

United States  
Department of  
Agriculture

Forest Service

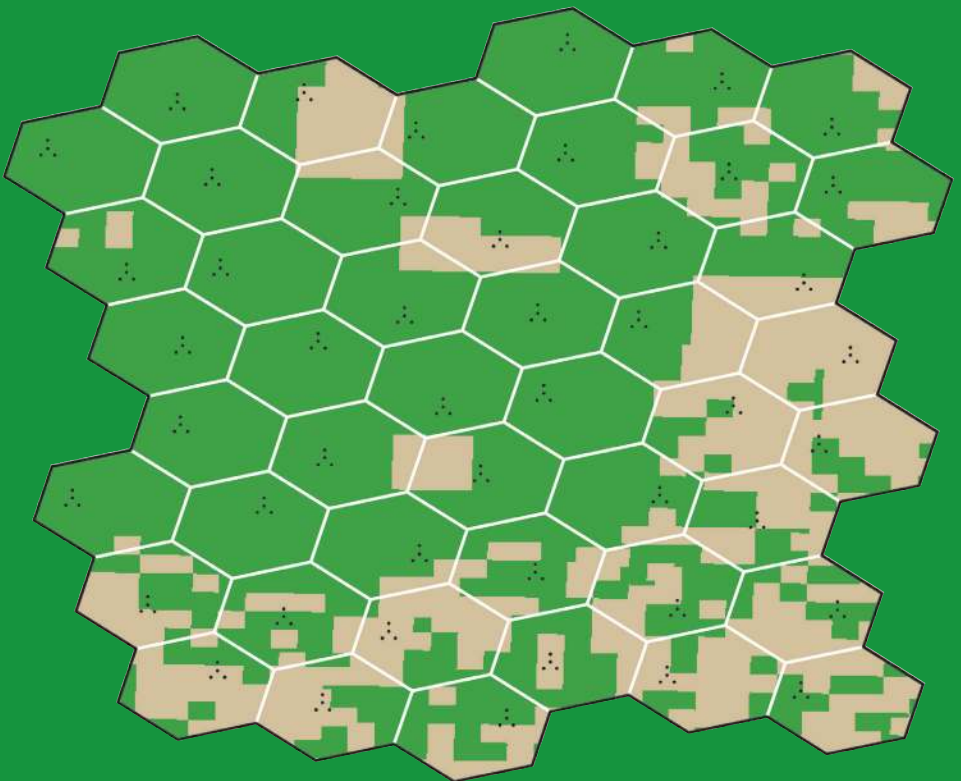


Southern  
Research Station

General Technical  
Report SRS-80

# The Enhanced Forest Inventory and Analysis Program—National Sampling Design and Estimation Procedures

William A. Bechtold and  
Paul L. Patterson, Editors



---

**The Authors:**

---

**William A. Bechtold**, Research Forester, USDA Forest Service, Southern Research Station, Asheville, NC 28802; and **Paul L. Patterson**, Mathematical Statistician, USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO 80526-1891.

**Cover:** The cover graphic illustrates FIA's sampling frame superimposed onto a landscape. The patchwork of green and brown under the hexagonal grid represents forest and nonforest areas. Each hexagon contains one ground plot symbolized by the cluster of four points located within each hexagon. The ground plots are not drawn to scale.

**Cover artwork:** Greg C. Liknes.

June 2005

Southern Research Station  
P.O. Box 2680  
Asheville, NC 28802

**The Enhanced Forest Inventory  
and Analysis Program—  
National Sampling Design and  
Estimation Procedures**

**William A. Bechtold and Paul L. Patterson, Editors**



# Table of Contents

Executive Summary .....	vii
<b>1 The Enhanced Forest Inventory and Analysis Program .....</b>	<b>1</b>
<i>Ronald E. McRoberts</i>	
1.1 Purpose and Scope of the Documentation .....	1
1.2 Overview of the Enhanced Forest Inventory and Analysis Program .....	2
1.2.1 Historical Perspective .....	2
1.2.2 Forest Health Monitoring .....	3
1.2.3 Annual Inventories .....	3
1.2.4 National Consistency .....	5
1.3 Overview of the Documentation .....	8
1.3.1 Sampling Frame .....	8
1.3.2 Plot Design .....	8
1.3.3 Sample-Based Estimators .....	9
1.3.4 Combining Data for Multiple Panels .....	9
1.3.5 Notation and Glossary .....	9
1.3.6 Web-Based Supplementary Documentation .....	9
1.4 Literature Cited .....	10
<b>2 The Forest Inventory and Analysis Sampling Frame .....</b>	<b>11</b>
<i>Gregory A. Reams, William D. Smith, Mark H. Hansen, William A. Bechtold, Francis A. Roesch, and Gretchen G. Moisen</i>	
2.1 Overview of Forest Inventory and Analysis Sampling Design .....	11
2.1.1 Forest Inventory and Analysis Populations .....	11
2.1.2 Phase 1 .....	12
2.1.3 Phase 2 .....	15
2.1.4 Phase 3 .....	16
2.2 Development of the Phase 2/Phase 3 Sampling Frame .....	18
2.2.1 Hexagonal Sampling Frame .....	18
2.2.2 Division of the Sampling Frame into Panels .....	20
2.2.3 Populating the Sampling Frame .....	22
2.2.4 Deviations from the Five-Panel Annual System .....	22
2.2.5 Theoretical Basis for the FIA Sampling Frame .....	23
2.3 Literature Cited .....	26

<b>3</b>	<b>The Forest Inventory and Analysis Plot Design</b> .....	27
	<i>William A. Bechtold and Charles T. Scott</i>	
3.1	Overview of the FIA Plot Design .....	28
3.1.1	Motivation Behind the FIA Plot Design .....	30
3.1.2	Field Protocols for Mapping Plots .....	31
3.1.3	Differences in Mapping Forest and Nonforest Plots .....	32
3.2	Condition Classification Based on Direct Field Observation .....	33
3.2.1	Discrete Variables That Trigger Recognition of a Unique Condition Class .....	33
3.2.2	Ancillary Condition-Class Variables .....	33
3.3	Computed Attributes .....	34
3.3.1	Computed Tree-Level Attributes .....	34
3.3.2	Condition Classification Based on Computed Attributes ....	35
3.3.3	Calculations of Area in Each Condition Class .....	38
3.4	Special Cases .....	39
3.4.1	Population Boundaries .....	39
3.4.2	Slivers .....	39
3.4.3	Nonsampled Plots and Plot Replacement .....	41
3.5	Literature Cited .....	42
<b>4</b>	<b>Sample-Based Estimators Used by the Forest Inventory and Analysis National Information Management System</b> .....	43
	<i>Charles T. Scott, William A. Bechtold, Gregory A. Reams, William D. Smith, James A. Westfall, Mark H. Hansen, and Gretchen G. Moisen</i>	
4.1	Panels and Estimation .....	43
4.2	Phase 1 .....	44
4.2.1	Satellite Classification and Known Strata Weights .....	44
4.2.2	Aerial Photography and Estimated Strata Weights .....	45
4.2.3	Combining Small Strata and Populations .....	46
4.3	Phase 2 .....	46
4.3.1	Estimation of Area by Stratum by Domain (Row and Column) Attributes .....	47
4.3.2	Estimation of Population Totals and Their Variances .....	50
4.3.3	Estimation of Other Attributes .....	52
4.3.4	Estimation of Ratios .....	55
4.3.5	Computation of Sampling Errors .....	59
4.3.6	Components of Change .....	59
4.4	Expansion Factors .....	65

4.5	Phase 3 .....	66
4.5.1	Estimation Procedures Used for FHM (Phase 3) Data .....	66
4.5.2	Combining with Phase 1 .....	66
4.5.3	Linking with Phase 2 .....	66
4.6	Literature Cited .....	67
<b>5</b>	<b>Combining Panels for Forest Inventory and Analysis</b>	
	<b>Estimation</b> .....	69
	<i>Paul L. Patterson and Gregory A. Reams</i>	
5.1	Single Panels vs. Multiple Panels .....	69
5.2	Combining Panels .....	71
5.2.1	The Moving Average Method .....	71
5.2.2	The Temporally Indifferent Method .....	73
5.3	Literature Cited .....	73
<b>6</b>	<b>Notation for Equations</b> .....	75
6.1	Indices .....	75
6.2	Double Sampling .....	76
6.3	Area Related .....	76
6.4	Tree and Plot Related .....	77
<b>7</b>	<b>Glossary of Terms</b> .....	79
<b>8</b>	<b>Web-Based Supplementary Documentation</b> .....	85

## Abstract

The Forest Inventory and Analysis (FIA) Program of the U.S. Department of Agriculture Forest Service is in the process of moving from a system of quasi-independent, regional, periodic inventories to an enhanced program featuring greater national consistency, annual measurement of a proportion of plots in each State, new reporting requirements, and integration with the ground sampling component of the Forest Health Monitoring Program. This documentation presents an overview of the conceptual changes, explains the three phases of FIA's sampling design, describes the sampling frame and plot configuration, presents the estimators that form the basis of FIA's National Information Management System (NIMS), and shows how annual data are combined for analysis. It also references a number of Web-based supplementary documents that provide greater detail about some of the more obscure aspects of the sampling and estimation system, as well as examples of calculations for most of the common estimators produced by FIA.

**Keywords:** Annual inventory, FIA, forest health monitoring, forest inventory, plot design, sampling frame.

## Acknowledgments

The authors extend special thanks to the nine subject-matter specialists who reviewed this manuscript. Their valuable feedback helped us construct a solid foundation for this important national program and improved the usefulness of the associated documentation for the wide range of customers and stakeholders served by Forest Inventory and Analysis. We are particularly grateful to Alan R. Ek, University of Minnesota, Department of Forest Resources, who acted as associate editor and coordinated a double-blind review process.



## Executive Summary

The Forest Inventory and Analysis (FIA) Program of the U.S. Department of Agriculture Forest Service has begun a move from quasi-independent, regional, periodic inventories to the national Enhanced FIA Program, which features greater national consistency, a complete and systematic annual sample of each State, new reporting requirements, and integration with the ground sampling component of the Forest Health Monitoring Program. The transition to Enhanced FIA has both resulted from and caused conceptual changes in approaches to conducting forest inventories, which, in turn, have required corresponding technical changes. This document provides background on the conceptual changes and articulates technical aspects of the Enhanced FIA Program as a means of ensuring a common understanding and practice among the regional programs. It facilitates further development of the national core program, provides a defensible statistical basis for the program's sampling and estimation components, and fosters greater credibility with users and stakeholders. Although this version of the documentation primarily is intended for an internal FIA audience, additional documentation of more general interest will be developed as the program matures.

The Enhanced FIA Program is conducted in three phases with objectives as follows:

1. In Phase 1, FIA personnel stratify land areas to increase the precision of estimates.
2. In Phase 2, FIA field crews obtain observations and measurements of the traditional FIA suite of variables.
3. In Phase 3, FIA field crews obtain observations and measurements of additional variables related to the health of forest ecosystems.

These three phases provide the basic framework for the program's structure.

The documentation emphasizes current, nationally consistent FIA practices where they exist. Examples include the glossary of terms, the plot configuration, sampling design and protocols, condition classes and mapped plots, computed variables, and estimation procedures. The documentation also notes aspects of the FIA Program in which regional options are not only practiced but are deemed necessary (e.g., the regional approaches to stratification). Although it is beyond the scope of this documentation to provide detail on all regional differences, such differences and options have been noted where essential to the topics discussed.

After completion of the first draft of this manuscript, an external editor coordinated technical reviews by subject-matter and statistical specialists from the university, government, and forest industry sectors. The editor solicited nine reviewers who participated in a double-blind review process (i.e., the identities of the authors and reviewers were not known to each other). After several rounds of reviews and revisions, the editor gave written approval to release this documentation as a peer-reviewed publication.

Finally, documentation of the Enhanced FIA Program is an ongoing, dynamic process. Additional documentation of other aspects of Enhanced FIA is currently planned. This includes such topics as Phase 3 estimation, model-based updating procedures, data quality assurance and quality control, long-term program direction, and more detail regarding regional differences. Regional options and enhancements to the prescribed national procedures discussed herein are currently documented at the discretion of the regional programs. Although an important focus of the Enhanced FIA Program is national consistency, it is through regional enhancements that innovations and new technologies will be introduced. As these demonstrate their value, stabilize, and become generally accepted, many will be incorporated into Enhanced FIA, and additional documentation will ensue.

# 1 The Enhanced Forest Inventory and Analysis Program

Ronald E. McRoberts<sup>1</sup>

## 1.1 Purpose and Scope of the Documentation

The Agricultural Research, Extension, and Education Reform Act of 1998 (Public Law 105–185), also known as the 1998 Farm Bill, prescribed conceptual changes in approaches to forest inventories conducted by the Forest Inventory and Analysis (FIA) Program of the U.S. Department of Agriculture (USDA) Forest Service. Realization of these conceptual changes required development and implementation of technical changes, some of which are substantial. The underlying purposes for documenting the conceptual and technical features of the resulting Enhanced FIA Program are fourfold:

1. To ensure a common understanding and practice among the regional FIA programs
2. To facilitate further development of the national core program, including FIA's National Information Management System (NIMS)
3. To provide a defensible statistical basis for the sampling and estimation components of the program
4. To promote credibility with users and stakeholders

The primary intended audience for this version of the documentation is the national FIA Program itself. Nevertheless, our users and stakeholders will also find it useful for understanding FIA methods and, as the Enhanced FIA Program matures and as internal issues are resolved, documentation specifically intended for external use will be published. To the extent possible at the present time, this documentation addresses the full range of conceptual issues, technical details, and statistical techniques for sample-based estimation.

---

<sup>1</sup> Ronald E. McRoberts, Mathematical Statistician, USDA Forest Service, North Central Research Station, St. Paul, MN 55108.

## 1.2 Overview of the Enhanced Forest Inventory and Analysis Program

### 1.2.1 Historical Perspective

Forest inventories conducted under the auspices of the national FIA Program historically have been commodity oriented, with emphasis on estimating the area and volume of the Nation's timber supply. These statewide inventories typically have been conducted on productive **timberland**<sup>2</sup> but not on **other forest lands**, not on **reserved forest lands**, and depending on the region, not on National Forest System lands. The design and implementation of FIA inventories have been the responsibility of the five regional FIA Programs that administer them; and **plot** configurations, sampling designs, measurement protocols, analytical techniques, and reporting standards frequently have been tailored to regional requirements. Those inventories were characterized as **periodic surveys** because field crews were concentrated in one or two States until the measurement of all plots was completed. States were selected for inventories on a rotating basis with time intervals between inventories for the same State ranging from 6 to 18 years (Gillespie 1999). The plot measurement component of periodic inventories for any given State required from 1 to 4 years, depending on the magnitude of the resource, the number of plots to be measured, and the number of field crews. The analysis component required an additional 2 to 5 years, during which time FIA staff analyzed the data and then compiled, published, and distributed reports.

The timeliness, quality, and usefulness of estimates obtained from periodic inventories came under scrutiny in the 1980s and 1990s. Estimates obtained from these inventories were degraded by the effects of conducting them over multiple years, while the bias and uncertainty of estimates increased over time due to factors such as change in land use, tree growth, tree mortality, and removals between inventories. The periodic nature of these inventories made consistent interstate estimation difficult, even within regions, while interregion estimation was even more difficult due to varying plot configurations, sampling designs, and measurement protocols. These factors, when compounded by the lack of measurements on all forested lands, caused national compilations to depend on a variety of ad hoc techniques. Finally, the environmental and forest ecosystem health interests of groups challenging the commodity focus of FIA inventories were difficult to address using only traditional FIA measurements.

---

<sup>2</sup> The first use of a glossary term in each chapter is in bold face.

FIA clients recognized the deficiencies inherent in these regional, periodic inventories. They registered their dissatisfaction and proposed solutions. Concerns related to **cycle length** led to proposals to increase the sampling intensity, reduce cycle lengths, and conduct midcycle updates. Clients also advocated consistency across regions and measurement of all vegetation on all forested lands. Various solutions were proposed to resolve these issues, but most were expensive to implement and represented only a piecemeal approach to dealing with problems inherent in the periodic inventories.

### **1.2.2 Forest Health Monitoring**

In response to user concerns regarding the health of forest ecosystems, the Forest Health Monitoring (FHM) Program was established in 1990. It was an independent, cooperative effort among multiple State and Federal Agencies focused on assessing and monitoring the health and sustainability of the Nation's forests using nationally standardized inventory procedures. The FHM Program consists of four primary activities. In the first—detection monitoring—field crews measure selected biotic and abiotic features of forests called “indicators” during a baseline period. The same features are remeasured at regular intervals to identify changes associated with natural forest succession and ecosystem disturbances. In the second activity—evaluation monitoring—teams of ecologists, entomologists, hydrologists, pathologists, silviculturists, and others conduct intensive field sampling and provide combined interpretations when the causes of detected changes are unknown. In the third activity—intensive site monitoring—long-term research is conducted on watershed-sized sites that have diverse forest types and biomes typical of those found in the United States. In the fourth activity—research on monitoring techniques—researchers focus on developing and refining indicator measurements to improve the efficiency and reliability of data collection and analysis. Together, these four FHM activities permit predictions of where and how future ecosystems might change under various environmental and management conditions.

### **1.2.3 Annual Inventories**

The impetus for the transition from regional, periodic inventories to nationally consistent, annual inventories came from two pilot studies in the 1990s and the reports of two Blue Ribbon Panels. In 1990, the North Central Research Station began a pilot study with the objective of producing annual, statewide inventory estimates that were no more costly and no less precise than those obtained from periodic inventories in the year of their completion. The 1992 report of the first Blue Ribbon Panel (BRP-I) recommended a nationally consistent approach to the collection, analysis, and reporting of forest inventory data (American Forest Council 1992). The first step toward

this goal was a directive issued in 1995 by the USDA Forest Service, Deputy Chief for Research instructing FIA to adopt the FHM plot configuration as a national replacement for the various regional plot configurations.<sup>3</sup> In the mid-1990s, the Southern Research Station began a second pilot study that featured annual inventories augmented with State support and based on a 5-year measurement **cycle**. The cumulative effect of the North Central and Southern Research Stations' pilot studies, the report of BRP-I, and the report of the 1998 Blue Ribbon Panel (American Forest and Paper Association 1998), which affirmed the recommendations of BRP-I, was passage of the 1998 Farm Bill [Agricultural Research, Extension, and Education Reform Act of 1998 (Public Law 105–185)].

The 1998 Farm Bill directed the Secretary of Agriculture to produce a strategic plan for forest inventory with several features—an annual forest inventory program; State reports every 5 years; a set of **prescribed core variables**, standards, and definitions; and integration of the ground sampling components of the FIA and FHM Programs. This legislation, together with the coalescing of the two pilot studies and national cooperation in standardizing inventories, resulted in an annual forest inventory program, designated Enhanced FIA, with identifiably new features:

1. A nationally consistent plot with four fixed-radius **subplots**
2. A systematic national sampling design for all lands
3. A complete, systematic, annual sample of each State
4. Reporting of data or data summaries within 6 months of completion of designated proportions of plot measurements
5. Provision for several estimators to combine data from multiple **panels**, some of which incorporate updating techniques
6. State inventory reports every 5 years
7. Integration of the FIA field component and the ground sampling component of the FHM detection monitoring activity

Implementation of the last feature was facilitated by the 1995 USDA Forest Service directive instructing FIA to adopt the FHM plot configuration, and takes advantage of efficiencies gained by consolidating the field components of these two forest inventories.

---

<sup>3</sup> SESCO, Jerry A. 1995. Letter dated March 30 to Station Directors and WO Research Staff Directors. On file with: FIA National Program Leader, USDA Forest Service, Forest Inventory and Analysis, Rosslyn Plaza, 1620 North Kent Street, Arlington, VA 22209.

## 1.2.4 National Consistency

Concern among FIA clients regarding the lack of program consistency was comparable to their concerns about the lack of timely FIA estimates and the need to monitor forest ecosystem health and sustainability. As a means of emphasizing national consistency in the Enhanced FIA Program, its technical aspects are described using the Ends-Ways-Means strategic planning model. Ends are the criteria that must be satisfied for the program to be characterized as nationally consistent; Ways are the procedures and protocols that lead to achieving the Ends; and Means are the resources that are committed to the effort. The following discussion focuses on selecting a set of Ends that lead to national consistency, and identifying technical Ways to achieve them. Some Ends require prescribed Ways to achieve them, while others require flexibility in the selection of Ways. The guiding principle is that Ways are to be facilitating, not inhibiting, and that they are to be prescribed only where necessary to achieve the Ends. Thus, the Ends leading to national consistency are achieved, while creativity and innovation to further enhance the FIA Program are encouraged.

The Enhanced FIA Program is described in terms of six Ends:

- End 1: a standard set of variables with nationally consistent meanings and measurements
- End 2: field inventories of all forested lands
- End 3: consistent estimation:
  - (a) the ability to obtain estimates for areas larger than the single county level for which FIA usually reports estimates
  - (b) the ability to obtain data and estimates for user-defined applications and areas of interest
- End 4: national precision guidelines
- End 5: consistent reporting and data distribution
- End 6: credibility with users and stakeholders

These Ends describe the major foci of the Enhanced FIA Program and provide direction for methodological research.

To ensure that the 6 Ends are achieved, 10 Ways have been prescribed. To achieve End 1, Ways 1 and 2 have been prescribed:

- Way 1: a national set of prescribed core variables with a national field manual that prescribes measurement procedures and protocols for each variable

Way 2: a national plot configuration

Core variables are identified in the national field manual and are measured on each plot in every State by field crews from each of the five regional FIA programs. Additional **enhanced prescribed core variables** and **nonprescribed optional variables** may be included, but no prescribed core variable may be deleted. Agreement on measurement protocols and procedures has required compromise among representatives of the five regional FIA programs, but that has led to a broad national consensus.

To achieve End 2, Way 3 has been prescribed:

Way 3: a national sampling design

The national sampling design emerged as a result of its development and implementation in one regional FIA program and its subsequent acceptance and implementation in two others. Both the national sampling design and the national plot configuration represent broad consensus, and both are described in greater detail in chapters 2 and 3.

