 2012. **A primer for nonresponse in the United States Forest Inventory and Analysis Program.** Environmental Monitoring and Assessment. 184: 1423–1433. https://www.fs.usda.gov/research/treesearch/40200 (accessed December 6, 2021).

Patterson, P.L.; Reams, G.A. 2005. **Combining panels for Forest Inventory and Analysis estimation.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 69–73.

Reams, G.A. 2000. **SAFIS area estimation techniques.** FIM: 32–36.

Scott, C.T.; Bechtold, W.A.; Reams, G.A.; [et al.]. 2005. **Sample-based estimators used by the forest inventory and analysis national information management system.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 53–77.

USDA Forest Service. 1992. **Forest Service resource inventories: An overview.** Washington, D.C.: USDA Forest Service, Forest Inventory, Economics, and Recreation Research Staff. 39 p. https://www.srs.fs.usda.gov/pubs/misc/fs_fia-overview.pdf (accessed December 6, 2021).

Westfall, J.A.; Patterson, P.L.; Coulston, J.W. 2011. **Post-stratified estimation: within strata and total sample size recommendations.** Canadian Journal of Forest Research. 41: 1130–1139. https://doi.org/10.1139/x11-031.

Wickham, J.; Stehman, S.V.; Gass, L.; [et al.]. 2017. **Thematic accuracy assessment of the 2011 National Land Cover Database (NLCD).** Remote Sensing of Environment. 191: 328341. https://doi.org/10.1139/x11-031.

Appendix 2: Ratio-to-Size Estimation

Charles T. Scott and James A. Westfall

The following describes an alternative approach to that currently used by Forest Inventory and Analysis (FIA) (Scott et al. 2005). The ratio-to-size estimator is presented as an alternative to estimation approaches already described in the Foundational Documentation chapter for means, totals, and ratios under a simple random sampling design. The ratio-to-size estimator is then extended to account for partially sampled plots in a post-stratified estimation framework. The material is drawn from an unpublished document, “Formulas for Estimators and Their Variances in NFI,” by Korhonen and Salmensuu (2014).

The ratio-to-size estimator is a ratio estimator that takes advantage of the proportional relationship between the total of the attribute observed and the size of the sampling unit (Thompson 2012, p. 160). Korhonen and Salmensuu (2014) used the estimator, in part, to account for the portion of the sampling unit that could not be measured, such as when a portion is too dangerous to measure or access is denied by the owner. As will be demonstrated, the estimator can be used to take advantage of the relationship between any two attributes measured. Specifically, this estimator is efficient when the attribute of interest is correlated with the size of the attribute in the denominator. A common example in forest inventory is when the tree volume or biomass on a plot is related to the portion of the plot that is forested. This situation occurs when plots are randomly placed on the landscape without regard to whether the location is forested in order to determine the total forest area and volume within the population. A given plot may straddle the boundary between forest and nonforest or between two different domains of interest, such as forest types. This may also happen when a portion of the plot cannot be measured for safety or other reasons. The estimator can also be used to estimate other ratios, such as the proportion of the total volume contributed by a certain tree species.

The following assumes a single fixed-area plot located within a single stratum based on the plot center. However, the results extend easily to clusters of plots or as used here, plots composed of subplots, such as is used by the Forest Inventory and Analysis program in the United States (Bechtold and Scott 2005). In that case, the attributes must be summarized to the plot level. Given the FIA plot configuration is fixed (no randomization of the subplots within the plot), the results given here apply (i.e., this is not two-stage random sampling).

