



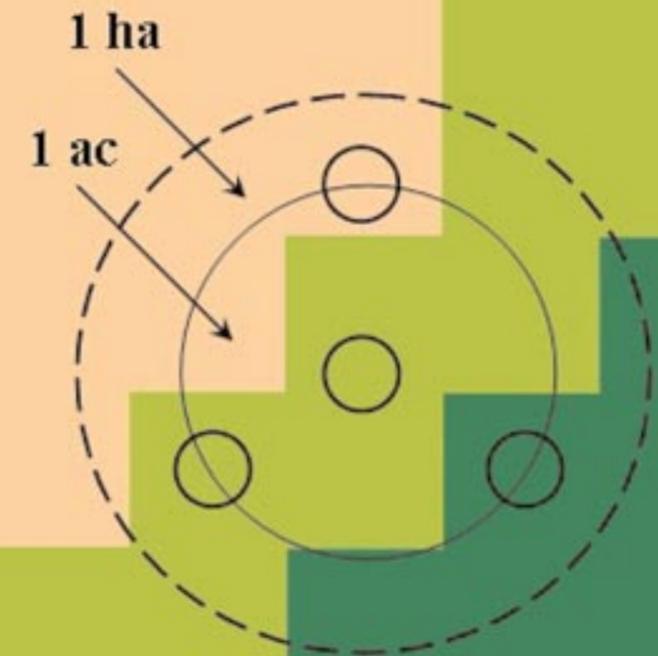
Proceedings of the Fifth Annual Forest Inventory and Analysis Symposium

New Orleans, Louisiana
November 18–20, 2003

United States
Department of
Agriculture
Forest Service
General Technical
Report WO-69
December 2005



Proceedings of the Fifth Annual Forest Inventory and Analysis Symposium





United States
Department of
Agriculture

Forest Service

Gen. Tech. Report
WO-69

December 2005



Proceedings of the Fifth Annual Forest Inventory and Analysis Symposium

Edited by

Ronald E. McRoberts

Gregory A. Reams

Paul C. Van Deusen

William H. McWilliams

New Orleans, LA

November 18–20, 2003

Sponsored by

Forest Inventory and Analysis, USDA Forest Service

National Council for Air and Stream Improvement

McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; McWilliams, William H., eds. 2005. Proceedings of the fifth annual forest inventory and analysis symposium; 2003 November 18–20; New Orleans, LA. Gen. Tech. Rep. WO–69. Washington, DC: U.S. Department of Agriculture Forest Service. 222 p.

Documents contributions to forest inventory in the areas of policy, assessments, statistics, remote sensing, information science, and forest health for the Forest Inventory and Analysis program of the USDA Forest Service.

KEY WORDS: annual forest inventory, assessments, forest health, information science, policy, remote sensing, statistics

Disclaimer

Papers published in these proceedings were submitted by authors in electronic media. Editing was done to ensure a consistent format. Authors are responsible for content and accuracy of their individual papers. The views and opinions expressed in this report are those of the presenters and authors and do not necessarily reflect the policies and opinions of the U.S. Department of Agriculture. The use of trade or firm names in this publication is for reader information and does not imply endorsement by the U.S. Department of Agriculture of any product or service.

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (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, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

Preface

The Fifth Annual Forest Inventory and Analysis Symposium was held in New Orleans, Louisiana, the second consecutive year at this location. Given the positive response to the 2002 symposium in New Orleans, we decided to return in 2003.

Each year of this symposium series the range of presentations has increased; 2003 was no exception, with several presentations related to forest health and several related to information science and distribution. of particular note for the 2003 symposium, we welcomed participation and a presentation from our Canadian

neighbors. The symposium organizers thank all participants and presenters and convey thanks to those who submitted their papers for these proceedings.

Ronald E. McRoberts
Gregory A. Reams
Paul C. Van Deusen
William H. McWilliams
St. Paul, Minnesota



Contents

Policy

Three Proposed Data Collection Models for Annual Inventories
Greg Reams, Bill Smith, Bill Bechtold, Ron McRoberts,
Frank Spirek, and Chuck Liff 1

Establishment of Canada's National Forest Inventory: Approach
and Issues

A.Y. Omule and Mark D. Gillis 9

Assessments

Northeastern Regional Forest Fragmentation Assessment:
Rationale, Methods, and Comparisons With Other Studies

Andrew Lister, Rachel Riemann, Tonya Lister, and Will
McWilliams 13

Southwestern Oregon's Biscuit Fire: An Analysis of Forest
Resources, Fire Severity, and Fire Hazard

David L. Azuma and Glenn A. Christensen 19

Assessment and Mapping of Forest Parcel Sizes

Brett J. Butler and Susan L. King 27

Incorporation of Precipitation Data Into FIA Analyses: A Case
Study of Factors Influencing Susceptibility to Oak Decline in
Southern Missouri, U.S.A.

W. Keith Moser, Greg Liknes, Mark Hansen,
and Kevin Nimerfro 33

Impact of Stream Management Zones and Road Beautifying
Buffers on Long-Term Fiber Supply in Georgia

Michal Zasada, Chris J. Cieszewski,
and Roger C. Lowe 41

Generating Broad-Scale Forest Ownership Maps: A Closest-
Neighbor Approach

Brett J. Butler 49

Impact of Definitions of FIA Variables and Compilation
Procedures on Inventory Compilation Results in Georgia

Brock Stewart, Chris J. Cieszewski,
and Michal Zasada 53

Statistics

K-Nearest Neighbor Estimation of Forest Attributes: Improving
Mapping Efficiency

Andrew O. Finley, Alan R. Ek, Yun Bai,
and Marvin E. Bauer 61

Proof of Concept for an Approach to a Finer Resolution
Inventory

Chris J. Cieszewski, Kim Iles, Roger C. Lowe,
and Michal Zasada 69

Can a Forest/Nonforest Change Map Improve the Precision of
Forest Area, Volume, Growth, Removals, and Mortality
Estimates?

Dale D. Gormanson, Mark H. Hansen,
and Ronald E. McRoberts 75

Stratum Weight Determination Using Shortest Path Algorithm
Susan L. King 83

A Three-Step Approach To Model Tree Mortality in the State
of Georgia

Qingmin Meng, Chris J. Cieszewski, Roger C. Lowe, and
Michal Zasada 91

Optimal Tree Increment Models for the Northeastern United
States

Don C. Bragg 97

Northeastern FIA Tree Taper Study: Current Status and Future
Work

James A. Westfall and Charles T. Scott 105

Mapped Plot Patch Size Estimates

Paul C. Van Deusen 111

FIA Quality Assurance Program: Evaluation of a Tree
Matching Algorithm for Paired Forest Inventory Data

James E. Pollard, James A. Westfall, Paul A. Patterson,
and David L. Gartner 117

Allocating Fire Mitigation Funds on the Basis of the Predicted Probabilities of Forest Wildfire Ronald E. McRoberts, Greg C. Liknes, Mark D. Nelson, Krista M. Gebert, R. James Barbour, Susan L. Odell, and Steven C. Yaddof.....	123	Forest Inventory and Analysis and Forest Health Monitoring: Piecing the Quilt Joseph M. McCollum and Jamie K. Cochran.....	175
Remote Sensing		Models for Estimation and Simulation of Crown and Canopy Cover John D. Shaw.....	183
Comparison of U.S. Forest Land Area Estimates From Forest Inventory and Analysis, National Resources Inventory, and Four Satellite Image-Derived Land Cover Data Sets Mark D. Nelson, Ronald E. McRoberts, and Veronica C. Lessard.....	129	Species Composition of Down Dead and Standing Live Trees: Implications for Forest Inventory Analysis Christopher W. Woodall and Linda Nagel	193
Accuracy Assessment of FIA's Nationwide Biomass Mapping Products: Results From the North Central FIA Region Geoffrey R. Holden, Mark D. Nelson, and Ronald E. McRoberts.....	139	Information Science and Distribution	
A Proposal for Phase 4 of the Forest Inventory and Analysis Program Ronald E. McRoberts.....	149	Access and Use of FIA Data Through FIA Spatial Data Services Elizabeth LaPoint	199
Forest Health		The Open-Source Movement: An Introduction for Forestry Professionals Patrick Proctor, Paul C. Van Deusen, Linda S. Heath, and Jeffrey H. Gove	203
Extending and Intensifying the FIA Inventory of Down Forest Fuels: Boundary Waters Canoe Area and Pictured Rocks National Lakeshore Christopher W. Woodall and Bruce Leutscher.....	159	A Knowledge Base for FIA Data Uses Victor A. Rudis	209
Detection Monitoring of Crown Condition in South Carolina: A Case Study William A. Bechtold and John W. Coulston	167	Comparison of Programs Used for FIA Inventory Information Dissemination and Spatial Representation Roger C. Lowe and Chris J. Cieszewski	215

Three Proposed Data Collection Models for Annual Inventories

Greg Reams, Bill Smith, Bill Bechtold¹, Ron McRoberts², Frank Spirek³, and Chuck Liff⁴

Abstract.—Three competing data collection models for the U.S. Department of Agriculture Forest Service Forest Inventory and Analysis (FIA) program's annual inventories are presented. We show that in the presence of panel creep, the model now in place does not meet requirements of an annual inventory system mandated by the 1998 Farm Bill. Two data-collection models that use subpaneling are defined, and the pros and cons of using those models are discussed.

The only data-collection model ensuring full compliance with the Farm Bill uses subpaneling with both spatial and temporal controls, resulting in the measurement of a single panel per year, nationally. The same field manual, portable data recorder, edit system, processing system, and estimation methods can be used within and among FIA regions. Such use will result in less duplication of effort and provide national consistency. The FIA program can produce, nationally, an annual database and annual estimates, as well as periodic reports based on 5-year-measurement requirements. Additional benefits will include the means to adjust measurement resources quickly and efficiently in order to measure resource availability by State. Additionally, the true sampling precision per fixed time-period is known, and intensification and detensifications are easy.

Introduction

The 1998 Farm Bill requires the Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture (USDA)

Forest Service to measure and process field plots at the rate of 20 percent per year, and to produce reports for each State at 5-year intervals. The legislation was designed to promote annual inventories based on a 5-year remeasurement cycle. Although the 20-percent per year requirement is explicit, the total number of plots by State or region on which the requirement is based was never specified—presumably to avoid micromanagement of the FIA sampling process. Optimistic that historic precision standards (3 percent per million acres of timberland and 5 percent per billion ft³ of growing stock volume) could be retained while implementing the new requirements, FIA established a systematic national plot network with an overall sampling intensity of 1 plot per 6,000 acres. The legislation required establishment of a database that could be used to produce annual or other estimates and publication of reports based on plots visited during the 5-year measurement periods. Also, advanced technologies such as remote sensing are to be developed and integrated into the program.

The FIA national plot network has been divided into five interpenetrating panels to accommodate the 20 percent per year requirement. Each panel uses overlapping samples (i.e., repeated observations on the same plots). Each panel of plots is characterized by complete and systematic spatial coverage across the population of interest (fig. 1). On completion of all panels, the process is then repeated with the next cycle of panel measurements. Ideally, all of the sample units in a panel are measured in the same way, and all sample units in a panel have the same revisit schedule. Panels can be divided into subpanels to accommodate decreases and increases of sampling intensity. When subpanels are selected systematically, such that each subpanel represents full spatial coverage, they are considered independent samples of the population, and population estimates can be calculated from the completed subpanels of an incomplete panel. Subpanels can be subdivided further into sub-subpanels, and sub-sub-subpanels as needed to accommodate planning and implementation of the survey program.

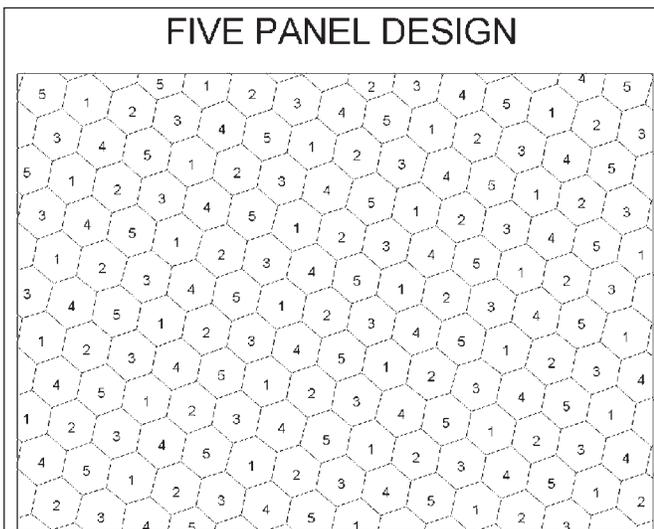
¹ Project Leader, Research Ecologist, and Research Forester, U.S. Department of Agriculture (USDA) Forest Service, Southern Research Station, Forest Sciences Laboratory, Forest Health Monitoring, 3041 Cornwallis Rd., Research Triangle Park, NC 27709.

² Mathematical Statistician, USDA Forest Service, North Central Research Station, 1992 Folwell Ave., St. Paul, MN 55108.

³ Computer Specialist, USDA Forest Service, Pacific Northwest Research Station, P.O. Box 1148; Corvallis, OR 97339.

⁴ Information Management Team Leader, USDA Forest Service, Rocky Mountain Research Station, 507 25th St., Ogden, UT 84401.

Figure 1.—National five panel design used by FIA. The national base sample has 1 plot per approximately 6,000-acre hex cell. All hexes labeled 1 are panel 1, hexes labeled 2 are panel 2, etc. Each panel comprises 20 percent of the entire FIA national base sample.



Although the five-panel system was designed to fulfill requirements specified in the 1998 Farm Bill, the FIA program was not adequately funded to measure the target number of plots (1 per 6,000 acres) at the specified rate of 20 percent per year. Even so, this seemed to have no serious negative effects, because several strategies could be used to preserve the integrity of the original design and still satisfy the legislated requirements. This article discusses those strategies and their relationship to the legislation, and it proposes other data-collection models that adequately address the legal stipulations.

Three Data Collection Strategies

The Farm Bill directly or indirectly mandates annual data collection, compilation, and inventory updates. Because FIA always has sought to retain the capacity for design-based estimation, we maintain that the optimal way to satisfy these requirements is by managing data collection efforts to produce temporally consistent panels.

Three data-collection strategies are proposed to mitigate the consequences of inability to measure plots at the rate of one complete panel per year. Under all of the methodologies described below, data collection rules are generally applied at the State

level and described as such. It is important, however, to note this is not absolutely necessary. FIA populations of interest, sometimes referred to as “estimation units,” usually are defined by political boundaries, i.e., counties or national forests. These estimation units are autonomous and additive, such that State-level estimates of inventory attributes are produced by aggregating data for all the estimation units that comprise a State. Complete and uniform spatial coverage is used to spread the samples evenly over the population to increase the likelihood of unbiased data processing and estimation (Reams *et al.*, in press). Because processing proceeds at the estimation-unit level, the sampling rules can be applied at this level. As long as the sampling rules result in complete and systematic spatial coverage for each estimation unit, there is no requirement that the sampling rules be uniform across estimation units within or among States.

Model 1: Creeping Panels

The creeping panel model removes the temporal restriction that each panel must be associated with exactly 1 year. Panels are started and finished based on the availability of funding and personnel. The lack of temporal restrictions allows the time required for panel completion to span several years, or it may proceed in the opposite direction, such that more than one panel is completed in a single year. Whatever the direction, any deviation from the measurement of exactly one complete panel of plots per year results in a situation that has been termed “panel creep.” The FIA management team proposed using and has now adapted the operational data-collection system known as the creeping-panel sampling strategy.

There are advantages to using this model. Of all the sampling strategies, creeping panels are the easiest to implement from a data-collection standpoint. Field logistics are simple because no special planning is required. Field crews are rarely required to backtrack over the same territory, except possibly to comply with the requirement that Phase 3 plots be measured during the growing season.

The disadvantages of this model, however, are numerous and inconsistent with the goals of an annual inventory. Because sampling rules vary by State, the lack of temporal control also implies a lack of spatial control such that no systematic annual coverage of a State or county can be assured. At a given time, different States may be measuring different panels, which complicates data retrievals and analyses across multiple States. Lack

of spatial and temporal controls create differences in sampling protocols that are difficult to track and manage. Two key inventory attributes, “panel number” and “manual version,” are allowed to vary by year. As field protocols change and new manual versions are released when field protocols change, there is no guarantee of consistency among States, regions, panels, or years regarding when specific manual versions are implemented. Several versions of data recorder software and processing systems must be maintained simultaneously.

Disparity between Phase 2 and Phase 3 sampling schedules may compromise FIA’s ability to combine data. For example, when panel creep is permitted for Phase 2 plots, but not for Phase 3 plots within the same panel, differences in the timing of panel measurements will be necessary. In some cases, field crews may have to use two different manual versions at the same time on a plot. These differences can quickly be exaggerated, as illustrated by the scheduling of panel measurements for the three data-collection models presented in table 1. For example, if data-collection is only funded at 80 percent, each and every Phase 2 panel is confounded with year after the first year, and Phase 2 and Phase 3 schedules are no longer synchronized (table 1).

Meeting significant Farm Bill standards cannot be guaranteed with this model. For example, data cannot be compiled annually, because panels are not scheduled for completion on an annual basis. Systematic spatial coverage across the population of interest, a prerequisite for standard processing, is not achieved using the creeping panel model. Required reports must be based on a variable number of panels if the data included represent a fixed, 5-year time interval. The 5-year reports will not be based on a 5-year interval if the data represent a complete set of panels.

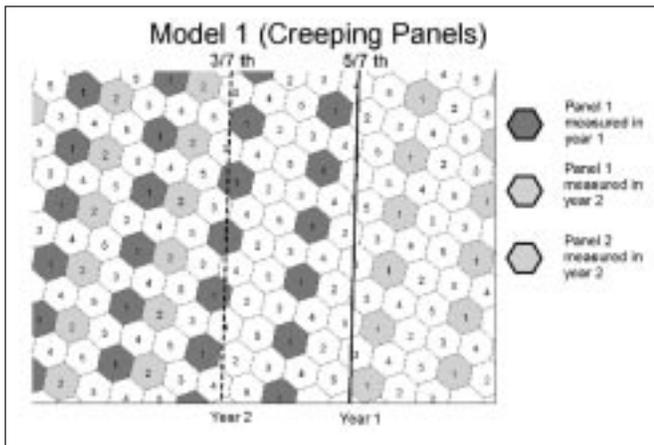
We have additional concerns about this model. If required 5-year reports are not based on a synchronic (5-year) interval, but instead include plots remeasured more than 5 years earlier, then precision estimates can be deceptive because the older data will artificially inflate the sample size. Old data are less reliable, and if substantial change has occurred in the intervening period lead to unknown bias in the inventory estimates. Usually such bias will not be reflected in the standard errors. The longer it takes to complete a panel, the greater the chance of spatially correlated measurement bias. Field crews usually start at one corner of the State, proceed until the field season ends, and then begin data collection where they finished in the next season. Figure 2 illustrates this “clumpy” approach to data collection. If a catastrophic event occurs between the first and second

Table 1.—Hypothetical data collection schedule for each of the three models. If about 80 percent can be measured, edited, and processed for an annual database, then measure P3 as subpanel 1 and P2 subpanels 2, 3, 4, and 5. Subpanel 1 is of size 1/16th and subpanels 2 through 5 are of size 3/16th. This results in measuring 13/16th or 81.25 percent of the entire full annual panel. For more exact matching of resources and data production, the P2 subpanels could be further subpaneled by size 1/16th.

Year	P3	P2 Data model		
		Model 1	Model 2	Model 3
1	1.1	begin 1	1.2, 1.3, 1.4, 1.5	1.2, 1.3, 1.4, 1.5
2	2.1	finish 1, begin 2	1.6, 2.2, 2.3, 2.4	2.2, 2.3, 2.4, 2.5
3	3.1	finish 2, begin 3	2.5, 2.6, 3.2, 3.3	3.2, 3.3, 3.4, 3.5
4	4.1	finish 3, begin 4	3.4, 3.5, 2.6, 4.2	4.2, 4.3, 4.4, 4.5
5	5.1	continue 4	4.3, 4.4, 4.5, 4.6	5.2, 5.3, 5.4, 5.5
6	1.1	begin 5	5.2, 5.3, 5.4, 5.5	1.2, 1.3, 1.4, 1.5
7	2.1	finish 5, begin 1	5.6, 1.2, 1.3, 1.4	2.2, 2.3, 2.4, 2.5
8	3.1	finish 1, begin 2	1.5, 1.6, 2.2, 2.3	3.2, 3.3, 3.4, 3.5
9	4.1	finish 2, begin 3	2.4, 2.5, 2.6, 3.2	4.2, 4.3, 4.4, 4.5
10	5.1	finish 3	3.3, 3.4, 3.5, 3.6	5.2, 5.3, 5.4, 5.5
11	1.1	begin 4	4.2, 4.3, 4.4, 4.5	1.2, 1.3, 1.4, 1.5
12	2.1	finish 4, begin 5	4.6, 5.2, 5.3, 5.4	2.2, 2.3, 2.4, 2.5

field seasons, population estimates derived from this panel would not accurately reflect the location or extent of resource damage. Remeasurement intervals may vary widely among panels and cycles. In the creeping panel example provided in table 1, panel 2 is measured during years 2 and 3. It is then remeasured in years 8 and 9. Thus, change will have to be computed using as many as 4 different intervals – years 2–8, 2–9, 3–8, 3–9. Because the data may have been collected in a spatially uneven manner, each interval could be associated with different environmental or cultural effects, yielding inaccurate or intractable trends for the panel as a whole.

Figure 2.—Collecting data from Panel 1 over 2 years—a spatially clumpy approach.



Model 2: Spatial Control

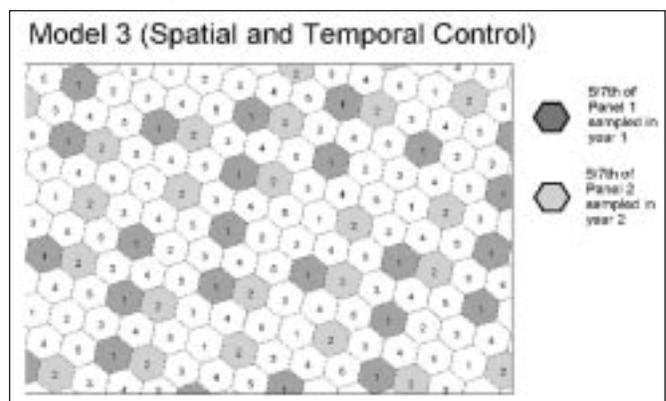
The spatial control model is an adaptation of the creeping panel model and was originally proposed by the FIA Statistics Band members and later advocated by the subteam (D-Team) of the FIA Information Band that is responsible for developing national data processing software. This model relies on subpanels to achieve systematic spatial coverage for the portion of a panel that can be measured in 1 year (Van Deusen 2003). A sufficiently large number of subpanels, each with systematic coverage of the population, are defined a priori. Phase 3 plots may be simply one of these subpanels, which would satisfy the requirement that a Phase 3 subpanel not be allowed to creep. Crews are assigned as many subpanels as can be measured in a field season. Measurement of the rest is postponed until the following year. Figure 3 illustrates a situation where five of seven subpanels are completed in year 1, the remaining two subpanels are done in year 2, and the result is systematic spatial coverage for both years.