Achieving the two components of End 3 is facilitated by the three Ways already described. However, to fully achieve End 3a, a fourth Way also has been prescribed:

Way 4: estimation using standardized formulae for sample-based estimators

End 3b was deemed necessary to accommodate the large number of non-FIA researchers, both within and outside the USDA Forest Service, who seek to use FIA data for their own applications and areas of interest. To achieve End 3b, two additional Ways have been prescribed:

Way 5: a national FIA database with core standards and user-friendly public access

Way 6: a national information management system

Ways 5 and 6 are necessary to make user friendly FIA data available to the general public while also yielding estimates that are consistent with those presented by the FIA Program. In addition, Ways 5 and 6 greatly facilitate Way 4, in particular, and estimation within the FIA Program, in general.

Compliance with the national FIA Program's precision guidelines (U.S. Department of Agriculture Forest Service 1970) associated with End 4 requires flexibility in prescribing Ways. The current guidelines primarily relate to estimates of forest area and inventory volume, and are formulated quantitatively as:



$$\frac{[Var(\hat{Y})]^{0.5}}{\hat{Y}} \left[ \frac{\hat{Y}}{S} \right]^{0.5} \leq PREC$$

where

$\hat{Y}$  = the estimate of the **attribute** of interest

$Var(\hat{Y})$  = the estimate of the variance of  $\hat{Y}$

$S$  = a scaling factor—one million acres for area estimates and one billion cubic feet for volume estimates

$PREC$  = the target precision per  $S$  units, which is 0.03 for area estimates or 0.05 for volume estimates in the Eastern United States and 0.10 in the Western United States

Guidelines for the precision of estimates of other attributes may be considered on a case-by-case basis as the need arises.

Budgetary constraints and natural variability among plots prohibit sample sizes sufficient to satisfy precision guidelines, unless the estimation process is enhanced using ancillary data. With FIA's traditional sample-based estimation, enhancement has been achieved via **stratified estimation** using remotely sensed data as the basis for **stratification**. However, regional differences in species diversity, topography, forest management practices, and other factors may require qualitatively different approaches to stratification to achieve End 4. Thus, no Ways regarding stratification are prescribed other than that the stratifications should be statistically defensible and feasible for incorporation into NIMS. However, assuming that the historical levels of sampling variability and the benefits of stratified estimation would continue, the national sampling intensity in terms of number of plots per unit area was selected so that, on average across much of the Nation, compliance with the precision guidelines would be achieved.

End 5 reflects FIA's response to users and stakeholders who desire consistency and temporal compatibility in cross-State and cross-region estimates of prescribed core variables.

Two Ways are prescribed to ensure that End 5 is achieved:

Way 7: a nationally consistent set of tables of estimates of prescribed core variables

Way 8: publication of statewide tables of estimates of prescribed core variables at 5-year intervals

Finally, achieving End 6, credibility with users, will require that the FIA Program not only develop and implement a nationally consistent program, but that technical details of the program be transparent and subject to stakeholder and public scrutiny. Thus, two additional Ways have been prescribed:

- Way 9: the technical aspects of the FIA Program, including procedures, protocols, and techniques, are documented
- Way 10: the technical documentation is peer-reviewed and published for general access

### 1.3 Overview of the Documentation

The documentation is divided into chapters, of which the following sections are brief overviews.

#### 1.3.1 Sampling Frame

Chapter 2 describes in detail the Enhanced FIA Program's three phases in the context of sample-based estimation. **Phase 1** is designed to produce stratifications of land area in the **population** of interest to reduce variance in the estimates. It entails the use of ancillary data, including remotely sensed imagery in the form of aerial photography and/or satellite imagery, to stratify the land area in a population of interest and to assign plots to **strata**. In **Phase 2**, field crews visit permanent ground plots and measure the traditional suite of FIA variables. The Phase 2 sample is based on a national array of approximately 6,000-acre hexagons containing one permanent ground plot each. It is designated the Federal base sample. In **Phase 3**, field crews measure additional variables related to the health of forest ecosystems. The Phase 3 sample comprises a 1/16th subset of the Phase 2 plots, resulting in a sampling intensity of one plot per approximately 96,000 acres. Because Phase 3 plots (previously denoted FHM plots) are also Phase 2 plots, they include all measurements made on Phase 2 plots, plus measurement of the FHM biotic and abiotic features associated with forest and ecosystem health. The chapter 2 documentation describes the three phases in greater detail, as well as the genesis of the national sampling frame and its theoretical basis.

#### 1.3.2 Plot Design

Chapter 3 describes the history, rationale, and configuration of Phase 2 and Phase 3 ground plots, as well as explanations of protocols used for mapping (or partitioning) plots by **condition class**. The basic FIA area and tree **attributes** are described, including those measured in the field and those computed from field measurements. Special situations that influence data collection and processing also are addressed.

### 1.3.3 Sample-Based Estimators

Chapter 4 focuses on the sample-based approach to estimation and describes the estimators used for area and forest attribute totals under assumptions for simple random estimation, stratified estimation, and **double sampling for stratification**. Variance estimators are provided so that sampling errors may be calculated for each cell in every output table. Estimators are derived to accommodate sample plots that contain multiple land uses and/or straddle **population boundaries**. **Ratio-of-means** estimators are described so that estimates may be computed on per-unit-area, per-tree, and per-stand bases. Methods are also described for estimating the **components of change** between measurements such as growth, mortality, and removals.

### 1.3.4 Combining Data for Multiple Panels

Chapter 5 provides a brief overview of considerations when selecting techniques to obtain multiple panel estimates of forest attributes for the required 5-year reports. Several strategies are noted including **moving average** estimation, a temporally indifferent method that ignores the multiyear nature of the multiple panel data, and model-based updating. Special attention is given to the moving average and the weighted moving average methods. As of the date of this publication, no national default estimator had been selected.

### 1.3.5 Notation and Glossary

Chapter 6 provides a reference for the consistent mathematical and statistical notation used throughout this documentation, and chapter 7 provides a comprehensive glossary of terms and expressions.

### 1.3.6 Web-Based Supplementary Documentation

The supplementary documents referenced in this manuscript are posted on the Web site [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

The purpose of these documents is to provide details about the algorithms, equations, and other specifics of the national FIA Program that are too technical for the chapter discussions. These documents are posted on the Internet because they are dynamic. They are currently in various stages of completion, and periodically will be updated and revised to accommodate changes in protocol and demand for technical detail. More information may be added to this Web site in the future as unresolved and new issues are presented and addressed.

## 1.4 Literature Cited

- American Forest Council. 1992. Report of the Blue Ribbon Panel on Forest Inventory and Analysis. Arlington, VA: American Forest Council. 14 p.
- American Forest and Paper Association. 1998. Second report of the Blue Ribbon Panel on Forest Inventory and Analysis. Arlington, VA: American Forest and Paper Association. 17 p.
- Gillespie, A.J.R. 1999. Rationale for a national annual forest inventory program. *Journal of Forestry*. 97: 16-20.
- U.S. Department of Agriculture Forest Service. 1970. Operational procedures. First report of the Blue Ribbon Panel on forest inventory and analysis. Forest Service Handb. 4809.11. Washington, DC: 11: 1-1. Chapter 10.

# 2 The Forest Inventory and Analysis Sampling Frame

Gregory A. Reams, William D. Smith,  
Mark H. Hansen, William A. Bechtold,  
Francis A. Roesch, and Gretchen G. Moisen<sup>1</sup>

## 2.1 Overview of Forest Inventory and Analysis Sampling Design

### 2.1.1 Forest Inventory and Analysis Populations

For purposes of sampling and estimation, Forest Inventory and Analysis (FIA) subdivides the total land area of the United States into mutually exclusive **populations**<sup>2</sup> and **subpopulations**. Populations are usually defined by county **boundaries** or by public ownerships that may or may not cross county boundaries (e.g., national forests). In cases where the sample size for individual counties is insufficient, groups of counties may be combined into a **super-county** to form a single population with adequate sample size. Based on user request, counties occasionally are split into subpopulations to accommodate enumerated (known) acreages supplied by public agencies (e.g., National Forest System and The Bureau of Land Management). This is done to ensure that FIA totals match the county-level acreages reported by the requesting agencies. Each FIA population and subpopulation has a known number of **plots** and a known area of land, obtained from the U.S. Census Bureau, from which population estimates are derived. Each is sampled and processed as a separate entity, so estimates of grand totals and their variances for groups of populations and subpopulations are additive. For example, State-level estimates are obtained by totaling the estimates from all populations and subpopulations bounded by the State.

Note that FIA estimation is based on land area which excludes **census water** (4.5 acres in size and at least 200 feet wide). Census water is thus subtracted

---

<sup>1</sup> Gregory A. Reams, FIA National Program Manager, USDA Forest Service, 1601 North Kent Street, Arlington, VA 22209; William D. Smith, Research Quantitative Ecologist, USDA Forest Service, Southern Research Station, Research Triangle Park, NC 27709; Mark H. Hansen, Research Forester, USDA Forest Service, North Central Research Station, St. Paul, MN 55108; William A. Bechtold, Research Forester, USDA Forest Service, Southern Research Station, Asheville, NC 28802; Francis A. Roesch, Mathematical Statistician, USDA Forest Service, Southern Research Station, Asheville, NC 28802; and Gretchen G. Moisen, Mathematical Statistician, USDA Forest Service, Rocky Mountain Research Station, Ogden, UT 84401.

<sup>2</sup> First use of a glossary term in each chapter is in bold face.

from the total area of land and water at the beginning of the estimation process. We anticipate that estimation eventually will be based on total area, including census water, when precise digitized census water boundaries become available from the U.S. Census Bureau. The capacity to tabulate the area of census water from digitized data will improve FIA's ability to generate forest statistics for user-defined polygons.

FIA engages in three types of sampling—**Phase 1**, **Phase 2**, and **Phase 3**. All three types are performed for each population or subpopulation of interest. The sample points associated with each phase are subsets of the previous phase, but from the descriptions that follow it should be clear that this is not intended to be an application of classical three-phase sampling.

### 2.1.2 Phase 1

Phase 1 is designed to reduce variance through **stratification**. Although the details have differed among FIA units, all have used **double sampling for stratification** since aerial photography became widespread in the 1950s. For a given population of interest, a supplemental grid of Phase 1 sample points (i.e., photo points or satellite **pixels**) is superimposed over the Phase 2 sample points, such that Phase 2 can be viewed as a subset of Phase 1. All sample points, both Phase 1 and 2, are then assigned to **strata** based on their classification from remote-sensing imagery.

The remote sensing medium selected to accomplish Phase 1 is left to the discretion of the FIA regions, but satellite imagery is replacing aerial photography as discussed in section 2.1.2.1. The number of photo points or pixels classified and the frequency of Phase 1 sampling are regional decisions. The number of strata, if any, and the definitions of these strata also are left to the discretion of the regions, but most recognize a minimum of two—**forest** and **nonforest**. Factors influencing the details of Phase 1 stratification include the homogeneity of the population; the timing, availability, and cost of remote imagery; and the availability of personnel available to perform the work. Nationally **prescribed core** methodology related to Phase 1 is limited to:

- If available, acquisition of new imagery at least once for each new cycle of panel measurements (e.g., every 5 years for States on a 5-year panel system)
- Application of the double sampling for stratification estimation techniques described in chapter 4

The only difference in estimation techniques associated with the details of Phase 1 sampling is whether the strata weights are treated as estimated or

known. With wall-to-wall satellite classification, strata weights are known. When photo points or satellite pixels are sampled, strata weights are treated as estimated. Even when wall-to-wall imagery is available, pixels are often sampled to ease the computational burdens of working with high-resolution imagery and multiple data layers. FIA units thus can choose between two approaches to stratified sampling:

1. The double sampling for stratification approach used when strata weights are estimated (sec. 4.2.2)
2. The **stratified estimation** approach used when the weights are known (sec. 4.2.1)

The choice is largely based on local efficiencies, but most units are moving toward the latter as satellite imagery replaces aerial photography.

### 2.1.2.1 Aerial Photography vs. Satellite Imagery

Since the 1950s and prior to satellite imagery, FIA used aerial photography to assign plots to strata at Phase 1 and in some cases to estimate forest area (Bickford 1952). The intensity of photo plots has varied over time and among FIA regions, ranging from one photo point per 230 acres in the South and Northeast, to one point per 248 acres in the Rocky Mountain region (assuming that each photo point has a radius of 50 feet). Photo points usually were established by overlaying a systematic grid on 1:40,000 black-and-white aerial photos, although other scales (e.g., 1:20,000) and media (e.g., color infrared) also have been used. Decisions regarding scale and media have been based on availability, timing, price, and coverage. A good historical overview of FIA Phase 1 sampling is provided by Frayer and Furnival (1999).

All FIA units have begun replacing photo-point classifications with satellite-based (pixel) classifications of land use. The primary source of FIA Phase 1 satellite imagery is the Landsat Thematic Mapper (TM) series. The TM sensor has a repeat cycle of 16 days and a swath width of 115 miles. This multispectral sensor has 6 nonthermal bands—three in the visible, one in the near infrared, and two in the midinfrared, all with 100-foot resolution. TM is the remote sensing platform of choice due to:

- Historic and planned continuity of wall-to-wall land cover classifications
- Moderate spatial and spectral resolution of the sensor
- A scale of resolution appropriate for matching ground-truth units to pixels for the computation of standard error estimates

Because TM satellite imagery has been used more often and with more success for forest assessments than any other satellite sensor, there is a

vast body of literature on classification algorithms using various analytical approaches including **unsupervised**, **supervised**, and various **hybrid classification** approaches. From this work it is known that land cover classification accuracies > 80 percent are difficult to achieve with satellite imagery, which is notably less than the 95-percent accuracies attained by experienced FIA photo interpreters. This difference in accuracy should not be disregarded because the gap is even wider when classification is attempted beyond forest and nonforest cover types. It also means that aerial photography will remain a useful tool even after the transition to satellite imagery is complete. Although photo classification is demonstrably more accurate than satellite classification on a point-by-point basis, satellite classification has several distinct advantages when compared to aerial photos (Wayman and others 2001, Wynne and others 1999):

- Satellite classification accuracy is expected to improve as classification algorithms and ancillary ground-truth data improve.
- The gain in precision from 80 percent accuracy with wall-to-wall satellite coverage offsets the 95 percent accuracy attained from a comparatively small sample of photo points.
- Satellite-derived thematic maps usually are generated from objective and consistent processes (although some human interpretation is needed to label classified cover types and other land features).
- Satellite imagery provides an opportunity for more frequent updates.
- Spectral change detection is relatively easy and particularly useful when analyzing change associated with timber removals, as well as catastrophic disturbances.
- Spatially explicit enumerations of the entire landscape (i.e., maps) can be automated.

The FIA Program has national precision standards of 3 percent per million acres of **timberland** and 5 percent per billion cubic feet of growing-stock volume in the Eastern United States. Recent Phase 1 applications using TM-based classifications for the National Land Cover Data (NLCD) indicate that FIA can come very close to meeting the precision standards (Hansen 2001). With a forest/nonforest stratification based on the most recent NLCD, the FIA North Central Research Station region produced sampling errors ranging from 2.83 to 3.71 percent per million acres of timberlands for four States (Indiana, Iowa, Minnesota, and Missouri). For these same States, sampling errors ranged from 6.03 to 6.73 percent per billion cubic feet of growing-stock volume.



Improved Phase 1 techniques offer an efficient opportunity to meet or exceed the stated precision goals, and the FIA Program plans to continue investigating alternative methods for improved stratification. TM image classification can be improved by auxiliary information from other sources (see Web page <http://www.fs.fed.us/ne/rsb>). Potentially useful auxiliary information currently under study includes the Gap Analysis Program, the Moderate Resolution Imaging Spectroradiometer, the Natural Resource Information System, topographic and ecological data layers, and high-resolution low-altitude photography and satellite images. Use of high-resolution imagery, with either visual or digital interpretation, may increase FIA's ability to classify highly fragmented landscapes more precisely, but it may not be cost effective for more general applications. One alternative to stratification, which has been used with moderate success in Alaska, is use of regression methods to correlate plot data with individual pixel values. This allows pixels to be summed to provide estimates for the area of interest and is actively being investigated for small-area applications. However, because it can become quite cumbersome operationally, this technique is not ready for general application at the State level (Scott 1986).

### 2.1.3 Phase 2

Phase 2 relates to FIA's network of permanent ground plots, which has a spatial sampling intensity of approximately one plot per 6,000 acres. Field crews install, monument, and measure ground plots if any portion of a plot contains a **forest land** use. Detailed field remeasurements of forest plots are repeated at regular intervals as long as the plots remain in forest (note that protocols for handling plots that cannot be sampled due to access restrictions are discussed in section 3.4.3). Nonforest plots are assigned a **nonforest** use code (**nonforest land**, census water, or **noncensus water**) and checked at each scheduled inventory for potential reversion to forest. Forest plots are installed if reversion occurs. Note that neither LANDSAT imagery nor stratification is used in the decision to visit a ground plot. Field crews physically visit all ground plots that have any chance of being forested. However, to avoid unnecessary costs in extensive areas of nonforest or in inner cities, some FIA regions use recent aerial photography to identify and assign land uses to plots that obviously have no chance of being forested. Phase 2 plots are assigned to strata based on their classification at Phase 1, which may or may not be consistent with the land use assigned by field crews at Phase 2. Discrepancies can result from misclassification or from changes since the imagery was obtained, and are factored into the estimation process described in chapter 4.

FIA's ongoing remeasurement process is designed to accommodate changes in protocols and plot design over time. This is accomplished by remeasuring

the previous plot installation at each new inventory. For example, as FIA moves from horizontal point samples to fixed-area **mapped plots**, the program is being careful to preserve change estimates that span the transition period. To complete this calibration over time, the horizontal point samples are remeasured for change estimates when new mapped plots are installed. The mapped plots will be remeasured at future visits.

To be classified as forest, an area must be at least 10 percent stocked with tree species, at least 1 acre in size, and at least 120 feet wide. **Stocking** protocols are further discussed in section 3.3.2.2.1, as well as the supplementary document “National Algorithms for Stocking Class, Stand Size Class, and Forest Type” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm). Previously forested land that is not stocked, and which has not been developed to another land use, is still considered forest (e.g., clearcuts). Land that meets the minimum requirements for forest, but is developed for a nonforest land use, is considered nonforest (e.g., city parks or campgrounds).

Discussions are underway within the FIA Program about presenting both use and cover estimates of land area. Researchers involved with remote sensing have been exploring the development of landcover estimates based on percent tree-crown cover, but this work has yet to be used operationally. The most significant impediment to estimating **attributes** of interest by cover class is the cost of increasing the scope of FIA such that field crews are required to measure trees and detailed area attributes on land that is simply classified as nonforest under current protocols.

#### 2.1.4 Phase 3

Phase 3 plots include all of the features of Phase 2, plus additional measurements such as tree-crown assessment, soil sampling, lichen communities, understory vegetation structure, ozone **bioindicators**, and **down woody material**. Every 16<sup>th</sup> Phase 2 plot is also a Phase 3 plot, so Phase 3 sample intensity is approximately one plot per 96,000 acres. All Phase 3 plots are combined with Phase 2 plots for Phase 2-based estimations of attributes common to both plot types (i.e., double sampling for stratification applies). Attributes unique to Phase 3 are estimated directly from the Phase 3 subset. Use of Phase 1 stratification and Phase 2 samples to enhance the estimation of attributes unique to Phase 3 is currently being studied. Because Phase 3 is a subsample of Phase 2, the use of double sampling with regression is being considered for estimating some Phase 3 attributes. Detailed estimation procedures for attributes specific to Phase 3 will be provided in future documentation.

A summary of the general attributes associated with Phases 1, 2, and 3 is provided in table 2.1.

**Table 2.1—Summary of general attributes associated with FIA Phase 1, Phase 2, and Phase 3 sampling**

Attribute	Phase 1	Phase 2	Phase 3
Sample type	Photo point or satellite pixel	Ground plot, subset of Phase 1	Ground plot, subset of Phase 2
Sample configuration	Point or pixel	Cluster of four 1/300-acre micro-plots, four 1/24-acre subplots, and optional four 1/4-acre macroplots	Same as Phase 2 <sup>a</sup>
Purpose	Stratification <sup>b</sup> of the landscape for the purpose of variance reduction	Samples FIA traditional attributes of interest, primarily related to tree species of all sizes	Samples FIA traditional attributes of interest, <sup>c</sup> plus additional attributes associated with forest health
Tessellation method	Supplemental regional grid superimposed over the population of interest <sup>d</sup>	Systematic national hexagonal cell grid	Systematic national hexagonal cell grid (subset of Phase 2 grid)
Base-grid intensity	At the discretion of each FIA unit	One plot per every 6,000-acre hexagonal cell	One plot per every 1/16 6,000-acre hexagonal cell (i.e., one per 96,000 acres)

FIA = Forest Inventory and Analysis.

<sup>a</sup> Note that additional sample designs associated with forest health indicators (to be described in a future document) are superimposed over the Phase 2 sample configuration on Phase 3 plots.

<sup>b</sup> Most FIA units recognize a minimum of two strata—forest and nonforest. Census water is currently subtracted from the total area prior to any stratification or estimation.

<sup>c</sup> Phase 3 plots also double as Phase 2 plots for estimation of attributes associated with Phase 2. Phase 3 plots are unique only when used to estimate attributes unique to Phase 3.

<sup>d</sup> Regional Phase 1 grids are systematic grids of varying density (up to wall-to-wall) that are not necessarily linked to the national hexagonal grid. The only prescribed requirement is that Phase 2 plot centers must be a subset of the Phase 1 points.

## 2.2 Development of the Phase 2/Phase 3 Sampling Frame

With passage of the 1998 Farm Bill [The Agricultural Research, Extension, and Education Reform Act of 1998 (Public Law 105–185)], Congress directed major changes in the way FIA conducts inventories. This legislation prescribes an annual inventory where a proportion of plots in each State must be measured every year. The switch from a variety of regional **periodic surveys** to a nationally standardized annual inventory required FIA to implement a new sampling frame. The 1998 law also precipitated the integration of Forest Health Monitoring (FHM) with FIA. When the two programs merged, FHM had already implemented a national sampling frame with a plot network that was systematically divided into panels measured on an annual basis. A national sampling grid was viewed as a more convenient and consistent method for tessellating the landscape and populating the sample frame than the county-by-county approach previously used by most FIA units; especially because county boundaries occasionally change, and counties may be divided into different subpopulations at different times. When the two programs integrated it was decided to build on the existing FHM sampling frame, where the FHM panels were redefined as subpanels of the larger FIA plot network.