Estimation

Simple Random Sample (SRS)

Estimates of forest proportion (or proportion of any category of interest) of the total population area, A_T, or of the density other attributes of interest, such as biomass, in domain d can be estimated by using Equation 59. The numerator is the total area measured that is in the domain of interest d (such as land use or cover class) for area proportions or as the total for other attributes (e.g., biomass) measured that is in the domain of interest d (such as a species or diameter class within a particular forest type). The denominator is the total area measured:

$$\bar{Y}_d = \frac{\sum_{i=1}^{n} y_{id}}{\sum_{i=1}^{n} a_i^m} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{K} \delta_{ikd} y_{ikd}}{\sum_{i=1}^{n}\sum_{k=1}^{K} a_{ik}^m} \quad (59)$$

Where

y_{id} = area or attribute total of plot i in domain d,

a_i^m = area measured on plot i (excluding portions that are inaccessible or out-of-population),

a_{ik}^m = the area measured on plot i in condition k (excluding portions that are inaccessible or out-of-population),

y_{ikd} = area or attribute total for condition k in plot i in domain d, and

δ_{ikd} = indicator variable which is 1 if condition k on plot i is in domain d, such as a specific forest type or species; 0 otherwise.

Note that although unmeasured areas are mapped to their own condition, they are not included in the k conditions being summed. The exception to this method is when area of nonresponse is estimated, in which case the unmeasured areas are included in the K conditions. Also, observation unit values y_{ikd} are simply summed and are not expressed on a per-unit-area basis, such as the sum of sampled tree biomass values. The sum of the area measured in the denominator takes the role of the sample size in the typical formula for calculating a mean. Population totals are estimated as

$$\hat{Y}_d = A_T \, \bar{Y}_d \tag{60}$$

The variance of the estimated mean (Eq. 59) is

$$v(\bar{Y}_d) = \frac{n}{n-1} \frac{\sum_{i=1}^{n}\left(y_{id} - a_i^m \bar{Y}_d\right)^2}{\left(\sum_{i=1}^{n} a_i^m\right)^2} = \frac{n}{n-1} \frac{\sum_{i=1}^{n} y_{id}^2 - 2\bar{Y}_d \sum_{i=1}^{n} y_{id} a_i^m + \bar{Y}_d^2 \sum_{i=1}^{n} (a_i^m)^2}{\left(\sum_{i=1}^{n} a_i^m\right)^2} \tag{61}$$

And the variance of the estimated total (Eq. 60) is

$$v\left(\hat{Y}_d\right) = A_T^2 \, v(Y_d) \tag{62}$$

Often in forest inventory, nested or concentric plots are used to sample trees of different sizes. Seedlings are sampled on small plots while large trees are sampled on large plots. A commonly desired estimate that spans more than one plot size is total number of trees. To address this, the largest plot size is used in the denominator and each size class, *j*, is rescaled to the largest size class, *J*. This approach ensures that estimates across plot sizes are equal to the sum of estimates from individual plot sizes. The plot attribute becomes

$$y_{id} = \sum_j^J y_{ijd} \frac{\sum_{i=1}^{n} a_{i_J}^m}{\sum_{i=1}^{n} a_{i_j}^m} \tag{63}$$

Where

y_{ijd} = sum of the attribute of interest on plot *i* having plot size *j* in domain d^m, and

a_{ij} = plot size $_{ij}$ area measured on plot *i*.

Ratio Estimates (SRS)

Often estimates of means are of more interest for a particular subcategory of land, rather than across the whole land area, such as biomass per hectare of forest land. It can be estimated as the ratio of the mean of the attribute across all land divided by the mean area proportion across all land:

$$\hat{R}_{dd'} = \frac{\bar{Y}_d}{\bar{Y}_{d'}} = \frac{\sum_{i=1}^{n} y_{id} \Big/ \sum_{i=1}^{n} a_i^m}{\sum_{i=1}^{n} y_{id'} \Big/ \sum_{i=1}^{n} a_i^m} = \frac{\sum_{i=1}^{n} y_{id}}{\sum_{i=1}^{n} y_{id'}} \tag{64}$$

Where

y_{id} = the attribute of interest in domain d on plot i where d is typically a subdomain of d', for example d is a species-specific biomass on forest land, and $y_{id'}$ = the area of interest in domain d' on plot i, where d' is typically the primary domain such as forest land.