This model offers significant improvements over the creeping panel. There is some guarantee that all States will have at least some subpanels measured and completed the same way in any given year, thus guaranteeing systematic coverage and simplifying data retrievals and analyses across multiple States. Differences in sampling protocols can be managed by establishing a rule that manual versions are linked specifically to year, to be implemented only at the beginning of a year.

More significantly, for meeting Farm Bill standards, data can be compiled on an annual basis, and required 5-year reports can be based on a fixed, 5-year interval. Also, precision is more accurately bound to sample size and less influenced by panel creep. The loss of precision caused by inadequate resources is immediately apparent and measurable, so that estimates of precision are not confounded by outdated, unreliable data. Also, annual data compilations reduce the potential for bias caused by catastrophic events. The matching of subpanels to the year in which a catastrophic event occurs eliminates the influence to an entire panel previously measured over a 2-year period.

Nonetheless, Model 2 does have drawbacks. More planning is required, which could complicate field logistics. Without careful prior planning, crews may not complete the prescribed number of subpanels or may have to make an additional pass through the State—if there is time to complete more subpanels. Inability to complete the prescribed number of subpanels might be overcome by revising the number of subpanels in the remaining estimation units to accommodate the shortage of resources. However, this complicates the tracking of which subpanels have been completed, and in which estimation units.

Figure 3.—Using subpanels to collect 5/7ths of Panel 1 in year 1 and 2/7ths in year 2.



Further, a less serious form of panel creep is still permitted. Although there is some guarantee that all States will have at least some subpanels completed the same way in a given year, the tracking of which panels and subpanels are measured when and where, is not straight forward. Although manual versions can be linked to years, the version used can still vary within a specific panel. Thus, disparities between Phase 2 and Phase 3 sampling schedules still exist (table 1). Assuming an 80 percent funding of data collection, as with Model 1, each and every panel is confounded with year after the initial year (table 1). Remeasurement intervals still may vary widely among panels, subpanels, and cycles. Remeasurement is not subject to any temporal control, so change estimation is not based on any fixed interval, thus complicating change analyses. Also, FIA may vary the sample size from year to year, which offers an advantage because it allows for annual database production.

Model 3: Spatial and Temporal Control

Model 3 has been proposed by members of the Statistics Band and advocated by members of the D-Team. It incorporates all of the spatial controls available in Model 2, but uses additional subpaneling to establish temporal limits as well, such that panel and year are perfectly coordinated. Model 3 differs from Model 2 in that the unfinished subpanels are simply skipped and will not be measured until the next cycle. Only plots in a single panel are measured in a given year, and the same panel is measured nationally. If a lack of resources makes it impossible to measure an entire panel for a given State, then the number of measured subpanels is adjusted accordingly. For example, the 2/7 subpanel scheduled for year 2 is postponed until the next-scheduled measurement of that panel during the next cycle (fig. 4).

Use of Model 3 reduces sampling intensity to more efficiently use limited resources. For example, the current Federal base intensity for FIA is 1 plot per 6,000 acres. If only 80 percent of that funding were available, then use of Model 3 would temporarily reduce plot intensity to one plot per 7,500 acres. When additional resources become available, the sample intensity can be increased. Intensification might be done with two strategies, or using a combination of those strategies:

1. The plot network for a given State is increased above the base sampling intensity of 1 plot per 6,000 acres.
2. Plots from future panels are temporarily assigned to the current panel.

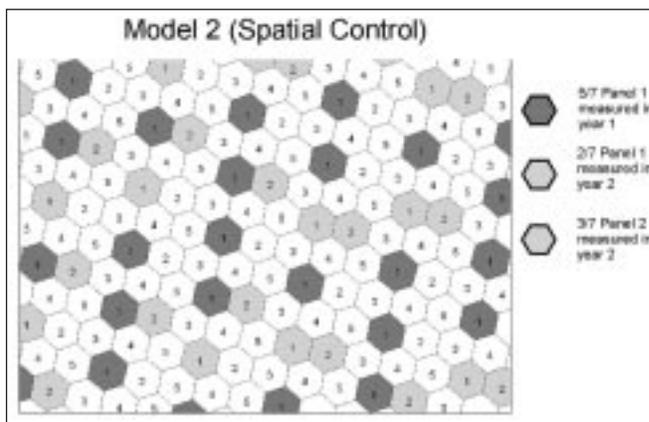
Strategy 2 has the advantage of accelerating change analysis. Suppose that a State had a short-term budget increase to measure plots in two panels for each of the next 3 years. Table 2 shows how panel 4 might be temporarily combined with panel 1, panel 5 with panel 2, and panel 1 with panel 3. This allows FIA to use the maximum number of plots for estimation of current inventory parameters, and makes change analyses possible in year 3. Choosing strategy 1 increases sample size and therefore is favored by those desiring more spatial information and precision, especially if the increased intensity can be maintained.

Model 3 has profound advantages with regard to fulfilling the Farm Bill standards. First, there is some guarantee that all States will have at least some subpanels measured and completed the same way in any given year, thus guaranteeing systematic coverage of the population and simplifying data retrievals and analyses across multiple States. Differences in sampling protocols can be managed by establishing a rule that manual versions are linked specifically to year, which automatically ties them to a specific panel. Disparities between Phase 2 and Phase 3 sampling

Table 2.—Hypothetical data collection schedule for panel acceleration under Model 3.

Year	Panel
1	1, 4
2	2, 5
3	3, 1
4	4
5	5

Figure 4.—Using subpanels to collect 5/7^{ths} of Panel 1 in year 1 and 5/7^{ths} of Panel 2 in year 2.



schedules are eliminated. Phase 3 plots are simply one of the subpanels scheduled for measurement in a given year. Most significantly, data can be compiled annually and 5-year reports can be based on a fixed, 5-year interval. Precision is more accurately bound to sample size and less influenced by panel creep. The reduced precision resulting from inadequate resources is immediately apparent and measurable; it is not confounded by outdated, unreliable data. Annual data compilations reduce the potential for bias introduced by the occurrence of catastrophic events. The catastrophic events are restricted to the year(s) in which such events occur, as opposed to influencing an entire panel over multiple years. No panel creep is permitted and no confounding of panel and year occur (table 1). Panels, subpanels, and manual versions are always linked to specific years. Remeasurement intervals are always based on a fixed interval (e.g., 5 or 10 years), if no borrowing of plots from future panels occurs. This model can readily accommodate the borrowing of plots from future panels to accelerate change analysis. The only accounting mechanism needed would be one code to designate temporary panel assignment.

The disadvantages of using Model 3 are fewer than for Model 2, but the two models share the following: More planning is necessary, which could complicate field logistics. In the absence of proper planning, crews may not complete the prescribed number of subpanels or may have to make an additional pass through the State, if there is time to complete additional subpanels. Inability to complete the prescribed number of subpanels could be overcome by revising the number of subpanels in the remaining estimation units to accommodate the shortage of resources. However, this complicates the accounting required to track which subpanels have been completed in which estimation units. Sample size is permitted to vary by year and panel, although this could be considered an advantage.

An FIA Precedent for Subpaneling

Within FIA the precedent for subpaneling, which is required when using Models 2 and 3, already has been established. To date, FIA has used Model 3 to create a Phase 3 subpanel (Subpanel 1) that is always measured annually—without creep. To accommodate the Farm Bill's requirements of annual surveys,

using Model 3 we only need to decide on a reasonable Phase 2 subpaneling strategy. To illustrate, consider defining six subpanels per panel. The Phase 3 subpanel is of size $1/16^{\text{th}}$, and is labeled Subpanel 1. Subpanels 2 through 6 are of size $3/16^{\text{th}}$ each and are labeled as Phase 2, Subpanels 2 through 6. If full funding for all six subpanels is available, FIA measures all six subpanels. If full funding is not available, it measures the Phase 3 Subpanel 1 and as many of the Phase 2 Subpanels 2 through 6 as possible. For example, if data from only about 40 percent of the plots can be measured, edited, and processed for an annual database, then the crew could measure all of Phase 3 (Subpanel 1), and two-fifths of Phase 2 (Subpanels 2 and 3). This results in measuring $7/16^{\text{th}}$ or 43 percent of the entire panel. If data from about 80 percent can be measured, edited, and processed for an annual database, crews could measure all of Phase 3 (Subpanel 1) and four-fifths of Phase 2 (Subpanels 2, 3, 4, and 5). This results in measuring $13/16^{\text{th}}$ or 81.25 percent of the entire annual panel. For more exact matching of resources and data production, the Phase 2 subpanels could be further subpaneled by size $1/16^{\text{th}}$ (table 1). Subpaneling in this manner guarantees production of annual databases, as well as spatially and temporally unbiased, design-based inventory estimates.

Conclusions

The FIA program is now using Model 1 in various regions, clearly in violation of the Farm Bill mandate. Moreover, Model 1 represents the worst possible compromise between annual and periodic inventories. It is not an annual inventory because annual databases and annual design-based estimates are not possible when panel creep occurs. When panel creep occurs, Model 1 is an inefficient periodic inventory where the only advantage gained from requiring crews to backtrack over the same area five or more times during an inventory cycle is the pretense of an annual inventory. Costs for plot production, training, and multiple versions of portable data recorders, as well as editing, processing, and estimation, are excessive.

Model 2 more closely meets requirements of the Farm Bill, although it creates numerous unnecessary challenges for data management and inventory estimation. By using spatial control, Model 2 results in a less serious form of panel creep, annual

databases are possible, and systematic coverage is assured, as well as annual and 5-year estimates. Although manual versions are linked to year, they can still vary by panel, and that will lead to a lack of national consistency. Also, disparities between Phase 2 and Phase 3 sampling schedules remain when using this model. Remeasurement intervals can vary widely among panels, subpanels, and cycles.

Model 3 is the only model that ensures full compliance with the Farm Bill. Subpaneling with spatial and temporal controls means one panel per year. The same panel is measured nationally. The same field manual, portable data recorder, edit system, processing system, and estimation methods can be used within and among FIA regions. This results in less duplication of effort and provides national consistency. The FIA program can produce an annual database nationally, annual estimates nationally, and periodic reports based on the required 5-year measurement period.

Additional benefits of Model 3 include the ability it gives FIA to quickly and efficiently adjust available measurement resources by State. Also, the true sampling precision per fixed time period of time is known, and intensification and detensifications are easy.

Recommendations

Data collection Model 3 is the only strategy that meets the requirements of the 1998 Farm Bill. The model provides for the greatest national consistency for sampling, database production, and inventory estimation. It will go a long way in helping FIA do its job.

Literature Cited

Reams, G.A.; Smith, W.D.; Hansen, M.H.; Bechtold, W.A.; Roesch, F.A.; Moisen, G.G. [In press.] The forest inventory and analysis sampling frame. In: Bechtold, W.A.; Patterson, P.L., eds. The enhanced forest inventory and analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.

Van Deusen, P.C. 2003. Issues related to panel creep. In: McRoberts, R.; Reams, G.; Van Deusen, P.; Moser, J., eds. Proceedings of the third annual forest inventory and analysis symposium; 2001 October 17–19; Traverse City, MI. Gen. Tech. Rep. NC-230. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 31–35.



Establishment of Canada's National Forest Inventory: Approach and Issues

A.Y. Omule¹ and Mark D. Gillis²

Abstract.—This paper describes Canada's National Forest Inventory (NFI) sampling design and implementation. It also describes issues related to annualizing the NFI using the approach of the U.S. Department of Agriculture Forest Service Enhanced Forest Inventory and Analysis program as a model. It concludes with an outline of plans to address the inventory annualization issues.

Canada's 10 Provinces and 3 Territories have a forest cover of approximately 418 million hectares, or nearly 50 percent of the country's total land mass. The forests are owned by the 10 Provinces (71 percent), the Federal Government and three Territories (23 percent), and private owners (6 percent). The Provinces and Territories are responsible for forest management, including management inventories. The Federal Government is responsible for the management and inventory of Federally administered forest lands and the compilation and reporting of the National Forest Inventory (NFI).

A new NFI has been designed (Gillis 2001) and is in various stages of implementation in partnership with the Provinces and Territories. The NFI provides accurate and timely national statistics on the extent, state, and changes over time of Canada's forests. This information is expected to support policy development, national and international reporting, and other emerging needs.

Sampling Design

The NFI sampling design is described in detail in the NFI documents available at the web site <http://www.nfi.cfs.nrcan.gc.ca> and published elsewhere (Gillis 2001, Wulder *et al.* 2002). Only a summary is provided here. The target population for the NFI is Canada's land mass, whether vegetated or not. This target

population, and thus the list of sampling units, is assumed to consist of infinitesimal points that are stratified for reporting purposes into 15 terrestrial ecozones (Ecological Stratification Working Group 1996). Ecozones are partitioned into subpopulations called NFI units for estimation purposes. An NFI unit is defined as an ecozone within the boundaries of a Province or a Territory. The partitioning of an ecozone into units is necessitated by differences in sampling schemes and different data collection methods that may otherwise occur within an ecozone.

The overall NFI sampling design is a probability sample of points in Canada, consisting of the following two components:

1. A single systematic sample of points across Canada, with a photo plot installed at each sample point (remote sensing).
2. A simple random subsample of the systematic sample within selected ecozones, with a ground plot established at or adjacent each subsample point (ground sampling).

The photo plots provide the primary source of the NFI data, and the ground plots provide additional information. Local variations in sampling design and data collection methods among the Provinces and Territories are permitted.

The remote sensing component involves a single systematic sample of locations (sampling points) located at the nodes of a national 20 km x 20 km fixed grid over the entire land mass. The grid sample locations are then grouped into ecozones (strata). A 2 km x 2 km permanent photo plot is installed at each sample location (fig. 1). The photo plot is partitioned into polygons or pixels using various classifiers (domains) based on aerial photo interpretation, other remotely sensed data such as the data from the Earth Observation for Sustainable Development (EOSD) project (Wulder 2002), and other data sources. Photo plots are the primary source of the NFI data. Attributes include area, biomass, carbon content, and volume by land cover class in a plot.

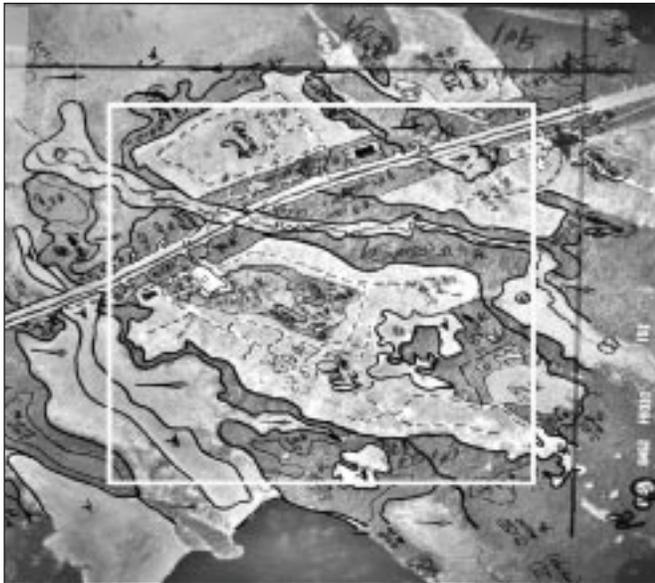
About 18,864 planned photo-plot locations exist in the country (table 1). Photo plot data are obtained from aerial photos in about one-half of these locations and from the EOSD data in

¹ Biometrician, National Forest Inventory, Canadian Forest Service, Natural Resources Canada, Pacific Forestry Centre, 506 Burnside Road, Victoria, BC V8Z 1M5. Phone: 250-363-0753; e-mail: magillis@pfc.forestry.ca.

² Manager, National Forest Inventory, Canadian Forest Service, Natural Resources Canada, Pacific Forestry Centre, 506 Burnside Road, Victoria, BC V8Z 1M5.

the remaining areas. In the three arctic ecozones and the prairies, only overall area totals (not broken down by classifier) will be obtained. The EOSD data may also be used as auxiliary information in the estimation of areas where both aerial photos and EOSD data exist.

Figure 1.—Example of a 2 km x 2 km NFI photo plot (1:20,000 midscale aerial photograph).



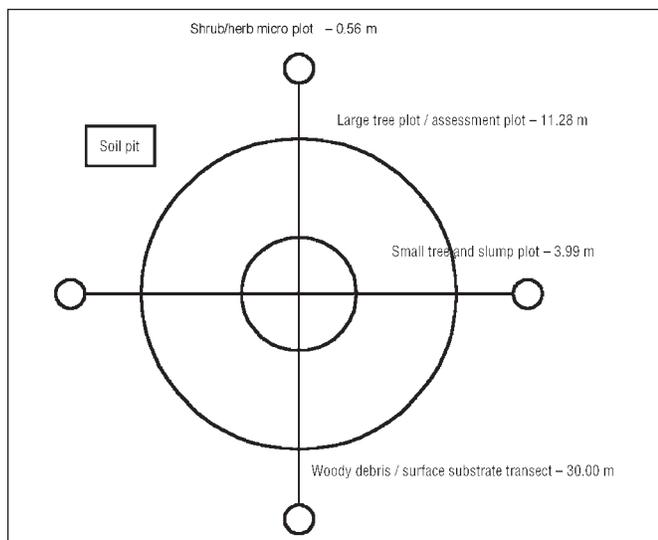
The ground-sampling component involves a simple random subsample of photo plot locations over a Province or a Territory or within an NFI unit. The nominal ratio of forested ground plots to photo plots is 1:10, with a minimum of 50 forested locations per ecozone, and no sampling in the three arctic ecozones. A permanent ground plot is established at or adjacent to the photo plot center, consisting of a cluster of nested circular plots, line transects, and a soil pit (fig. 2). The variables of interest include measurements or descriptions of trees, shrubs and herbs, woody debris, soils, and site. The nominal number of forested ground plots range from 0 to 163 per ecozone and from 0 to 268 per Province/Territory, for a national total of about 1,139 (table 1).

Ground plots are only established in forested locations that are classified as either “Vegetated Treed,” as locations with the potential to be classified as Vegetated Treed, or as locations that have been harvested. Nonforested ground plot locations are tracked over time and plots will be established when these locations become forested. Inaccessible plot locations are replaced with suitable subjectively selected matches, and difficult-access plot locations are subsampled.

Table 1.—Number of planned and outstanding (to March 2004) NFI photo and ground plots by province and territory.

Province or Territory	Photo plots					
	Ground plots		Aerial photos		Satellite imagery	
	Design total	Outstanding (03/04)	Design total	Outstanding (03/04)	Design total	Outstanding (03/04)
Alberta	166	166	1,656	1,656	0	0
British Columbia	268	50	2,414	100	0	0
Manitoba	96	28	950	698	665	482
New Brunswick	19	0	194	0	0	0
Newfoundland and Labrador	40	23	391	234	588	398
Northwest Territory	136	104	0	0	1,945	1,500
Nova Scotia	14	4	141	0	0	0
Nunuvut	0	0	0	0	287	287
Ontario	203	193	1,523	1,123	970	237
Prince Edward Island	2	0	12	0	0	0
Quebec	150	150	1,494	1,494	1,781	1,781
Saskatchewan	45	30	561	219	1,067	600
Yukon	0	0	0	0	1,225	225
Total	1,139	748	9,336	5,524	9,528	5,510

Figure 2.—Layout of the NFI ground plot (the dimensions refer to plot radius or transect length).



Implementation

The NFI is an interagency partnership project with decentralized implementation. The Provinces and Territories define the number and distribution of photo plots and ground plots in their jurisdictions, and collect and provide data to the Canadian Forest Service (CFS). The plots are installed and assessed following national photo interpretation and ground sampling guidelines. The program is implemented through bilateral agreements between the Federal Government and the Provincial/Territorial Governments. The Federal Government develops and maintains standards, procedures, and data infrastructure; provides data on Federal land; and conducts data analysis and reporting. The NFI project is coordinated by the CFS with guidance from Canadian Forest Inventory Committee (CFIC), a grouping of forest inventory managers and experts representing the Provinces, Territories, and the Federal Government (CFS, Parks, Agriculture and Agri-foods, and Environment).

To date, approximately 40 percent and 34 percent of the photo plots and ground plots, respectively, covering eight Provinces and one Territory, have been established. A pilot project is underway in one Province, and planning activities are ongoing in two Territories. The NFI plot establishment phase is expected to be completed by December 2006, allowing for a remeasurement to start in 2007.

The NFI plot establishment activities (over 5 years), including design, data collection, and systems development and management, are estimated to cost CAN\$13.7 million. The annual maintenance, including remeasurement, systems management, compilation, analysis, and reporting, is estimated to cost CAN\$2.4 million.

Annualization

The first NFI report based on the new design should be completed in 2006. It will include statistical tables and maps for key attributes by terrestrial ecozone. The key attributes include those identified by the CFIC and related to the Canadian Council of Forest Ministers Criteria and Indicators reporting. To meet the NFI objective of providing information in a timely fashion, annual NFI reports will be produced that also include estimates of change over time. The following issues related to annualizing Canada's NFI in a cost-effective and statistically defensible way must be addressed:

1. **Re-inventory sampling strategy.** How applicable to the NFI is a rotating panel model, such as that of the Enhanced Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture Forest Service? What are the implications of adding more plots or variables at a later time? How do we capture changes due to disturbances in the population over time?
2. **Plot remeasurement.** Can remeasurement intervals be varied by geographic areas (e.g., North versus South)? Is it acceptable to measure only some attributes in only some plots? How will remeasurement of destructively sampled plots (e.g., microplots clipped for shrub/herb biomass measurement) be performed?
3. **Change estimation.** How applicable is FIA's moving average estimator to our NFI? Should growth models, the Kalman filter, or other mode-based estimation be incorporated into the estimation? What is the impact of changes in strata definitions and boundaries over time on the estimation? What is the best procedure for estimation in NFI units with few plots or no plots?