### 2.2.1 Hexagonal Sampling Frame

The U.S. Environmental Protection Agency’s Environmental Monitoring and Assessment Program originally developed the sampling frame used by FHM (Overton and others 1990, White and others 1992). This framework is actually based on a triangular grid, but the cells surrounding each point on a triangular grid form a hexagonal shape, so the sampling frame can also be viewed as a network of hexagonal cells. The hexagonal frame was projected onto the landscape by centering a large base hexagon over the continental United States (fig. 2.1). Similar hexagons were then extended from the base hexagon to tessellate the planet. The result is described as a truncated icosahedron (White and others 1992) made up of 20 hexagon faces and 12 pentagon faces, which give the framework a “soccer ball” appearance (fig. 2.2). To achieve the desired sample intensity for FHM, the base hexagon was then subdivided into approximately 28,000 smaller hexagons with centers about 17 miles apart. To avoid alignment with property boundaries that follow the public land survey system, the hexagon configuration was randomly offset from cardinal directions. To accommodate the sampling intensity and frequency desired by FIA, the hexagonal sampling frame was further modified as described in the next section 2.2.2.

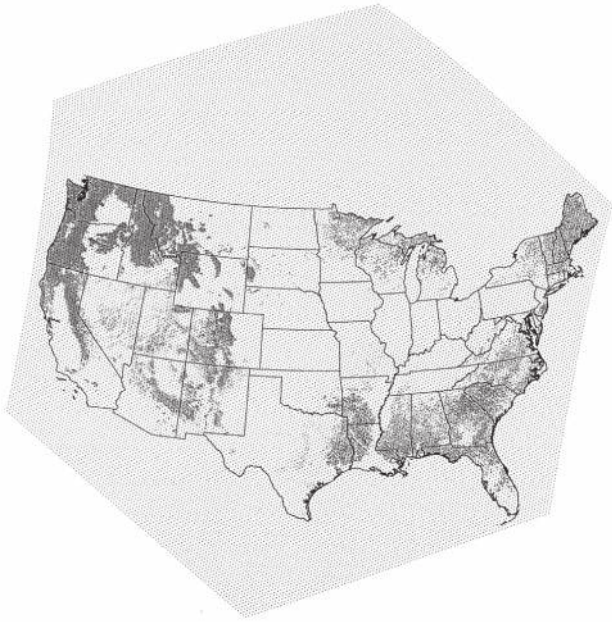


Figure 2.1—Base hexagon positioned over the conterminous United States.

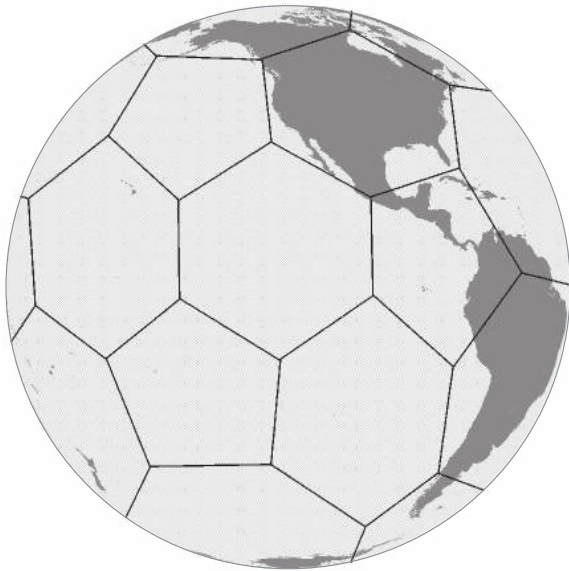


Figure 2.2—Truncated icosahedron made up of 20 hexagon faces and 12 pentagon faces.

## 2.2.2 Division of the Sampling Frame into Panels

The original FHM sampling frame conveniently accommodated 3-, 4-, 7-, 9-, and 11-panel rotations, and multiples of these. In other words, the centers of the hexagons in a given panel formed a triangular pattern of equidistant points for these panel rotations. Figure 2.3 shows the triangular pattern of the four-panel system originally used for FHM. FIA requires a five-panel rotation to accommodate the measurement frequency mandated by the 1998 Farm Bill (20 percent per year). Unfortunately, the five-panel system does not conform to an equidistant triangular configuration. To satisfy the desired sampling frequency for FIA, the program used a parallelogram-shaped pattern of hexagon centers to assign hexagons to panels (fig. 2.4). Although hexagons within a given panel are no longer

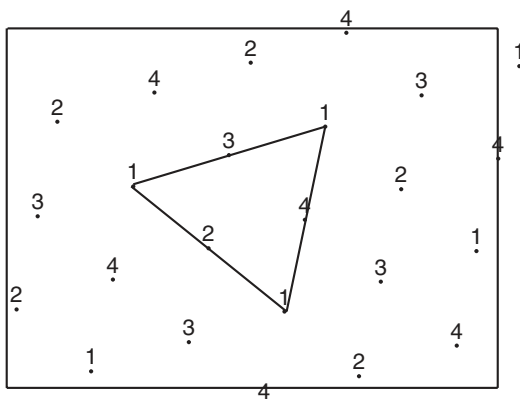


Figure 2.3—Hexagon panel assignments illustrating the triangular pattern of a four-panel rotation.

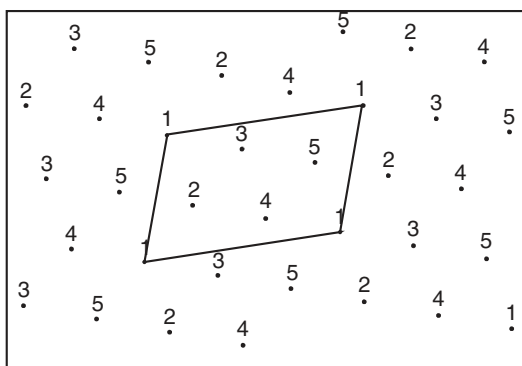


Figure 2.4—Hexagon panel assignments illustrating the parallelogram pattern of a five-panel rotation.

equidistant, the parallelogram configuration provides the most uniform spatial arrangement possible for five-panel rotations, and multiples thereof.

To satisfy the desired sampling intensity for FIA, FHM hexagons (approximately 160,000 acres) were subdivided in 27 smaller hexagons, resulting in hexagons of 5,937 acres. Figure 2.5 [from Brand and others (2000)] shows the spatial arrangement of the FIA hexagons relative to the original FHM hexagons. Figure 2.6 from Brand and others (2000) details the systematic coverage resulting from the panel assignment process. Again, note the parallelogram pattern that results from connecting the hexagons in any given panel.

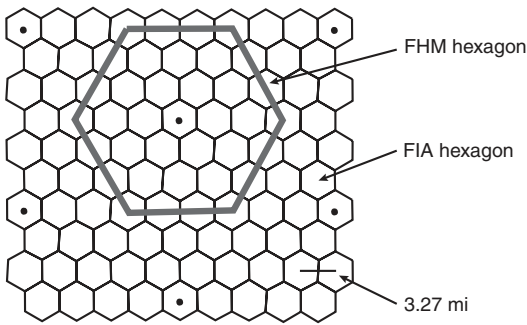


Figure 2.5—The FIA hexagon lattice (each black dot is at the center of an FHM hexagon).

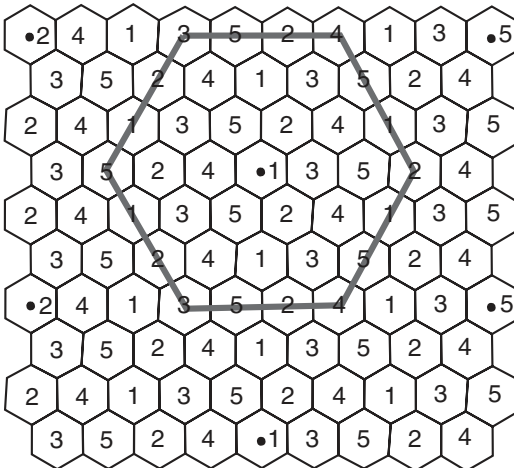


Figure 2.6—Assignment of hexagons to one of five panels (shown by number).

### **2.2.3 Populating the Sampling Frame**

Once the FIA hexagon frame was established, and hexagons were assigned to panels, one field plot was allotted to each hexagon as follows:

1. If the FIA hexagon contained an FHM plot, the existing FHM plot was selected.
2. If not, then an existing FIA plot was selected.
3. If there were multiple FIA plots in the hexagon, the one closest to hexagon center was selected and the others were abandoned.
4. If there were no FHM or FIA plots in the hexagon, a new sample location was selected based on a random azimuth and distance from hexagon center.

Because FHM plots originally were measured on a four-panel annual system, some additional constraints were necessary when reassigning those plots to FIA panels. The following constraints resulted in minor perturbations of the parallelogram pattern, which were accepted so that historic measurement sequences and cohorts would remain unchanged:

1. No existing FHM plots were dropped.
2. FHM panels retained their historic measurement sequence, so colocated Phase 2 and Phase 3 plots kept their preexisting FHM panel number (this constraint was relaxed in States that had intensified FHM sampling frames).
3. The subset of Phase 3 plots was increased by 20 percent to accommodate a fifth panel (to preserve the same annual sampling intensity established under the four-panel FHM system).

Additional technical details related to panel assignments and population of the sampling frame are available in the supplementary document “The Hexagon/Panel System for Selecting FIA Plots Under an Annual Inventory” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

### **2.2.4 Deviations from the Five-Panel Annual System**

Panels and their associated plots, are scheduled for measurement based on their panel assignment. Panels are measured in sequence, one at a time. After all five panels have been completed the process is repeated. Ideally, exactly one panel per year would be completed in each State. However, the realities of budgetary constraints and logistical problems (e.g., forest fires) prevent some States from being inventoried at the prescribed rate of one



panel per year. This situation can lead to “panel creep,” where the length of time to complete an inventory panel exceeds 1 year. This situation is most common in States that do not have additional resources to move from the federally financed 7-year **cycle length** to a 5-year cycle length.

The concept of subpaneling the five-panel system is an alternative that will be implemented if the measurement cycle becomes too protracted. A number of subpaneling schemes could be developed to yield timelier inventory results and still retain uniform spatial coverage. For example, the FIA Western Pacific Northwest and Rocky Mountain regions are now funded to collect data on a 10-year cycle. To accommodate the funding disparity, those two regions are using the five-panel design where each panel has been divided into two subpanels (each with complete spatial coverage); one subpanel is scheduled for measurement each year. This is analogous to a 10-panel system.

### **2.2.5 Theoretical Basis for the FIA Sampling Frame**

It is clear from the previous discussion that the current FIA sampling frame was forged from a variety of preexisting regional FIA and national FHM sampling frames. The goal of this approach was to maintain linkage with historical data to the extent possible (to preserve temporal consistency and continuity for trend estimation), and to smooth the transition from the numerous variations of periodic systems. This approach relates to established sampling theory in a number of ways. In this section we give one general description of the joint distribution resulting from the marriage of various periodic designs with the common annual design.

Sample plots are linked to a systematic triangular grid with time-interpenetrating panels. In a triangular grid, the cells surrounding each grid point are hexagonal and the grid is systematically divided into panels. Assuming one panel per year is measured for  $T$  consecutive years, then every  $T$  years the panel measurement sequence begins again. If panel 1 were measured in 1998, it also would be measured in  $1998+T$ ,  $1998+2T$ , and so on. Panel 2 would be measured in 1999,  $1999+T$ , and  $1999+2T$ . Of the numerous methods that might have been used to choose existing sample-point locations for retention in the new design, the preferred option was to assign existing plots to the nearest triangular grid point (i.e., hexagon center). Extra plots in each grid cell (hexagon) were subsequently deleted, and new plots were randomly added to empty grid cells. Although the methodology does not produce a regular grid of sample points at a fine scale (i.e., grid-point intersections); it does at a coarse scale (i.e., grid-point cells). This feature has the advantage of masking the exact location of ground plots, which is required by law.

Assume that the sample points from the entire collection of periodic inventories constitute a random sample from the infinite set of points contained within the boundaries of the United States. Panel assignments are made to hexagons in a systematic fashion. Although panel assignments are not random with respect to the triangular grid, they are random with respect to the underlying area-based population, due to the random establishment of the grid combined with a scale-dependent assumption of randomly arranged population elements. The entire sampling frame is a three-dimensional cube—two dimensions incorporate the land area and the third represents time.

Assume that random tessellation of the land area into identical, mutually exclusive hexagons ( $H$ ) defines two samples:

$S_1$  = A selection from the previous randomly chosen plot locations, where each chosen point is assigned to the hexagon from within which it was selected. The individual element of  $S_1$  for each hexagon  $j$  is denoted  $s_{1j}$ .

$S_2$  = A sample of random points resulting from a random tessellation of the land area into identical, mutually exclusive hexagons. A random point is chosen for the sample from the infinite set of points within each hexagon. The individual element of  $S_2$  for each hexagon  $j$  is denoted  $s_{2j}$ .

Let:

$$I_j = \begin{cases} 1 & \text{iff a previous sample point was selected from within hexagon } j \\ 0 & \text{otherwise} \end{cases}$$

Then a single sample point is chosen for each hexagon  $j$  such that

$$s_j = I_j s_{1j} + (1 - I_j) s_{2j}, \quad j = 1, \dots, N.$$

$I_j$  randomly selects an element from 1 of 2 random samples. We also assume that  $H$  randomly assigns one of the  $T$  panels to each sample element.

Adding the dimension of time to the two dimensions that constitute the land area of the United States produces a population which is a three-dimensional cube. The primary **sampling unit** (PSU) is a series of line segments, linear in time. That is, when the time dimension is collapsed down onto the area dimensions, each series of line segments collectively appear as a single point on the area. When the area dimensions are collapsed down to the time dimension, each line segment within a series is of an approximate length of 1 day. Individual segments occur every  $T \pm 1$  years within each series.

Within a sufficiently small segment of time, all points within the land area dimensions of the volume common to each area segment created by the overlapping inclusion areas of all possible subsets of trees occurring on the land area [in the sense of Roesch and others (1993)] could be viewed as a temporally specific sampling unit. However, because these segments change over time, the PSU appears as a point in the temporally specific land area dimensions of the cube. That is, if we slice the population into, say, annual volumes, such that land area constitutes the base and time constitutes the height, and then view the annual subpopulation from the top, we'll observe a set of  $1/T$  points on the land area base. Each point exists within a temporally specific inclusion area for a specific subset of trees. The temporal slices actually could be of any height; however, the thinner the slice, the smaller the sample per land area of interest. The wider the slice, the fuzzier the segment boundaries, as the subsets of trees change. For most of FIA's purposes, annual slices will constitute the minimum height that forms a reasonable compromise between temporal specificity and land-area generality.

The plot measurements provide support to the point (line) from which they were chosen. The plot measurements for an individual sample point (sample line) are multiplied by the inverse of the land area (land area/temporal volume) upon which they were based, resulting in a value per acre for each sample point (a value per acre per temporal unit for each sample line). The collection of sample points per area of interest (sample lines per area/temporal volume) contributes to the estimates for that area (area/time volume). The sampling units have known inclusion probabilities, which are used in the estimation equations.

This discussion supports the detailed estimation procedures described in chapter 4, which assumes that the FIA **systematic sample** for Phase 2 and 3 can be treated as a **simple random sample**. The systematic coverage provided by the hexagonal grid eliminates the clumping of samples and loss of precision that would occur with a purely random assignment of plots. The use of the hexagonal grid also increases the chances of sampling infrequent forest types. Given that plot locations are randomly assigned within hexagons, the chance of the sample network coinciding with a systematic land feature or spatially periodic phenomenon is greatly reduced. Research on the periodicity concern indicates that the hypothetical has little chance of occurring (Milne 1959). Cochran (1977) provides the following justification for the use of simple random-sample-based estimates for systematic samples:

Consider all  $N!$  finite populations which are formed by the  $N!$  permutations and any set of numbers  $y_1, y_2, \dots, y_N$ . Then, on the average over these finite populations,  $E(V_{sy}) = V_{ran}$ . Note that  $V_{ran}$  is the same for all permutations.

Madow and Madow (1944) state that if the order of the items in a specific population can be regarded as drawn at random from the  $N!$  permutations, systematic sampling is (on average) equivalent to simple random sampling.

### 2.3 Literature Cited

- Bickford, C.A. 1952. The sampling design used in the forest survey of the Northeast. *Journal of Forestry*. 50(4): 290-293.
- Brand, G.J.; Nelson, M.D.; Wendt, D.G.; Nimerfro, K.K. 2000. The hexagon/panel system for selecting FIA plots under an annual forest inventory. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C., eds. Proceedings of the first annual forest inventory and analysis symposium. Gen. Tech. Rep. NC-213. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 8-13.
- Cochran, W.G. 1977. Sampling techniques. 3<sup>rd</sup> ed. New York: John Wiley. 428 p.
- Frayser, W.E.; Furnival, G. M. 1999. Forest survey sampling designs: a history. *Journal of Forestry*. 97(12): 4-10.
- Hansen, M.H. 2001. Remote sensing precision requirements for FIA estimation. In: Reams, G.A.; McRoberts, R.E.; Van Deusen, P.C., eds. Proceedings of the second annual forest inventory and analysis symposium. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station: 43-51.
- Madow, W.G.; Madow, L.H. 1944. On the theory of systematic sampling. *Annals of Mathematical Statistics*. 15: 1-24.
- Milne, A. 1959. The centric systematic area sample treated as a random sample. *Biometrics*. 15: 270-297.
- Overton, W.S.; White, D.; Stevens, D.L. 1990. Design report for EMAP (environmental monitoring and assessment program). EPA/600/3-91/053. Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development. 43 p.
- Roesch, F.A., Jr.; Green, E.J.; Scott, C.T. 1993. An alternative view of forest sampling. *Survey Methodology*. 19(2): 199-204.
- Scott, C.T. 1986. An evaluation of sampling with partial replacement. In: Oderwald, R.G.; Burkhart, H.E.; Burk, T.E., eds. Use of auxiliary information in natural resource inventories. SAF 86-01. Blacksburg, VA: Society of American Foresters: 74-79.
- Wayman, J.P.; Wynne, R.H.; Scrivani, J.A.; Reams, G.A. 2001. Landsat TM-bases forest area estimation using iterative guided spectral class rejection. *Photogrammetric Engineering and Remote Sensing*. 67(10): 1155-1166.
- White, D.; Kimerling, A.J.; Overton, W.S. 1992. Cartographic and geometric components of a global sampling design for environmental monitoring. *Cartography and Geographic Information Systems*. 19: 5-22.
- Wynne, R.H.; Oderwald, R.G.; Reams, G.A.; Scrivani, J.A. 1999. Optical remote sensing for forest area estimation. *Journal of Forestry*. 98(5): 31-36.

# 3 The Forest Inventory and Analysis Plot Design

William A. Bechtold and Charles T. Scott<sup>1</sup>

This chapter describes the **prescribed core plot**<sup>2</sup> design currently used by Forest Inventory and Analysis (FIA) for **Phase 2** and **Phase 3** ground sampling. FIA ground plots relate to the sampling frame discussed in the previous chapter as follows:

- One plot has been randomly located within each 6,000-acre hexagon.
- Each plot has been assigned to one of five **panels** as described in section 2.2.2.
- Each plot has been designated Phase 2 or Phase 3 based on the rules outlined in section 2.2.3.
- The center point of each plot constitutes the primary **sampling unit** (PSU) described in section 2.2.5.
- The area and vegetation data gathered on each plot serve to support and quantify the information associated with each PSU.

The plot design characteristics, field protocols, and calculations discussed in this chapter are intended to provide additional background to the estimation procedures outlined in chapter 4; some explanation of the most important derived values produced by FIA; and discussion of sampling and estimation issues associated with the plot design. More detail is provided in the referenced supplementary documentation, and a complete description of all field measurements can be found in the FIA Phase 2 field guide<sup>3</sup> available on the Web site <http://fia.fs.fed.us/library.htm#manuals>.

All of the measurements described herein likewise apply to Phase 3, because Phase 3 plots are a subset of Phase 2. Additional detailed measurements associated with Phase 3 forest health “indicators” (e.g., tree crowns, soils,

---

<sup>1</sup> William A. Bechtold, Research Forester, USDA Forest Service, Southern Research Station, Asheville, NC 28802; and Charles T. Scott, FIA Program Manager, USDA Forest Service, Northeastern Research Station, Newtown Square, PA 19073.

<sup>2</sup> First use of a glossary term in each chapter is in bold face.

<sup>3</sup> U.S. Department of Agriculture Forest Service. 2004. Forest inventory and analysis national core field guide: field data collection procedures for phase 2 plots, Version 2.0. [Not paged]. Vol. 1. Intern. Rep. On file with: USDA Forest Service, Forest Inventory and Analysis, Rosslyn Plaza, 1620 North Kent Street, Arlington, VA 22209.

lichens, downed woody material, and understory vegetation) are described in the Phase 3 field guides, also posted on the above Web site. Some of these indicators have specialized plot designs superimposed over the basic Phase 2 plot design. Specialized designs unique to Phase 3 indicators are beyond the scope of this manuscript but are covered in detail in the field guides. Separate manuscripts that document specialized Phase 3 plot designs and associated estimation procedures have been written and are currently in review.

The prescribed core plot design originated with the Forest Health Monitoring (FHM) Program in 1990 and was selected as the national standard for FIA in 1995. Shortly thereafter, FIA began converting its various regional plot designs to the national standard. Most FIA units had been using 5- or 10-point clusters of prism points arranged in a variety of patterns. While all FIA units change to the national plot design, previously installed plot configurations are being remeasured to provide estimates of change (growth, removals, and mortality). As earlier designs are remeasured to estimate change, the new design is simultaneously installed to yield current inventory estimates and to provide the basis for change estimation upon future remeasurement.