The variance of the ratio estimate can be computed by using the same approach as in (Eq. 61) for $\bar{Y}_d$, but replacing a_i with $y_{id'}$. That is, instead of summing across all measured areas, only those areas in the domain d are summed.

$$v\left(\hat{R}_{dd'}\right) = \frac{n}{n-1} \frac{\sum_i^n \left(y_{id} - \hat{R}_{dd'} y_{id'}\right)^2}{\left(\sum_i^n y_{id'}\right)^2} = \frac{n}{n-1} \frac{\sum_i^n y_{id}^2 - 2\hat{R}_{dd'} \sum_i^n y_{id} y_{id'} + \hat{R}_{dd'}^2 \sum_i^n y_{id'}^2}{\left(\sum_i^n y_{id'}\right)^2} \tag{65}$$

This is a much simpler estimator than is shown in Equation 11 in the chapter on Foundational Documentation, but the results are identical when there are no partial plots due to inaccessibility. When there are partial plots, then Equation 65 more accurately reflects the variance.

Post-stratification

In a post-stratified estimation framework, the mean across the entire population area is estimated as

$$\bar{Y}_d = \sum_{h=1}^{H} W_h \bar{Y}_{hd} = \sum_{h=1}^{H} W_h \frac{\sum_{i=1}^{n_h} y_{hid}}{\sum_{i=1}^{n_h} a_{hi}^m} \tag{66}$$

Where

W_h = weight for stratum h,

y_{hid} = the attribute of interest in domain d on plot i in stratum h, and

a_{hi}^{m} = area measured on plot i in stratum h (excluding portions that are inaccessible or out-of-population).

The total is estimated by multiplying by the total area:

$$\hat{Y}_d = A_T\, \bar{Y}_d \tag{67}$$

The ratio is computed by dividing the total of the attribute in domain $d\left(\hat{Y}_d\right)$ by the total of the attribute in domain $d'\left(\hat{Y}_{d'}\right)$:

$$\hat{R}_{dd'} = \frac{\hat{Y}_d}{\hat{Y}_{d'}} = \frac{A_T \sum_{h=1}^{H} W_h \bar{Y}_{hd}}{A_T \sum_{h=1}^{H} W_h \bar{Y}_{hd'}} = \frac{\sum_{h=1}^{H} W_h \bar{Y}_{hd}}{\sum_{h=1}^{H} W_h \bar{Y}_{hd'}} \tag{68}$$

Note: It is important to first estimate the totals $\hat{Y}_d$ and $\hat{Y}_{d'}$ across strata then create the ratio estimate, rather than doing the ratio estimate by strata and subsequently constructing a weighted average. The first method is unbiased, the second is not.

Post-Stratified Variance Estimation

In post-stratification, a sample of the population is selected first then the stratification is applied (Cochran 1977, p. 134). This is typical of long-term National Forest Inventories that use permanent plots. This results in within-stratum sample sizes being random variates, i.e., the sample sizes are not predetermined and would vary among samples. This source of variation is accounted for in the second term of the variance estimator. The variance estimator of the mean for post-stratification is

$$v(\bar{Y}_d) = n^{-1}\left(\sum_{h=1}^{H} W_h\, n_h v(\bar{Y}_{hd}) + \sum_{h=1}^{H}(1-W_h) n_h n^{-1} v(\bar{Y}_{hd})\right) \tag{69}$$

Where the stratum variance is

$$v(\bar{Y}_{hd}) = \frac{n_h^2}{n_h - 1}\, \frac{\sum_{i=1}^{n_h} y_{hid}^2 - 2\bar{Y}_{hd} \sum_{i=1}^{n_h} y_{hid} a_{hi}^{m} + \bar{Y}_{hd}^2 \sum_{i=1}^{n_h} a_{hi}^{m\,2}}{\left(\sum_{i=1}^{n_h} a_{hi}^{m}\right)^2} \tag{70}$$