These and other questions were discussed at a 2-day CFS workshop in Victoria, British Columbia, in March 2004. Workshop participants included biometricians from the FIA program, the CFS scientists and NFI staff, and representatives from the Provinces and Territories. The workshop concluded with several options and recommendations for consideration by the CFIC at its next meeting in June 2004 in Dawson City, Yukon. Methods for photo plot and ground plot remeasurements will then be developed in the light of the recommendations from the CFIC.

Acknowledgment

The authors thank Dr. Steen Magnussen, Canadian Forest Service, for reviewing an earlier draft of this paper.

Literature Cited

- Ecological Stratification Working Group. 1996. A National Ecological Framework for Canada. Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources Research and Environment Canada, State of Environment Directorate, Ottawa, Canada. (<http://sis.agr.gc.ca/cansis/publications/ecostrat/intro.html>).
- Gillis, M.D. 2001. Canada's national forest inventory, responding to current information needs. *Environmental Monitoring and Assessment* 67. 121–9.
- Wulder, M. 2002. Mapping the land cover of the forested areas of Canada with Landsat data. In: Proceedings, international geoscience and remote sensing symposium (IGARSS) and 24th Canadian remote sensing symposium; 2002 June 24–28; Toronto, Canada. Piscataway, NJ: IEEE.
- Wulder, M.; Kurz, W.A.; Gillis, M. 2003. National level forest monitoring and modeling in Canada. *Progress in Planning*. (61) 4.

Northeastern Regional Forest Fragmentation Assessment: Rationale, Methods, and Comparisons With Other Studies

Andrew Lister, Rachel Riemann, Tonya Lister, and Will McWilliams¹

Abstract.—Forest fragmentation is thought to impact many biotic and abiotic processes important to ecosystem function. We assessed forest fragmentation in 13 Northeastern States to gain a greater understanding of the trends in and status of this region's forests. We reclassified and then statistically filtered and updated classified Landsat imagery from the early 1990s, and devised analysis routines that allowed for automated processing of large areas. We discuss the rationale for the study and the choices made in data set preparation and analysis routines, describe the methods used, and compare our methods with those of other coarse-scale fragmentation studies.

Introduction

The U.S. Department of Agriculture Forest Service's Northeastern Forest Inventory and Analysis unit (NE-FIA) collects data relating to quantity, quality, distribution, and health of forests from a network of ground plots distributed uniformly across 13 Northeastern States. These data are summarized and used to produce annual reports of the trends in and status of the region's forest resources. In addition to tabular summaries (e.g., McWilliams *et al.* 2002), analytical reports are produced that integrate contextual information, social data, and historical perspectives to help users interpret the numerical data. Data on fragmentation provide contextual information. For example, two counties with similar forest-area percentages can have different landscape configurations. Interpreting tabular data within the context of landscape configuration will help us gain a better understanding of the status of the forest resource and aid regional planners and decisionmakers.

Forest fragmentation also is an important issue in the ecology community. The partitioning of large, homogeneous landscape units into smaller patches by human activities and other processes influences animal behavior, plant-seed dispersal, hydrological processes, and local weather conditions (Forman 1995), all of which affect our forests. Analyzing NE-FIA forestry data through the prism of forest fragmentation can help ecologists understand regional ecological patterns.

Our objective was to design an efficient, scalable process that would produce contextual data on forest fragmentation. Specifically, we wanted to (1) provide a rationale for assessing regional forest fragmentation, (2) describe the methods used in the assessment, and (3) compare our methods with those of other coarse-scale fragmentation studies.

The protocol we developed was tailored to NE-FIA's reporting needs. Past efforts entailed manually interpreting points on a grid superimposed over aerial photography. At each point location, fragmentation metrics were recorded (Riemann and Tillman 1999). Disadvantages of this approach include high labor and materials costs and a great dependence on the quality of the photointerpreter. With the completion of the National Land Cover Data (NLCD) (Vogelmann *et al.* 2001), a 30-m Landsat-based land use/land cover classification, and the development of APACK, an efficient software application for calculating fragmentation metrics (Mladenoff and DeZonia 2001), new opportunities have emerged for measuring landscape patterns over large areas.

Before designing the procedures used in the assessment, we developed the following rationale for the analysis: to provide information for analysts and others interested in interpreting NE-FIA data with respect to patch features that are commonly reported as having a direct or indirect influence on biological systems, e.g., the average size of contiguous forest patches, their degree of isolation from other patches, shape characteristics, and length of interface between the patches and other land cover types (Forman 1995).

¹ Research Forester, Research Forester, Research Forester, and Project Leader, respectively, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, FIA, 11 Campus Blvd., Ste. 200, Newtown Square, PA 19073. Phone: 610-557-4038; e-mail: alister@fs.fed.us.

We also developed a definition of forest patch that matched as closely as possible NE-FIA's definitions of "forest." For land use on an NE-FIA plot to be classified as forest, it must be at least 0.4 ha (1 acre) in extent and nearly devoid of human development (except for silvicultural treatments). For example, agricultural fields with trees or recreation areas with paths and undergrowth control would not be considered forest. On the basis of these criteria, a forest patch was defined as a contiguous area of forest cover that is at least 0.4 ha in size and differs sharply from its surroundings due to land use change, bisection by a road, or interface with a water feature such as a large river or lake. Including characteristics of forest structure in our definition might be preferable, but the data do not allow for finer distinctions beyond broad land use/land cover categories.

We are aware of only two other regional or superregional forest fragmentation assessments in the landscape ecology literature: studies by Riitters *et al.* (2002) and Heilman *et al.* (2002). Both used raster data from NLCD and devised algorithms for segmenting large images and calculating metrics. After reviewing their methods, we chose a different analytical approach, primarily because of the manner in which the NLCD data were preprocessed. Riitters *et al.* excluded roads in their analysis, and Heilman *et al.* included only major roads. We believe that the ecological effects of all road sizes are too important to ignore. Also, we wanted to correct the situation in which NLCD forest is over-predicted in areas with high tree cover but a nonforest land use, for example, a residential area with an extensive tree canopy. Because any metrics calculated depend on the accuracy of the source data set used (e.g., Riemann *et al.* 2003), we believe that this correction was critical. Finally, both Heilman *et al.* and Riitters *et al.* used subcounty-scale analysis units. NE-FIA produces statistical summaries at the county or multicounty scale, which requires different procedures than those used in the other two studies. We partitioned the landscape into political units (counties) to more closely match the reporting needs of NE-FIA.

Methods

Combining Imagery and Roads

We obtained NLCD data from the U.S. Geological Survey for the 13 Northeastern States under the purview of NE-FIA (fig. 1)

Figure 1.—The States that make up the study region: Ohio, West Virginia, Maryland, Pennsylvania, Delaware, New York, New Jersey, Connecticut, Rhode Island, Massachusetts, Vermont, New Hampshire, and Maine.

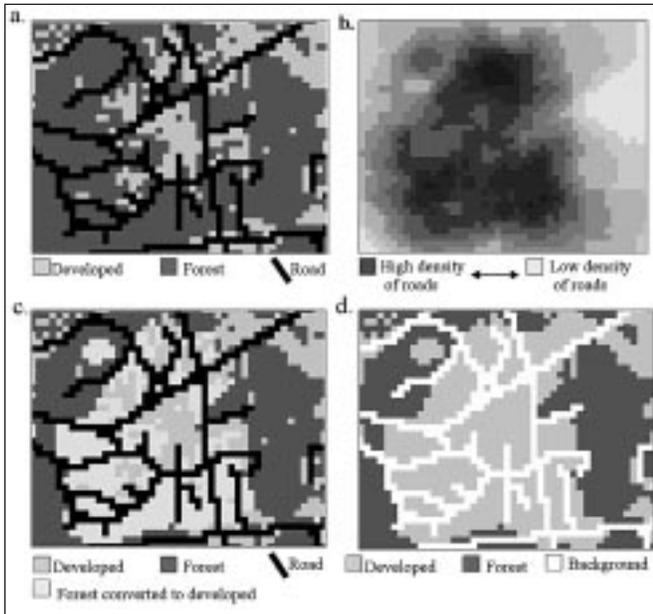


and then merged these data to create a contiguous, regional raster data set. We collapsed the original 21 NLCD classes into six new classes representing the land uses we were willing to consider together as a single patch (table 1) to create a new mosaic (*M*). We combined Geographic Information System (GIS) coverages of roads from the U.S. Census Bureau's TIGER/Line Files (U.S. Department of Commerce 2002) with *M* to create a new data set (*M+R*) in which each pixel of *M* that co-occurred with a road became a background or "no data" pixel in *M+R* (fig. 2a). In addition to boundaries created by roads, water and the edges of analysis units did not contribute to the edge measurements. We did encounter registration errors in various areas between *M* and the roads' data, but ignored them, assuming that the false patches created by these errors generally represent a marginal proportion of the total area and number of patches.

Updating and "Correcting" the NLCD Data Set for Missing Development

We had previously noted that NLCD overrepresented forest pixels in areas that include both development and high levels of

Figure 2.—(a) roads overlaid on the NLCD image in a simplified region that includes two classes (forest and developed) ($M+R$); (b) results of a convolution filter that provides an index of road density (RD); (c) results of a Boolean expression that replaces forest with high road density with developed ($M+R+F$); (d) the final map in which patches smaller than 0.4 ha have been removed, and roads have been converted to background.



tree canopy cover (Riemann *et al.* 2003). For these forested areas with higher road density, we applied a convolution filter (moving window) with a circular, seven-pixel-radius kernel to $M+R$ so that the count of road pixels within the kernel was calculated and attached to each pixel in the output (RD) (fig. 2b). This output was then evaluated using Boolean logic of the form “If a pixel in $M+R$ is forest, and the co-occurring pixel in RD has a value greater than 35, then update that pixel with the class ‘developed’ (table 1); otherwise, leave it with the original value of $M+R$.” We decided on a threshold of pixel values greater than 35 in RD as indicative of high road density through a heuristic approach using different thresholds and different areas of the study region.

Approximating NE-FIA’s Minimum Area Definition for Forest Land

To approximate NE-FIA’s area requirements for forest classification, we eliminated all isolated patches of pixels from the map updated in the previous step ($M+R+F$) (fig. 2c) that contained

fewer than four pixels of the same land cover type and replaced them with the majority land cover surrounding each updated pixel. This, in effect, defined the minimum mapping unit of $M+R+F$ as 3,600 m² (0.9 acres) (fig. 2d).

Analysis of Reporting Units and Automation

We defined several scales of reporting unit based on the interests of NE-FIA analysts and data consumers: county, watershed, ecoregion, and State. We obtained GIS layers for county and State boundaries from the U.S. Census Bureau (U.S. Department of Commerce 2002). We designed a series of GIS-based software programs that used these GIS layers to clip $M+R+F$ and process each resulting analysis unit using APACK software, as well as Environmental Systems Research Institute’s ArcInfo GIS. Fragmentation metrics for each land use class from table 1 and the landscape as a whole were compiled in tabular form for each analysis unit (table 2). We do not address our choice of metrics in this article.

Table 1.—Collapsing scheme used to convert NLCD or derived classes into our classification scheme (see text); background classes form patch boundaries but do not form patches.

Our class	NLCD or derived class
Developed	Residential, commercial, high road density forested
Barren	Quarries, gravel, bare earth, transitional
Forest	Deciduous, conifer, mixed, woody wetlands
Natural vegetation	Shrubs, grasslands, herbaceous wetlands
Agriculture	Pasture, row crops, grains, orchards
Background	Water, roads, areas outside of the analysis region

Discussion of Methods

Riitters *et al.* (2002) did not preprocess the NLCD data other than recoding them to forest/nonforest. This approach did not meet our objectives. During our initial analyses, we determined that a single string of pixels can connect two isolated forest patches, creating a “super patch” that constitutes a large portion of the land area of the analysis unit. The methods of Riitters *et al.* (2002) are based on a sliding window and were not meant to produce patch-based measurements, whereas our requirements dictated a patch-based approach. Also, we wanted

Table 2.—Examples of fragmentation metrics calculated for each analysis unit.

Percent land use in	Shared edge between forest and	Other metrics
Developed	Developed	Forest edge density
Barren	Barren	Avg. corrected patch perimeter-area ratio
Forest	Natural vegetation	Avg. normalized patch area
Natural vegetation	Agriculture	Patch size summary statistics
Agriculture		Patch size histograms
		Patch connectivity metrics

to retain information on different categories of nonforest land because the nonforest land use type bordering forests is believed to affect the ecology of that forest (Forman 1995).

Heilman *et al.* (2002) preprocessed the NLCD data using a subset of the TIGER/Line roads data—U.S. interstates and routes and State and county highways—to account for the super patch problem and in recognition of the ecological impacts of roads on terrestrial ecosystems (Trombulak and Frissell 2000). They also recoded the NLCD data to forest/nonforest, and, like Riitters *et al.*, lost information by grouping all nonforest land use types in a single category.

The work of Heilman *et al.* (2002) did not meet our objectives because they omitted road classes such as rural, neighborhood, and vehicular trails, and because the boundaries of their analysis units were formed by roads rather than by political boundaries. We believed that including all available roads was important because even unpaved forest roads strongly affect the local ecology (Haskell 2000). Further, Heilman *et al.* eliminated urban areas from their analysis. We included these areas because some urban areas have significant tree cover or marginal forest (Riemann 2003).

Our methods and those of Riitters *et al.* and Heilman *et al.* (2002) share several weaknesses. First, the NLCD data have varying accuracies (Yang *et al.* 2001). By collapsing the 21 NLCD classes (table 1), we no doubt raised the overall accuracy of the data set, although measuring this directly would be difficult. At best, the NLCD forest/nonforest accuracy rates tend to range from 80–95 percent across the study region (Yang *et al.* 2001).

Second, the spatial mismatch between the TIGER/Line roads data and the NLCD image can be substantial. We experimented with several ways to address this, e.g., buffering the roads, but believed that the additional inaccuracy introduced by using the

roads was offset by the ability to delineate meaningful forest patches in a way that met our definitions.

Third, the NLCD classification is driven by land cover. If an area is completely covered with tree canopy but is mowed beneath the tree canopy, e.g., in a town park, the NLCD might classify that area as forest, while the NE-FIA classification would be nonforest. This definitional mismatch is inherent in most satellite-based land cover classifications of vegetation. In our analysis, we assumed that all tree-covered areas greater than 0.4 ha are forest, although this supposition is not true. We made this assumption because no other consistently classified, national, land use/land cover maps exist at relatively fine scales. Provided that these deficiencies are recognized and understood, we believe that our method can effectively assess forest fragmentation at the regional scale.

Conclusions

One strength of our approach is the type of automation we developed. We were able to quickly and efficiently use a combination of GIS, spreadsheets, and C (programming language) programs to partition, preprocess, calculate metrics for, and compile tabular summaries of data in our analysis units. This flexibility allows us to generate metrics for any attribute of interest for which a GIS data source exists. Also, our preprocessing of the NLCD data adds an indicator of below-canopy fragmentation to areas that are tree covered on the NLCD image but replete with roads below the canopy. By including all available roads as patch-creating entities, we are in agreement with the prevailing view that any road size affects forests in numerous ways (Trombulak and Frissell 2000). Finally, by

eliminating patches that do not approximate NE-FIA's forest definition, we arrive closer to the point where we can mitigate the distinction between tree cover and NE-FIA's definition of forest.

Literature Cited

- Forman, R.T.T. 1995. Land mosaics: the ecology of landscapes and regions. Cambridge, MA: Cambridge University Press. 632 p.
- Haskell, D.G. 2000. Effects of forest roads on macroinvertebrate soil fauna of the southern Appalachian Mountains. *Conservation Biology*. 14: 57–63.
- Heilman, G.E.; Strittholt, J.R.; Slosser, N.C.; Dellasala, D.A. 2002. Forest fragmentation of the conterminous United States: assessing forest intactness through road density and spatial characteristics. *BioScience*. 52: 411–422.
- McWilliams, W.H.; Alerich, C.A.; Devlin, D.A.; *et al.* 2002. Annual inventory report for Pennsylvania's forests: results from the first two years. Resour. Bull. NE-156. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 71 p.
- Mladenoff, D.J.; DeZonia, B. 2001. APACK 2.17 user's guide. Madison, WI: University of Wisconsin, Forest Landscape Ecology Lab. 52 p.
- Riemann, R. 2003. Pilot inventory of FIA plots traditionally called "nonforest." Gen. Tech. Rep. NE-312. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 44 p.
- Riemann, R.; Lister, A.; Hoppus, M.; Lister, T. 2003. Fragmentation statistics for FIA: designing an approach. In: Proceedings, 3rd annual Forest Inventory and Analysis symposium; 2001 October 17–19; Traverse City, MI. Gen. Tech. Rep. NC-230. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 146–155.
- Riemann, R.; Tillman, K. 1999. FIA photointerpretation in southern New England: a tool to determine forest fragmentation and proximity to human development. Res. Pap. NE-709. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 12 p.
- Riitters, K.H.; Wickham, J.D.; O'Neill, R.V.; *et al.* 2002. Fragmentation of continental United States forests. *Ecosystems*. 5: 815–822.
- Trombulak, S.C.; Frissell, C. 2000. Review of ecological effects of roads on terrestrial and aquatic communities. *Conservation Biology*. 14: 19–29.
- U.S. Department of Commerce, Census Bureau. 2000. Census 2000 TIGER/Line Files [machine-readable data files]. Washington, DC: U.S. Department of Commerce, Census Bureau.
- Vogelmann, J.E.; Howard, S.M.; Yang, L.; *et al.* 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering & Remote Sensing*. 67(6): 650–662.
- Yang, L.; Stehman, S.V.; Smith, J.H.; Wickham, J.D. 2001. Thematic accuracy of MRLC land cover for the Eastern United States. *Remote Sensing of the Environment*. 76: 418–422.



Southwestern Oregon's Biscuit Fire: An Analysis of Forest Resources, Fire Severity, and Fire Hazard

David L. Azuma and Glenn A. Christensen¹

Abstract.—This study compares pre-fire field inventory data (collected from 1993 to 1997) in relation to post-fire mapped fire severity classes and the Fire and Fuels Extension of the Forest Vegetation Simulator growth and yield model measures of fire hazard for the portion of the Siskiyou National Forest in the 2002 Biscuit fire perimeter of southwestern Oregon. Post-fire severity classes are related to pre-fire torching indexes, and torching indexes seem to be correlated with the pre-fire volume of stands. Our analysis represents an initial look at fire severity, hazard, and forest attributes.

Fire is one of the most serious disturbances to occur on western landscapes (Agee 1993), and the 2002 fire season was one of the worst in recorded history in Oregon. Total acreage burned exceeded 800,000 acres. The largest fire was the Biscuit fire, which burned for 55 days before being contained. The Biscuit fire perimeter encompassed more than 500,000 acres in Oregon and California, with 460,000 acres in Oregon. Suppression costs in 2002 for southwestern Oregon, including the Apple, Tiller, Timber Rock, and Biscuit fires, were estimated to be more than \$200 million, with approximately \$150 million spent on the Biscuit fire alone.

The Biscuit fire offers a unique opportunity to study the effects of and recovery from a major wildfire on reserved and nonreserved lands. The reserved area includes the 180,000-acre Kalmiopsis Wilderness Area, which is entirely within the fire boundary. Nonreserved lands are all other areas and are considered potentially available for active forest management such as timber harvest. Forest types range from moist coastal Douglas fir to knobcone pine to a variety of hardwood types.

In the aftermath of the fire, land managers at the Siskiyou National Forest have had to deal with policy debates concerning salvage logging and whether or how to implement reforestation

measures. The size of the Biscuit fire has focused public attention on fire risk, forest health impacts, and salvage logging.

The U.S. Department of Agriculture Forest Service's Forest Inventory and Analysis (FIA) program collects and analyzes data that directly contribute information to the debate on post-fire recovery options. Our report describes the Biscuit fire study, presents several inventory results from analyzing the pre-fire data, and explains the data collection that occurred during the 2003 field season. We used the pre-fire information to describe each plot's characteristics, and then related these to post-fire severity estimates derived from the Burn Area Emergency Recovery (BAER) team's severity map. To evaluate how well post-fire severity related to pre-fire measures of fire hazard, we used the Fire and Fuels (FFE) (Reinhardt and Crookston 2003) of the Forest Vegetation Simulator (FVS).

Methods

For the Biscuit fire, we arranged the remeasurement in 2003 of all the Continuous Vegetation Survey (CVS) plots (Max *et al.* 1996) installed by the national forest system in the mid to late 1990s. Along with the standard plot remeasure, we added various post-fire descriptive parameters to help us make inferences about fire severity and its effects on trees and soils. Post-fire tree parameters included percent crown colors (percent of crown that was brown, green, and black), two scorch heights and azimuths, percent stem black, and cause of death. Additional ground variables included percent cover of new litter and depth, previous litter and depth, previous humus and depth, and percent charring levels on previous litter, humus, soil, rock, moss, brown cubical rot, liverworts, and lichens.

The pre-fire data set consists of a systematic grid of 180 field plots in the fire perimeter (fig. 1). Tree parameters include, but are not limited to diameter at breast height (d.b.h.), height, age, species, compacted crown ratio, and insect and damage

¹ Research Forester and Forester, respectively, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, P.O. Box 3890, Portland, OR 97208-3890; e-mail: dazuma@fs.fed.us.

information. Plot variables include standard site descriptors and coarse woody debris and understory vegetation cover transects. Some variables are used as inputs to the FFE model to estimate fire-related plot attributes such as torching index, crowning index, canopy bulk density, and canopy base height.

We used the BAER team's severity map and compared pre-fire data among four severity classes: very low or unburned, low, moderate, and high (Parsons and Orlemann 2002) (table 1). The map (fig. 1) was based on automated classification of 30-meter pixel Landsat 7 satellite imagery and was field-validated and

calibrated using ground crew and helicopter reconnaissance. The minimum mapping size for the severity classes was 50 acres. Using a geographic information system (GIS), the FIA field plot centers were overlaid on the severity map to assign a severity class to each plot.