### 3.1 Overview of the FIA Plot Design

Phase 2 and Phase 3 ground plots are clusters of four points arranged such that point 1 is central, with points 2 through 4 located 120 feet from point 1 at azimuths of 0, 120, and 240 degrees (fig. 3.1). Each point in the cluster is surrounded by a 24-foot fixed-radius **subplot** where trees 5.0 inches diameter at breast height (**d.b.h.**) and larger are measured. All four subplots total approximately 1/6 acre. Each subplot contains a 6.8-foot fixed-radius **microplot** where saplings (1.0 to 4.9 inches d.b.h.) and seedlings are measured. All four microplots total approximately 1/75 acre. Microplots are offset from subplot centers (12.0 feet at an azimuth of 90 degrees) to minimize trampling. Each subplot is surrounded by a **prescribed optional** 58.9-foot fixed-radius **macroplot**, which can be useful for sampling rare occurrences such as large trees (e.g., 40.0 inches d.b.h. and greater) or mortality. Macroplots encompass subplots, as well as the additional area from 24.0 to 58.9 feet beyond the subplot circumference. All four macroplots total approximately 1 acre. When used together, microplots, subplots, and macroplots constitute a **tri-areal plot** design for sampling trees in three different tree-diameter ranges. In regions where the optional macroplots are not used, the plot design is **bi-areal**.

For attributes such as large trees that are always measured within subplots, whether or not macroplots are utilized, it is sometimes useful to describe

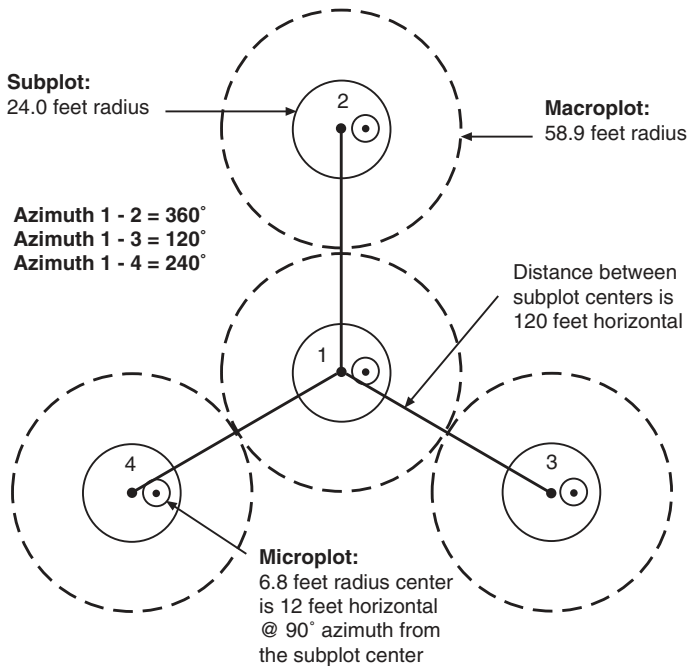


Figure 3.1—FIA plot design.

the trees between 24.0 and 58.9 feet on tri-areal plots as being located in an annular ring. For example, this description avoids the need to redefine the range of diameters sampled on the subplot if the diameter threshold for sampling large trees is changed. Theoretically, distinctions between macroplot trees in the annular and inner portions of the macroplot do not have any implications for the estimation procedures described in chapter 4.

In addition to the trees measured on FIA plots, data are also gathered about the area or setting in which the trees are located. Area classifications are particularly useful for partitioning the **forest** into meaningful categories (i.e., **domains**) for analysis. Some of these area **attributes** are measured (e.g., percent slope), some are assigned by definition (e.g., ownership group), and some are computed from tree data (e.g., percent **stocking**).

To enable division of the forest into various domains of interest for analytical purposes, it is important that the tree data recorded on these plots are properly associated with the area classifications. To accomplish this, plots are mapped by **condition class**. Field crews assign an arbitrary number to the first condition class encountered on a plot. This number is then defined by a series of predetermined discrete variables attached to it (i.e., land use,

forest type, stand size, regeneration status, tree density, stand origin, ownership group, and disturbance history). Additional conditions are identified if there is a distinct change in any of the condition-class variables on the plot. Further details are provided in section 3.2.

Sometimes a plot straddles two or more distinct condition classes. **Boundaries** of condition classes can bisect the subplots, or may occur between subplots. When they bisect a subplot, condition boundaries are mapped using two or three azimuths as described in section 3.1.2. Similarly, microplots and macroplots, if used, also are mapped. So for each ground plot, the microplot, subplot, and macroplot area in each condition class are known, as is the location and condition class of every tree tallied. Because FIA primarily is concerned with the classification and monitoring of **forest land**, no tree data are recorded for **nonforest land** uses.

At first glance, an unwieldy number of condition-class permutations seems likely at the regional scale, especially because condition classes from the same dataset must be processed in different combinations from one inventory summary table to the next depending on the domain of interest. However, most plots have only one or two condition classes and data summarizations are easily managed with **indicator functions** as described in chapter 4.

### 3.1.1 Motivation Behind the FIA Plot Design

FIA has historically used cluster plots, primarily because they reduce between-plot variance and, therefore, the total number of plots necessary to achieve a given accuracy standard (Scott 1993). The 4-point cluster was chosen because experience with FHM and FIA pilot studies showed that on average, crews can complete one 4-point cluster plot per day. It is conceivable that the number of points comprising the national standard may be revised in the future if the field workload changes such that a different number of cluster points is more efficient.

The **mapped-plot** feature of the design arose from the need to correctly match tree data with area classifications when plots straddle multiple conditions. Before the advent of mapped plots, some FIA units moved plots into a single, uniform condition. This generated a bias by altering the selection probabilities of trees, especially those near condition edges (Williams and others 1996). Other FIA units addressed the problem by prohibiting the movement of plots, but then blended area data from distinctly different conditions. Although unbiased for area and volume totals, this procedure resulted in domain misclassifications. For example, a plot that straddled a pure oak forest type and a pure pine forest type might be classified as a mixed oak/pine forest type.



In 1991, FIA project leaders and inventory specialists met with a panel of university and forest industry biometricians to discuss the problem and explore a variety of potential alternatives. A committee was subsequently appointed by the FIA project leaders to review the alternatives and recommend a solution (Hahn and others 1995). The tri-areal, fixed-radius mapped design was ultimately selected because it solved both the bias and classification problems, it had the flexibility needed to satisfy a growing FIA customer base, and it permitted greater use of the data for such nontraditional purposes as forest health monitoring.

The tri-areal design is a departure from the polyareal plot sampling (pps) approach (Husch and others 1982) originally implemented by FIA in the early 1960s. The pps design is more efficient for sampling timber and estimating volume, but fixed-radius plots add versatility by preserving information about the spatial relationships among trees. Fixed-sized plots also are more compatible with mapped-plot designs, because the area that must be mapped is constant. In addition, pps sampling in conjunction with mapped plots often leads to a situation where the full range of tree sizes is not sampled in all conditions (Scott and Bechtold 1995). This has negative consequences that are difficult to correct when area attributes such as forest type and stand size are computed from the tree data. This problem rarely occurs with the tri-areal design and is much easier to manage when it does occur.

### 3.1.2 Field Protocols for Mapping Plots

Field crews specify and define (if not previously defined) the condition class at each subplot center, as described in section 3.2. If a subplot straddles two or more conditions, they then specify the condition class that contrasts with the condition at subplot center. Standing at subplot center and facing the **contrasting condition**, they record the two azimuths where the condition-class boundary crosses the subplot perimeter. A third azimuth (with a distance) is permissible if the boundary contains a sharp curve or a corner (fig. 3.2). All trees tallied are then assigned to the condition class in which they occur. Horizontal distance and azimuth to the center of each tree are recorded for remeasurement purposes and to establish spatial relationships among the sampled trees. Microplots (and macroplots, if used) are mapped in a similar fashion. It is not necessary to match boundaries at the edge of each plot type, so microplots, subplots, and macroplots all are mapped independently. Microplot, subplot, and macroplot areas in each condition class are computed from field measurements when the data are processed (sec. 3.3.3).

Field crews are trained to recognize and map only those boundaries that are distinct and obvious. A variety of logic checks are programmed into field

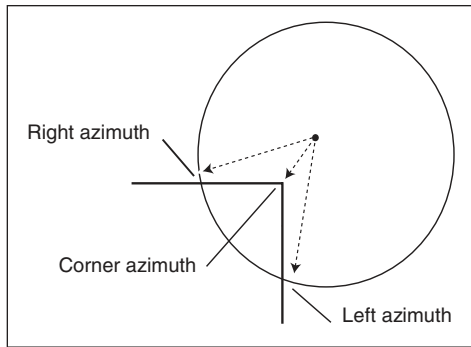


Figure 3.2—Using azimuths to reference a boundary to a subplot center.

data recorders to check boundary data for errors and to verify that the tree location and condition observations are consistent with condition-class and boundary observations (Scott and Bechtold 1995). Most condition-class boundaries do not require mapping because they occur between subplots, or are indistinct. The frequency of mapped subplots depends on the homogeneity of the landscape, but is typically < 5 percent of the total number of subplots in a region. There is some concern over the level of detail to which plots are mapped, as well as the repeatability of boundary recognition and placement. We expect that mapping protocols will be evaluated as part of the FIA quality control program and adjusted as necessary.

### 3.1.3 Differences in Mapping Forest and Nonforest Plots

Any plot that intersects with a forest land use is designated as a forest plot. Otherwise, the plot is classified as **nonforest**, **census water**, **noncensus water**, or (if inaccessible) nonsampled. In order to reduce the field costs associated with nonforest plots, mapping is initiated only in the presence of accessible forest. For those plots that contain no accessible forest, only the condition status (e.g., nonforest, census water) at the center of subplot 1 is recorded, and that condition is assigned to the entire plot.

Similarly, forest plots may include individual subplots or macroplots where no accessible forest is present. There, only the condition status at subplot center is recorded and that condition is assigned to the entire subplot. Thus, when two or more condition classes occur within a subplot or macroplot, boundaries between them are mapped only when one or both conditions are classified as accessible forest. Boundaries between adjacent nonforest, census water, and nonsampled conditions are mapped only on subplots containing some accessible forest.

## 3.2 Condition Classification Based on Direct Field Observation

Some condition-class variables trigger the mandatory identification of a distinct condition class in the field; others are ancillary, recorded only after a new condition class is recognized. Both mandatory and ancillary condition-class variables typically are used to specify the domains of interest (e.g., a specific forest type and physiographic class) for which **population** estimates are generated for some attribute of interest (e.g., acres or volume).

### 3.2.1 Discrete Variables That Trigger Recognition of a Unique Condition Class

There are seven discrete condition-class variables that require recognition of a unique condition in the field: condition status (land use), reserve status, owner group, regeneration status, tree density, forest type, and stand size. If one of these variables changes during plot measurement, a new condition is defined and mapped if necessary. All are subjective field calls, some of which have guidelines and/or subsampling protocols to assist crews make a determination.

### 3.2.2 Ancillary Condition-Class Variables

Ancillary condition-class variables are recorded in the field whenever unique conditions are defined, but these variables do not trigger the recognition of new condition classes. These ancillary variables, obtained for all forested conditions include detailed owner class, private owner industrial status, artificial regeneration species, stand age, disturbance history, treatment history, and physiographic class.

#### 3.2.2.1 Site Index Equations and Site Productivity Classes

In addition to ancillary condition-class variables, one or more site trees are measured in each unique condition class if qualified trees are available. If there is no reason to suspect a difference in site quality among condition classes, the same site tree(s) may be used for multiple conditions on a plot. Site trees are used in determining site quality (i.e., the capacity of forest land to grow trees). A **site index** or **site productivity class** is thus associated with each forested condition class.

Site index is the average total height that the dominant and codominant trees in fully stocked, even-aged stands will obtain at key ages (Husch and others 1982). Site productivity class, also known as site class or yield capacity, is the maximum mean annual increment in cubic feet per acre that can be expected in fully stocked, natural, even-aged stands. Using regionally specific equations, site index is computed as a function of the stand age and the

average height of dominant or codominant trees as determined from the species, d.b.h., and total height of qualifying site trees on or near the plot. The following selection criteria are preferred for site trees: acceptable species, free of damage, dominant or codominant crown position, between 15 and 120 years old, and at least 5.0 inches d.b.h. Resulting site-index values are then applied to site-productivity equations, or look-up tables, to determine the maximum mean annual increment for a given condition class. The site index and site class equation references used by the various FIA units are Brickel (1970), Clendenen (1977), and Edminster and others (1985) in the Interior West; Carmean and others (1989) in the North Central region;<sup>4</sup> Scott and Voorhis (1986) in the Northeast; Hanson and others (2002) in the Pacific Northwest; and Vissage and Greer in the South.<sup>5</sup>

### **3.3 Computed Attributes**

In contrast to attributes that are observed and classified directly in the field such as the condition-class variables in the previous sections; others are computed. Some attributes are computed at the tree level; some are computed at the condition-class or plot level. Computed attributes can be the measures upon which population estimates are based (e.g., acres, numbers of individuals, and volume) or they might be used to specify domains. An example of the latter would be placement of continuous variables into discrete classes (e.g., volume-per-acre class) in order to estimate the area in each class. Some attributes are computed in addition to being observed directly (i.e., forest type and stand size), so these have both field-assigned and computed values.

#### **3.3.1 Computed Tree-Level Attributes**

Tree-level attributes are variables associated with the individual trees tallied on FIA ground plots. Expressions of tree volume and weight are among the most basic statistics reported by FIA. The functions used to compute these values are typically statistical models developed or calibrated by State or region. The most commonly reported volume and weight statistics are described in the supplementary document “FIA Volume Calculations” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

---

<sup>4</sup> For site-index to site-class conversions used by the North Central FIA unit see the supplementary document “Site Productivity Assignment for the North Central FIA unit” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm). [Date accessed: December 9, 2004].

<sup>5</sup> Vissage, John S.; Greer, Travis R., Jr. Site class and site index – two estimates of site quality for the Southern Research Station Forest Inventory and Analysis unit. 12 p. Internal document. On file with: J.S. Vissage, USDA Forest Service, North Central Forest Experiment Station, 1992 Folwell Ave., St. Paul, MN 55105.

### 3.3.2 Condition Classification Based on Computed Attributes

Computed condition-class attributes further describe condition classes and the domains they represent. These attributes usually are derived from tree-level (vegetation) data. They can be discrete (e.g., forest type) or continuous (e.g., percent stocking).

#### 3.3.2.1 Condition-Level Per-Acre Ratios

For modeling, or for summarizations of area data by discrete per-acre classes, tree-level statistics (e.g., volume or basal area) can be used to compute per-acre ratios for individual plots, or per-acre ratios for specific condition classes within a plot. Condition-level ratios are computed by summing the tree-level attribute of interest (e.g., basal area) for all trees in the condition class and then dividing by the area of the plot in that condition:

$$y_{ik} = \frac{\sum_j^4 \sum_t y_{ijkt}}{\sum_j^4 a_{ojk}} \quad (3.1)$$

where

$y_{ijkt}$  = the attribute of interest associated with tree  $t$  on microplot, subplot, or macroplot  $j$  covering condition  $k$  on plot  $i$

$a_{ojk}$  = area used to observe the attribute of interest (microplot, subplot, or macroplot  $j$ ) covering condition  $k$  on plot  $i$

When combining subplot and microplot values, condition-level ratios are:

$$y_{ik} = \frac{\sum_j^4 \sum_t y_{ijkt}}{\sum_j^4 a_{ojk}} + \frac{\sum_j^4 \sum_t y'_{ijkt}}{\sum_j^4 a'_{ojk}} \quad (3.2)$$

where

$y'_{ijkt}$  = the attribute of interest associated with tree  $t$  on microplot  $j$  covering condition  $k$  on plot  $i$

$a'_{ojk}$  = area of condition  $k$  on microplot  $j$  on plot  $i$

### 3.3.2.2 Condition-Level Attributes Based on Stocking

Stocking class, stand-size class, and forest type are important condition-level attributes of interest that are calculated from the stocking contributions of individual trees. Although forest type and stand-size class are assigned in the field, these attributes are also computed from the tree tally when the field data are processed. The primary purpose of the field assignments is to delineate the unique condition classes encountered on each plot, and to supply an alternative value in case the area sampled is too small to derive a calculated value (i.e., calculated values for forest type and stand size are not produced for conditions that occupy < 25 percent of a plot).

#### 3.3.2.2.1 Algorithm to Assign Stocking Values to Individual Trees

FIA uses a complicated algorithm to assign stocking percentages to individual trees, the details of which are provided in the supplementary document “National Algorithms for Stocking Class, Stand-Size Class, and Forest Type” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm). A simplified explanation follows to summarize the concept. The algorithm is based on “A-line” values, which are described by Gingrich (1967) as the number of trees per acre where average maximum stocking occurs in undisturbed stands. A-line values are negatively correlated with mean stand diameter, so they vary by diameter class as well as by species. The formula used to assign percent-stocking values to individual tally trees is:

$$s_{ijkt} = (100) \frac{1}{A_{max}} \quad (3.3)$$

where

$s_{ijkt}$  = the percent stocking assigned to tree  $t$  on (microplot, subplot, or macroplot  $j$ ) covering condition  $k$  on plot  $i$

$A_{max}$  = the A-line value (trees per acre) associated with the species and diameter class of tree  $t$

Once  $s_{ijkt}$  is assigned, division by  $a_{oijk}$  adjusts the sample tree to a per-acre basis as shown in equation 3.1 when  $s_{ijkt}$  is substituted for  $y_{ijkt}$ . Equation 3.1 then yields the total percent stocking for condition class  $k$  on plot  $i$ . For reporting purposes, condition-level stocking percentages commonly are grouped into the classes listed below:

Stocking class	Class stocking range
	<i>percent</i>
Nonstocked	0 – < 10
Poorly stocked	10 – < 35
Moderately stocked	35 – < 60
Fully stocked	60 – ≤ 100
Overstocked	> 100

### 3.3.2.2 Stand-Size Class of Each Condition Class Based on Stocking Algorithm

Tree-level stocking values also are used to categorize each condition by stand-size class. Each tree is first assigned to one of the following size classes based on its d.b.h.:

Stand-size class	Stand-size class d.b.h. range
Seedling-sapling	d.b.h. < 5.0 inches
Poletimber	5.0 inches ≤ d.b.h. < 9.0 inches for softwoods 5.0 inches ≤ d.b.h. < 11.0 inches for hardwoods
Sawtimber	9.0 inches ≤ d.b.h. for softwoods 11.0 inches ≤ d.b.h. for hardwoods

For a given condition class, stocking values for each tree (i.e.,  $s_{ijkt} / a_{oijk}$ ) are then summed across all trees in each stand-size class, and for all stand-size classes combined. Stand-size class is then assigned on the basis of which of the following stocking requirements is satisfied first:

Stand-size class	Stocking requirement
Nonstocked	Total stocking across all size classes < 10 percent
Seedling-sapling	Seedling-sapling stocking > 50 percent of total stocking
Poletimber	Poletimber stocking > sawtimber stocking
Sawtimber	Poletimber stocking ≤ sawtimber stocking

### 3.3.2.2.3 Forest Type of Each Condition Class Based on Stocking Algorithm

Tree-level stocking values also are used to categorize each condition by forest type. The forest type assignment algorithm is quite complicated and still undergoing evaluation. Details are provided in the supplementary document “National Algorithms for Stocking Class, Stand-Size Class, and Forest Type” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

### 3.3.3 Calculations of Area in Each Condition Class

#### 3.3.3.1 Mathematical Functions

Geometric and trigonometric functions can be used to calculate the area within each condition class as specified by Scott and Bechtold (1995).

#### 3.3.3.2 Computer Simulation

Mathematical functions are useful for calculating areas at any one time, or when performing data recorder logic checks; but they can be unwieldy for some applications—particularly the calculation of **area change matrices** between inventories. Change matrices are necessary to quantify land-use and condition-class change and to enable the partitioning of growth, removals, and mortality by condition-class attributes at either the initial or terminal inventory of a measurement **cycle**.

Change matrices are produced by overlaying a computer-generated map of each subplot at time  $t$  (the previous inventory) with a similar map of the same subplot at time  $t+1$  (the current inventory) (fig. 3.3). The area of the

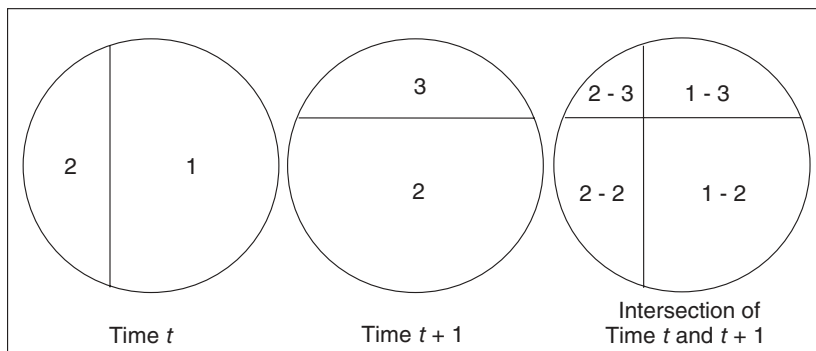


Figure 3.3—Condition-class change matrix between two points in time.



intersection of all combinations of initial and terminal condition classes is then calculated by using a computer to count dots on an electronically generated grid superimposed onto the intersected maps (Bechtold and others 2003). For each cell of the intersection matrix, an observation is created that includes area percent, as well as all of the condition-class variables at both time  $t$  and time  $t+1$ . Subplot-level change data are then combined to produce plot-level matrices. Similar matrices are produced for microplots and macroplots. Note that it is not necessary for field crews to retain specific condition-class numbers over time. Numbers assigned to conditions remain arbitrary and are defined by the series of condition-class variables attached to them.

### 3.4 Special Cases

#### 3.4.1 Population Boundaries

Plots (and portions of plots) are assigned to the population in which they are located with indicator functions. Plots (and portions of plots) in the population of interest are then pooled to compute population estimates as described in chapter 4.

FIA sampling protocols recognize four types of population boundaries: (1) County, (2) National, (3) Federal agency (e.g., National Forest System, Bureau of Land Management), and (4) Census water.

All except county boundaries are currently mapped in the field. County boundaries are often not observable in the field, so whole plots are assigned to the county in which the center of subplot 1 is located. The inability to recognize population boundaries is not considered a problem because this implies that forest conditions are the same on both sides of the boundary, so no bias is introduced by sampling area outside the population of interest (i.e., the numerators and denominators of **ratio-of-mean** (ROM) estimators are incremented proportionately).