The variance of the total is

$$v\left(\hat{Y}_d\right) = A_T^2\, v(\bar{Y}_d) \tag{71}$$

The general form for the variance of the ratio estimate is

$$v\left(\hat{R}_{dd'}\right) = \frac{1}{\hat{Y}_{d'}^2}\left[v\left(\hat{Y}_d\right) + \hat{R}_{dd'}^2\, v\left(\hat{Y}_{d'}\right) - 2\,\hat{R}_{dd'}\, cov\left(\hat{Y}_d, \hat{Y}_{d'}\right)\right] \quad (72)$$

The $cov\left(\hat{Y}_d, \hat{Y}_{d'}\right)$ under a post-stratified design is

$$cov\left(\hat{Y}_d, \hat{Y}_{d'}\right) = \frac{A_T^2}{n}\left[\sum_{h=1}^{H} W_h\, cov\left(\hat{Y}_{hd}, \hat{Y}_{hd'}\right) + \sum_{h=1}^{H}(1 - W_h)\frac{cov\left(\hat{Y}_{hd}, \hat{Y}_{hd'}\right)}{n}\right] \quad (73)$$

Where the stratum covariance for ratio-to-size estimation is

$$cov\left(\hat{Y}_{hd}, \hat{Y}_{hd'}\right) = \frac{n_h^2}{(n_h - 1)} \frac{\sum_{i=1}^{n_h}\left(y_{hid} - a_{hi}^m \bar{Y}_{hd}\right)\left(y_{hid'} - a_{hi}^m \bar{Y}_{hd'}\right)}{\left(\sum_{i=1}^{n_h}\left(a_{hi}^m\right)\right)^2}$$

$$= \frac{n_h^2}{(n_h - 1)} \frac{\sum_{i=1}^{n_h} y_{hid} y_{hid'} - \bar{Y}_{hd'} \sum_{i=1}^{n_h} a_{hi}^m y_{hid} - \bar{Y}_{hd} \sum_{i=1}^{n_h} a_{hi}^m y_{hid'} + \left(\sum_{i=1}^{n_h} a_{hi}^{m^2}\right) \bar{Y}_{hd} \bar{Y}_{hd'}}{\left(\sum_{i=1}^{n_h} a_{hi}^m\right)^2} \quad (74)$$

Literature Cited

Bechtold, W.A.; Scott, C.T. 2005. **The Forest Inventory and Analysis plot design.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 27–42.

Cochran, W.G. 1977. **Sampling techniques.** New York: John Wiley & Sons. 428 p.

Korhonen, K.T.; Salmensuu, O. 2014. **Formulas for estimators and their variances in NFI.** Revised by Scott, C.T. An internal paper for the Forest and Agricultural Organization of the United States. https://www.dropbox.com/s/lhtpnom97p3a0gt/Formulas_KKor_2015-09-08%20Chip%20edits.docx?dl=0 (accessed May 10, 2019).

Scott, C.T.; Bechtold, W.A.; Reams, G.A.; [et al.]. 2005. **Sample-based estimators used by the forest inventory and analysis national information management system.** In: Bechtold, W.A.; Patterson, P.L., eds. 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: 53–77. https://doi.org/10.2737/SRS-GTR-80.

Thompson, S.K. 2012. **Sampling.** 3rd edition. Hoboken, NJ: John Wiley & Sons, Inc. 472 p.

Appendix 3. Metric Equivalents

When you know:	Multiply by:	To find:
Inches (in)	2.54	Centimeters
Feet (ft)	0.305	Meters
Acres (ac)	0.4047	Hectares
Cubic feet	0.0283	Cubic meters
Pounds (lb)	0.4536	Kilograms
Tons (ton)	907	Kilograms

FOREST SERVICE U.S. DEPARTMENT OF AGRICULTURE

Northern Research Station

www.nrs.fs.usda.gov