We used the southcentral Oregon and northeastern California (SORNEC) FVS variant to estimate pre-fire conditions. The SORNEC variant includes the FFE fire model used to calculate canopy base height, torching, and crowning index for each plot. Torching index (TI), representing the wind speed at which fire could be expected to move from surface fuels into crown fuels, is highly influenced by vertical stand structure (ladder fuels). The higher TI values indicate that a higher wind speed is required to move the fire into the crowns, giving these TI values a reduced fire hazard compared with plots that have a lower TI. Crowning index (CI), the wind speed at which a crown fire could be expected to be sustained, is heavily influenced by crown bulk density (Van Wagner 1977). Both TI and CI are used as indexes of potential fire hazard. FFE model defaults are used for pre-fire estimates of all plots. These defaults include these assumptions: 70° Fahrenheit temperature, a 20-foot wind speed of 6 miles per hour, and moisture assumptions that depend on predicted fuel loadings. Fuel loading predictions are estimated from cover type and structural stage. After fuel loading was determined for a particular plot, the plot was assigned one of 13 stylized fuel models dependent on the specific FFE variant. Potential fire estimate calculations were done using Rothermel's 1972 fire behavior prediction system (Reinhardt and Crookston 2003).

Figure 1.—Map of Biscuit fire by BAER burn severity class. Continuous vegetation survey (CVS) inventory plots included in the fire boundary are on a 3.4-mile grid for reserved lands and a 1.7-mile grid for nonreserved lands.

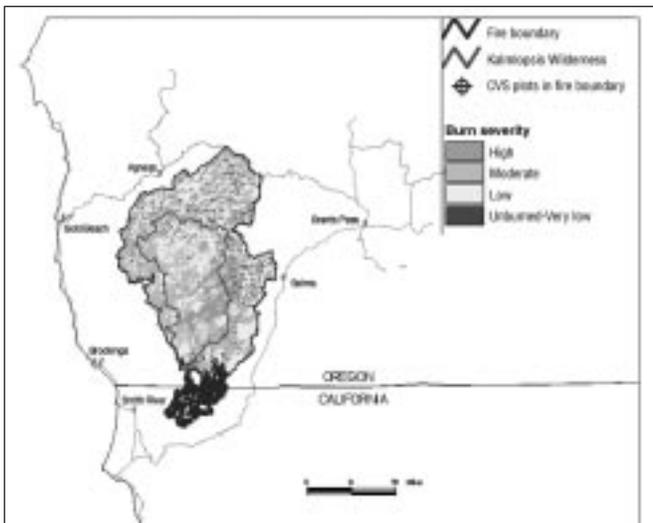


Table 1.—Description of Burn Area Emergency Recovery (BAER) severity classes.

Collapsed class	BAER burn severity	Description of fire effects
Low/unburned	Very low or unburned	Mosaic of unburned and very low severity ground fire. Consumption of ground cover and vegetation mortality is minimal. Canopy remains vigorous and green. Mortality of trees and shrubs is slight.
	Low	Vegetation is lightly scorched, large trees are mostly not killed, and very small diameter fuels consumed.
	Moderate	Much of the litter has been consumed. Fine fuels close to the ground may be all consumed, and trees may exhibit 40- to 80% mortality.
High/moderate	High	Complete consumption of tree crowns has occurred, few to no leaves or needles remain on trees, and mortality can be assumed to be close to 100%.

Site Description

The Siskiyou is one of the most florally diverse national forests. The complex geology, transverse nature of the Siskiyou Range, and extreme climate conditions combine to provide a variety of niches. The inventory plots contained 14 coniferous and 12 hardwood species. The Forest has 100 different sensitive species and 15 different plant series that can be divided into 92 different plant associations.

Inventory Overview

Within the Biscuit fire perimeter, six major forest types comprised more than 85 percent of the land area. Nonstocked area and forest types with less than 5 percent of the area accounted for the remaining land area. Douglas fir was the most prevalent forest type inside and outside the reserved area, followed by tanoak and western white pine. Generally, few differences existed between the percentage of area by forest type when comparing the reserved and nonreserved areas (table 2).

We estimated pre-fire volume to be 7.02 billion board feet, of which 2.4 billion occurred in the Kalmiopsis Wilderness Area and 4.6 billion on nonreserved lands. The proportion of the total volume by species is constant between the reserved and nonreserved areas and is presented as a combined statistic in table 3. The major species regardless of land status is Douglas fir, with around 70 percent of the volume, followed by sugar pine and tanoak. The average torching index, crowning index, and canopy base height are also presented as a combined statistic for reserved and nonreserved lands by forest type in table 3.

Results

Although the fire area is divided by reserved and nonreserved status, the forest statistics are similar. The percentage of area in the four major softwood types (Douglas fir, sugar pine, western white pine, and Jeffery pine) is the same for the reserved and nonreserved (table 2). The greatest discrepancy appears in the age class distribution, in which the reserved and nonreserved have 53 percent versus 37 percent of the area in the 100–200-year age class. Surprisingly, the nonreserved area has a greater percentage of area in the 200 and older year age class, with 29 percent, versus 15 percent in the reserved (table 4).

The relationship between burn severity and land status is presented in table 5 by area and percentage. The percent of acres burned in the various severity classes are relatively close between the reserved (45 percent) and nonreserved (35 percent) areas in the high and moderate severity classes combined.

The plots that have higher TI tend to be in areas classed as low or very low severity and to have more volume per acre than in areas classed as high severity (table 6). Figure 2 shows a similar relationship between TI wind speed and average plot volume by graphing the average TI by average volume per acre for both the high/moderately burned stands and the low/unburned stands. TI tends to be highest for plots with more volume per acre than those with less volume. Also, as figure 2 shows, the TI tends to be lower for the plots that were most severely burned when compared to those with little or no burn damage.

Table 2.—Estimated pre-fire area and percent by reserved status by forest type for the Oregon portion of the 2002 Biscuit fire, Siskiyou National Forest.

Forest type	Nonreserved area		Reserved area	
	Acres	Percent	Acres	Percent
Douglas fir	132,000	46	72,800	42
Jeffery pine	14,400	5	—	—
Sugar pine	6,100	2	12,600	7
Western white pine	28,000	10	23,900	14
Canyon live oak	15,500	5	10,100	6
Tanoak	43,200	15	33,300	19
Nonstocked	14,800	5	7,500	4
Others	31,500	11	14,400	5
Total	285,500	100	174,600	100

Table 3.—Estimated pre-fire gross board foot volume, torching index, crowning index, and canopy base height by forest type, 2002 Biscuit fire, Siskiyou National Forest.

Forest type	Total volume	Torching index	Crowning index	Canopy base height
	Million board feet, Scribner rule	Average (m.p.h.)	Average (m.p.h.)	Average (feet)
Douglas fir	4,971	7.9	42.2	13.9
Jeffrey pine	113	4.4	105.1	13.4
Sugar pine	804	4.7	71.4	11.5
Western white pine	163	2.3	50.8	6.5
Canyon live oak	18	8.5	104.5	21.9
Tanoak	335	12.0	76.3	25.8
Others	614	5.9	84.5	16.2
All species	7,018	6.9	60.6	14.5

Table 4.—Estimated pre-fire area and percent of total fire area by reserved status and stand age class for the 2002 Biscuit fire, Siskiyou National Forest.

Age class Years	Nonreserved area		Reserved area	
	Acres	Percent	Acres	Percent
20–39	7,500	3	0	0
40–59	9,600	3	—	0
60–79	32,600	11	17,000	10
80–99	28,800	10	32,500	19
100–119	25,400	9	23,400	13
120–139	12,500	4	24,500	14
140–159	30,700	11	6,300	4
160–179	18,600	6	31,400	18
180–199	20,800	7	6,300	4
200–299	68,600	24	25,800	15
300 and older	14,000	5	0	0
Nonstocked	14,400	5	0	0
Total^a	283,600	100	174,600	100

^a One plot was not assigned an age class, accounting for the discrepancy in area totals.

Table 5.—Estimated pre-fire area, standard error, and percent by reserved status and BAER burn severity class for the 2002 Biscuit fire, Siskiyou National Forest.

Burn severity	Nonreserved area		Reserved area	
	Acres	Percent	Acres	Percent
High	38,000	13	33,000	19
Moderate	62,000	22	46,000	26
Low	117,000	41	56,000	32
Very low	68,000	24	39,000	22
Total^a	285,000 (4)	100	175,000 (5)	100

^a Standard errors for the totals are presented.

Table 6.—Average volume, large to small tree ratio, down woody material volume, percent understory cover, torching index, crowning index, and canopy base height with standard errors by BAER burn severity class.

	BAER burn severity			
	High	Moderate	Low	Very low
Volume per acre (1,000 board feet, Scribner rule)	10.6	7.8	17.6	24.7
Standard error	2.5	1.7	1.8	2.6
Small tree to large tree ratio ^a (number of trees)	94.5	102.6	76.8	30.0
Standard error	24.1	26.3	13.3	4.6
Down woody material > 5 inches in diameter (tons/acre)	4.23	4.46	9.14	5.58
Standard error	1.84	1.27	1.97	1.45
Cover in 1–5-foot height class (%)	43	43	31	23
Standard error	5	3	4	4
Torching index (miles/hour)	4.2	4.2	7.0	10.2
Standard error	1.0	1.2	1.4	1.6
Crowning index (miles/hour)	74.8	66.6	48.3	59.6
Standard error	12.2	7.9	4.2	5.1
Canopy base height (feet)	11.2	9.2	14.2	20.9
Standard error	2.2	1.5	2.1	3.2

^a Computed as the number of nonsawtimber trees to sawtimber trees.

In a chart of average TI by burn severity and site class (fig. 3), two important trends become apparent. The average torching index tends to decrease (representing an increased hazard) as site class declines, regardless of burn severity, and the most severely burned areas tended to have lower torching indexes than areas burned only slightly or not at all. Because a positive correlation between a stand's potential volume per acre and site quality exists, figures 2 and 3 both show a decrease in torching index with declining site quality or stand volume. The areas in the Biscuit fire that had more standing volume, and which also tended to be on the more productive sites, suffered less fire impact as measured by BAER burn severity.

The relationship between severity class and crowning index is not as strong. The smallest mean crowning indexes were in the low and very low severity classes, but the difference in crowning indexes between low/very low and moderate/high burn severity is small. Perhaps one explanation for this result is that when the fire is in the crowns, it takes about the same wind to move it through the crowns. These data suggest that the plots that burned the most severely are those that had lower torching indexes, as indicated by the FFE estimates. The results listed in table 3 show how TI, CI, and canopy base height differ by the major forest types in the burn area. The average CI for all forest types is high and indicates a lower probability of a fire moving through the crowns than the chance of it moving from a ground fire up into the canopy.

Some interesting differences in torching index were found among forest types. Pine forest types (Jeffery pine, sugar pine, and western white pine) tended to be more susceptible to torch-

ing than the other major forest types (fig. 4). We found that the forest types with the highest pre-fire TI (lowest hazard) are the two major hardwood types, canyon live oak and tanoak. The percentage of BAER severity by forest type results support the TI by forest type results. As with TI, the three pine types tended to have the greatest percentage of area in the two highest severity categories, moderate and high (fig. 5). The BAER mapping also showed that the two major hardwood types, canyon live oak and tanoak, had the least percentage of burn in the two most severe classifications.

Discussion

Data from the FIA plots in the Biscuit fire boundary enable exploring how pre-fire plot characteristics relate to post-fire severity estimates based on the BAER map. One of our most important findings is that plots with a higher timber volume and on more productive sites appear to have lower torching hazard; as the trees increase in size, they tend to dominant the stand, excluding the brush component and smaller trees (table 6), thereby reducing the ladder fuels and reducing the incidence of torching. The low and very low severity plots also had the highest canopy base heights, although only the very low severity plot base height was significantly different. The generally higher canopy base heights and lower ladder fuels appear to be related to the lower burn severity. After the ground-based measurements of severity are analyzed, this hypothesis can be further explored.

Figure 2.—Average torching index by BAER burn severity and stand volume class. As average volume 1,000 board feet/acre increases, torching index tends to also increase suggesting the possibility of decreased torching potential in the stand.

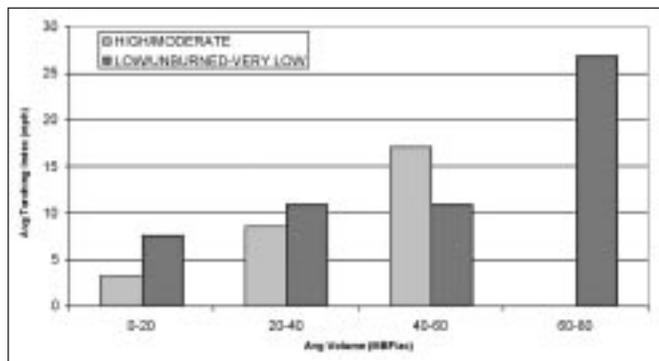
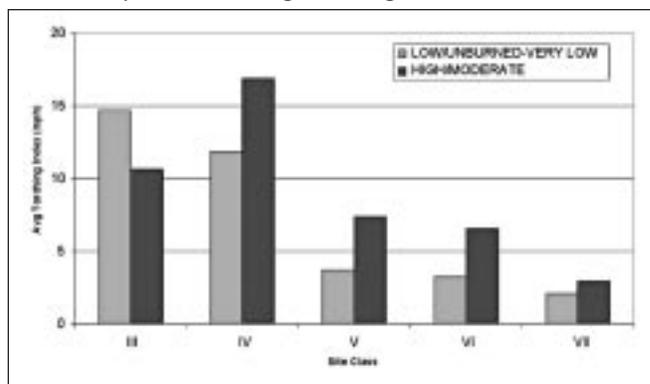


Figure 3.—Average torching index by BAER burn severity by average torching index and site class. High productivity is indicated by lower site classes. High probability of torching is indicated by a lower average torching index.

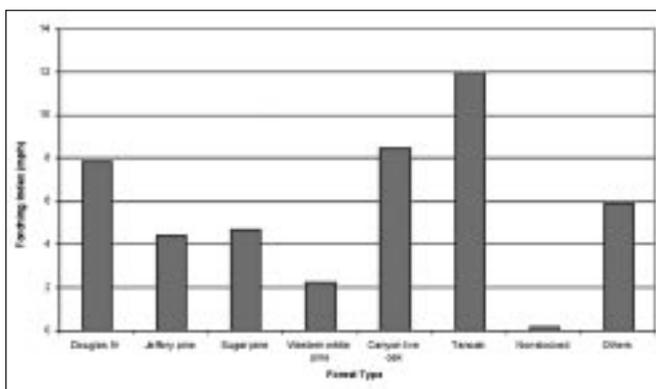


For our analysis, we found that using FFE to provide pre-fire estimates of torching index appears to be a useful approach to identify areas that may be severely impacted by wildfire. Further analysis of post-fire ground-based measurements of severity, however, will provide better estimates of the nature and strength of this relationship. Crowning index did not provide the same utility in terms of predicting burn severity, but this may be due to factors specific to this fire, such as forest types and stand structures, topography, and the intensity of this fire.

The differences in average torching index by forest type is another important finding from this analysis. A clear difference exists between the torching index prevalent for the three major pine types (Jeffery, western white, and sugar) and the high TI prevalent for the two major hardwood types (fig. 4). Like our findings of the relationship between torching index and burn severity, these results can only be confirmed definitively with ground-based measurements.

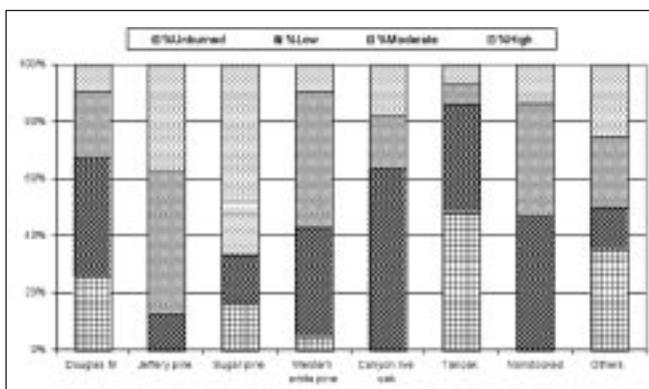
Our study is the first analysis relating various fire modeling parameters to burn severity and pre-fire stand characteristics. The reserved and nonreserved areas are similar in terms of area burned by BAER severity class but differ by stand age class distribution. The nonreserved acres had the largest area in the oldest age class. We found that the plots classed as low or very-low burn severity tended to have more volume per acre than the more severely burned plots. These plots also tended to have less cover in the stand's small tree and brush component and have a higher average canopy base height. Using the FFE model to

Figure 4.—Average torching index by forest type. Pine forest types tend to have a lower torching index wind speed, indicating a potentially greater probability of torching occurring in these types.



estimate specific fire related parameters such as torching and crowning index, we found BAER burn severity was closely related to torching index, but much less related to crowning index. Torching index was less for those plots that were the most severely burned, suggesting that this may be a useful variable to assess current fire hazard. Torching index also tended to drop when volumes per acre and site quality decline. Azuma *et al.* (in press) provides a more in-depth explanation of various resource parameters and their relationship to burn severity. All the results presented here use the BAER team's remote sensing based method of classifying fire severity and may not reflect fine-scale, on the ground differences in severity. Future research will include analyzing post-fire measures of severity on stands and the relationships between pre-fire stand characters (topographic and vegetative) and fire weather.

Figure 5.—Percentage of major forest types by BAER severity class. Pine forest types tended to exhibit an increased occurrence of the more severe BAER classes, moderate and high.



Literature Cited

- Agee, J.K. 1993. Fire ecology of Pacific Northwest forests. Washington, DC: Island Press. 493 p.
- Azuma, D.L.; Donnegan, J.; Gedney, D. [In press.] Southwest Oregon Biscuit fire: an analysis of forest resources and fire severity. Res. Pap. PNW-RP-560, Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 32 p.

Max, T.A.; Schreuder, H.T.; Hazard, J.W.; *et al.* 1996. The Pacific Northwest Region vegetation and inventory monitoring system. Res. Pap. PNW-RP-493. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 22 p.

Parsons, A.; Orlemann, A. 2002. BAER (burned area emergency rehabilitation) burn severity mapping: methods and definitions. http://www.biscuitfire.com/burn_severity.htm. [Date accessed unknown.]

Reinhardt, E.; Crookston, N., tech. eds. 2003. The fire and fuels extension to the forest vegetation simulator. Gen. Tech. Rep. RMRS-GTR-116. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 209 p.

Van Wagner, C.E. 1977. Conditions for the start and spread of crown fire. *Canadian Journal of Forest Research*. 7: 23–34.

Assessment and Mapping of Forest Parcel Sizes

Brett J. Butler and Susan L. King¹

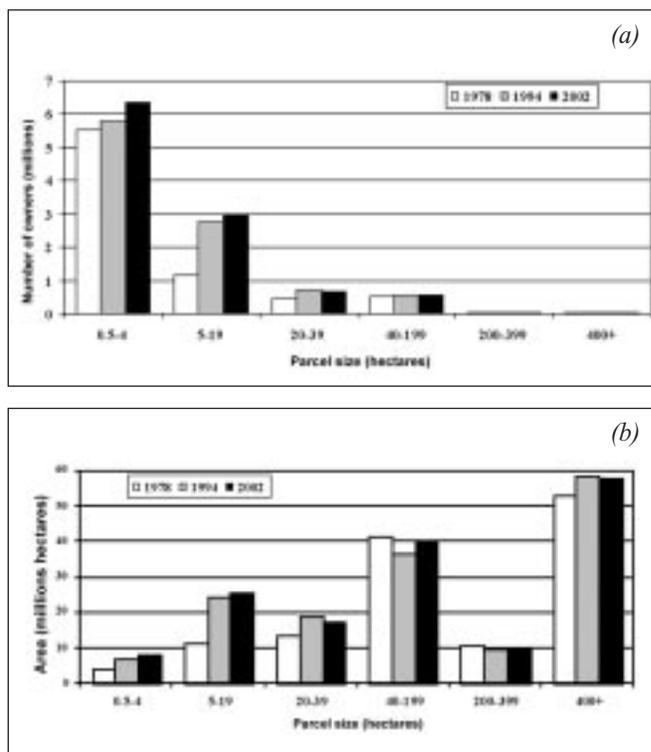
Abstract.—A method for analyzing and mapping forest parcel sizes in the Northeastern United States is presented. A decision tree model was created that predicts forest parcel size from spatially explicit predictor variables: population density, State, percentage forest land cover, and road density. The model correctly predicted parcel size for 60 percent of the observations in a validation data set (weighted kappa = 0.45). This decision tree model was used to create a map representing the average forest parcel size across the region.

Introduction

Our Nation's forest resources are becoming increasingly constrained by social factors. Wear *et al.* (1999) found a negative relationship between population density and harvesting probability, with harvesting probabilities approaching zero as population densities approached 58 people/km² (150 people/mi²). One factor correlated with harvesting activity is the area of forest land owned by individual forest landowners. Preliminary results from the National Woodland Owner Survey (NWOS) (Butler and Leatherberry 2003) suggest that parcel size also is correlated with ownership objectives, management activities, and future land use intentions.

Over the past 25 years, the number of private forest landowners in the United States increased from 7.8 million to 10.6 million (Birch *et al.* 1982, Butler and Leatherberry 2003). Most of this increase comes from landowners who own less than 20 ha (50 ac) of forest land (fig. 1). This increase in the number of landowners with smaller parcels is known as forest parcelization. As Sampson and DeCoster (2000) stated, forest parcelization is a major factor leading to “a growing crisis in maintaining sustainable private forests” (Samson and DeCoster 2000, 6). The financial difficulties of managing smaller parcels

Figure 1.—Size of forest landholdings as a function of (a) number of owners and (b) area of forest land owned in the United States in 1978, 1994, and 2002.



of forest land (i.e., economies of scale) (Row 1978) are partially responsible for the loss of working forest lands; the changing characteristics of the landowners (Egan and Luloff 2000), including changing ownership objectives and management practices, is another contributing factor.

To date, most information about forest parcel size across broad geographic regions has been produced in a tabular format (e.g., Birch 1996). Information about forest parcel size could be improved by providing the data in a spatially explicit (i.e., map) format to enable visual inspection of spatial patterns and combining with other spatial data sources to conduct further analyses. Estimation techniques must be employed because no nationally available spatially explicit data sources on forest parcel size exist.

¹ Research Forester and Operations Research Analyst, respectively, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, 11 Campus Boulevard, Suite 200, Newtown Square, PA 19073. Phone: 610-557-4045; fax: 610-557-4250; e-mail: bbutler01@fs.fed.us.