#### 3.4.2 Slivers

**Slivers** are defined as conditions that occupy less than one full subplot and not encountered on any other subplots of a given plot. Slivers have the potential to create data processing or analysis issues, but in most cases the difficulties are minor. Alternative designs that do not require mapping create different or more serious problems such as classification anomalies and tree selection probabilities that result in biased estimators (Hahn and others 1995).

### **3.4.2.1 Slivers and Continuous Variables**

One concern regarding slivers is that extremely small values in the denominators of per-acre ratios might unreasonably inflate estimates of per-acre values. This is not a problem because the ROM estimators prescribed in chapter 4 avoid the use of plot-level ratios. Slivers are pooled with similar conditions from other plots in the calculation of population means and totals. Although users of FIA data should be aware that a sliver may yield an unrealistic mean per-acre ratio for a rare domain in a small population, that possibility is usually of little consequence for standard FIA estimations. A domain that rarely would not be isolated in any standard output and, for reporting purposes, would be pooled with other domains.

Modeling is the only application where slivers and potentially inflated per-acre values are isolated and used as individual observations. The modeler has a variety of options to deal with this problem, such as accepting the increased variance, pooling slivers with other conditions on the same plot, pooling slivers with similar conditions from different plots, or deleting slivers from the analysis.

### **3.4.2.2 Slivers and Classification Variables Based on Tree Data**

Slivers have the potential to inflate per-acre continuous variables that are computed for individual condition classes and then grouped into discrete classes for presentation in summary tables (e.g., area by volume-per-acre class). Such inflated values are rare and never stand alone. They are simply grouped into the highest class presented in the summary table. The most serious consequence is increased within-class variance caused by estimates from plots of different sizes.

Slivers can pose a slightly different problem for computed classification variables that are not per-acre estimates (e.g., forest type). When the tree tally on a given plot falls below a certain threshold, sufficient data may not be available to make an accurate classification. In such cases it is necessary to accept the computed classification at face value, revert to a subjective field classification, or engage in auxiliary sampling to obtain enough field data to compute the classification. The amount of field data required for reliable area classifications depends on the spatial scale of the vegetation upon which the classification is based (Williams and others 2001). FIA is still evaluating the minimal areas required for classifications most commonly computed from tree data, forest type and stand size, so subjective field calls are available in addition to computed values for these. Preliminary analyses suggest that computed values for forest type and stand size are unreliable for conditions that occupy areas smaller than one full subplot.

A related problem involves attributes, usually indices, for which it is important to base the classification on plots of equal size (e.g., species diversity index). This is not a common application of FIA data and there is no prescribed method for handling this situation, but an analyst has the option to adjust the index (e.g., species/area curves), delete partial plots from the analysis, or pool data from other conditions on the plot. It is noteworthy that pooling vegetation data is not an option for plots that contain nonforest land uses, because FIA protocols ignore vegetation on land uses defined as nonforest.

### 3.4.3 Nonsampled Plots and Plot Replacement

For various reasons, some plots (or portions of plots) within a given population cannot be sampled at the time they are scheduled for measurement. Such plots are classified as “nonsampled”, a “nonsampled reason” is assigned, and no additional data are recorded. The magnitude of the problem has not been fully evaluated, but nonsampled plots have the potential to be a significant factor in populations with relatively few forested plots (e.g., Plains States). FIA currently assumes these plots are randomly distributed, so nonsampled plots assume the **strata** means for all estimated values as discussed in chapter 4, which ensures that estimates are produced for the entire population—not just the accessible portion. This approach presumes that conditions on nonsampled plots are missed at random, which may not be valid under some circumstances. More precise methods of assigning attributes to nonsampled plots are being considered, and will be implemented if they yield better results.

Access refusal by landowners is the most common reason for nonsampled plots, but plots occasionally are inaccessible due to hazardous situations encountered by field crews. To avoid altering the sampling network such that it becomes nonrepresentative of the population, inaccessible plots usually are not replaced. However, inaccessible plots may be replaced where nonreplacement causes inadequate sample size, or where there is evidence that replacement results in estimators that are less biased.

Less often, missing data or corrupted plot files are discovered after the field measurements for a panel are completed. When it is impractical to correct the situation, such plots also are classified as nonsampled. They are then resampled at their next scheduled measurement.

Field crews occasionally fail to relocate previously established plots. Upon verification that a plot is truly lost, a replacement is installed at the approximate location of the original, and the lost plot is retired from the panel.

### 3.5 Literature Cited

- Bechtold, W.A.; Heravi, N.E.; Kinkenon, M.E. 2003. A simulation algorithm to approximate the area of mapped forest inventory plots. Gen. Tech. Rep. SRS-67. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 8 p.
- Brickell, J.E. 1970. Equations and computer subroutines for estimating site quality of eight Rocky Mountain species. Res. Pap. INT-75. Ogden, UT: U.S. Department of Agriculture Forest Service, Rocky Mountain Research Station. 22 p.
- Carmean, W.H.; Hahn, J.T.; Jacobs, R.D. 1989. Site index curves for forest species in the Eastern United States. Gen. Tech. Rep. NC-128. St. Paul, MN: U.S. Department of Agriculture Forest Service, North Central Forest Experiment Station. 142 p.
- Clendenen, G.W. 1977. Base-age conversion and site index equations for Engelmann spruce stands in the central and southern Rocky Mountains. Res. Note INT-223. Ogden, UT: U.S. Department of Agriculture Forest Service, Rocky Mountain Research Station. 6 p.
- Edminster, C.B.; Mowrer, T.H.; Shepperd, W.D. 1985. Site index curves for aspen in the central Rocky Mountains. Res. Note RM-453. Ogden, UT: U.S. Department of Agriculture Forest Service, Rocky Mountain Research Station. 4 p.
- Gingrich, S.F. 1967. Measuring and evaluating stocking and stand density in upland hardwood forests in the Central States. *Forest Science*. 1: 38-53. Vol. 13.
- Hahn, J.T.; MacLean, C.D.; Arner, S.L.; Bechtold, W.A. 1995. Procedures to handle inventory cluster plots that straddle two or more conditions. *Forest Science Monograph*. 31: 12-25.
- Hanson, E.J.; Azuma, D.L.; Hiserote, B.A. 2002. Site index equations and mean annual increment equations for Pacific Northwest Research Station Forest Inventory and Analysis inventories. Res. Note PNW-RN-533. Portland, OR: U.S. Department of Agriculture Forest Service, Pacific Northwest Research Station. 24 p.
- Husch, B.; Miller, C.I.; Beers, T.W. 1982. *Forest Mensuration*. 3<sup>rd</sup> ed. New York: John Wiley. 402 p.
- Scott, C.T.; Bechtold, W.A. 1995. Techniques and computations for mapping plot clusters that straddle stand boundaries. *Forest Science Monograph*. 31: 46-61.
- Scott, C.T. 1993. Optimal design of a plot cluster for monitoring. In: Rennolls, K.; Gertner, G., eds. *The optimal design of forest experiments and forest surveys: Proceedings, IUFRO S.4.11 conference*. London, UK: University of Greenwich. 233-242 p.
- Scott, C.T.; Voorhis, N.G. 1986. Northeastern forest survey site index equations and site productivity classes. *Northern Journal of Applied Forestry*. 3: 144-148.
- Williams, M.S.; Schreuder, H.T. Reich, R.M. 1996. The extent of bias caused by substituting points in forest survey: A simulation study. Res. Pap. RM-327. Fort Collins, CO: U.S. Department of Agriculture Forest Service, Rocky Mountain Forest and Range Experiment Station. 7 p.
- Williams, M.S.; Schreuder, H.T.; Czaplowski, R.L. 2001. Accuracy and efficiency of area classifications based on tree tally. *Canadian Journal of Forest Research*. 31: 556-560.

# 4 Sample-Based Estimators Used by the Forest Inventory and Analysis National Information Management System

Charles T. Scott, William A. Bechtold,  
Gregory A. Reams, William D. Smith,  
James A. Westfall, Mark H. Hansen, and  
Gretchen G. Moisen<sup>1</sup>

## 4.1 Panels and Estimation

This chapter outlines **prescribed core**<sup>2</sup> procedures for deriving **population** estimates from **attributes** measured in conjunction with the **Phase 1** and **Phase 2** samples. These estimation procedures also apply to those **Phase 3** attributes in common with Phase 2. Given the sampling frame and **plot** design described in the previous two chapters, many estimation approaches can be applied. In fact, one goal of the overall design is to maximize flexibility, so Forest Inventory and Analysis (FIA) data can be used to address a variety of analytical needs.

Much of the flexibility inherent in the Enhanced FIA design is derived from the way **panels** are combined for analysis. This topic is addressed in chapter 5. For estimations involving a single panel or a **periodic survey**, the approach to estimation would proceed directly as outlined in chapter 4. When the estimation is for some combination of annual panels, then the estimation procedures discussed herein may require modification depending on the method used to combine the panels. Related modifications are discussed in conjunction with the two specific methods for combining panels presented in chapter 5 (secs. 5.2.1 and 5.2.2).

---

<sup>1</sup> Charles T. Scott, FIA Program Manager, USDA Forest Service, Northeastern Research Station, Newtown Square, PA 19073; William A. Bechtold, Research Forester, USDA Forest Service, Southern Research Station, Asheville, NC 28802; Gregory A. Reams, FIA National Program Manager, USDA Forest Service, 1601 North Kent Street, Arlington, VA 22209; William D. Smith, Research Quantitative Ecologist, USDA Forest Service, Southern Research Station, Research Triangle Park, NC 27709; James A. Westfall, Research Forester, USDA Forest Service, Northeastern Research Station, Newtown Square, PA 19073; Mark H. Hansen, Research Forester, USDA Forest Service, North Central Research Station, St. Paul, MN 55108; and Gretchen G. Moisen, Mathematical Statistician, USDA Forest Service, Rocky Mountain Research Station, Ogden, UT 84401.

<sup>2</sup> First use of a glossary term in each chapter is in bold face.

## 4.2 Phase 1

**Stratification** is a statistical tool used to reduce the variance of attributes of interest by partitioning the population into homogenous **strata**, such as **forest** and **nonforest**. It may also involve partitioning out a highly variable but small portion of the population.

To use stratified sampling methods, the strata sizes (weights) must be either known or estimated. As discussed in chapter 2, strata weights commonly are estimated by classifying points on aerial photographs or **pixels** on satellite imagery. Full enumeration of an entire landscape (population) is possible with wall-to-wall satellite imagery, resulting in strata weights that can be treated as known values in the variance formulations. Whether the strata weights are known or estimated, strata classifications (e.g., forest and nonforest) from remote sensing are never perfectly accurate. Although this increases the variance of the resulting population estimates, it does not introduce any bias due to the interaction between Phases 1 and 2 in the estimation process.

### 4.2.1 Satellite Classification and Known Strata Weights

Using wall-to-wall satellite imagery and computer-aided classification, the population can be divided into strata of known size—typically forest, **noncensus water**, nonforest, or combinations of these. In this case the classified imagery divides the total area of the population ( $A_T$ ) into pixels of equal and known size (e.g., 30 m square with Landsat TM), and the classification assigns one of  $H$  strata to each of these pixels. Here, the stratum weight,  $W_h$  ( $h = 1, \dots, H$ ), typically is determined as the number of pixels classified as stratum  $h$  divided by the total number pixels in the population of interest. The weights are known quantities that are fixed by the classification system and selected strata. As used here, “pixel” indicates either an individual pixel, or a non-overlapping block of multiple pixels. Each stratum,  $h$ , then contains  $n_h$  ground plots, each selected with known probability, where the Phase 2 attributes of interest are observed. Note that the strata sample sizes,  $n_h$ , are random because ground-plot locations are chosen prior to stratification.

Satellite classification systems separate the reflectance values from each pixel into a set of  $H$  distinct values (i.e., ranges). Such systems can range from very complicated functions dependant on a number of different reflectance band values and ancillary data layers, to very simple step functions of a single reflectance band. A variety of automated classification schemes are available, but these rarely match FIA stratification requirements. These must usually be modified or additional classifications must be performed to produce strata that are relevant to FIA estimation needs.

When using **classified satellite imagery** for stratification, one must know the location of each plot and each pixel, so that plots and pixels can be linked in the estimation process. FIA assigns each plot to one and only one stratum using the pixel corresponding to plot center. Typically this is done using a global positioning system (GPS) instrument for ground plots and geo-reference information associated with the imagery for pixels. If GPS field data are not available, plot locations can be digitized manually from aerial photos or maps.

Most of the estimation presented here is based on the assumption of equal probability sampling, where all elements of the population have the same probability of being sampled by a ground plot. This assumption can be violated if ground-plot information is used to help classify satellite imagery. Inevitably, ground-plot data will be used in the development of some classification algorithms. Breidt and Opsomer (2004) show that for algorithms based on general linear models, this approach provides valid results and variances. Other types of classification algorithms (e.g., nonparametric) have not been similarly evaluated.

It is also important to remember the importance of the pixel-plot link when changing or considering a change in the stratification algorithm. Two similar algorithms may divide a population into very similar strata that have similar or identical names. The application of a different classification system will not only change the  $W_h$  (even when they have identical names), but also the strata assignments of ground plots.

#### **4.2.2 Aerial Photography and Estimated Strata Weights**

In some cases, especially when using aerial photos, it is not practical or cost effective to divide imagery into strata of known sizes. Strata sizes then may be estimated using photo-plot sampling with manual interpretation, and applying **double sampling for stratification** as opposed to the **stratified estimation** that occurs when strata weights are known. When double sampling for stratification is used in conjunction with satellite imagery or other spatial layers, only a sample of pixels is classified, as opposed to the wall-to-wall estimation described in the previous section. This relieves some of the computational burden of more complex computer-aided classification methods. The loss of precision when such methods are applied to a subset of pixels is minimal when large Phase 1 sample sizes are used (Moisen and Edwards 1999).

Under double sampling for stratification, a set of sample points (i.e., Phase 1 plots) is placed on aerial photo or satellite coverage of the population. Phase 1 plots are then assigned to strata by photo interpretation or computer-aided

classification algorithm. Phase 1 plots are usually placed on a grid, typically every 200 to 300 acres. The Phase 2 and Phase 3 plots usually are a subset of the Phase 1 photo plots. Any Phase 2 and Phase 3 plots that are not must also be classified, but their values are not used in developing the estimates of strata sizes. Here:

$n'$  = total number of Phase 1 plots (pixels) sampled in the population

$n'_h$  = number of Phase 1 plots (pixels) classified as belonging to stratum  $h$   
( $h = 1, \dots, H$ ,  $H$  = total number of strata)

### 4.2.3 Combining Small Strata and Populations

Because sample sizes are random with respect to strata, some may contain too few plots to compute a reliable mean and variance. Experience indicates that a minimum of four Phase 2 plots per stratum are required. If less than four, then similar strata must be collapsed (combined) until the minimum is attained. Because stratification schemes may vary regionally to accommodate resource differences, there are no prescribed rules by which strata are combined.

The minimum-sample-size rule also affects some small populations or populations with little forest (e.g., some counties in the Eastern or Great Plains regions of the United States). Populations with fewer than 12 forested Phase 2 plots should thus be combined with adjoining counties (i.e., populations) forming what is termed as a “**super-county**.” The use of super-counties will be especially important when using only a subset of panels.

## 4.3 Phase 2

This chapter builds on chapter 3 by specifying the population estimators used for calculation of area and other attributes of interest associated with **mapped plots**. Generally, these attributes are summed for each plot (after adjustment to correct for any plots that overlap the population boundary), and then averaged across plots in the stratum. The strata means are then combined using the Phase 1 strata weights to form a population mean. This mean is then expanded to a total using the total land area in the population. This approach to estimation is first described for area attributes, and then other attributes, because the formulas are somewhat different. Examples of the estimation procedures discussed below are provided in the supplementary documents “Examples of FIA Point-in-Time Estimation Procedures for Several Common Cases” and “Examples of FIA Change-Component Estimation Procedures for Several Common Cases” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).



### 4.3.1 Estimation of Area by Stratum by Domain (Row and Column) Attributes

For a given attribute of interest (e.g., forest area), estimates of population totals and **domain** totals (e.g., area by forest type and stand size class) are produced similarly. The only difference is that different restrictions (filters) are placed on each cell of an output table depending on which domain (row or column) the cell is in. Mapped **condition classes** (e.g., forest type) are often used to specify domains, but other attributes also can be used (e.g., species).

Row and column attributes must be discrete categorical variables. Continuous attributes can be converted into discrete categories by dividing them into classes (e.g., diameter classes). The attribute of interest is “summed” into a cell only when it satisfies the row and column requirements (i.e., when it is in the domain of interest). Each plot and tree has an associated **indicator function**,  $\delta_{hijkd}$  (or  $\delta_{hijtd}$ ), which is 1 if the attribute is in the domain  $d$  of interest, or 0 if not. For example, when estimating the area in northern hardwood (row) sapling stands (column), the indicator function is 1 when the specified forest type and stand-size requirement are satisfied; otherwise it is 0. This method is described in Cochran (1977) for estimating domain (cell) means.

#### 4.3.1.1 For Each Table Cell, Compute the Attribute of Interest for Each Plot

Each plot is assigned to only one stratum based on the Phase 1 stratification of the plot center. For area estimation, the attribute of interest is the proportion of the plot in the **domain** of interest:

$$P_{hid} = \frac{\sum_j^4 \sum_k^{K_{hj}} a_{mhijk} \delta_{hijkd}}{a_m \bar{P}_{mh}} \quad (4.1)$$

where

$P_{hid}$  = proportion of plot  $i$  in the domain of interest  $d$ , for plots assigned to stratum  $h$ , adjusted for stratum  $h$  plots that overlap the population **boundary**

$a_{mhijk}$  = mapped area (acres) of **subplot (macroplot)  $j$**  covering condition  $k$  on plot  $i$  assigned to stratum  $h$  (The area is computed using the largest area mapped, which is the subplot except in the Pacific Northwest (PNW) where the macroplot is used.)

$\delta_{hijkd}$  = zero-one domain indicator function, which is 1 if condition  $k$  on subplot (macroplot)  $j$  of plot  $i$  assigned to stratum  $h$  belongs to the domain of interest  $d$

$K_{hij}$  = the number of conditions that exist on subplot (macroplot)  $j$  of plot  $i$  assigned to stratum  $h$

$a_m$  = total area of the largest-sized plot on which area attributes are mapped (i.e., four times the subplot or macroplot area)

$\bar{p}_{mh}$  = mean proportion of stratum  $h$  mapped plot areas falling within the population (see equation 4.2)

Table margins (row and column totals) are treated like any other cell. Additivity is a property of the table construction, not the estimators. Most tables will be additive (i.e., the cells in the table body will add to totals in the margins such as area by forest type and stand size). However, a table of the area containing combinations of species and diameter classes will not be additive, because each cell represents the number of acres on which the particular species and diameter class combination occurs.

#### 4.3.1.2 Adjustment for Partial Plots Outside the Population

Equation 4.1 essentially treats all areas sampled on stratum  $h$  plots equally—every square foot is expanded equally whether it is part of a partial plot or not. For a given population, dividing by  $\bar{p}_{mh}$  in equation 4.1 adjusts the plot observations to account for any portions of stratum  $h$  plots falling outside the population. Note that  $\delta_{hijkd} = 0$  for condition classes outside the population in equation 4.1, because conditions outside the population are never conditions of interest. Reasons for condition classes being outside the population include:

- Partial plots that straddle an international boundary (i.e., Canada or Mexico)
- Partial plots that straddle a mapped ownership population boundary (e.g., national forest)
- Whole or partial plots within the population that are nonsampled (e.g., denied access or hazardous conditions)
- Whole or partial plots in **census water** (unless census water is estimated, i.e., included in  $A_7$ ) (Note that FIA currently subtracts census water from  $A_7$ , but anticipates that census water will be estimated when precise digitized census water boundaries become available from the U.S. Census Bureau. If and when census water is estimated from the FIA sample, a Phase 1 stratum will likely be created for it.)

If all plots are entirely within the population boundaries, then  $\bar{p}_{mh}$  would be 1. Otherwise, the average mapped-plot area actually sampled within the population is divided by the standard plot area,  $a_m$ . This approach was taken as a way of handling the potential bias introduced by ignoring portions of plots straddling population boundaries. Essentially, this creates a buffer around the population to ensure that plots which are only partially inside the population of interest are included in the estimation. More area is sampled, but more plots are also taken into the sample. Multiplying the totals by the larger area is on average equivalent to the adjustment made by using  $\bar{p}_{mh}$ :

$$\bar{p}_{mh} = \sum_i^{n_h} \sum_j^4 \sum_k^{K_{hij}} \frac{a_{mhijk} \delta_{hijk}}{a_m n_h} \quad (4.2)$$

where

$n_h$  = number of ground plots with Phase 1 assignments to stratum  $h$  (For initial area tables, this includes all plots sampled with any portion of a subplot (macroplot) in the population. For subsequent tables, any plots that are entirely nonsampled are excluded.)

$\delta_{hijk}$  = zero-one in-sample indicator function, which is 1 if condition  $k$  on subplot (macroplot)  $j$  of plot  $i$  assigned to stratum  $h$  is within the boundaries of the population (Nonsampled areas are included in initial area tables in order to estimate their areas, but are zero otherwise. Missing values are also treated as zero values.)

Clearly, there are cases where  $\bar{p}_{mh}$  will not be 1, meaning that  $\bar{p}_{mh}$  is a constant for a given stratum, but varies between strata and populations. The variation due to nonsampled plots and plots extending beyond population boundaries is expected to be small enough to be ignored. One particular area of concern is the checkerboard ownership pattern in the West where National Forest System (NFS) boundaries may be treated as population boundaries. FIA is evaluating the frequency of plots that straddle NFS boundaries and may switch to the **ratio-of-means** estimators described by Zarnoch and Bechtold (2000), if necessary.

### 4.3.1.3 Compute Strata Means and Variances

Plot values are averaged within each stratum. In the case of **simple random sampling**, this is the final estimate because simple random sampling is just stratified sampling with a single stratum.