By combining point-based estimates of forest parcel size with spatially explicit ancillary data, we model forest parcel size across the Northeastern United States and generate a forest parcel map. In previous research (King and Butler, in press), we explored linear regression, neural network, and decision tree techniques for generating this model. We found the linear regression models had low predictive power (i.e., poor R^2 values) and required transformations of multiple variables, including the dependent variable. The model building process associated with the neural networks was complex and difficult to repeat, and the results were difficult to interpret. The decision tree models were easy to implement, produced results that were easy to interpret, and appeared to have relatively high predictive power. Because of these factors, we opted to use decision tree models for this study. The accuracy and goodness-of-fit of the decision tree modeling technique are tested using a validation data set and quantified using a confusion matrix and a weighted kappa statistic. Potential to improve these methods and future research directions also are discussed.

Methods

To generate a forest parcel map for the Northeastern United States, we created a decision tree model that predicted parcel size as a function of a set of independent variables. The Northeastern United States was selected because of data accessibility. Similar data have been collected for all the contiguous States, and the methods employed in this study can be expanded to this broader area.

The dependent variable was derived from the U.S. Department of Agriculture Forest Service's Forest Inventory and Analysis program's NWOS program (Butler and Leatherberry 2003). The NWOS randomly selects private forest landowners from across the United States and collects information, including size of landholding. The landowners were selected by assessing the forest cover at a random set of points across the United States; for those points determined to be located in forested areas, the ownership of record was determined through tax records. For this study, we used 764 points representing landowners surveyed in 2002 in the Northeastern United States.

The sizes of the forested landholdings were grouped into three categories: 0.5–19 ha (1–49 ac), 20–399 ha (50–999 ac),

and greater than or equal to 400 ha ($\geq 1,000$ ac). These categories are a simplified version of the categories used in previously published research (e.g., Birch *et al.* 1982, Birch 1996, Butler and Leatherberry 2003). The more detailed groups reported in earlier reports are collapsed to aid in producing results that are more accurate, more consistent, and easier to interpret.

Predictor variables needed to have a theoretical relationship to forest parcel size, be spatially explicit, and be publicly available. The independent variables were percentage forest land cover, percentage urban land cover, percentage agricultural land cover, population density, housing density, change in population levels, distance to roads, distance to U.S. Census-defined cities and places, road density, a population gravity model, and State. Land cover data were summarized over 6.25 ha (15.5 ac) areas based on the Multi-Resolution Land Characteristics Consortium's National Land Cover Data (Vogelmann *et al.* 2001). Population, housing, city, and place data were derived from U.S. Census data (U.S. Commerce Department 2000). Population and housing density were summarized for each U.S. Census-defined block group. Road data were obtained from Bureau of Transportation Statistics (U.S. Department of Transportation 1998). Parcels closer to urban areas are more likely to be influenced by urban land use processes; we included a gravity model that quantifies the "influence" of these population centers across the landscape (equation 1) (Kline and Alig 2001).

$$Gravity\ Index_i = \sum_{j=1}^3 \frac{(Population_k)^{0.5}}{Distance_{ij}} \quad (1)$$

where:

i = sample point; and

j = the three cities having the largest influences, as defined by this equation.

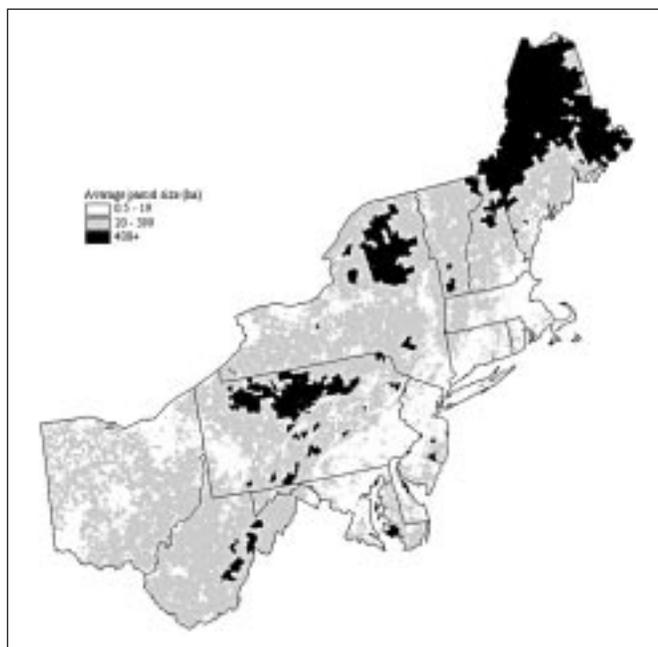
Due to high Pearson correlation coefficient values with other predictor variables, the percentage agricultural land cover and housing density variables were excluded in favor of other variables that had higher predictive powers and more direct interpretations.

Each independent variable not in the proper format was converted into a surface or grid layer and reprojected to an Albers projection using Geographic Information System (GIS) software (Environmental Systems Research Institute 2001). The value for each independent variable at each sample point with

Table 1.—Confusion matrix representing observed and predicted ownership categories based on a closest-neighbor estimation technique (numbers in parentheses represent column percentages).

Predicted	Observed						
	Federal	State	Local	Forest industry	Other corporate	Family	Total
Federal	42 (56.8)	3 (1.0)	2 (2.7)	1 (0.2)	7 (1.5)	23 (1.1)	78 (2.3)
State	0 (0.0)	139 (45.1)	3 (4.1)	20 (4.9)	26 (5.7)	102 (4.9)	290 (8.5)
Local	1 (1.4)	8 (2.6)	6 (8.1)	4 (1.0)	11 (2.4)	25 (1.2)	55 (1.6)
Forest industry	0 (0.0)	13 (4.2)	3 (4.1)	251 (61.5)	36 (7.9)	65 (3.1)	368 (10.8)
Other corporate	2 (2.7)	21 (6.8)	11 (14.9)	40 (9.8)	157 (34.6)	204 (9.7)	435 (12.7)
Family	21 (28.4)	111 (36.0)	35 (47.3)	80 (19.6)	183 (40.3)	1391 (66.1)	1821 (53.2)
Nonforest	8 (10.8)	13 (4.2)	14 (18.9)	12 (2.9)	34 (7.5)	293 (13.9)	374 (10.9)
Total	74	308	74	408	454	2103	3421

Figure 3.—Average size of forest parcels in the Northeastern United States as predicted by a decision tree model.



Conclusions

The final decision tree model (fig. 2) makes sense intuitively. This model found population density to be the most powerful predictor of parcel size and shows the expected negative relationship. Having larger parcel sizes predicted in Maine and Pennsylvania versus the other States is logical considering the strong forest industries in those States and the States' development and population patterns. Percentage forest had the expected relationship with parcel size with areas of more forest being more likely to have larger forested parcels at a given population level. Road distance, although only of tertiary importance in the model, also exhibited the expected relationship to parcel size; areas of higher road density were more likely to have smaller forested parcels.

The resulting forest parcel size map (fig. 3) also seems reasonable. The areas with the highest probabilities of having larger forested parcels are located in areas with higher concentrations of forest industry or large public landholdings.

Northern Maine is an example of the former, and north-central Pennsylvania and the Adirondack region of New York are examples of the latter.

In addition to refining the decision tree model, we want to explore some geospatial modeling techniques, such as indicator kriging, before we attempt to implement this study on a broader geographic scale. Further refinement of accuracy assessment techniques will also be explored, including error/probability mapping.

Forest parcel size is an omnipresent factor influencing how and why land is used. Modeling of forest parcel size will yield more insight into the forest parcelization phenomenon and the impact of these trends on forest fragmentation, harvesting practices, and other trends.

Literature Cited

- Birch, T.W. 1996. Private forest-land owners of the United States, 1994. Resour. Bull. NE-134. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 183 p.
- Birch, T.W.; Lewis, D.G.; Kaiser, H.F. 1982. The private forest-land owners of the United States. Resour. Bull. WO-1. Washington, DC: U.S. Department of Agriculture, Forest Service. 64 p.
- Butler, B.J.; Leatherberry, E.C. 2003. Preliminary results from the national woodland owner survey, 2002. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. <http://www.fs.fed.us/woodlandowners>.
- Congalton, R.G.; Green, K. 1999. Assessing the accuracy of remotely sensed data: principles and practices. New York: CRC Press. 137 p.
- Egan, A.F.; Luloff, A.E. 2000. The exurbanization of America's forests: research in rural social science. *Journal of Forestry*. 98(3): 26–30.
- Environmental Systems Research Institute, Inc. 2001. ArcInfo, version 8.3. Redlands, CA: Environmental Systems Research Institute.
- Kass, G.V. 1980. An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*. 29(2): 119–127
- King, S.L.; Butler, B.J. [In press.] Generating a forest parcel map for Madison County, NY. In: Beavers, M., Barrett, T.M., eds. Proceedings, 2003 systems analysis in forest resources symposium; 2003 October 7–9; Stevenson, WA. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Kline, J.D.; Alig, R.J. 2001. A spatial model of land use change in western Oregon and western Washington. Res. Pap. PNW-RP-528. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 24 p.
- Row, C. 1978. Economies of tract size in timber growing. *Journal of Forestry*. 76(9): 576–582.
- Sampson, R.N.; DeCoster, L. 2000. Forest fragmentation: implications for sustainable private forests. *Journal of Forestry*. 98(3): 4–8.
- SPSS, Inc. 2002. SPSS Answer Tree, version 3.1. Chicago, IL: SPSS.
- U.S. Department of Commerce, Census Bureau. 2000. Census 2000 TIGER/Line Data. Washington, DC: U.S. Department of Commerce, Census Bureau. <http://www.census.gov/geo/www/tiger/>. [Date accessed unknown.]
- U.S. Department of Transportation, Bureau of Transportation Statistics. 1998. Geographic data technology Dynamap/1000 street shapefiles, version 9.0. Washington, DC: U.S. Department of Transportation, Bureau of Transportation Statistics. http://www.bts.gov/programs/geographic_information_services/download_sites/gdt/maindownload.html. [Date accessed unknown.]

Vogelmann, J.E.; Howard, S.M.; Yang, *et al.* 2001. Completion of the 1990s national land cover dataset for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering & Remote Sensing*. 67: 650–652.

Wear, D.N.; Liu, R.; Foreman, J.M.; Sheffield, R.M. 1999. The effects of population growth on timber management and inventories in Virginia. *Forest Ecology and Management*. 118: 107–115.

Incorporation of Precipitation Data Into FIA Analyses: A Case Study of Factors Influencing Susceptibility to Oak Decline in Southern Missouri, U.S.A.

W. Keith Moser¹, Greg Liknes², Mark Hansen³, and Kevin Nimerfro⁴

Abstract.—The Forest Inventory and Analysis program at the North Central Research Station focuses on understanding the forested ecosystems in the North Central and Northern Great Plains States through analyzing the results of annual inventories. The program also researches techniques for data collection and analysis. The FIA process measures the above-ground vegetation and the site (soils) factors, but not climatic data. This pilot study, centered around three inventory units in southern Missouri, assigned weather data obtained from the National Oceanic and Atmospheric Administration to particular forest inventory plots, based on nearest distance. We incorporated precipitation and maximum and minimum temperatures into a temporary database, then analyzed the growth and forest health data for the plots for any relationships among the climate data. We found an apparent relationship between precipitation and the hypothesized relationship between the variables believed to predispose the stand toward oak decline and mortality variables, particularly in larger, older trees. Adding precipitation as an independent variable helped increase the quality of the predictions of the mortality models in situations where we concentrate on size/age groups more prone to forest health problems. Finally, we found evidence of spatial patterns of precipitation across the Ozark Plateau in southern Missouri that appear to be correlated with landscape-level patterns of mortality. Management activities need to address the role of the predisposing variables in influencing susceptibility to oak decline.

As the level of precipitation seems to exacerbate the predisposing variables' effects, historical patterns of rainfall and soil moisture retention need to be taken into account when regenerating and managing oak forests in the Missouri Ozarks.

The Central Hardwood forest ranges from eastern Oklahoma northeast to southern New England (Hicks 1998). Oak-hickory forests constitute the vast majority of acreage in the Eastern United States (Powell 1993) and in the Central Hardwood forest region. In the State of Missouri, oak-hickory forests constituted almost three-fourths of the total forest land area, and oaks made up 66 percent of all growing stock removals on timberland between 1999 and 2002 (Moser *et al.* 2004). Forest health problems affecting oak growth and survival could have a significant impact on Missouri's forest ecosystem and economy.

Oak decline is considered a "complex": a suite of pathogens and insects that together contribute to reduced growth, quality defects, and mortality (Manion 1981) for trees species in the red oak (*Erythrobalanus*) group, particularly black oak (*Quercus velutina* Lam.) and scarlet oak (*Quercus coccinea* Muenchh.). Consisting of the two-lined chestnut borer, the red oak borer, *Armillaria* fungus, and *Hypoxylon* canker (with additional impacts caused by four other insects [Wargo *et al.* 1983]), oak decline is native to the Central Hardwoods region and has long been endemic to oak forests (Starkey *et al.* 1989). Although evidence of oak decline has been observed in the Eastern United States since the 19th century (Millers *et al.* 1989), the complex has had an increasing impact on the forests of the Ozark Plateau of Missouri and Arkansas with evidence of crown dieback, growth reduction, and mortality in oak forests since the 1980s far exceeding historic levels. The severe drought of the late 1990s, combined with the advancing age of the Ozark forests,

¹ Research Forester, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108. Phone: 651-649-5155; e-mail: wkmoser@fs.fed.us.

² Research Forester, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

³ GIS Specialist, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

⁴ Programmer, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

has intensified the spread and severity of the effects (Lawrence *et al.* 2002). Most red oaks in the Ozarks are at least 70 to 80 years old, and grow in rocky soil on broad ridges or south- and west-facing slopes. Typical oak decline symptoms include branch dieback from the tips, sparse foliage, and reduced growth. Mature red oaks with more than 30 percent dead limbs and branches are considered to have a high mortality rate.

This study on oak decline uses data from Forest Inventory and Analysis (FIA) plots, Forest Health Monitoring (FHM) plots, and other sources to describe the past, present, and future condition of oak forests on the Mark Twain National Forest (MTNF) and other forest land in southern Missouri.

Missouri Oaks: Many Potential Victims

Oak species dominate Missouri forests (Moser *et al.* 2004), particularly those in the Ozark region. Oak-hickory and oak-pine forests constitute 78 percent of the forest land of the State. Of the 16.3 billion cubic feet of volume of all live trees on forest land that are in hardwood species groups, 36.2 percent were in white oak species groups and 31.1 percent were in red oak species groups (Moser *et al.* 2004). Forest health problems have long been identified for Missouri oaks (Missouri Conservation Commission 1976). According to Kathy Kromroy of the North Central Research Station, the percentage of dead oaks is increasing, from 1.3 percent in 1972 to 9.6 percent in 1989). Mortality has reached epidemic proportions in Arkansas and Missouri (Lawrence *et al.* 2002) and has severely affected parts of the Mark Twain National Forest. While planning for the future, the MTNF asked the U.S. Department of Agriculture, Forest Service, North Central Research Station FIA, and Northeastern Area State and Private Forestry for data and trends based on the FIA data. In response, both units formed the Mark Twain Oak Decline (MTOD) collaborative study. The investigation sought to provide the answers the MTNF needed for their planning process, and to present the results of the research so that other interested groups may benefit from the effort.

In the process of putting together their forest plan, MTNF managers identified the following specific objectives for the study group:

- Determine if any changes occurred in growth and mortality between the inventories in 1989 and 1999–2002 attributable to oak decline.

- Determine the relevant indications of oak decline and which species, age class, crown closure, and/or crown position is most affected.
- Determine the distribution of these oak decline effects.

Five Factors Influencing Susceptibility to Oak Decline

Previous work developed interim management guidelines for forests susceptible to oak decline (Moser and Melick 2002). Underlying the recommendations were assumptions, based on personal observations and input from many field managers and other researchers, about the impact of five stand and site factors present in all susceptible forests (Millers *et al.* 1989, Moser and Melick 2002, Starkey *et al.* 1989, Nebeker *et al.* 1992). Moser and Melick postulated that the five site factors influenced the likelihood of attack by oak decline in the following ways:

1. **Site.** Ridgetops and south-west aspects frequently have poor nutrients and/or water availability.
2. **Age.** Stands more than 70 years old are susceptible.
3. **Species.** Scarlet and black oaks are the most prone to oak decline, particularly on poor sites.
4. **Density.** Trees in stands with higher densities are more stressed than those in lower density stands.
5. **Lack of diversity.** Stands with high proportions of susceptible oaks are more prone to oak decline.

Local climate, particularly precipitation, impacts productivity and stand dynamics, and unfavorable weather can exacerbate the effects of the five factors on already stressed trees. Many references describe the importance of climate to vegetation. Water is a source of oxygen used in photosynthesis (Nobel 1991). Zimmerman and Brown (1980, p. 162) state “the availability of water is the most singly important environmental factor limiting growth and distribution in trees.” After light, water is the most prominent limit to growth (Oliver and Larson 1995).

This paper attempts to determine the following:

- If weather, particularly precipitation, significantly influenced mortality.
- If adding precipitation as an independent variable increases the quality of predictions of mortality models.
- If spatial patterns of precipitation exist, and, if so, to determine if they show any relationship with landscape-level patterns of mortality.

Management Implications

Given the widespread nature of oak decline on the Ozark Plateau, not enough foresters, loggers, or markets are available to deal with all the areas potentially requiring management action. As an intermediate measure, Moser and Melick (2002) suggest some management guidelines focused on manipulating the five factors to avoid or reduce the opportunity for oak decline, or to at least mitigate its effects. The guidelines suggest increasing diversity, both in species and tree size (age), before oak decline is present. Species should be matched to the most appropriate sites: pines on south- and west-facing slopes and ridgetops, white oaks on midslopes, northeast slopes for northern red oaks, and dryer (north and east aspect) sites for scarlet and black oaks. Species that are especially susceptible, such as scarlet and black oaks, would be aggressively thinned to increase vigor and harvested by the time they reach 70 to 80 years of age.

Methodology

Incorporating Rainfall

We obtained daily precipitation records, generally dating back to 1948, for all National Weather Service weather stations in southern and central Missouri. We assigned a weather station to the nearest FIA inventory plot. The distance from plots to the nearest weather station varied from a few hundred yards to 17 miles. We then parsed and transferred these files into temporary tables in the North Central FIA database, with assignment values (such as State, county, plot) that we joined with standard plot data. In some cases, many plots were assigned to a particular weather station. Because oak decline events appear to be influenced by recent patterns of drought (Starkey *et al.* 1989), for the purposes of this study we limited our use of the climatic data to the average annual rainfall from 1990 to 1999.

Data Analysis

Table 1 lists the independent variables we examined and each corresponding factor. We used 1989 and 2002 data collected on FIA plots on the Mark Twain National Forest. In this case, the independent variables were 1989 FIA variables, except for the 1990–99 average annual rainfall; the dependent variables were the six mortality values for 2002, the remeasurement period. In

addition to the basal areas of scarlet oak (species code 806) and black oak (species code 837), we also examined northern red oak (*Quercus rubra* L., species code 833) and white oak (*Q. alba* L., species code 802). We evaluated the correlations between the independent variables representing the five factors, average annual rainfall, and mortality and estimated the significance of adding precipitation to the model using equation 1.

Equation 1 is the formula for appraising the “value” of adding a variable to evaluate certain weather data as increasing model predictability. SS_1 = the sum of squares residual without the rainfall variable; SS_2 = the sum of squares residual with rainfall; p_1 = number of coefficients estimated without rainfall variable; p_2 = number of coefficients estimated with rainfall; n = total number of observations. We refer to this statistic as the “value” F-statistic, to distinguish it from the “model” F-statistic in Draper and Smith (1981).

$$F = \frac{(SS_1 - SS_2) / (p_2 - p_1)}{SS_2 / (n - p_2)} \quad (1)$$

Table 1.—*Hypothetical factors and the Forest Inventory Analysis (FIA) dataset variables (“factor variables”) examined in the overall oak decline study.*

Category	FIA dataset variable
Age	Stand age 1989
Density	Total BA 1989 Oak BA 1989
Site	Site Condition 1989 Aspect 1989 Slope 1989
Species mix and diversity	BA 802 (white oak) 1989 BA 806 (scarlet oak) 1989 BA 833 (northern red oak) 1989 BA 837 (black oak) 1989
Weather	Average Annual Rainfall 1990–99

Results

Correlations Between the Five Factors and Mortality

The hypothesis that the five factors and rainfall influence tree mortality, particularly that of oaks, is supported by the data in the correlation table (table 2). All mortality variables were negatively correlated with average annual rainfall. Most the variables

representing the five factors (table 1) were positively correlated with the mortality variables, except for aspect. This last result probably is caused by the cardinal nature of aspect, with north and east (315° to 135°) sites, all other things being equal, being less susceptible to oak decline and the south and west aspects (135° to 315°) being more prone to oak decline. An ordinal or binary variable might have been a better choice for evaluation. Using equation 1, we calculated the F-statistic, examining the statistical significance of adding average annual precipitation from 1990–99 to the 10 site-specific variables.

We focused on mortality as evidence of forest decline. Several measures of per acre mortality reside in the FIA database, so we concentrated on measures of total biomass, numbers of trees, and growing stock volume. Total biomass mortality (expressed as dry weight) more closely mirrors the total site productivity. Mortality trees per acre represents the product of the number of trees found dead since the last inventory and the tree expansion factor. The minimum diameter of this category is 5 inches, probably representing middle-aged and older trees. Mortality trees per acre, combined with basal area, give some indication of density-dependent mortality. Finally, growing stock mortality focuses more on larger trees and separates those segments of the total tree population less likely to suffer from normal competitive pressures than trees 1 to 2 inches in diameter, for example. This last category would have to have a higher proportion of mortality resulting from oak decline than trees less than 5 inches in diameter.

Was Precipitation a Significant Addition to the Model?

We calculated $F = 4.215$, greater than the 95-percent threshold value of 3.92 which was therefore statistically significant. Mortality trees per acre (oak species only) exhibit $F = 5.880$, again indicated that precipitation significantly improved the quality of fit of the model to the data.

Biomass mortality per acre for all species and for oak species alone was evaluated in the same manner. With $F = 2.127$ and $F = 3.505$, adding precipitation had no significant impact on estimating either mortality variable. Inserting precipitation into the model for growing stock mortality per acre for all species resulted in $F = 4.460$, indicating a significant addition. For growing stock mortality for oaks only, $F = 3.307$ suggested that the precipitation did not significantly improve the quality of fit of the model to the data.

How Much Did Including Precipitation Add to the Quality of Model Prediction? Table 3 also lists the model R^2 and F statistics with and without precipitation. For all six variables, R^2 decreased when precipitation was removed from the model. Mortality trees per acre for all species had $R^2 = 0.0678$ and $R^2 = 0.0579$ with and without average annual rainfall from 1990 to 1999. Mortality trees per acre, oak species only, had similar results, with $R^2 = 0.1113$ and $R^2 = 0.09813$ with and without rainfall. Dry biomass mortality per acre for all species had $R^2 = 0.1745$ with and $R^2 = 0.1701$ without the precipitation variable. Dry biomass mortality per acre, oaks, had $R^2 = 0.2481$ with rainfall and $R^2 = 0.2443$ without. Rainfall was not a signif-

Table 2.—Direction of correlation (sign) between the factor variables and rainfall vs. three mortality variables.

Mortality variable	Average rainfall 1990–99	Stand age 1989	Total BA 1989	Oak BA 1989	Site condition 1989	Aspect 1989	Slope 1989	BA 802 1989	BA 806 1989	BA 833 1989	BA 837 1989
Mortality trees per acre	-	-	+	+	+	-	+	+	-	+	-
Mortality trees per acre—oaks	-	-	+	+	-	+	-	-	-	-	+
Growing stock mortality per acre	-	+	+	+	+	-	+	+	+	+	+
Growing stock mortality per acre—oaks	-	+	+	+	+	-	+	+	+	+	+
Dry biomass mortality per acre	-	+	+	+	+	-	+	+	+	+	+
Dry biomass mortality per acre—oaks	-	+	+	+	+	+	+	+	+	+	+

icant addition to the variable mix for dry biomass for all species or oaks only. Dry biomass mortality per acre for all species and for oaks had R^2 of 0.1745 and 0.2481, respectively, with precipitation and 0.1701 and 0.2443 without precipitation. Finally, growing stock mortality per acre for all species had an $R^2 = 0.1503$ with and $R^2 = 0.1407$ without rainfall, whereas growing stock mortality per acre, oaks, had an $R^2 = 0.1697$ with and $R^2 = 0.1628$ without.

As a measure of the increase in quality of fit of the model to the data when including precipitation, the F-statistic was significant for mortality trees per acre and growing stock mortality per acre, but was not significant for dry biomass mortality per acre. Yet, dry biomass mortality exhibited the highest quality of model prediction. One conclusion might be that dry biomass more completely represents the accumulated total site productivity and is perhaps less sensitive to minor fluctuations in rainfall. The other two measures, mortality trees per acre and growing stock mortality per acre, focused on trees 5 inches in diameter and greater, and perhaps have higher percentages of trees in stressed situations and thus more sensitivity to fluctuations in precipitation. On the other hand, growing stock mortality does not include rough and rotten trees, so the demise of these supposedly healthy trees might be a more profound indicator of the effects of a lack of precipitation.

Why Is R^2 Not Approaching 100 Percent? Statistical models of biological systems rarely achieve perfection in their predictive ability. The more complex the system, the more opportunity for influence by unforeseen or nonsystematic variables such as individual land owner management, site-specific edaphic and microclimatic influence, or a host of potential

interactions among trees and other biota or among the trees themselves. Has oak decline really kicked in? Much evidence exists of oak decline, such as growth reduction, crown dieback, evidence of pathogens and insects, but, unlike in earlier periods, mortality has not yet responded. Finally, while our weather data is more precise than regional assessments, the relatively few number of stations vis-à-vis our plots still leaves us with some landscape-level generalizations.

“Optimal” Number of Variables

As a further test of the value of rainfall in understanding mortality data on FIA plots, we constructed an algorithm to evaluate all combinations of independent variables with mortality trees per acre in 2002 as the dependent variable. The results of this run (Ron McRoberts, North Central Research Station, pers. comm.) show that the fewest number of variables that were still statistically significant as predictors of mortality trees per acre ($F > 3.92$, $n = 408$) were the following:

- Total basal area 1989.
- Basal area of scarlet oak 1989.
- Basal area of black oak 1989.
- Site condition 1989.
- Annual rainfall 1990–99.

Spatial Arrangement of Rainfall and Mortality Across the Landscape

Finally, we examined the spatial arrangement of rainfall and mortality across the landscape. Although the study is ongoing, early indications suggest some interesting patterns. Figure 1 shows a bubble plot of average annual rainfall across all plots

Table 3.—Analysis of mortality measurements as a function of independent factor variables.

Mortality Measure	R^2 (all variables)	Model F-statistic (all variables)	R^2 (no precipitation)	Model (no precipitation)	Value F-statistic for annual precipitation variable	Significant at 95 percent?
Mortality trees per acre	.0678	2.625	.0579	2.446	4.215	Yes
Mortality trees per acre—oaks	.1113	4.520	.09813	4.330	5.880	Yes
Dry biomass mortality per acre	.1745	7.630	.1701	8.158	2.127	No
Dry biomass mortality per acre—oaks	.2481	11.91	.2443	12.86	3.505	No
Growing stock mortality per acre	.1503	6.382	.1407	6.516	4.460	Yes
Growing stock mortality per acre—oaks	.1697	7.375	.1628	8.305	3.307	No

in the three Ozark units. Larger circles represent higher levels of precipitation. Note the band of low-rainfall plots across the center of the map. This figure should be compared to figure 2, where we graphed mortality over roughly the same area, illustrating mortality with shading (light for high mortality and dark for low mortality). Note that many of the areas of high mortality in figure 2 occur in the same locations as areas of low rainfall in figure 1, suggesting that the areas of high mortality might be influenced by the relative lack of rain. Of course, other factors such as site and age demand investigation, but these early results are promising.

Conclusions

The first question addressed in this study was whether precipitation significantly improved the estimated relationship between the five factors variables and the mortality variables. The answer appears to be yes, particularly in situations where the trees were potentially stressed to begin with, such as larger, older trees.

We also sought to discover if, by adding precipitation as an independent variable, we increased our understanding and the quality of the predictions of the mortality models. Here again, the answer is yes, in situations where we are able to concentrate on size/age groups prone to forest health problems, as opposed to smaller/younger trees where normal competition-induced mortality played a role.

Finally, we were curious whether spatial patterns of precipitation existed across the Ozark Plateau in southern Missouri and if they might show any relationship with landscape-level patterns of mortality? The answer is yes, but we need more detailed further investigations to confirm it.

Management activities need to take into account the role of the five factors in influencing susceptibility to oak decline. As low levels of precipitation seems to exacerbate the five factor's effects, historical patterns of rainfall and soil moisture retention need to be taken into account when regenerating and managing oak forests in the Missouri Ozarks.

Figure 1.—Bubble plot of average annual rainfall, in inches, attributed to each FIA plot; larger circles represent higher levels of precipitation. Inset shows location of study area in southern Missouri, U.S.A.

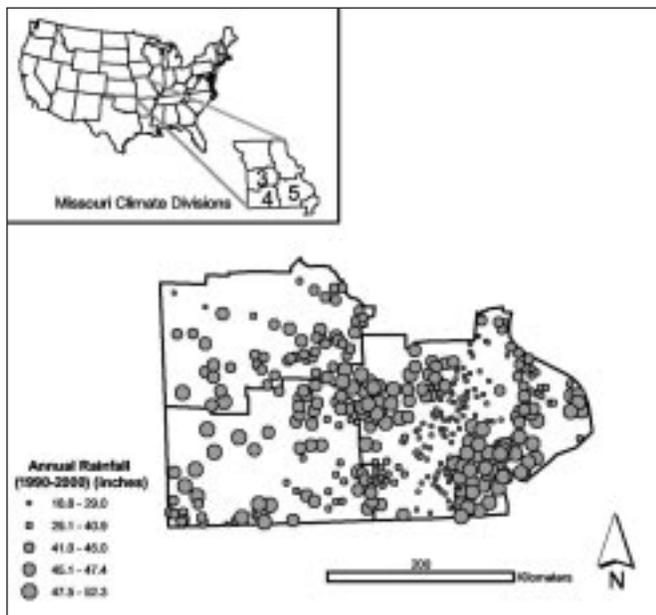
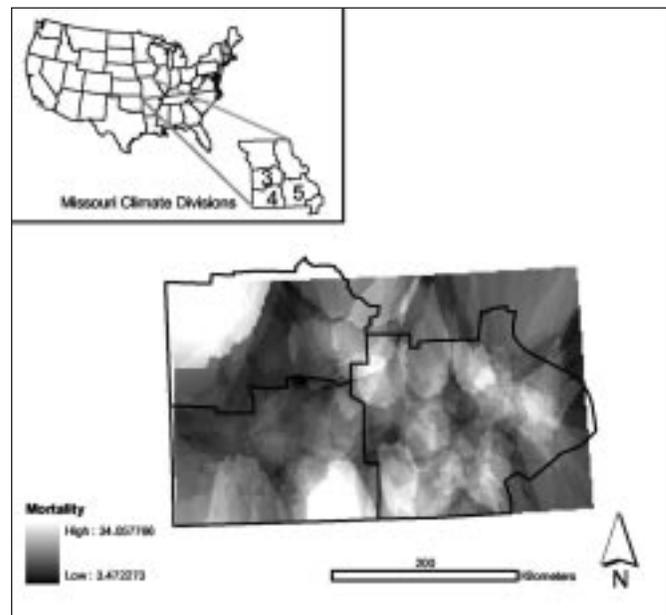


Figure 2.—Two-dimensional interpolation of tree mortality, where the x-y axes denote location and the shading represents annual mortality, with white or light gray being higher mortality and the darker gray representing less mortality.



Acknowledgments

Ron McRoberts, Will McWilliams, and John Shaw provided valuable editorial suggestions.

Literature Cited

Draper, N.R.; Smith, H. 1981. Applied regression analysis. New York: John Wiley and Sons. 709 p.

Hicks, R.R. 1998. Ecology and management of central hardwood forests. New York: John Wiley and Sons. 412 p.

Lawrence, R.; Moltzan, B.; Moser, W.K. 2002. Oak decline and the future of Missouri's forests. *Missouri Conservationist*. 63(7): 11–18.

Manion, P.D. 1981. Tree disease concepts. Englewood Cliffs, NJ: Prentice-Hall. 399 p.

Millers, I.; Shriner, D.S.; Rizzo, D. 1989. History of hardwood decline in the Eastern United States. Gen. Tech. Rep. NE-126. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 75 p.

Moser, W.K.; Brand, G.J.; Treiman, T.; *et al.* 2004. Missouri's forest resources in 2002. Resour. Bull. NC-233. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 42 p.

Moser, W.K.; Melick, R. 2002. Management recommendations for oak decline. Unpublished memo to the Missouri Department of Conservation and the Mark Twain National Forest. 3 p.

Nebeker, T.E.; Ammon, V.D.; Boyle, C.R.; *et al.* 1992. Oak decline: a comparison of study results from different areas of the South. Tech. Bull. 183. Mississippi State, MS: Mississippi Agricultural and Forestry Experiment Station. 19 p.

Nobel, P.S. 1991. Physicochemical and environmental plant physiology. New York: Academic Press. 635 p.

Oliver, C.D.; Larson, B.C. 1995. Forest stand dynamics. New York: John Wiley and Sons. 520 p.

Powell, D.S.; Faulkner, J.L.; Darr, D.R.; *et al.* 1993. Forest resources of the United States, 1992. Gen. Tech. Rep. RM-234. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 133 p.

Starkey, D.A.; Oak, S.W.; Ryan, G.W.; *et al.* 1989. Evaluation of oak decline areas in the South. Southern Region Protection Report R8-PR 17. Atlanta, GA: U.S. Department of Agriculture, Forest Service, Southern Region. 36 p.

Wargo, P.M.; Houston, D.R.; LaMadeleine, L.A. 1983. Oak decline. Forest Insect and Disease Leaflet 165. Washington, DC: U.S. Department of Agriculture, Forest Service. 4 p.

Zimmerman, M.H.; Brown, C.L. 1980. Trees: structure and function. New York: Springer-Verlag. 336 p.



Impact of Stream Management Zones and Road Beautifying Buffers on Long-Term Fiber Supply in Georgia

Michal Zasada¹, Chris J. Cieszewski², and Roger C. Lowe³

Abstract.—Streamside management zones (SMZs) and road beautifying buffers (RBBs) in Georgia have had an unknown impact on the available wood supply in the state. We used Forest Inventory and Analysis data, Landsat Thematic Mapper imagery, Gap Analysis Program and other geographic information system data to estimate the potential impact of SMZs and RBBs in the current and future Georgia forest inventories. The analyzed scenarios are based on long-term simulations of wood supply in the State under various assumptions of regulatory constraints, expected harvesting, and intensities of management practices. The results are expressed in the form of affected areas and volumes. Obtained results suggest that introducing only some of the harvesting constraints would not drastically affect sustainable fiber supply in the State, even in the presence of increased future harvesting. The cumulative impact of obligatory SMZs, RBBs, and other anticipated factors, such as potential loss of forested land to urban expansion, could have a strong negative impact on the level of sustainable harvesting, reducing the future fiber supply in Georgia.

A streamside management zone (SMZ) is a mandated protection zone around a stream, lake, or other water body, usually containing the bank vegetation and strip of forest. This zone must be protected because of its special importance for water quality. Riparian zones help maintain water quality, buffer rivers from adjacent pollution sources, filter sediments, absorb nutrients, stabilize stream banks, provide habitat and food for some animals and plants, and moderate stream temperature (Welsch 1991).

In 1976, the U.S. Environmental Protection Agency recommended using Best Management Practices (BMPs) as a primary method for controlling nonpoint source pollution (NPSP). The State of Georgia chose a nonregulatory system of voluntary compliance, which now is based on “Georgia’s Best Management Practices for Forestry” issued by the Georgia Forestry Commission in 1999.

Although a large number of studies on riparian/streamside management zones have been conducted in the South (Wenger 1999, for example), the literature on their extent assessment and other statistics is scarce. For perennial streams, BMPs currently recommend leaving evenly distributed 50 square feet of basal area per acre or at least 50 percent of the canopy cover after a harvest. If the stream is classified as a trout stream, BMPs recommend creating an additional no-harvest zone around the stream’s bank. For intermittent streams, requirements include leaving 25 square feet of basal area per acre or at least 25 percent of canopy cover after a harvest (GFC 1999). The impact of these potential harvesting limitations on long-term wood supply in the State remained unknown. In the future, Georgia may face the possibility of introducing mandatory BMPs for all forested areas. The current standards for BMPs may also change to more closely meet demands of environmental organizations calling for widening of the required buffers around streams and further restricting the forest management inside of them (Wenger 1999), having an unknown impact on the State’s wood production capability.

Objectives

The primary objective of this study was to evaluate, based on available data, the impact of harvest constraints in the protective zones on long-term wood supply in Georgia. We used large-

¹ Postdoctoral Fellow, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA; Assistant Professor at Division of Dendrometry and Forest Productivity, Faculty of Forestry, Warsaw Agricultural University, 02-776 Warsaw, Poland, and Corresponding Author. E-mail: mzasada@dendro.sggw.waw.pl.

² Associate Professor, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA.

³ GIS Analyst, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA.

scale estate simulation software and a spatially explicit Georgia forest inventory database developed from the Forest Inventory and Analysis (FIA) inventory data, Landsat Thematic Mapper (LTM) images, and various geographic information system (GIS) data available for the State.

Methods and Assumptions

Zasada *et al.* (2005) provided a detailed review of the literature on SMZs and RBBs, as well as a preliminary assessment of the potential extent of the SMZs and RBBs in Georgia using 1997 FIA inventory data, various GIS data, and Landsat TM images. They used LTM-image-based polygons populated using forest industry ground inventories to create a detailed spatial database with forest types, species groups, basal areas, volumes, and site productivities. The resulting spatially explicit database was used together with an estate management simulation model (OPTIONS from D.R. Systems, Inc.). Because this research was a continuation of the studies described in Zasada *et al.* (2005), we applied the assumptions described there to define the current simulation scenarios.

OPTIONS can be used to examine different forest management scenarios including financial, industrial, and policy-related decisions and sustainability analysis. The simulator is based on forecasting information for individual polygons (without optimization). All the records used by the program (including spatial data) are processed annually.

A detailed setup of OPTIONS runs was similar to that described by Cieszewski *et al.* (2003). Because of the focus on stream and road buffers in this study, we added additional management regimes attributed to species groups occupying the analyzed buffers. The major difference between management of various species within and outside of the buffers was that selective harvesting was performed in buffer stands with minimum required residual basal area defined by the BMPs, while clearcutting was allowed only on the nonbuffer areas.

We defined “primary stream buffers” as those created according to BMPs, with widths depending on stream classification and slope. “Primary road buffers” were assumed to have a width of 40 feet, and “secondary stream and road buffers” a width of the widest buffer anticipated in BMPs (100 feet). We followed

the BMP recommendations, allowing buffers to be selectively cut with appropriate minimum residual basal area left after harvesting. We considered the following three options of buffer combinations and widths showing various levels of regulatory restrictions:

- Only primary stream buffers.
- Primary stream and road buffers together.
- Secondary stream and road buffers together.

Next, we supplemented all above-mentioned assumptions by two harvesting levels at the State scale. First, we assumed that harvesting in Georgia would remain unchanged in the future, and we set it at 1.5 billion cubic feet per year according to the most recent FIA report on the State’s forest resources (Thompson 1998). Because it is likely that wood utilization may increase in the future (e.g., Wear and Greis 2002) we considered also an increasing statewide harvesting level. We assumed that from the current level of 1.5 billion cubic feet per year, harvesting would gradually increase to 2.25 billion cubic feet per /year in 2040, which means that we expect harvesting in the near future to increase by 50 percent over the 1997 harvesting level.

We also considered various intensities of management on the State level. The first variant assumed that about 30 percent of all pine plantations in the State are managed intensively, and no additional intensively managed plantations will be established in the future (Zasada *et al.* 2003). In the second variant we assumed that the intensity of management will increase, and that half of newly established plantations will be managed intensively, which means a transition rate to intensive management plantations of 50 percent.

We ran all the simulations for a 200-year prediction period. By using such a long simulation period we achieved a certain equilibrium between the forest productivity and its harvesting, which changes with forest age structure and regeneration practices and is likely to require more than two rotation periods. In most scenarios, the 200-year simulation period was sufficient to stabilize wood availability on a certain level, which could be assumed to reflect the resource production versus harvesting balance in the distant future. We do not believe that we can predict the state of forests into such a remote future, but instead we intend to determine the impact of different actions on forest resources in the State.

Results

Detailed assessment of the stream and road buffers is summarized in table 1. Narrow stream buffers (40 feet) established according to Georgia's BMPs occupy about 980,000 acres, which makes up 4.01 percent of total forested area of the state. Assuming all buffer widths of 100 feet, the stream buffers would occupy 8.65 percent of forested land. Forests in the determined stream buffers maintain 4.32 and 9.27 percent of total inventoried Georgia's wood volume, respectively. Similar results were obtained for road buffers. Primary (40 feet) road buffers occupy almost 890,000 acres, which makes up 3.64 percent of forested area and 3.52 percent of total volume. Secondary (100 feet) buffers would occupy 2,120,000 acres of forests (8.68 percent of area and 8.40 percent of total volume). These results reveal reasonable proportions. For example, the share of broadleaf species in stream buffers is 2 to 3 times higher than in road buffers. This can be attributed to specific forest types usually occupying riparian area.

We present the results of the simulations graphically by means of changes in inventory volume, and changes in volume available for harvesting. Figure 1 demonstrates results for the conservative scenario assuming no changes in harvesting and intensive management of southern pine stands. In this scenario, even combined wide (100 feet) stream and road buffers do not seem to have a dramatically negative impact on wood availability in the future, allowing for sustainable harvesting in the considered timeframe.

Figure 2 demonstrates results based on the assumption of harvesting in the State gradually increasing during the next 40 years to 150 percent of its current level. In this scenario, it is impossible to maintain sustainable harvesting without increased intensive management, which could compensate for increased demand on wood (fig. 3).

Discussion and Conclusions

Our results showed that introduction of SMZs and RBBs could affect wood supply in the future by excluding more than 17 percent of forest areas from harvesting. The magnitude of this impact depends on the extent of potential buffers, future wood demand, and intensity of management. Considering "the most probable" scenario, however, this impact should be moderate. Other elements of introduction of stream and road buffers also could affect forestry operations, such as an increased cost of management in the protective buffers as suggested by Cabbage and Woodman (1993).

We performed our study using the most commonly available data on streams and roads. Yet, the available data on streams omit many small intermittent and ephemeral streams. According to various researchers in different regions of the country, especially in the west, riparian zones were identified on as much as 60 percent of forested area, and in some cases in Georgia we suspect that stream lengths can be as much as double that reported (mapped) in available sources. Clearly, the knowledge in this area is incomplete and we recommend that the issue be further studied.

Table 1.—Detailed summary of primary (according to BMP) and secondary (100-foot wide) stream buffers (left) and primary (40 feet) and secondary (100-foot wide) road buffers (right) in Georgia.

Forest type	Buffer regime	Area [x10 ³ ac]	%	Volume [x10 ⁶ ft ³]	%
Evergreen	Primary	226	0.93	272	0.81
	Secondary	542	2.15	631	1.88
Mixed	Primary	141	0.58	166	0.49
	Secondary	291	1.19	344	1.02
Deciduous	Primary	613	2.51	1,015	3.01
	Secondary	1,296	5.31	2,147	6.38
Total	Primary	980	4.01	1,453	4.32
	Secondary	2,112	8.65	3,122	9.27

Forest type	Buffer regime	Area [x10 ³ ac]	%	Volume [x10 ⁶ ft ³]	%
Evergreen	Primary	401	1.64	482	1.43
	Secondary	964	3.95	1,160	3.45
Mixed	Primary	225	0.92	265	0.79
	Secondary	521	2.13	615	1.83
Deciduous	Primary	264	1.08	436	1.30
	Secondary	635	2.60	1,052	3.12
Total	Primary	889	3.64	1,184	3.52
	Secondary	2,120	8.68	2,827	8.40

Figure 1.—Changes in inventory volume (left) and volume available for harvesting (right) for scenarios assuming steady harvesting of 1.5 billion cf/year and the current level of intensive management (30 percent of pine plantations). First row: no buffers; second: BMP stream buffers; third: both narrow buffers; and fourth: both wide buffers.