The stratum mean is the sum of the plot observations,  $P_{hid}$ , divided by the number of plots in the stratum,  $n_h$ :

$$\overline{P}_{hd} = \frac{\sum_i^{n_h} P_{hid}}{n_h} \quad (4.3)$$

with variance estimator:

$$v(\overline{P}_{hd}) = \frac{\sum_i^{n_h} P_{hid}^2 - n_h \overline{P}_{hd}^2}{n_h(n_h - 1)} \quad (4.4)$$

Note that  $\overline{P}_{hd}$  is the mean of field-based observations in domain  $d$  within each Phase 1 stratum. This means that some plots classified as forest on the ground may have been assigned to a nonforest stratum and vice versa. These are not viewed as “misclassifications”, but as strata with less than ideal homogeneity. The estimators remain unbiased.

### 4.3.2 Estimation of Population Totals and Their Variances

Generally, individual counties are the populations of interest (i.e., the basic building blocks for estimation). Counties may be divided into **subpopulations** that are processed independently. This is the case when a portion of a county has an intensified Phase 2 sampling grid, has enumerated acreages, or has a measurement **cycle** that differs from the rest of the county. These scenarios are not uncommon when sampling land owned by public agencies [i.e., NFS, Bureau of Land Management (BLM), and National Park Service (NPS)]. Because populations and subpopulations are mutually exclusive, estimated totals are additive. Likewise, because different populations and subpopulations are independent, the associated variance estimates are also additive. Thus, totals from groups of counties can be combined to formulate State and regional totals; or segments of NFS land, by county, can be combined to yield totals for a specific national forest. County areas provided by the U.S. Census Bureau, which are used in the estimation of population totals, are updated at least every 10 years. NFS and BLM provide similar area totals for their lands, and totals by county if they have intensified or otherwise altered the sampling effort on their lands.

As first noted in equation 4.2, nonsampled plots present an estimation problem that requires more attention than plots that are otherwise out of the population. Because field crews cannot reliably ascertain the actual land use of such plots, initial area tables in FIA reports will report the proportion of total area designated as nonsampled. However, in all subsequent tables, these plots (or the nonsampled portions of them) will be processed as if they

were out of the population. This approach either reduces the sample size ( $n_h$ ), or the mean proportion of stratum  $h$  observed plot areas falling within the population, or both. The result is to adjust upward the number of acres represented by the accessed plots, or portions of them, in each stratum. Thus the area that could not be accessed is redistributed based on the accessible plots within each stratum. This essentially replaces nonsampled plots with the strata means and increases the strata variances due to the reduction in sample size. This approach has the advantage of simplicity, but has the potential to incur bias if the nonsampled plots are not representative of the rest of the population. Other methods of accounting for nonsampled areas are under investigation, including remote sensing both for direct measurement of a subset of attributes and for use in identifying similar plots for imputation purposes.

In the simple random sampling case, there is only one stratum, so totals are estimated by multiplying the population mean by the total area in the population. FIA rarely uses simple random sampling, but it has been used for Forest Health Monitoring (FHM) and is useful in responding to special requests where stratification data may be lacking. Processing protocols for the FIA plot design under a simple random sampling approach using ratios of means are discussed in detail by Zarnoch and Bechtold (2000).

To estimate the total area in each domain when the population has been stratified, the strata means are averaged using the strata weights and then multiplied by the total land area in the population. The estimated total is given for the stratified estimation and double sampling for stratification cases, respectively, as:

$$\hat{A}_d = A_T \sum_h^H W_h \overline{P}_{hd} = A_T \overline{P}_d \tag{4.5}$$

and

$$\hat{A}_d = A_T \sum_h^H \frac{n'_h}{n'} \overline{P}_{hd} = A_T \overline{P}_d$$

where

$A_T$  = total area in the population in acres

$\overline{P}_d$  = estimated proportion of the population in the domain of interest  $d$

$W_h$  = weight for stratum  $h$  (i.e., the proportion of the population area,  $A_T$ , that is in stratum  $h$ )

An approximation of the variance of the total area in the stratified estimation (and the simple random sampling) case, where strata sizes are known, is adapted from equation 5A.42 in Cochran's (1977) finite sampling framework. An estimate of this population variance (finite population correction factors ignored) is given by:

$$v(\hat{A}_d) = \frac{A_T^2}{n} \left[ \sum_h^H W_h n_h v(\overline{P}_{hd}) + \sum_h^H (1 - W_h) \frac{n_h}{n} v(\overline{P}_{hd}) \right] \quad (4.6)$$

The first term is for stratification, assuming proportional allocation, and the second term reflects the fact that the sample sizes are not fixed by strata in advance.

Double sampling for stratification applies when the strata weights are estimated. The variance of total area in this case is adapted from equation 12.32 in Cochran (1977), again ignoring finite population correction factors. The second term accounts for the fact that strata sizes are estimated:

$$v(\hat{A}_d) = A_T^2 \left\{ \sum_h^H \left( \frac{n'_h - 1}{n' - 1} \right) \frac{n'_h}{n'} v(\overline{P}_{hd}) + \frac{1}{n' - 1} \sum_h^H \frac{n'_h}{n'} (\overline{P}_{hd} - \overline{P}_d)^2 \right\} \quad (4.7)$$

As noted above, totals and their variances from different populations are additive. Thus the variance of a total across multiple populations or subpopulations is the sum of their variances. Examples of how to apply equations 4.1 through 4.7 are given in spreadsheet form in the supplementary document "Examples of FIA Point-in-Time Estimation Procedures for Several Common Cases" at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm) (see sections "plot summary", "problem number 1", and "problem number 2").

### 4.3.3 Estimation of Other Attributes

Population totals for attributes other than area usually are calculated by summing attributes to the plot level and then averaging at the stratum level. Indicator functions are used to identify the attribute of interest (e.g., total volume of white oak) in the domain of interest (e.g., oak-pine stands). The attribute of interest is summed for each plot and then divided by the observed plot area and the mean proportion of stratum  $h$  observed plot areas falling within the population, yielding an estimate of the attribute of interest on a per-unit-area basis:

$$y_{hid} = \frac{\sum_j \sum_t^4 y_{hijt} \delta_{hijt}}{a_o \bar{p}_{oh}} \quad (4.8)$$

where

$y_{hijt}$  = attribute of interest for tree  $t$  on macroplot, subplot, or **microplot**  $j$  of plot  $i$  assigned to stratum  $h$

$\delta_{hijt}$  = zero-one domain indicator function, which is 1 if tree  $t$  on subplot  $j$  of plot  $i$  assigned to stratum  $h$  belongs to the domain of interest  $d$

$a_o$  = total area normally used to observe the attribute of interest on a plot (i.e., four times the microplot, subplot, or macroplot area)

$\bar{p}_{oh}$  = mean proportion of stratum  $h$  observed-plot areas falling within the population (see equation 4.9)

In equation 4.8, dividing by  $\bar{p}_{oh}$  adjusts the plot observations to account for any portions of stratum  $h$  plots falling outside the population:

$$\bar{p}_{oh} = \frac{\sum_i^{n_h} \sum_j^4 \sum_k^{K_{hij}} a_{ohijk} \delta_{hijk}}{a_o n_h} = \frac{1}{a_o n_h} \sum_i^{n_h} \sum_j^4 \sum_k^{K_{hij}} a_{ohijk} \delta_{hijk} \quad (4.9)$$

where

$a_{ohijk}$  = area normally used to observe the attribute of interest (microplot, subplot, or macroplot  $j$ ) covering condition  $k$  on plot  $i$  assigned to stratum  $h$

In equation 4.8, dividing by the observed plot area and by the proportion of plots outside the population allows attributes such as number of trees across all diameter classes to be summed across plot types while accounting for any differences in the proportion of the various plot types that are outside the population. For example, if  $y_{hid}$  is the number of trees 1.0-inch diameter at breast height (**d.b.h.**) and larger sampled on microplots and subplots, then equation 4.8 should include a term for each plot type. This adjustment is the main difference between the estimators described herein and the ratio-of-means estimators outlined by Zarnoch and Bechtold (2000), where population totals are computed on the basis of each plot size separately, then summed for all plot sizes. Because totals for the latter alternative are not independent, the variance is complicated by the need to include a covariance term among plot sizes. We opted for a simpler approach, where all estimates are combined at the plot level and then treated as a single attribute of interest when calculating population totals. Using this method, the variance is much easier to

compute, and the resulting estimate (equation 4.10) is equivalent to the separate approach. When combining subplot and microplot values, the attribute of interest is computed as:

$$y_{hid} = \frac{\sum_j^4 \sum_t y_{hijt} \delta_{hijt}}{a_o \bar{p}_{oh}} + \frac{\sum_j^4 \sum_t y'_{hijt} \delta'_{hijt}}{a'_o \bar{p}'_{oh}} \quad (4.10)$$

where

$y'_{hijt}$  = attribute of interest for tree  $t$  on microplot  $j$  of plot  $i$  in stratum  $h$

$\delta'_{hijt}$  = zero-one domain indicator function, which is 1 if tree  $t$  on microplot  $j$  of plot  $i$  assigned to stratum  $h$  belongs to the domain of interest  $d$

$a'_o$  = total microplot area

$\bar{p}'_{oh}$  = mean proportion of stratum  $h$  microplot areas falling within the population (equation 4.9)

The plot attributes from either equation 4.8 or equation 4.10 are then summed across all plots within each stratum and divided by the total number of plots in the stratum to yield the stratum mean of the  $y_{hid}$ :

$$\bar{Y}_{hd} = \frac{\sum_i^{n_h} y_{hid}}{n_h} \quad (4.11)$$

with estimated variance:

$$v(\bar{Y}_{hd}) = \frac{\sum_i^{n_h} y_{hid}^2 - n_h \bar{Y}_{hd}^2}{n_h(n_h - 1)} \quad (4.12)$$

As was the case for area, the strata means are averaged using the strata weights, then multiplied by the total land area. The estimated total is given for the stratified estimation and double sampling for stratification cases, respectively, as:

$$\hat{Y}_d = A_T \sum_h^H W_h \bar{Y}_{hd} = A_T \bar{Y}_d \quad (4.13)$$



or

$$\hat{Y}_d = A_T \sum_h \frac{n'_h}{n'} \overline{Y_{hd}} = A_T \overline{Y}_d$$

where

$\overline{Y}_d$  = population mean of the attribute of interest in the domain of interest  $d$

As with the area estimate when strata weights are known, an approximation to the variance of the attribute total in the stratified estimation (and the simple random sampling) case was developed in the same way as for equation 4.6:

$$v(\hat{Y}_d) = \frac{A_T^2}{n} \left[ \sum_h W_h n_h v(\overline{Y_{hd}}) + \sum_h (1 - W_h) \frac{n_h}{n} v(\overline{Y_{hd}}) \right] \quad (4.14)$$

The estimated variance of the attribute total in the double sampling for stratification case is again adapted from equation 12.32 in Cochran (1977):

$$v(\hat{Y}_d) = A_T^2 \left\{ \sum_h \left( \frac{n'_h - 1}{n' - 1} \right) \frac{n'_h}{n'} v(\overline{Y_{hd}}) + \frac{1}{n' - 1} \sum_h \frac{n'_h}{n'} (\overline{Y_{hd}} - \overline{Y}_d)^2 \right\} \quad (4.15)$$

As noted in section 4.3.2, totals and their variances from different populations are additive for a given domain. Thus the variance of a total across multiple populations or subpopulations is the sum of their variances. Although the additivity property is also true for means, sums of means generally are not useful. To obtain a mean over populations, the totals should be added and divided by the total area of the populations. The result is a mean weighted by the population areas. An example of how to apply equations 4.8 through 4.15 is given in problem number 3 of the supplementary document “Examples of FIA Point-in-Time Estimation Procedures for Several Common Cases” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

### 4.3.4 Estimation of Ratios

Often, interest is not in the totals but in the attribute of interest expressed on a per-acre, per-condition (stand), or per-tree basis. An approach that is also compatible with the aforementioned estimates of population totals is the ratio-of-means estimator, wherein the numerator is the estimated attribute total and the denominator depends on the ratio to be estimated. The three cases can all be estimated using one of the following general formulas. The first is for stratified estimation where strata weights are known and the second for double sampling where strata weights are estimated:

$$\hat{R}_{dd'} = \frac{\hat{Y}_d}{\hat{X}_{d'}} = \frac{\sum_h W_h \overline{Y_{hd}}}{\sum_h W_h \overline{X_{hd'}}} \quad (4.16)$$

or

$$\hat{R}_{dd'} = \frac{\hat{Y}_d}{\hat{X}_{d'}} = \frac{\sum_h \frac{n'_h}{n'} \overline{Y_{hd}}}{\sum_h \frac{n'_h}{n'} \overline{X_{hd'}}$$

The strata means for the denominator,  $\overline{X_{hd'}}$ , are computed in the same manner as the numerator. Note that the numerator and denominator have different domains of interest, with  $d$  being a subset of  $d'$  (e.g., 12-inch oaks in oak-hickory stands). If the denominator is an area attribute, then  $x_{hid'}$  is estimated using equation 4.1 and the area attribute replaces  $P_{hid}$ . If the denominator is a tree or other attribute, then  $x_{hid'}$  is estimated using equation 4.8 and the attribute replaces  $y_{hid}$ . The choice of the individual plot observations,  $x_{hid'}$ , is described in sections 4.3.4.1 through 4.3.4.3 for three common situations.

The variance estimator from equation 5.6.10 in Särndal and others (1992) is:

$$v(\hat{R}_{dd'}) = \frac{1}{\hat{X}_{d'}^2} \left[ v(\hat{Y}_d) + \hat{R}_{dd'}^2 v(\hat{X}_{d'}) - 2\hat{R}_{dd'} \text{cov}(\hat{Y}_d, \hat{X}_{d'}) \right] \quad (4.17)$$

In the stratified estimation case, the right-hand-side variances are computed using equation 4.6 for area attributes and equation 4.14 for other attributes. The covariance is estimated as:

$$\text{cov}(\hat{Y}_d, \hat{X}_{d'}) = \frac{A_d^2}{n} \left[ \sum_h W_h n_h \text{cov}(\overline{Y_{hd}}, \overline{X_{hd'}}) + \sum_h (1 - W_h) \frac{n_h}{n} \text{cov}(\overline{Y_{hd}}, \overline{X_{hd'}}) \right] \quad (4.18)$$

where

$$\text{cov}(\overline{Y_{hd}}, \overline{X_{hd'}}) = \frac{\sum_i^{n_h} y_{hid} x_{hid'} - n_h \overline{Y_{hd}} \overline{X_{hd'}}}{n_h (n_h - 1)} \quad (4.19)$$

In the double sampling for stratification case, the variance is approximated by assuming the covariance can be computed similarly to the area attributes (equation 4.7) and other attributes (equation 4.15)

$$\begin{aligned} \text{cov}(\hat{Y}_d, \hat{X}_{d'}) = A_T^2 \left\{ \sum_h^H \left( \frac{n'_h - 1}{n' - 1} \right) \frac{n'_h}{n'} \text{cov}(\overline{Y}_{hd}, \overline{X}_{hd'}) \right. \\ \left. + \frac{1}{n' - 1} \sum_h^H \frac{n'_h}{n'} (\overline{Y}_{hd} - \overline{Y}_d) (\overline{X}_{hd'} - \overline{X}_{d'}) \right\} \end{aligned} \quad (4.20)$$

Equations 4.16 and 4.17 can be used for all three estimation of ratio cases—to express values on a per-acre, per-tree, or per-condition basis.

#### 4.3.4.1 Estimation on a Per-Acre Basis

Often interest is in expressing the attribute of interest on a per-acre basis. This can be estimated by dividing  $\hat{Y}_d$  in equation 4.13 by the total surface area in the population,  $A_T$ . Because the area is known, the variance of the ratio is simply the variance of the total (equation 4.14) divided by the square of the total area,  $A_T$ .

However, interest is more commonly in the attribute total expressed on a per-forested-acre basis—a ratio estimate. The denominator of equation 4.16 can be the estimate of total forest area ( $\hat{A}_F$ ) computed using equation 4.5, or the area of another domain such as the area in a specific forest type. When the denominator is derived from equation 4.5, the  $x_{hd'}$  is equal to  $P_{hid}$  in equation 4.1 which is then used to compute the  $X$  values in equations 4.16 through 4.20.

Indicator functions are used to specify the domains and attributes of interest that define  $y_{hi}$  and  $x_{hi}$ . These may or may not change for various cells in the tables, depending on the ratio of interest. For example, the value of  $x_{hd'}$  can be the same for all values in a table, such as those used in computing  $\hat{A}_F$ . This is useful for estimating additive tables, such as stand tables with numbers of trees per acre by species and diameter class. Alternatively, the value of  $x_{hd'}$  might change for each cell, such as the area by forest type and stand size, so that the attribute average can be estimated for each combination (domain). An example of how to apply equations 4.16 through 4.20 for estimation on a per-acre basis is given in problem number 4 of the supplementary document “Examples of FIA Point-in-Time Estimation Procedures for Several Common Cases” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

#### 4.3.4.2 Estimation on a Per-Condition (Per-Stand) Basis

A few attributes are only observed at the condition (or stand) level, such as stand age. To compute average stand age requires a slight modification of the approach to the  $y_{hid}$  attribute. The approach is to weight the condition attribute,  $y_{hik}$ , by the area in the condition of interest on the plot:

$$y_{hid} = \frac{\sum_j^4 \sum_k^{K_{hj}} y_{hik} a_{mhijk} \delta_{hijkd}}{a_m \bar{p}_{mh}} \quad (4.21)$$

The denominator reflects the proportion of the area sampled in the condition of interest:

$$x_{hid} = \frac{\sum_j^4 \sum_k^{K_{hj}} a_{mhijk} \delta_{hijkd}}{a_m \bar{p}_{mh}} \quad (4.22)$$

The result is the average of the attribute of interest weighted by the area in which it occurs. These values of  $y_{hid}$  and  $x_{hid}$  are then used in equations 4.16 through 4.20. An example of how to apply equations 4.16 through 4.22 for estimation on a per-stand basis is given in problem number 5 of the supplementary document “Examples of FIA Point-in-Time Estimation Procedures for Several Common Cases” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

#### 4.3.4.3 Estimation on a Per-Tree Basis

For some applications, the attribute of interest is an individual tree attribute, such as average tree height or conks per tree, where the ratio of interest is expressed on a per-tree basis. The ratio estimator then becomes the estimate of the population total for the attribute of interest,  $\hat{Y}_{d'}$ , divided by the total number of trees,  $\hat{X}_{d'}$ , in the population. Thus,  $y_{hid}$  is the sum of tree attributes observed on plot  $i$  in stratum  $h$  in the domain of interest; and  $x_{hid}$  is the number of trees observed on plot  $i$  in stratum  $h$  in the domain of interest. Those values are then used in equations 4.16 through 4.20. An example of how to apply equations 4.16 through 4.20 for estimation on a per-tree basis is given in problem number 6 of the supplementary document “Examples of FIA Point-in-Time Estimation Procedures for Several Common Cases” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

### 4.3.5 Computation of Sampling Errors

Sampling errors are used by FIA to reflect the accuracy of the estimates. Expressed on a percentage basis in order to enable comparisons between the precision of different estimates, sampling errors are computed by dividing the estimate into the square root of its variance:

$$S.E.\% = 100 \frac{\sqrt{v(\hat{Y}_d)}}{\hat{Y}_d} \quad (4.23)$$

Assuming normality of the distribution of estimates, the percent sampling error can be used to compute an approximate 67 percent confidence interval. If the sampling error is doubled, then an approximate 95 percent confidence interval can be formed. A better approximation is achieved by using the appropriate Student's *t*-values.

### 4.3.6 Components of Change

FIA inventories are designed to measure net change over time, as well as the individual **components of change** that constitute net change (e.g., growth, removals, mortality). Change estimates are computed for two sequential measurements of each inventory panel. Upon remeasurement, a new initial inventory is established for remeasurement at the next scheduled inventory. As such, computation of change components is not intended to span more than one inventory cycle. Rather, the change estimation process is repeated cycle by cycle. This simplifies field protocols and ensures that change estimation is based on short and relatively constant time intervals (e.g., 5 years). Change estimates for individual panels are combined across multiple panels in the same manner as panels are combined to obtain current inventory parameters such as total standing volume.

FIA recognizes the following components of change as prescribed core variables; they usually are expressed in terms of growing-stock or all-live volume, where *t* is the initial inventory of a measurement cycle, and *t*+1 is the terminal inventory:

$G_s$  = survivor growth—the growth on trees tallied at time *t* that survive until time *t*+1.

$I$  = ingrowth—the volume of trees at the time that they grow across the minimum d.b.h. threshold between time *t* and time *t*+1. The estimate is based on the size of trees at the d.b.h. threshold which is 1.0 inch for all-live trees and 5.0 inches for growing-stock trees. This term also includes trees that subsequently die (i.e., ingrowth mortality), are cut (i.e., ingrowth

cut), or diverted to nonforest (i.e., ingrowth diversion); as well as trees that achieve the minimum threshold after an area reverts to a forest-land use (i.e., reversion ingrowth).

$G_I$  = growth on ingrowth—the growth on trees between the time they grow across the minimum d.b.h. threshold and time  $t+1$ .

$R$  = reversion—the volume of trees on land that reverts from a **nonforest land** use to a **forest land** use (or, for some analyses, land that reverts from any source to **timberland**) between time  $t$  and time  $t+1$ . The estimate is based on tree size at the midpoint of the measurement interval. Tree size at the midpoint is modeled from tree size at time  $t+1$ .

$G_R$  = reversion growth—the growth of reversion trees from the midpoint of the measurement interval to time  $t+1$ . Tree size at the midpoint is modeled from tree size at time  $t+1$ . This term also includes the subsequent growth on ingrowth trees that achieve the minimum diameter threshold after reversion.

$M$  = mortality—the volume of trees that die from natural causes between time  $t$  and time  $t+1$ . The estimate is based on tree size at the midpoint of the measurement interval (includes mortality growth). Tree size at the midpoint is modeled from tree size at time  $t$ .

$G_M$  = mortality growth—the growth of trees that died from natural causes between time  $t$  and the midpoint of the measurement interval. Tree size at the midpoint is modeled from tree size at time  $t$ . This term also includes the subsequent growth on ingrowth trees that achieve the minimum diameter threshold prior to mortality.

$C$  = cut—the volume of trees cut between time  $t$  and time  $t+1$ . The estimate is based on tree size at the midpoint of the measurement interval (includes cut growth). Tree size at the midpoint is modeled from tree size at time  $t$ . Trees felled or killed in conjunction with a harvest or silvicultural operation (whether they are utilized or not) are included, but trees on land diverted from forest to nonforest (diversions) are excluded.