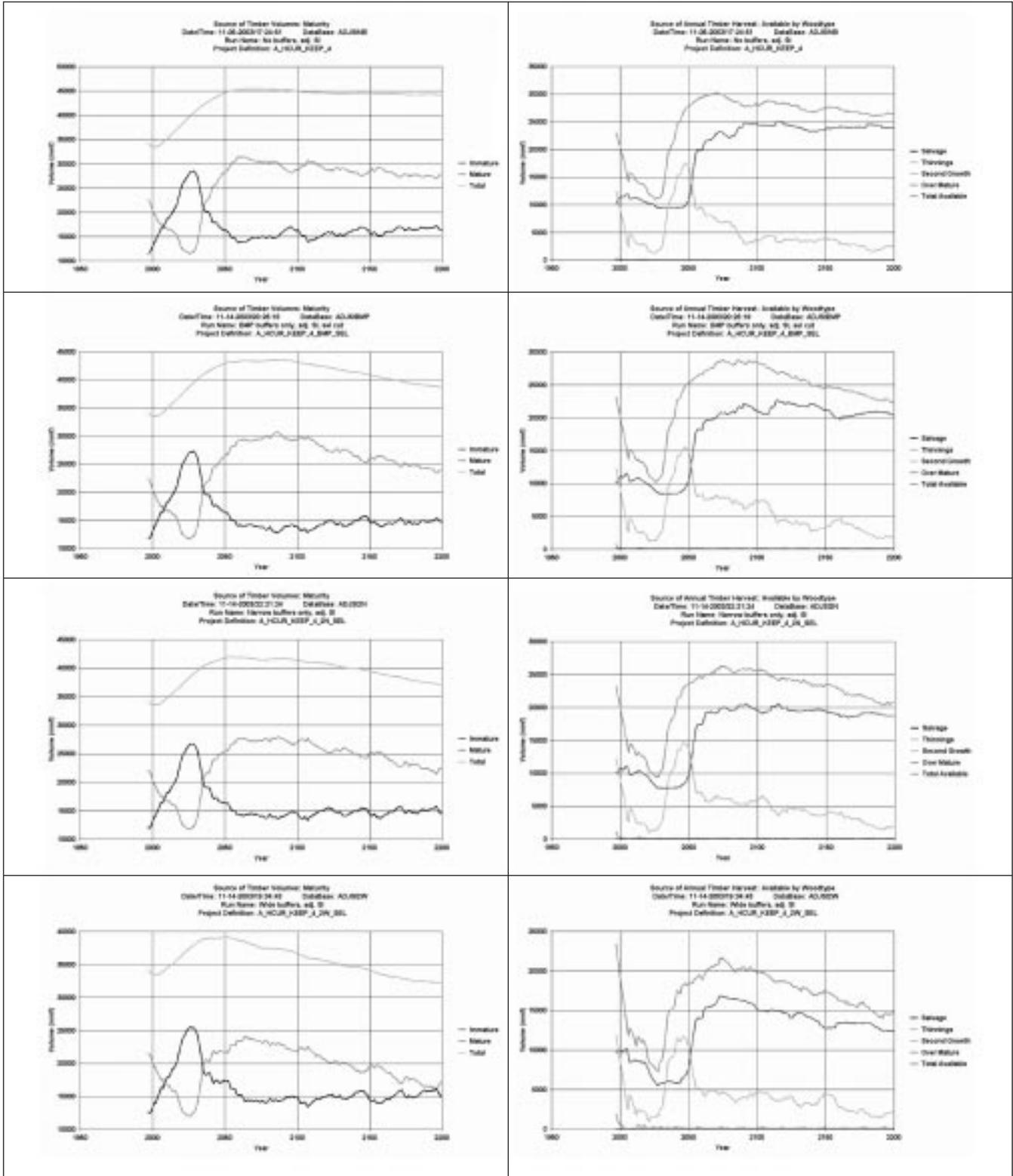


Figure 2.—Changes in inventory volume (left) and volume available for harvesting (right) for scenarios assuming harvesting increased from 1.5 billion cf/year in 1997 to 2.25 billion cf/year in 2040 and the current level of intensive management (30 percent of pine plantations). First row: no buffers; second: BMP stream buffers; third: both narrow buffers; and fourth: both wide buffers.

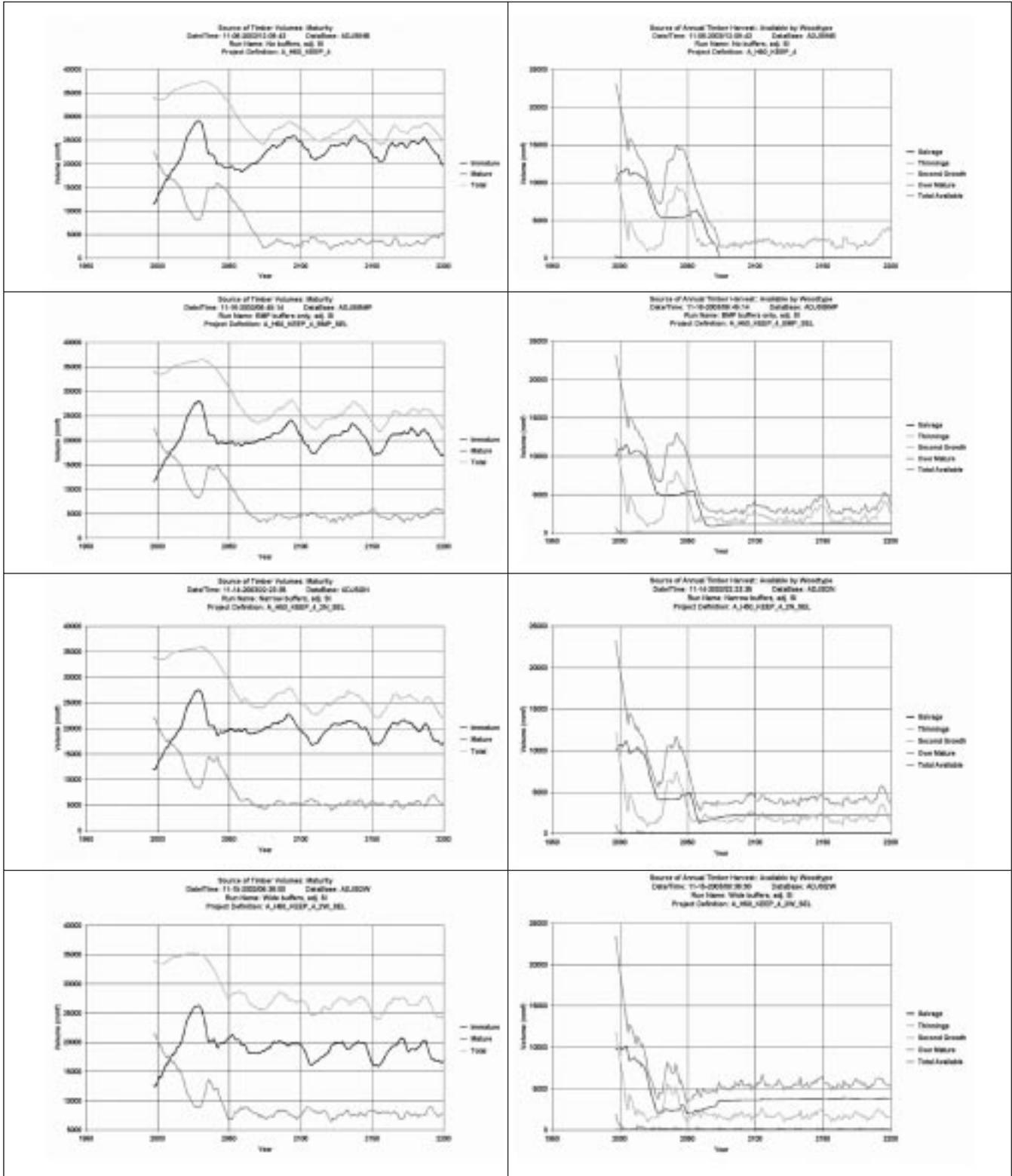
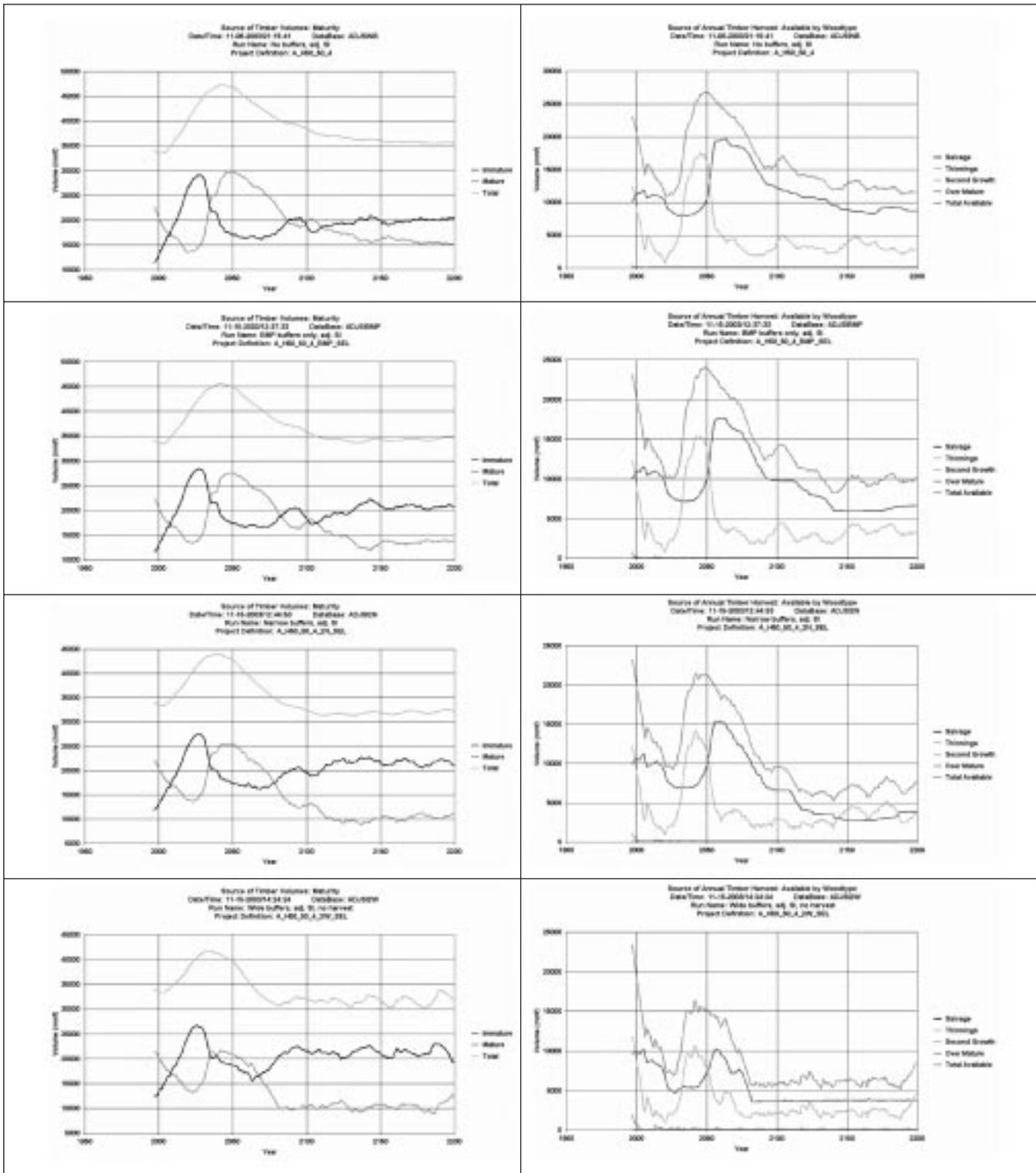


Figure 3.—Changes in inventory volume (left) and volume available for harvesting (right) for scenarios assuming harvesting increased from 1.5 billion cf/year in 1997 to 2.25 billion cf/year in 2040 and an increased intensity of management (50 percent newly established pine plantations are to be managed intensively). First row: no buffers; second: BMP stream buffers; third: both narrow buffers; and fourth: both wide buffers.



Increased wood demand, together with a large area of land reserved for protective uses, could significantly decrease volume available for harvesting in the future. Our results showed that allowing for protection of natural areas of special interest and maintaining the region's competitive status in the world market might require other supplementary measures, such as increasing the extent of intensive management practices in commercial forests (Sedjo and Botkin 1997).

In this study we have not considered any analyses of impact of urban expansion on long-term fiber supply in the State. In all probability, progressive urbanization will further contribute negatively to availability of forest areas and volumes available for harvesting.

Literature Cited

- Cieszewski, C.J.; Zasada, M.; Borders, B.E.; *et al.* 2003. Spatially explicit sustainability analysis of long-term fiber supply in Georgia, USA. *Forest Ecology and Management*. 187(2-3): 345-359.
- Cubbage, F.W.; Woodman, J.N. 1993. Impacts of streamside management zones on timber supply in Georgia. Southern Forest Economics Workers Meeting; 1993 April 23-24; Durham, NC: 130-137.
- Georgia Forestry Commission (GFC). 1999. Georgia's best management practices for forestry. Georgia Forestry Commission.
- Hansen, M.H.; Frieswyk, T.; Glover, J.F.; Kelly, J.F. 1992. The eastwide forest inventory database: users manual. Gen. Tech. Rep. NC-151. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 48 p.
- Miles, P.D.; Brand, G.J.; Alerich, C.L.; *et al.* 2001. The forest inventory and analysis database: database description and user's manual, version 1.0, revision 8. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station.
- Ruefenacht, B.; Vanderzanden, D.; Morrison, M.; Golden, M. 2002. New technique for segmenting images. RSAC program manual. U.S. Department of Agriculture, Forest Service, Remote Sensing Application Center. 14 p.
- Sedjo, R.A.; Botkin, D. 1997. Using forest plantations to spare natural forests. *Environment*. 39(10): 15-20, 30.
- Thompson, M.T. 1998. Forest statistics for Georgia, 1997. Resour. Bull. SRS-36. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 92 p.
- Wear, D.N.; Greis, J.G. 2002. Southern forest resource assessment. Gen. Tech. Rep. SRS-53. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 635 p.
- Welsch, D.J. 1991. Riparian forest buffers: function and design for protection and enhancement of water resources. NA-PR-07-91. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Area, State and Private Forestry.
- Wenger, S.J. 1999. A review of the scientific literature on riparian buffer width, extent, and vegetation. Athens, GA: Public Service & Outreach, Institute of Ecology, University of Georgia. 59 p.
- Zasada, M.; Cieszewski, C.J., Borders, B.E. 2003. Impact of intensive management practices on long-term simulations of forest resources in Georgia. In: Proceedings, Society of American Foresters 2002 National Convention; 2002 October 5-9; Winston-Salem, NC. SAF Publication. 03-01: 64-70.
- Zasada, M.; Cieszewski, C.J.; Lowe, R.C.; Zawadzki, J. 2005. Using FIA and GIS data to estimate areas and volumes of potential stream management zones and road beautifying buffers. In: Proceedings of the Fourth Annual Forest Inventory and Analysis Symposium; 2002 November 19-21; New Orleans, LA: U.S. Department of Agriculture, Forest Service, North Central Research Station: 253-257.



Generating Broad-Scale Forest Ownership Maps: A Closest-Neighbor Approach

Brett J. Butler¹

Abstract.—A closest-neighbor method for producing a forest ownership map using remotely sensed imagery and point-based ownership information is presented for the Northeastern United States. Based on a validation data set, this method had an accuracy rate of 58 percent.

Introduction

The ownership of America's forest resources can be divided into Federal, State, and local governments; forest industry; other corporate; and family and individual ownerships. Although large variability can exist in these categories, they have proven effective in understanding how forest land is used (e.g., Haynes 2003), who receives the goods and services produced, and how private or public entities influence these trends (e.g., Sampson and DeCoster 1997). Owners are the critical link between forests and society, and a full understanding of forest resources necessitates an understanding of forest ownership patterns.

The distribution of forest owners across the country is far from uniform. Eastern forests are dominated by private owners, while western forests are dominated by public owners (Smith *et al.* 2001). The distribution of forest owners at finer scales also varies greatly due to historic land distribution policies and local economic and social forces. Most data on forest ownership have been summarized in tabular format by geographic unit, e.g., State (Smith *et al.* 2001) or county (Griffith and Widmann 2003). Although these data constitute important information, they are limited with respect to geographic resolution, do not allow for visual examination of subcounty spatial patterns, and are not conducive for combining with other spatially explicit information. A forest ownership map can overcome many of these shortcomings.

Ownership maps (i.e., plat books) are available from county or municipal tax offices. Ideally, these maps would be

accurate, publicly accessible, and in compatible digital formats. Combined with a forest map, the ownership map should depict forest ownership throughout the Nation. Unfortunately, detailed ownership records have not been assembled in electronic format for multistate regions. An exception is the Managed Area Database (McGhie 1996) that provides boundaries for major Federal and State ownerships in the United States. Other ownership maps are available at finer scales (e.g., counties) or at the State level with specific limitations, for example, only holdings greater than 200 ha (500 ac). Also, such maps are often proprietary.

Because no national maps or data sources exist for forest ownership patterns, estimation procedures are necessary. In this article, I describe a method for producing a forest ownership map for the Northeastern United States. Data from remotely sensed imagery and ground-based forest inventories are combined using a closest-neighbor approach. The accuracy of this technique is assessed using validation data, and directions for future research are discussed.

Methods

The study area for this project is the 13 Northeastern States stretching south from Maine to Delaware and west to Ohio. This area was selected because forest inventory data were readily accessible. Geographic information system layers for this study were derived from Moderate Resolution Imaging Spectroradiometer imagery (MODIS) (Hansen *et al.* 2002) and U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis (FIA) plots. MODIS is satellite-based imagery. For the data used herein, the spatial resolution or pixel size is 500 m (1,640 ft) with 2000 and 2001 acquisition dates. The MODIS product used is Vegetation Continuous Fields percent forest cover data, which is produced and distributed by the Global Land Cover Facility at the University of Maryland. This information represents the percentage of each 25-ha (62-ac) pixel that is covered by tree canopies. A forest/nonforest map

¹ Research Forester, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, 11 Campus Boulevard, Suite 200, Newtown Square, PA 19073. Phone: 610-557-4045; fax: 610-557-4250; e-mail: bbutler01@fs.fed.us.

was generated by assigning a value of 1 (forest land) to all pixels with a tree cover of at least 55 percent and a value of 0 (nonforest land) to all other pixels. The 55-percent minimum was selected because it generated the same regional forest percentage (61) as an independent data source (Smith *et al.* 2001).

From the FIA data, 13,686 forested plots or sample points were identified in the Northeast. The footprint of these plots covers less than 1/10 of 1 percent of the region's forest area, but the random selection procedure and large sample size ensures that the sample is representative of the broader region. For each identified plot, data on ownership of record were obtained from tax records, and each ownership was categorized as Federal, State, or local government; forest industry; other corporate; or family and individual. Forest industry includes all private holders who own a primary wood-processing facility. Other corporate includes all other businesses, associations, and tribal lands with no primary processing facilities. Family ownership includes all forest land owned by individuals or families that are not incorporated.

The coordinates of the plot locations were fuzzed to mask exact plot coordinates and prevent disclosure of landowners' identities. For each coordinate at each point, a randomly selected value of + 3,250 to - 3,250m (+/- 10,663 ft) was added.

The FIA sample points were divided into training and validation sets. One in four randomly selected sample points were reserved for the validation set. For every forested pixel on the forest/nonforest map, an ownership category was assigned based on the ownership of the closest FIA plot in the training set to generate the forest ownership map. Euclidean distances were used to find the closest plots (Environmental Systems Research Institute 2001).

For each plot in the validation set, the modeled or estimated ownership category was obtained from the forest/ownership map generated. Observed and predicted values for each validation point were used to create a confusion matrix and assess model accuracy. Although the 25 percent of plots used for validation were selected at random, the effect of this specific subset of the points on the accuracy estimate is unknown. In future efforts, we will use multiple iterations of the validation selection process to assess accuracy variability.

Results

The resulting ownership map (fig. 1) is a fair approximation of the forest ownership pattern in the Northeastern United States. The most striking feature is the vast amount of family and individual forest land in the region. As would be expected, this feature is supported by FIA tabular estimates (Smith *et al.* 2001). The map shows the major holdings by the forest industry in Maine, northern New Hampshire and northern Vermont, north-central Pennsylvania, the Adirondack region of New York, and south-central West Virginia. Large State-owned forest holdings are evident in Pennsylvania.

Of the 3,421 validation points, 58 percent were classified correctly (table 1). The most common category for a misclassified plot was family and individual ownership. This error was related to the fact that family and individual ownerships represent a plurality of forest owners in the Northeast. Part of the misclassification was due to the use of fuzzed coordinates; 26 percent of the errors were forested plots being assigned to areas classified as nonforest on the map. Other errors were attributable to the degree of accuracy and resolution of the forest/nonforest map.

Figure 1.—This forest ownership map for the Northeastern United States was generated using a closest-neighbor approach.

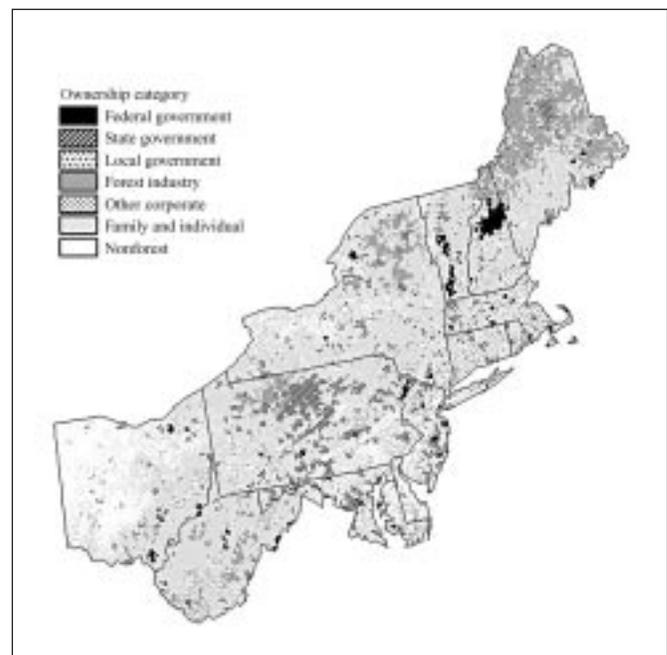


Table 1.—*Confusion matrix representing observed and predicted ownership categories based on a closest-neighbor estimation technique (numbers in parentheses represent column percentages).*

Predicted	Observed						
	Federal	State	Local	Forest industry	Other corporate	Family	Total
Federal	42 (56.8)	3 (1.0)	2 (2.7)	1 (0.2)	7 (1.5)	23 (1.1)	78 (2.3)
State	0 (0.0)	139 (45.1)	3 (4.1)	20 (4.9)	26 (5.7)	102 (4.9)	290 (8.5)
Local	1 (1.4)	8 (2.6)	6 (8.1)	4 (1.0)	11 (2.4)	25 (1.2)	55 (1.6)
Forest industry	0 (0.0)	13 (4.2)	3 (4.1)	251 (61.5)	36 (7.9)	65 (3.1)	368 (10.8)
Other corporate	2 (2.7)	21 (6.8)	11 (14.9)	40 (9.8)	157 (34.6)	204 (9.7)	435 (12.7)
Family	21 (28.4)	111 (36.0)	35 (47.3)	80 (19.6)	183 (40.3)	1,391 (66.1)	1,821 (53.2)
Nonforest	8 (10.8)	13 (4.2)	14 (18.9)	12 (2.9)	34 (7.5)	293 (13.9)	374 (10.9)
Total	74	308	74	408	454	2,103	3,421

Data disclosure is a concern with the spatial display of FIA plot data. Because this method produces a relatively high rate of misclassification, data disclosure likely is not a significant issue.