$G_C$  = cut growth—the growth of cut trees between time  $t$  and the midpoint of the measurement interval. Tree size at the midpoint is modeled from tree size at time  $t$ . This term also includes the subsequent growth on ingrowth trees that achieve the minimum diameter threshold prior to being cut.

$D$  = diversion—the volume of trees on land diverted from forest to nonforest (or, for some analyses, this may also include land diverted to **reserved forest land** and **other forest land**), whether utilized or not, between time  $t$  and time  $t+1$ . The estimate is based on tree size at the midpoint of the measurement interval (includes diversion growth). Tree size at the midpoint is modeled from tree size at time  $t$ .

$G_D$  = diversion growth—the growth of diversion trees from time  $t$  to the midpoint of the measurement interval. Tree size at the midpoint is modeled from tree size at time  $t$ . This term also includes the subsequent growth on ingrowth trees that achieve the minimum diameter threshold prior to diversion.

FIA recognizes the following components of change as **prescribed optional variables**:

$CI$  = cull increment—the net reduction in growing-stock volume due to reclassification of growing stock trees to cull trees between two surveys. Cull increment is the volume of trees that were growing stock at time  $t$ , but cull at time  $t+1$ . The estimate is based on tree size at the midpoint of the measurement interval (includes cull increment growth). Tree size at the midpoint can be modeled from tree size at time  $t$ , time  $t+1$ , or both.

$G_{CI}$  = cull increment growth—the growth to the midpoint of the measurement interval between time  $t$  and  $t+1$  of trees that were growing stock at time  $t$ , but cull trees at time  $t+1$ . Tree size at the midpoint can be modeled from tree size at time  $t$ , time  $t+1$ , or both.

$CD$  = cull decrement—the net gain in growing-stock volume due to reclassification of cull trees to growing stock trees between two surveys. Cull decrement is the volume of trees that were cull at time  $t$ , but growing stock at time  $t+1$ . The estimate is based on tree size at the midpoint of the measurement interval. Tree size at the midpoint can be modeled from tree at time  $t$ , time  $t+1$ , or both.

$G_{CD}$  = cull decrement growth—the growth from the midpoint of the measurement interval to time  $t+1$  on trees that were cull at time  $t$ , but growing stock at time  $t+1$ . Tree size at the midpoint can be modeled from tree size at time  $t$ , time  $t+1$ , or both.

Except for  $R$ ,  $D$ ,  $G_D$ , and  $G_R$ , all components listed above are computed from plot areas where land use is defined as forest at both time  $t$  and time  $t+1$ . Note that it is not possible to measure the terminal d.b.h. of all trees that were cut, died, or diverted to a nonforest land use. To minimize potential bias associated with the growth of these trees, estimates of  $G_M$ ,  $G_C$ ,  $G_D$ ,  $D$ ,  $R$ , and  $G_R$  are modeled on the basis of the measurement-interval midpoint. The midpoint is calculated as  $\Delta_i / 2$ , where  $\Delta_i$  is the time in years (rounded to the nearest 10<sup>th</sup>) between measurements for an individual plot. Models to predict midpoint tree sizes are developed regionally and may include a variety of factors, including terms to account for slowed growth on mortality trees.

The use of midpoint tree sizes creates special situations where careful attention is needed to account for all change components. Particularly notable are ingrowth trees that assume a status other than survivor. For instance, between time  $t$  and time  $t+1$ , a tree may cross the 5.0-inches d.b.h. threshold (ingrowth), grow to 5.4-inches d.b.h., and then die. Under those circumstances, there would be three components of change: (1) Ingrowth ( $I$ ), the volume at 5.0-inches d.b.h.; (2) Mortality growth ( $G_M$ ), the volume growth from 5.0-inches d.b.h. to 5.4-inches d.b.h.; and (3) Mortality ( $M$ ), the volume at 5.4-inches d.b.h. Similar circumstances occur where ingrowth trees are associated with cutting, reversions, and diversions. Such situations implicitly require that midpoint tree sizes be modeled for all trees 1.0-inch d.b.h. and larger in order to check for trees that may have crossed the tree diameter threshold before removal or death.

For reporting growth and change, the individual components are usually combined as follows, and expressed either in terms of growing-stock or all-live volume:

$$\text{Gross ingrowth} = I + R$$

$$\text{Accretion} = G_s + G_i + G_R + G_M + G_C + G_D$$

$$\text{Gross growth} = \text{gross ingrowth} + \text{accretion}$$

$$\text{Mortality} = M$$

$$\text{Removals} = C + D$$

$$\text{Net growth} = \text{gross growth} - \text{mortality}$$

$$\text{Net change} = \text{net growth} - \text{removals}$$

The above terms for accretion and net growth are modified as follows for FIA regions that elect to produce additional output containing optional expressions of cull increment and decrement. Note that these optional terms are relevant only when components of change are expressed in terms of growing-stock volume:

$$\text{Accretion} = G_s + G_i + G_R + G_M + G_C + G_D + G_{CI} + G_{CD}$$

$$\text{Net growth} = \text{gross growth} - \text{mortality} + \text{CD} - \text{CI}$$

In addition to volume, all change components may also be expressed in terms of basal area or weight. More commonly, some ( $I$ ,  $R$ ,  $M$ ,  $C$ ,  $D$ ,  $CI$ , and  $CD$ ) are occasionally expressed as numbers of trees.

A variety of estimators have been proposed for the various components of change (Beers and Miller 1964, Gregoire 1993, Roesch and others 1989, Van Deusen and others 1986). When only one fixed-size plot is involved,



estimation of change components is very straightforward and most estimators are equivalent. This is the case for most FIA reporting purposes because volumes generally are reported in terms of growing stock, which is based on trees 5.0-inches d.b.h. and larger that are only recorded on the subplot (except PNW which also uses macroplots).

For regions using the macroplots, or for expressions of growth involving microplot trees (e.g., growth of trees 1.0-inch d.b.h. and larger), change estimation is complicated by trees that grow from one plot size to the next. This requires techniques designed for variable-radius plots. Historically, FIA units have used one of two methods for calculating components of change for variable-radius plots—Beers and Miller (1964) or Van Deusen and others (1986). Because the Van Deusen estimator is more appropriate for prism sampling and FIA has moved away from prism sampling, we have decided to use the simpler Beers-Miller approach.

The Beers-Miller estimator weights all survivor growth ( $G_s$ ) on the basis of plot size at time  $t$ :

$$G_s = s_2 - s_1 \quad (4.24)$$

where

$s_2$  = tree size at time  $t+1$  weighted on the basis of plot size at time  $t$

$s_1$  = tree size at time  $t$  weighted on the basis of plot size at time  $t$

Ingrowth,  $I$ , is defined as those trees on the microplot that grew across the 1.0-inch threshold:

$$I = s_{dbh=1.0} \quad (4.25)$$

where

$s_{dbh=1.0}$  = the size of an ingrowth tree at the 1.0-inch d.b.h. threshold and growth on ingrowth as

$$G_I = s_2 - s_{dbh=1.0} \quad (4.26)$$

Both  $s_2$  and the size of the tree at 1.0-inch d.b.h. ( $s_{dbh=1.0}$ ) are weighted on the microplot basis for trees on the microplot that were < 1.0-inch d.b.h. at time  $t$  but greater than 1.0 inch at time  $t+1$ . Note that the Beers-Miller estimator ignores trees that grow onto the subplot from outside the microplot.

When estimating change, individual trees are placed into the appropriate change-component category(s). Tree attributes associated with the change-component of interest are then summarized to the plot level using equations

4.8 and 4.10, annualized by dividing plot-level periodic values by the number of years between the initial and terminal measurements of each plot, and summarized to the population level as specified in equations 4.11 and 4.13. The components of change are converted to average annual values as follows, where  $\Delta_t$  is the time in years (rounded to the nearest 10<sup>th</sup>) between measurements for an individual plot:

$$\text{Annual gross ingrowth} = (I + R) / \Delta_t$$

$$\text{Annual accretion} = (G_S + G_I + G_R + G_M + G_C + G_D) / \Delta_t$$

$$\text{Annual gross growth} = \text{annual gross ingrowth} + \text{annual accretion}$$

$$\text{Annual mortality} = M / \Delta_t$$

$$\text{Annual removals} = (C + D) / \Delta_t$$

$$\text{Annual net growth} = \text{annual gross growth} - \text{annual mortality}$$

$$\text{Annual net change} = \text{annual net growth} - \text{annual removals}$$

Annualized values for accretion and net growth for regions that include the optional expressions of cull increment and decrement would be:

$$\text{Annual accretion} = (G_S + G_I + G_R + G_M + G_C + G_D + G_{CI} + G_{CD}) / \Delta_t$$

$$\text{Annual net growth} = \text{annual gross growth} - \text{annual mortality} + (CD - CI) / \Delta_t$$

Observe that some of the change components pertain to trees on conditions that remained in forest for an entire inventory cycle; some are based on trees and areas that become forest between the initial and terminal inventory of a cycle (reversions); and some pertain to trees on conditions removed from the forest land base (diversions). Improved estimates of change can be obtained by stratification on the basis of both initial and terminal land use: (1) Forest to forest, (2) Forest to nonforest, (3) Nonforest to forest, and (4) Nonforest to nonforest. This might also be expanded to include the finer subsets of forest recognized by FIA such as timberland, reserved forest land, and other forest land.

This requires classification of the same Phase 1 points at both time  $t$  and time  $t+1$ —either photo interpretation of the same photo plots if photography was used or classification of the same pixels if satellite imagery is used. Note that the same collapsed strata should be used for both current and change estimates if the results are to be compatible.

Once the strata weights are assigned, estimation by domains within strata follows the same procedure outlined for current inventory parameters. For

standard reporting purposes, domains identified or partitioned by forest attributes (e.g., owner group) usually are defined on the basis of initial classifications for forest-nonforest and forest-forest parameters; and domains for nonforest-forest inventory attributes are defined on the basis of terminal classifications. Examples of computational procedures for change estimation are provided in the supplementary document “Examples of FIA Change-Component Estimation Procedures for Several Common Cases” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

#### **4.4 Expansion Factors**

When periodic inventories and flat files were the FIA standards, it was convenient to calculate a small set of expansion factors by which individual plot-level or tree-level observations could be converted to their population-level equivalents. This allowed population totals to be obtained via summation, which greatly simplified the estimation process. Expansion factors were popular with external FIA clients, many of whom used this concept to build their own processing systems. The tradeoff for such simplicity is that the use of expansion factors precludes the ability to calculate variances. At best, the variances of estimators derived from expanded values can only be approximated, and these approximations are known to be poor (Alegria and Scott 1991).

Expansion factors are less practical with panelized inventory systems, which are designed to increase analytical flexibility by allowing panels to be combined in a variety of ways. Each different panel combination produces a unique set of expansion factors, rendering expansion factors associated with panel systems less stable than those produced by periodic systems.

The use of expansion factors is discouraged because they prohibit accurate variance estimation and they no longer have the advantage of simplicity. However, there is still a demand for them, and it may take a while to convert processing systems to the estimation procedures specified in this chapter. Therefore, FIA will continue to offer expansion factors until a demand is no longer apparent. The derivation of expansion factors is described in the supplementary document “Computation of FIA Plot Expansion Factors” at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

## 4.5 Phase 3

### 4.5.1 Estimation Procedures Used for FHM (Phase 3) Data

The FHM Program did not utilize any Phase 1 stratification prior to merging with FIA. In order to avoid reporting regional and State totals that conflicted with those reported by FIA, FHM statistics usually were presented as population-level means based on either simple random sampling or Generalized Least Squares (GLS). Zarnoch and Bechtold (2000) developed one such approach based on simple random sampling and ratio-of-means estimators. Smith and Conkling (2004) used GLS estimation procedures, where population-level current mean values and annual change estimates are obtained from linear models for repeated measurements (Gregoire and others 1995, Urquhart and others 1993, Van Deusen 1996). The estimate of current value is:

$$y_t = b_{t-1} + b_1(\text{Year}_t - \text{Year}_{t-1}) \quad (4.29)$$

where

$y_t$  = predicted value of the attribute at year  $t$

$b_{t-1}$  = the mean value of attribute at year  $t-1$

$b_1$  = the annual change in  $y$  from Year  $t-1$  to Year  $t$

Both  $b_{t-1}$  and  $b_1$  are computed by estimated generalized least squares using Proc MIXED® (SAS Institute 1999).

### 4.5.2 Combining with Phase 1

Because the FIA and FHM inventory systems have merged, it is now possible to combine Phase 1 and Phase 3 data using stratified random estimation and double sampling for stratification. We are studying use of Phase 1 stratification to enhance the estimation of attributes unique to Phase 3 and plan to document estimation procedures for attributes specific to Phase 3 in subsequent manuscripts.

### 4.5.3 Linking with Phase 2

For inventory parameters common to both Phase 2 and Phase 3, usually there is no advantage in generating population estimates from the smaller Phase 3 subset, because the estimates will not match and the Phase 3 variances will be larger due to a reduced  $n$ . Phase 3 plots, therefore, should be combined with Phase 2 plots when estimating inventory parameters common to both phases.

For those inventory parameters that are unique to Phase 3, it is possible to model relationships between unique Phase 3 parameters and parameters common to both Phase 2 and Phase 3. Such models can then be used to extrapolate estimates of Phase 3 parameters onto the more intensive Phase 2 sampling network.

## 4.6 Literature Cited

- Alegria, J., Scott, C.T. 1991. Generalized variance function applications in forestry. Res. Note NE-345. Radnor, PA: U.S. Department of Agriculture Forest Service, Northeastern Forest Experiment Station. 4 p.
- Beers, T.W.; Miller, C.I. 1964. Point sampling: research results, theory, and application. Res. Bull. 786. Lafayette, IN: Purdue University. Agricultural Experiment Station. 56 p.
- Breidt, F.J.; Opsomer, J.D. 2004. Endogenous and post-stratification in surveys: classifying with a sample-fitted model. Preprint Series #02-23. Ames, IA: Department of Statistics, Iowa State University. 25 p.
- Cochran, W.G. 1977. Sampling techniques. New York: John Wiley. 428 p.
- Gregoire, T.G. 1993. Estimation of forest growth from successive surveys. *Forest Ecology and Management*. 56: 267-278.
- Gregoire, T.G.; Schabenberger, O.; Barret, J.P. 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Canadian Journal of Forest Research*. 25: 137-156.
- Moisen, G.G.; Edwards T.C., Jr. 1999. Use of generalized linear models and digital data in a forest inventory of northern Utah. *Journal of Agricultural, Biological, and Environmental Statistics*. 4: 372-390.
- Roesch, F.A.; Green, E.J.; Scott, C.T. 1989. New compatible estimators for survivor growth and ingrowth from remeasured horizontal point samples. *Forest Science*. 35: 281-293.
- Särndal, C.E.; Swensson, B.; Wretman, J. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag. 694 p.
- SAS Institute. 1999. SAS/STAT® user's guide. Version 8. Cary NC: SAS Institute Inc. 2552 p. Vol. 2.
- Smith, W.D.; Conkling, B.L. 2004. Analyzing forest health data. Gen. Tech. Rep. SRS-077. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 33 p.
- Urquhart, N.S.; Overton, W.S.; Birkes, D.S. 1993. Comparing sampling designs for monitoring ecological status and trends: impact of temporal patterns. In: *Statistics for the Environment*. London: John Wiley: 71-85.
- Van Deusen, P.C. 1996. Incorporating predictions into an annual forest inventory. *Canadian Journal of Forest Research*. 26(10): 1709-1713.
- Van Deusen, P.C.; Dell, T.R.; Thomas, C.E. 1986. Volume growth estimation from permanent horizontal points. *Forest Science*. 32: 415-422.
- Zarnoch, S.J.; Bechtold, W.A. 2000. Estimating mapped-plot attributes with ratio of means. *Canadian Journal of Forest Research*. 30: 688-697.



# 5 Combining Panels for Forest Inventory and Analysis Estimation

Paul L. Patterson and Gregory A. Reams<sup>1</sup>

## 5.1 Single Panels vs. Multiple Panels

The term **panel**<sup>2</sup> denotes a set of samples where the same elements are measured on two or more occasions. Historically, Forest Inventory and Analysis (FIA) has used a single panel to conduct **periodic surveys**. Annual panels, however, allow greater flexibility because they can be combined in a variety of ways. Note that FIA assumes complete spatial coverage for each panel across the **population** of interest. When estimating inventory **attributes** for a single panel, the estimation approach proceeds as outlined in chapter 4. When estimating inventory attributes for combined panels, however, such procedures may require modification, depending on how the panels are combined. Related modifications are discussed in conjunction with the two specific methods presented in sections 5.2.1 and 5.2.2.

Dividing a single large periodic survey into a series of smaller surveys by measuring panels, one at a time, has several noteworthy advantages:

1. Individual panels can yield information about variations that occur within a measurement **cycle**; they can estimate year to year as well as long-term cycles and trends. This greatly improves our ability to understand the causes and timing of changes in the resource, as opposed to assuming linear trends.
2. Successive measurement of panels can provide quicker feedback to facilitate decisions that depend on knowledge of fluctuations in the survey attributes. If necessary, field protocols can be modified at the next scheduled panel, rather than waiting for a full inventory cycle to be completed.
3. Panels are highly responsive to widespread catastrophic events. The impact of a catastrophic event that occurs in a single year (e.g., fire or hurricane) can be gauged immediately. In the past, alternative methods such as interim periodic surveys were used to deal with catastrophic events (Sheffield and Thompson 1992).
4. Panels provide a natural, temporal link to other annual ancillary data.

---

<sup>1</sup> Paul L. Patterson, Mathematical Statistician, USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO 80526-1891; and Gregory A. Reams, FIA National Program Manager, USDA Forest Service, 1601 North Kent Street, Arlington, VA 22209.

<sup>2</sup> First use of a glossary term in each chapter is in bold face.

Although annual inventories are considered superior to periodic inventories for FIA applications, conversion comes at a price. Some advantages of periodic surveys include:

1. Travel cost is minimized. Annual inventories require field crews to travel across the entire population each time a panel is measured.
2. Change estimates apply to just two points in time (although field measurements often take 2 to 3 years to complete). With multiple panels, change estimates are staggered over two inventory cycles, rather than just one.
3. The length of time required to measure individual panels can be inconsistent and difficult to manage. Budget constraints, regional issues, and logistical problems all influence the time needed to complete individual panels and sets of panels. As a result, the time required to complete a panel (or **subpanel**) typically does not equal exactly 1 year. From State to State, the time needed to finish one complete set of inventory panels can range from 3 to 10 years due to panel acceleration or panel creep as discussed in chapter 2.
4. The sample size is nearly always sufficient. A single panel has sampling errors that are  $\sqrt{P}$  times larger than when using all  $P$  panels. However, as FIA moves to the annual system this will cause a short-term problem for analysts who must report results from only one or two panels of data. The problem also may make some of the more sophisticated methods for combining panels inappropriate for small samples. For example, recent data from 220 counties in Indiana, Missouri, and Illinois, where two panels have been completed, showed an average of only two forested **plots** per county per panel.

Additional discussion of the advantages and disadvantages of multiple or single panels is provided by Köhl and Scott (2000).

FIA uses panels to measure both current inventory and change. Change can be estimated in a multitude of ways. One method uses the net difference between two sequential, but different, panels. Assuming this approach involves independent samples, the variance of the difference is the sum of the variances, roughly  $2s^2/n$ . Measuring different panels over time yields estimates of net change, but only remeasured panels can provide information about specific **components of change** behind the net change. The latter are particularly useful for researching the dynamics of causation and associated relationships. There, change is directly observed, so the variance is reduced by the correlation,  $R$ , between measurements, roughly  $(1 - R^2)s^2/n$ . Alternatively, this can be expressed as the reduction in the variance of the difference due to the covariance between occasions:



$v(y_t - y_{t-1}) = v(y_t) + v(y_{t-1}) - 2\text{cov}(y_t, y_{t-1})$ . The remeasured panels approach is generally preferred for its robustness, efficiency, and ability to isolate individual components of change.

## 5.2 Combining Panels

There is currently no **prescribed core** procedure for combining panels. Due to different spatial, temporal, and forest characteristics within and among regions, it is not clear if any single technique will work for all. Whatever estimation strategy is used to estimate current conditions, variance reduction usually can be attained by combining current data with earlier data from previous panels. Several estimation strategies have been devised to take advantage of previous data (Czaplewski 1995, Reams and Van Deusen 1999); those now being investigated by FIA include:

1. The **moving average** (MA)
2. The temporally indifferent (TI) method
3. Modeling [updating plots, mixed estimators (Van Deusen 2002), Kalman filters (Brockwell and Davis 1996), and various time series models (Johnson and others 2003)]

The first two are relatively straightforward, highly compatible with the estimators presented in chapter 4, and discussed in further detail in sections 5.2.1 and 5.2.2. The third technique, modeling, has so many possible variations that potential approaches are beyond the scope of this chapter.

### 5.2.1 The Moving Average Method

Let  $P$  denote the number of panels to be combined for analysis. Let  $Y_p$  denote the true quantity for panel  $p$ , where  $p = 1, \dots, P$ ; and let  $\hat{Y}_p$  denote the estimate of  $Y_p$  obtained using the appropriate technique from chapter 4. Note that each panel is treated as an independent estimate, which permits:

1. The weighting of individual panels
2. **Phase 1 stratification** instruments to differ among panels (i.e., different maps may be used to stratify different panels)

Using the above notation, the MA estimator is given by:

$$\hat{Y}_{MA,P} = \sum_{p=1}^P w_p \hat{Y}_p \quad (5.1)$$

where

$\{w_p\}_{p=1,\dots,P}$  is a set of constant positive weights that sum to 1 across all combined panels (Roesch and Reams 1999)

The variance formula for  $\hat{Y}_{MA,p}$  is:

$$V\left[\hat{Y}_{MA,p}\right] = \sum_{p=1}^P w_p^2 V\left[\hat{Y}_p\right], \quad (5.2)$$

where

$V\left[\hat{Y}_p\right]$  for each panel is calculated as specified in chapter 4

The MA estimator is appealing because it is simple and the use of previous panels can lead to a substantial reduction in the variance over the variance of an estimate based on an individual panel (Gillespie 1999).

Roesch and Reams (1999) suggest that equal weighting of all panels ( $w_p = \frac{1}{P}$ , for  $p = 1, \dots, P$ ), or heavier weighting of more recent panels would be appropriate in equation 5.1. Johnson and others (2003) have shown, with simulation based on FIA data, that in most situations the moving average with equal weights has the smallest mean squared error.

Because the MA is a weighted sum of estimates across all panels of interest, it can be viewed as an estimate of the attribute of interest at some time between the first and last years of the time period from which the panels were drawn. The specific point in time depends on the weights used, as well as the direction and magnitude of change that have influenced that attribute from panel to panel. Also, moving averages and related techniques result in estimators that dampen trends by obscuring annual fluctuations, and in that sense do not measure the current status of a finite population, but rather a temporal average of that population. Such estimators will make changes appear smaller than they are, and the use of older panels potentially creates a lag bias when estimating current conditions. However, in the absence of some widespread catastrophic event, the smoothing and lag effects of moving averages usually will be inconsequential and more than offset by the reduction in variance acquired from using the maximum number of available panels (Johnson and others 2003). Still, there is some concern that potential lag bias may mask time trends (Roesch and Reams 1999). FIA is now researching whether lag bias significantly influences trends associated with the attributes of interest occurring on **forest lands**. Obviously, in the presence of a widespread catastrophic event, lag bias cannot be ignored. The best way to adjust methods for such situations is also an area of ongoing research.

Finally, the MA approach does not require separate Phase 1 stratification for each panel. Thus, the weighting feature of the MA estimator may still be used when applying the same Phase 1 stratification to any or all panels.

### **5.2.2 The Temporally Indifferent Method**

The temporally indifferent (TI) method differs from the MA method in that all panels of interest are pooled into the equivalent of one large periodic inventory, and the same Phase 1 stratification is applied across all panels. Although this approach lacks some of the flexibility offered by alternative methods of combining panels, it does have advantages over periodic inventories because individual panels can be used to produce spatially unbiased estimates before the results of a complete periodic inventory are available. Note that in the simple random sampling case, the TI method is equivalent to the MA with weights proportional to the number of plots in each panel.

The TI method is simpler than the MA approach in that estimation proceeds directly as specified in chapter 4, without the added complication of weighting (i.e., equations 5.1 and 5.2 are not used). In addition to simplicity, use of the TI method may be advantageous when sample sizes per panel are small. For example, when the MA approach is used in conjunction with stratification, the variance estimates for individual panels may be inflated by small sample sizes within each stratum. This could offset the variance reduction obtained through the MA estimator's weighted sum. When this is the case, the larger sample sizes per stratum attained with the TI method could reduce the variance considerably more than the MA alternative.

Finally, potential smoothing of temporal trends and lag bias associated with catastrophic disturbances present the same problems described for the MA estimator, with the added disadvantage that no weights are used to adjust for these effects. If weighting is necessary to overcome lag bias or to adjust for catastrophic events, the MA method is preferred.

## **5.3 Literature Cited**

- Brockwell, P.J.; Davis, R.A. 1996. Introduction to time series and forecasting. New York: Springer Verlag. 420 p.
- Czaplewski, R.L. 1995. Continuous adaptive monitoring of status and trends in ecosystem conditions. Proceedings of the 1995 Society of American Foresters conference. Bethesda, MD: Society of American Foresters. 482 p.
- Gillespie, A.J.R. 1999. Rationale for a national annual inventory program. *Journal of Forestry*. 97: 16-20.
- Johnson, D.S.; Williams, M.S.; Czaplewski, R.L. 2003. Comparison of estimators for rolling samples using forest inventory and analysis data. *Forest Science*. 49(1): 50-63.

- Reams, G.A.; Van Deusen, P.C. 1999. The southern annual forest inventory system. *Journal of Agricultural Biological and Environmental Statistics*. 4: 346-360.
- Köhl, M.; Scott, C.T. 2000. A comparison of periodic versus permanent surveys. In: Burk, T.E.; Hansen, M.H., eds. *Integrated tools for natural resources inventories in the 21<sup>st</sup> century: Proceedings of the IUFRO Conference*. GTR-NC-212. St. Paul, MN: U.S. Department of Agriculture Forest Service, North Central Research Station: 94-103 p.
- Roesch, F.A.; Reams, G.A. 1999. Analytical alternatives for an annual inventory system. *Journal of Forestry*. 97: 33-37.
- Sheffield, R.M.; Thompson, M.T. 1992. Hurricane Hugo: effects on South Carolina's forest resource. Res. Pap. SE-284. Asheville, NC: U.S. Department of Agriculture Forest Service, Southeastern Forest Experiment Station. 51 p.
- Van Deusen, P.C. 2002. Comparison of some annual forest inventory estimators. *Canadian Journal of Forest Research*. 32: 1992-1995.

# 6 Notation for Equations

## 6.1 Indices

$H$  = number of strata

$P$  = number of complete panels being combined for estimation

$T$  = number of years necessary to complete  $P$  panels

$N$  = number of sampling units in the population

$n$  = number of plots

$h$  = stratum index,  $h=1, \dots, H$

$W_h$  = weight for stratum  $h$  (i.e., proportion of the population that is in stratum  $h$ )

$n_h$  = number of ground plots with Phase 1 assignments to stratum  $h$   
(For initial area tables, includes all plots sampled with any portion of a subplot (macroplot) in the population. For subsequent tables, any plots that are entirely nonsampled are excluded.)

$d$  = domain of interest index, typically referring to a cell in a table, such as a combination of stand size class (column) and species (row)

$i$  = plot index,  $i=1, \dots, n_h$

$j$  = subplot index,  $j=1, \dots, 4$

$t$  = tree index,  $t=1, \dots,$

$K_{hij}$  = number of condition classes on subplot (macroplot)  $j$  of plot  $i$  assigned to stratum  $h$

$k$  = condition-class index,  $k=1, \dots, K_{hij}$

$p$  = panel index,  $p=1, \dots, P$

$w_p$  = weight assigned to panel  $p$

## 6.2 Double Sampling

$n'$  = total number of Phase 1 plots (pixels) sampled in a population

$n'_h$  = number of Phase 1 plots (pixels) classified as belonging to stratum  $h$ ,  
 $h = 1, \dots, H$

## 6.3 Area Related

$A_T$  = total acres in a population

$\hat{A}_d$  = estimated acres within a domain of interest  $d$

$\delta_{hijkd}$  = zero-one domain indicator function, which is 1 if condition  $k$  on subplot (macroplot)  $j$  of plot  $i$  assigned to stratum  $h$  belongs to the domain of interest  $d$

$\delta_{hijk}$  = zero-one in-sample indicator function, which is 1 if condition  $k$  on subplot (macroplot)  $j$  of plot  $i$  assigned to stratum  $h$  is within the boundaries of the population (Nonsampled areas are included in initial area tables in order to estimate their areas, but are zero otherwise. Missing values are also treated as zero values.)

$a_m$  = total area of the largest size plot on which area attributes are mapped (i.e., four times the subplot or macroplot area)

$a_{mhijk}$  = mapped area (acres) of subplot (macroplot)  $j$  covering condition  $k$  on plot  $i$  assigned to stratum  $h$  (The area is computed using the largest area mapped, which is the subplot except in the Pacific Northwest (PNW) where the macroplot is used.)

$\bar{p}_{mh}$  = mean proportion of stratum  $h$  mapped plot areas falling within the population

$a_o$  = total area normally used to observe the attribute of interest on a plot (i.e., four times the microplot, subplot, or macroplot area)

$a_{ohijk}$  = area normally used to observe the attribute of interest (microplot, subplot or macroplot  $j$ ) covering condition  $k$  on plot  $i$  assigned to stratum  $h$

$\bar{p}_{oh}$  = mean proportion of stratum  $h$  observed-plot areas falling within the population

$\alpha'_o$  = total microplot area

$\bar{p}'_{oh}$  = mean proportion of stratum  $h$  microplot plot areas falling within the population

$P_{hid}$  = proportion of plot  $i$  in the domain of interest  $d$ , for plots assigned to stratum  $h$ , adjusted for stratum  $h$  plots that overlap the population boundary

$\bar{P}_{hd}$  = mean of the plot proportions in the domain of interest  $d$  assigned to stratum  $h$

$\bar{P}_d$  = estimated proportion of the population in the domain of interest  $d$

## 6.4 Tree and Plot Related

$y_{hijt}$  = attribute of interest for tree  $t$  on macroplot, subplot, or microplot  $j$  of plot  $i$  assigned to stratum  $h$

$\delta_{hijt}$  = zero-one domain indicator function, which is 1 if tree  $t$  on subplot  $j$  of plot  $i$  assigned to stratum  $h$  belongs to the domain of interest  $d$

$y'_{hijt}$  = attribute of interest for tree  $t$  on microplot  $j$  of plot  $i$  assigned to stratum  $h$

$\delta'_{hijt}$  = zero-one domain indicator function, which is 1 if tree  $t$  on microplot  $j$  of plot  $i$  assigned to stratum  $h$  belongs to the domain of interest  $d$

$y_{hid}$  = the estimate of the attribute of interest for plot  $i$  in stratum  $h$  in domain of interest  $d$

$\bar{Y}_{hd}$  = stratum  $h$  mean of the plot estimates of the attribute of interest in the domain of interest  $d$

$\bar{Y}_d$  = mean of the attribute of interest in domain of interest  $d$

$\hat{Y}_d$  = estimated total for the attribute of interest in domain of interest  $d$

$\hat{R}_{dd'}$  = ratio of means estimator (per acre, per condition, or per tree) in domains of interest  $d$  and  $d'$

$\hat{Y}_p$  = the estimate for the attribute of interest for panel  $p$

$\hat{Y}_{MA,p}$  = the moving average estimator for the attribute of interest for  $P$  panels





# 7

## Glossary of Terms

**area change matrix:** the area of the intersection of all combinations of initial and terminal condition classes between two points in time, compiled for a microplot, subplot, macroplot, or plot.

**attribute:** a discrete or continuous variable, usually associated with the classification or measurement of area or vegetation.

**bi-areal plot:** a plot design that incorporates two different plot sizes at each sample location for the purpose of measuring trees in two different tree-diameter ranges.

**bioindicator:** the use of a biological entity's condition, frequency, and abundance as an indicator of ecosystem quality.

**boundary (condition class):** the border between two distinctly different condition classes.

**boundary (population):** the border of a population or subpopulation.

**census water:** areas of permanent water  $\geq 4.5$  acres or  $\geq 200$  feet wide.

**classified satellite imagery:** a map (satellite image) that defines and displays the spatial arrangement of each classified stratum on a pixel basis.

**components of change:** the different subdivisions of the changes that can occur to a tree between measurements, such as growth, mortality, and removals.

**condition class (or condition):** the combination of discrete attributes that describe the area associated with a plot. These attributes include condition status (land use), forest type, stand origin, stand size, owner group, reserve status, and stand density, as well as other ancillary and computed attributes.

**contrasting condition:** The condition class that differs from the condition class located at the subplot center (for boundaries on the subplot or macroplot) or at the microplot center (for boundaries on the microplot), i.e., the condition class present on the other side of a boundary.

**cycle:** one sequential and complete set of panels.

**cycle length:** the period of time required to measure a complete set of panels (synonymous with measurement cycle).

**diameter at breast height (d.b.h.):** the diameter of a tree stem, located at 4.5 feet above the ground (breast height) on the uphill side of a tree. The point of diameter measurement may vary on abnormally formed trees.

**diameter at root collar (d.r.c.):** the diameter of a shrub-like “woodland” tree species, measured outside bark at the ground line or stem root collar.

**domain:** a class (or combination of classes) for which a population estimate is made for some attribute of interest. Domains are typically the row and column margins of tabular output tables (e.g., saw timber stands on publicly owned timberland).

**double sampling for stratification:** a sampling method whereby a large sample of plots are stratified in Phase 1, then a subsample are measured for all attributes in Phase 2. When the strata are homogeneous with respect to the attribute, then the estimators are more accurate versus simple random sampling.

**down woody material:** dead pieces of wood > 3.0 inches in diameter. Down woody material includes downed, dead tree and shrub boles, large limbs, and other woody pieces that are severed from their original source of growth or are leaning more than 45 degrees from vertical.

**enhanced prescribed core variable:** all FIA units produce a value for these variables and there is a prescribed national protocol for measuring or calculating these variables. However, a given FIA unit is collecting data in greater detail than national protocol requires, and the detailed data can be aggregated to core specifications. Examples: fifth micro-plot in NE, additional disturbance codes beyond the prescribed codes.

**forest (or forest land):** land that is at least 10 percent stocked by forest trees of any size, or land formerly having such tree cover, and not currently developed for a nonforest use. The minimum area for classification as forest land is one acre. Roadside, stream-side, and shelterbelt strips of timber must be at least 120 feet wide to qualify as forest land. Unimproved roads and trails, streams and other bodies of water, or natural clearings in forested areas are classified as forest, if less than 120 feet in width or one acre in size. Grazed woodlands, reverting fields, and pastures that are not actively maintained are

included if the above qualifications are satisfied. Forest land includes three subcategories: timberland, reserved forest land, and other forest land.

**hybrid classification:** a combination of supervised and unsupervised classification (e.g., guided classification).

**indicator function:** a variable with a value of 0 or 1 that is used to specify attributes of interest (e.g., white-oak growing-stock volume), and domains of interest (e.g., northern hardwood forest types) in the estimation process.

**macroplot:** a circular area with a fixed horizontal radius of 58.9 feet (1/4 acre). Macroplot centers are co-located with subplot centers. Macroplots are used in the optional tri-areal design, primarily for sampling relatively rare events.

**mapped plot:** a plot that has been partitioned into unique and distinct condition classes by establishing the boundaries between them.

**microplot:** a circular area with a fixed horizontal radius of 6.8 feet (1/300 acre), primarily used to sample trees less than 5.0 inches at d.b.h./d.r.c.

**moving average:** a weighted average of the estimates for distinct panels.

**noncensus water:** bodies of water from 1 to 4.5 acres in size and water courses from 30 feet to 200 feet in width.

**nonforest:** areas defined as nonforest land, census water, or noncensus water.

**nonforest land:** land that does not support, or has never supported, forests, and lands formerly forested where use for timber management is precluded by development for other uses. Includes areas used for crops, improved pasture, residential areas, city parks, improved roads of any width and adjoining rights-of-way, power line clearings of any width, and noncensus water. If intermingled in forest areas, unimproved roads and nonforest strips must be more than 120 feet wide, and clearings, etc., more than 1 acre in size, to qualify as nonforest land.

**nonprescribed core variable:** all FIA regions must produce a value for the variable; but there is no prescribed protocol for measuring or calculating the variable. Examples: tree volume, site index.

**nonprescribed optional variable:** a value is produced for the variable at the discretion of the FIA regions, and there is no prescribed national protocol for measuring or calculating the variable.

**other forest land:** forest land other than timberland and reserved forest land. It includes available and reserved low-productivity forest land, which is incapable of producing 20 cubic feet of growing stock per acre annually under natural conditions because of adverse site conditions such as sterile soil, dry climate, poor drainage, high elevation, steepness, or rockiness.

**panel:** a sample in which the same elements are measured on two or more occasions. FIA divides plots into five panels that can be used to independently sample the population.

**periodic survey:** a noncontinuous inventory system. A survey strategy whereby a set of inventory panels is measured simultaneously over a short time frame, often 1 to 3 years in the case of FIA, and there is a time lag, often many years, before the panels are remeasured.

**phase 1:** FIA activities related to remote-sensing, the primary purpose of which is to obtain strata weights for population estimates.

**phase 2:** FIA activities conducted on the network of ground plots. The primary purpose is to obtain field data that enable classification and summarization of area, tree, and other attributes associated with forest land uses.

**phase 3:** a subset of Phase 2 plots where additional attributes related to forest health are measured.

**pixel:** picture elements—the elements of a digitized picture. The resolution of a picture is dependent on the size and number of elements of which it consists.

**plot:** a cluster of 4 points arranged such that point 1 is central, with points 2, 3, and 4 located 120 feet from point 1 at azimuths of 360, 120, and 240 degrees, respectively. Each point includes a microplot, a subplot, and an optional macroplot.

**population:** a basic building block of land area for which the number of plots and the land area being sampled are known. Typically, this is the county, but some counties may be grouped into super-counties due to small numbers of forested plots or to mask a large landowner.

**prescribed core variable:** all FIA regions produce a value for these variables and there is a prescribed national protocol for measuring or calculating these variables. Examples: d.b.h., azimuth, distance, species.

**prescribed optional variable:** a value is produced for these variables at the discretion of the FIA regions; but, when measured, the protocol must conform to prescribed national standards. Examples: magnetic declination, subplot condition list, sapling damage, percent rough cull.

**ratio of means:** an estimator which is computed as the ratio of the means of two random variates (attributes), such as the volume per acre of forested land.

**reserved forest land:** land permanently reserved from wood products utilization through statute or administrative designation.

**sampling unit:** the sampling unit is the basic unit of selection and observation. All FIA units use the center point of the 4-point cluster of subplots as the primary sampling unit.

**simple random sample:** a method of selecting  $n$  units out of the  $N$  such that every one of the samples has an equal chance of being chosen.

**site index:** the average total height that dominant and codominant trees in fully-stocked, even-aged stands will obtain at key ages, usually 25 or 50 years.

**site productivity class (or site class):** the maximum mean annual increment in cubic feet per acre that can be expected in fully-stocked, natural even-aged stands.

**sliver:** a condition class that occupies less than 25 percent of a plot (less than one full subplot and not encountered on other subplots).

**stocking:** at the tree level, stocking is the density value assigned to a sampled tree, usually in terms of numbers of trees or basal area per acre, expressed as a percent of the total tree density required to fully utilize the growth potential of the land. At the stand level, stocking refers to the sum of the stocking values of all trees sampled.

**strata:** nonoverlapping subdivisions of the population such that each primary sampling unit is assigned to one and only one subdivision (or stratum). The relative sizes of these strata are used to compute strata weights.

**stratification:** a statistical tool used to reduce the variance of the attributes of interest by partitioning the population into homogenous strata. It may also involve partitioning a highly variable but small portion of the population.

**stratified estimation:** estimation of population attributes using the total area of the population, strata means, and known strata weights. Strata means and weights are obtained from the stratification of the population either before or after the selection of sampling units.

**subplot:** a circular area with a fixed horizontal radius of 24.0 feet (1/24 acre), primarily used to sample trees at least 5.0 inches at d.b.h./d.r.c.

**subpopulation:** a subdivision of a population for which the area sampled is known and controlled for, such as the area within a county in national forest ownership. Sub-populations are not necessarily a subset of one single population.

**super-county:** a group of counties that have been combined to form a single population. Counties are combined into super-counties when the sample size for individual counties is too small.

**supervised classification:** training sites with known properties are used to extract spectral statistics from an image data set by interactively identifying sites in the image. These statistics are used to establish starting values for cluster means, and a clustering algorithm is used to classify the image.

**systematic sample:** a method of selecting  $n$  units out of the  $N$  such that restrictions are placed on the samples chance of being chosen.

**timberland:** Forest land that is producing or capable of producing in excess of 20 cubic feet per acre per year of wood at culmination of mean annual increment (MAI). Timberland excludes reserved forest lands.

**tri-areal plot:** a plot design that incorporates three different plot sizes at each sample location for the purpose of measuring trees in three different tree-diameter ranges.

**unsupervised classification:** radiance values of an image data set are used in a statistical clustering algorithm. The clusters are labeled after the classification.

## 8 Web-Based Supplementary Documentation

The supplementary documents referenced in this manuscript are posted on the Web site at [http://srsfia2.fs.fed.us/publicweb/statistics\\_band/stat\\_documents.htm](http://srsfia2.fs.fed.us/publicweb/statistics_band/stat_documents.htm).

The purpose of these documents is to provide details about the algorithms, equations, and other specifics of the FIA National Program that are too technical for the chapter discussions. These documents are posted on the Internet because they are dynamic. Currently they are in various stages of completion; periodically they will be updated and revised to accommodate changes in protocol and demand for technical detail. More information may be added to this Web site in the future as unresolved and new issues are presented and addressed. Supplementary documents available at the time of publication include:

- “The Hexagon/Panel System for Selecting FIA Plots Under an Annual Inventory”
- “Site Productivity Assignment for the North Central FIA Unit”
- “FIA Volume Calculations”
- “National Algorithms for Stocking Class, Stand-Size Class, and Forest Type”
- “Examples of FIA Point-in-Time Estimation Procedures for Several Common Cases”
- “Examples of FIA Change-Component Estimation Procedures for Several Common Cases”
- “Computation of FIA Plot Expansion Factors”





**Bechtold, William A.; Patterson, Paul L., Editors.** 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.

The Forest Inventory and Analysis (FIA) Program of the U.S. Department of Agriculture Forest Service is in the process of moving from a system of quasi-independent, regional, periodic inventories to an enhanced program featuring greater national consistency, annual measurement of a proportion of plots in each State, new reporting requirements, and integration with the ground sampling component of the Forest Health Monitoring Program. This documentation presents an overview of the conceptual changes, explains the three phases of FIA's sampling design, describes the sampling frame and plot configuration, presents the estimators that form the basis of FIA's National Information Management System (NIMS), and shows how annual data are combined for analysis. It also references a number of Web-based supplementary documents that provide greater detail about some of the more obscure aspects of the sampling and estimation system, as well as examples of calculations for most of the common estimators produced by FIA.

**Keywords:** Annual inventory, FIA, forest health monitoring, forest inventory, plot design, sampling frame.



The Forest Service, United States Department of Agriculture (USDA), is dedicated to the principle of multiple use management of the Nation's forest resources for sustained yields of wood, water, forage, wildlife, and recreation. Through forestry research, cooperation with the States and private forest owners, and management of the National Forests and National Grasslands, it strives—as directed by Congress—to provide increasingly greater service to a growing Nation.

The USDA prohibits discrimination in all its programs and activities on the basis of race, color, national origin, sex, religion, age, disability, political beliefs, sexual orientation, or marital or family status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD).

To file a complaint of discrimination, write USDA, Director, Office of Civil Rights, Room 326-W, Whitten Building, 1400 Independence Avenue, SW, Washington, D.C. 20250-9410 or call (202) 720-5964 (voice and TDD). USDA is an equal opportunity provider and employer.