Conclusions

The map resulting from the closest-neighbor approach is useful for displaying broad forest ownership patterns. This type of product would be appropriate for inclusion in a State forest inventory report or other reports concerned with forest resources across wide-ranging areas.

Our accuracy assessment, however, highlighted several underlying shortcomings. Although the general location and distribution of forest ownership may be correct, the exact locations of specific ownerships are not modeled accurately, in part because fuzzed coordinates were used to generate the map.

This shortcoming limits the utility of the map for inclusion in spatial modeling projects. Thus, a percentage-based method may be more appropriate, although such a product might be inferior to a discrete map that is easier to display and interpret.

In addition to using actual coordinates, one can test other techniques for improving the accuracy of the map. The first step should be to include ancillary information, for example, data from the Managed Area Database (McGhie 1996) that depict the actual boundaries of forest owners. These data would be particularly useful for increasing accuracy in the depiction of large public ownerships. Incorporating other ancillary data and using spatial and/or nonspatial modeling techniques also could increase accuracy.

The method presented here is a relatively simple approach to generating a forest ownership map using the best available data. After additional methods are tested, this project should be expanded to include the entire Nation.

Literature Cited

- Environmental Systems Research Institute, Inc. 2001. ArcInfo, version 8.3. Redlands, CA: Environmental Systems Research Institute.
- Griffith, D.M.; Widmann, R.H. 2003. Forest statistics for West Virginia: 1989 and 2000. Resour. Bull. NE-157. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 119 p.
- Hansen, M.; Defries, R.; Townshend, J.R.G.; *et al.* 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. Remote Sensing of the Environment. 83: 303–319.
- Haynes, R.W. 2003. An analysis of the timber situation in the United States: 1952–2050. Gen. Tech. Rep. PNW-560. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 254 p.
- McGhie, R.G. 1996. Creation of a comprehensive managed areas spatial database for the conterminous United States: summary project technical report. NCGIA Tech. Rep. 96-4. Santa Barbara, CA: University of California at Santa Barbara. 53 p.
- Sampson, R.N.; DeCoster, L.A. 1997. Public programs for private forestry: a reader on programs and options. Washington, DC: American Forests. 95 p.
- Smith, W.B.; Vissage, J.S.; Sheffield, R.M.; Darr, D.R. 2001. Forest resources of the United States, 1997. Gen. Tech. Rep. NC-219. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 190 p.

Impact of Definitions of FIA Variables and Compilation Procedures on Inventory Compilation Results in Georgia

Brock Stewart¹, Chris J. Cieszewski², and Michal Zasada^{3,4}

Abstract.—This paper presents a sensitivity analysis of the impact of various definitions and inclusions of different variables in the Forest Inventory and Analysis (FIA) inventory on data compilation results. FIA manuals have been changing recently to make the inventory consistent between all the States. Our analysis demonstrates the importance (or insignificance) of different variations of the compilation procedures on the statistical summaries regarding volume and area distributions.

The U.S. Department of Agriculture Forest Service Forest Inventory and Analysis (FIA) program, providing nationwide information about forest resources, has changed rapidly during the past several years. Following the recommendations of the second Blue Ribbon Panel and the Agricultural Research, Extension, and Education Reform Act of 1998 (Section 253c), the periodic inventory system, providing information for individual States every 5 to 10 years, has switched to an annual system in which 20 percent of the total number of sample plots (a “panel”) is measured annually. The FIA also has emphasized eliminating differences between inventory systems and database designs in the program regions and introducing a consistent system using the same database format.

During the transition from a periodic to annual system and the adaptation of the regional systems, many changes were made to the inventory design, manuals, and definitions. The process of the database conversion from Eastwide Forest Inventory database (Hansen *et al.* 1992) to the common FIA database introduced additional inconsistencies in data, causing a few changes in the calculation algorithms (Miles *et al.* 2001).

Georgia was one of the first southern States to introduce the annual forest inventory system. In 1997 FIA finished the last periodic inventory in the State (Thompson 1998), and then reorganized its inventory grid to match the national scheme, measuring single panels on an annual basis.

Problem Definition and Objective

Between 1998 and 2004, three panels were measured in Georgia, and data were made publicly available on the FIA server (see table 1). Because data were collected over a few years, the official manual changed during measurements of the particular panel.

This project sought to identify and describe the consequences of differences in definitions of several variables collected by the inventory crews in the field on results obtained during the data processing. We chose the following variables:

- Timberland area.
- Volume of all live trees on timberland.
- Growing stock volume on timberland.

For each variable, we show the definition according to the FIA manuals, present equations used for data compilation, and provide requirements (filters) used in the data processing algorithms. We compared results and validated them, if possible, using values in the official FIA publications or on FIA Web sites.

Definitions

Timberland

Timberland is defined as “forest land capable of producing 20 cubic feet of industrial wood per acre per year and not withdrawn from timber utilization” (Thompson 1998, p.10).

¹ Graduate Student Assistant, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602.

² Associate Professor, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602. Phone: 706-542-8169; fax: 706-542-8356; home page: www.growthandyield.com/chris.

³ Postdoctoral Fellow, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602.

⁴ Assistant Professor, Department of Dendrometry and Forest Productivity, Faculty of Forestry, Warsaw Agricultural University, Nowoursynowska 159, 02-776, Warsaw, Poland.

Table 1.—Source of publication for the last periodic inventory and three annual panels in Georgia—the 1997, 2000, 2001, and 2002 data sets as of 2003.

Timberland area		
Year	Published estimate (thousand acres)	Source
1997	23,796.1	Miles, <i>et al.</i>
2000	23,893.6	http://66.147.25.28/publicweb/individual_states/html/ga_info_statement.htm
2001	23,890.7	http://66.147.25.28/html/Panel%202.doc
2002	23,894.8	http://66.147.25.28/html/Panel7.doc
Volume of all live trees on timberland		
Year	Published estimate (million cubic feet)	Source
1997	33,661.4	Miles, <i>et al.</i>
2000	35,001.9	http://66.147.25.28/publicweb/individual_states/html/ga_info_statement.htm
2001	34,659.6	http://66.147.25.28/html/Panel%202.doc
2002	35,649.6	http://66.147.25.28/html/Panel7.doc
Volume of growing stock trees on timberland		
Year	Published estimate (million cubic feet)	Source
1997	31,704.0	Miles, <i>et al.</i>
2000	31,206.5	http://66.147.25.28/publicweb/individual_states/html/ga_info_statement.htm
2001	31,151.8	http://66.147.25.28/html/Panel%202.doc
2002	32,185.3	http://66.147.25.28/html/Panel7.doc

The raw data, originally downloaded from http://srsfia2.fs.fed.us/html/ga_data_disclaimer.htm, has been moved to http://66.147.25.28/publicweb/individual_states/ga/ga_data_disclaimer.htm.

To calculate the estimate of total timberland acreage for a State from the FIA data, the number of acres that each condition represents is calculated and these values are summed over all conditions meeting the definition of being a timberland condition (Miles *et al.* 2001, p.104).

The number of acres that a condition represents is calculated as the product of the following variables:

- *expcurr*; “The number of acres the sample plot represents for making current estimates of area” (Miles *et al.* 2001, p.30).
- *condprop*; “Proportion of the plot that is in the condition” (Miles *et al.* 2001, p.42).

A condition is a timberland condition if it meets the following requirements:

- *landclcd*=1; classified as accessible forest land— “Land at least 10 percent stocked by forest trees of any size, or formerly having had such tree cover, and not currently developed for nonforest use. The minimum area considered for classification is 1 acre. Forested strips must be at least 120 feet wide” (Thompson 1998, p.43).
- *reservcd*=0; not “withdrawn by law(s) prohibiting the management of the land for the production of wood products” (Miles *et al.* 2001, p.44).
- *siteclcd* in (1,2,3,4,5,6); capable of producing at least 20 cubic feet per acre per year (Miles *et al.* 2001).

Volume of All Live Trees

The volume of a live tree is “the cubic-foot volume of sound wood” in a live tree that is “at least 5.0 inches d.b.h. [diameter at breast height] from a 1-foot stump to a minimum 4.0-inch

top DOB [diameter outside bark] of the central stem” (Thompson 1988, p.10).

To calculate the estimate of volume of all live trees on timberland for a State from the FIA data, each tree’s expanded net cubic foot volume is calculated and these values are summed over all live trees that are on timberland conditions (Miles *et al.* 2001).

The expanded net cubic foot volume of a tree is calculated as the product of the following variables:

- expvol; the number of acres the tree’s plot represents for making volume estimates (Miles *et al.* 2001).
- tpacurr; “Current number of trees per acre that the tree represents for calculating number of trees on forest land” (Miles *et al.* 2001, p.81).
- volcfnet; “The net volume of wood in the central stem of a sample tree 5.0 inches diameter or larger, from a 1-foot stump to a minimum 4-inch top DOB, or to where the central stem breaks into limbs all of which are less than 4.0 inches DOB” (Miles *et al.* 2001, p.82).

A tree is identified as a live tree from the data by statuscd=1; “Identifies whether the sample tree is live, cut, or dead” (Miles *et al.* 2001, p.69).

Growing Stock Volume

Growing stock trees are defined as “living trees of commercial species classified as sawtimber, poletimber, saplings, and seedlings. Trees must contain at least one 12-foot or two 8-foot logs in the saw-log portion, currently or potentially (if too small to qualify), to be classed as growing stock. The log(s) must meet dimension and merchantability standards to qualify. Trees must also have, currently or potentially, one-third of the gross board-foot volume in sound wood” (Thompson 1998, p.8).

To calculate the estimate of volume of growing stock trees on timberland for a State from the FIA data, each tree’s expanded net cubic foot volume is calculated and these values are summed over growing stock trees that are on a timberland conditions (Miles *et al.* 2001).

A tree is identified as a growing stock tree from the data by treeclcd=2; “All trees of commercial species, except rough or rotten cull trees” (Miles *et al.* 2001, p.73).

For the three above variables, we developed the equations below to make calculations on the condition-level (for area

estimates) or tree-level (for volume estimates) data sets and summed over conditions or trees meeting the requirement filters below.

Equations

For TIMBERLAND, we used the following equation:

Equation 1: (Miles *et al.* 2001, table 2)

$$\text{(The number of acres a condition represents) =}$$

$$\text{expcurr*condprop} = \frac{\text{ac. ac}_{\text{COND}}}{\text{Plot ac}_{\text{PLOT}}}$$

For VOLUME OF ALL LIVE TREES ON TIMBERLAND and GROWING STOCK VOLUME ON TIMBERLAND, the equations were as follows:

Equation 1: (Miles *et al.* 2001, table 4)

$$\text{expvol*tpacurr*volcfnet} = \frac{\text{ac. Trees}}{\text{Plot ac.}} \frac{\text{ft}^3}{\text{Tree}_i} = \frac{\text{ft}^3}{\text{Plot}}$$

Equation 2: Same as equation 1 except for the inclusion of condprop in the multiplication.

$$\text{expvol*tpacurr*volcfnet*condprop}$$

$$= \frac{\text{ac. Trees}}{\text{Plot ac.}} \frac{\text{ft}^3}{\text{Tree}_i} \frac{\text{ac}_{\text{COND}}}{\text{ac}_{\text{PLOT}}} = \frac{\text{ft}^3}{\text{Plot}} \frac{\text{ac}_{\text{COND}}}{\text{ac}_{\text{PLOT}}}$$

Requirements (Filters)

For TIMBERLAND area, we tested the following two filters:

Filter 1: This filter defines a timberland condition (Miles *et al.* 2001, Thompson 1998)

Condition-level requirements	
landclcd=1	Accessible forest
reserved=0	Not reserved land
siteclcd in (1,2,3,4,5,6)	Land capable of producing more than 19 cubic feet/acre/year

Filter 2: Same as Filter 1 except for summing over conditions where siteclcd is 2 to 7 instead of 1 to 6, or for land capable of producing no more than 224 cubic feet/acre/year.

For VOLUME OF ALL LIVE TREES ON TIMBERLAND, we tested the following three filters:

Filter 1: (Miles *et al.* 2001, table 4)

Condition-level requirements	
landclcd=1	Accessible forest
reservcd=0	Not reserved land
siteclcd in (1,2,3,4,5,6)	Land capable of producing more than 19 cubic feet/acre/year
Tree-level requirements	
statuscd=1	Live trees

Filter 2: Same as Filter 1 except for summing over trees where statuscd = 1 or 4 instead of 1 only, or for trees having status as live trees or missed live trees.

Filter 3: Same as Filter 1 except for summing over trees in conditions where siteclcd is 2 to 6 instead of 1 to 6, or for land capable of producing between 19 and 224 cubic feet/acre/year. Finally, for GROWING STOCK VOLUME ON TIMBERLAND the three filters applied were as follows:

Filter 1: (Miles *et al.* 2001, table 4)

Condition-level requirements	
landclcd=1	Accessible forest
reservcd=0	Not reserved land
siteclcd in (1,2,3,4,5,6)	Land capable of producing more than 19 cubic feet/acre/year
Tree-level requirements	
statuscd=1	Live trees
treeclcd=2	Tree class code

Filter 2: Same as Filter 1 except for summing over trees where statuscd = 1 or 4 instead of 1 only, or for trees having status as live trees or missed live trees.

Filter 3: Same as Filter 1 except for summing over trees in conditions where siteclcd is 2 to 7 instead of 1 to 6, or for land capable of producing no more than 224 cubic feet/acre/year.

Results

When we state in our report that a compiled estimate “is the same as” or “matches” the published estimate, we mean that no difference existed between the two at the precision of the published estimate. For example, for timberland the difference was less than 100 acres.

Results obtained for the two TIMBERLAND algorithms are presented in table 2.

When timberland area is compiled using the suggested algorithm (equation 1 and filter 1 for timberland area) (Miles *et al.* 2001, table 2), the compiled estimate is the same as the published estimate for all 4 years except 2001, which is 17,900 acres less than the published estimate, and less than one-tenth of a percent different (-0.07 percent).

The difference in compiled estimates of timberland area for 2001 from algorithm 1 to algorithm 2 is explained in terms of the number of conditions in site index classes in table 3, and in terms of the number of acres represented by all conditions in Georgia’s 2001 data in each siteclcd in table 4. The difference between summing over siteclcd is 1 to 6 (algorithm 1) and siteclcd is 2 to 7 (algorithm 2) is 17,000 acres (22,000 and 5,000 acres with algorithms 1 and 2, respectively) (see tables 3 and 4).

Results obtained for various algorithms for the VOLUME OF ALL LIVE TREES ON TIMBERLAND calculation are presented in table 5.

Table 2.—Results obtained for various TIMBERLAND algorithms.

Algorithm	Equation	Filter	Difference between estimates compiled using the indicated algorithm and the published estimate (1,000 acres)			
			1997	2000	2001	2002
1	1	1	0.0	0.0	-17.9	0.0
2	1	2	0.0	-11.4	-0.9	-3.8
Algorithm	Equation	Filter	Differences above expressed as a percent			
			1997 (%)	2000 (%)	2001 (%)	2002 (%)
1	1	1	0.0	0.0	-0.07	0.0
2	1	2	0.0	-0.04	-0.00	-0.02

Table 3.—Difference in compiled estimates of the timberland area for 2001 from algorithm 1 to algorithm 2 (the number of conditions in each site index class).

siteclcd	Number of conditions	1,000 acres
1	1	5.0
2	15	107.9
3	138	1,072.6
4	664	5,584.9
5	1,451	12,225.8
6	585	4,876.7
7	2	22.0

Number of acres represented by conditions where landclcd=1, reservcd=0 and by the 7 siteclcd instances, from the 2001 data

Table 4.—Difference in compiled estimates of the timberland area for 2001 from algorithm 1 to algorithm 2 (number of acres represented by all conditions in Georgia's 2001 data).

siteclcd in	Sum	Difference between the sum and the 2001 published estimate for timberland area
1 to 6	23,872.8	-17.9
1 to 7	23,894.8	4.1
2 to 6	23,867.9	-22.8
2 to 7	23,889.8	-0.9

Sum of (expcurr*condprop) over conditions in 2001 data where landclcd=1 and reservcd=0 and siteclcd instances

Table 5.—Results obtained for various algorithms for the VOLUME OF ALL LIVE TREES ON TIMBERLAND calculation.

Algorithm	Equation	Filter	Difference between estimates compiled using the indicated algorithm and the published estimate (million cubic feet)			
			1997	2000	2001	2002
1	1	1	0.0	23,404.0	24.0	59.2
2	1	3	0.0	23,385.6	20.4	56.4
3	2	2	-4,821.8	0.0	-6,233.9	-6,095.3
Differences above expressed as a percent						
Algorithm	Equation	Filter	1997 (%)	2000 (%)	2001 (%)	2002 (%)
1	1	1	0.0	66.86	0.07	0.17
2	1	3	0.0	66.81	0.06	0.16
3	2	2	-14.32	0.0	-17.99	-17.10

The suggested algorithm for volume of all live trees on timberland (equation 1 and filter 1 for volume of all live trees on timberland area) (Miles *et al.* 2001, table 4) resulted in the compiled estimate equaling the published estimate for only the 1997 data.

The compiled estimate of volume of all live trees on timberland from the 2000 data matched the published estimate when we modified the suggested algorithm by including the condprop variable as a product in the equation and including summing over missed live, as well as live, trees.

No algorithms made the compiled estimate of volume of all live trees on timberland match the published estimate for 2001 and 2002. Algorithm 2 resulted in the least difference for these 2 years, which is the same as the suggested algorithm with the modification of excluding trees on conditions where siteclcd=1 (table 6).

Results obtained for various algorithms for the calculations of GROWING STOCK VOLUME ON TIMBERLAND are presented in table 7.

As for VOLUME OF ALL TREES ON TIMBERLAND, using the suggested algorithm for GROWING STOCK VOLUME ON TIMBERLAND resulted in the compiled and published estimates matching for only the 1997 data, although they are close for 2001 (table 7). Similarly, when modifying the suggested algorithm by including condprop as a product in the equation and including summing over missed, as well as live trees, the compiled estimate matched the published estimate for the 2000 data.

No algorithms used for GROWING STOCK VOLUME ON TIMBERLAND provide the published estimates for 2001 and 2002. The closest estimates came from using equation 1 and filter 3, the suggested algorithm with the modification of excluding siteclcd=1 and including siteclcd=7. These differences can be explained in terms of the volume represented by each siteclcd (table 8).

Table 6.—Volume of live trees represented by each statuscd, siteclcd combination from the data for 2001 and 2002.

2001				2002			
statuscd	siteclcd	No. trees	Million Cu Ft	statuscd	siteclcd	No. trees	Million Cu Ft
1	1	19	3.6	1	1	19	2.8
1	2	415	318.1	1	2	2,262	1,244.9
1	3	3,811	2,506.4	1	3	9,259	5,125.9
1	4	19,507	9,015.1	1	4	30,233	10,833.9
1	5	39,091	14,215.7	1	5	51,324	14,338.2
1	6	15,586	8,624.7	1	6	10,538	4,163.1
1	7	41	7.5				
4	3	59	15.0	4	2	16	5.6
4	4	205	114.4	4	3	111	47.2
4	5	457	179.8	4	4	284	124.8
4	6	217	104.2	4	5	523	171.9
				4	6	130	44.8

Only combinations where volume is not equal to zero are shown.

Table 7.—Results obtained for various algorithms for GROWING STOCK VOLUME ON TIMBERLAND calculations.

Algorithm	Equation	Filter	Difference between estimates compiled using the indicated algorithm and the published estimate (million cubic feet)			
			1997	2000	2001	2002
1	1	1	0.1	20,888.1	2.0	38.2
2	1	3	0.1	20,869.8	-0.3	35.4
3	2	2	-4,523.3	0.0	-5,554.5	-5,458.5
Algorithm	Equation	Filter	Differences above expressed as a percent			
			1997 (%)	2000 (%)	2001 (%)	2002 (%)
1	1	1	0.0	66.94	0.01	0.12
2	1	3	0.0	66.88	0.0	0.11
3	2	2	-14.27	0.0	-17.83	-16.96

Table 8.—Volume of growing stock trees represented by each treeclcd, statuscd, siteclcd combination from the data for 2001–02.

2001					2002				
treeclcd	statuscd	siteclcd	No. trees	Million Cu Ft	treeclcd	statuscd	siteclcd	No. trees	Million Cu Ft
2	1	1	15	3.6	2	1	1	15	2.8
2	1	2	250	307.6	2	1	2	1,442	1,169.4
2	1	3	2,410	2,362.6	2	1	3	6,033	4,797.5
2	1	4	12,681	8,230.7	2	1	4	19,471	9,940.2
2	1	5	23,201	12,519.2	2	1	5	30,828	12,723.2
2	1	6	8,233	7,730.1	2	1	6	5,437	3,590.6
2	1	7	7	1.3					
2	4	3	42	11.2	2	4	2	11	5.4
2	4	4	140	101.7	2	4	3	71	39.5
2	4	5	303	141.9	2	4	4	184	108.2
2	4	6	129	85.9	2	4	5	355	138.4
					2	4	6	76	34.8
					4	4	6	8	0.6

Only combinations where volume is not equal to zero are shown.

Discussion

Analysts working with the FIA data may want to replicate the results published by the FIA. Users of FIA data must know which algorithms can be used on which sets of data. In this paper, we explain examples of potential problems and solutions that may be experienced while working with the new annual FIA forest inventory data. For instance, we used the data for the “Forest Maps” and “SAFIS vs. FIA” sections of the Fiber Supply Assessment about the forestry growth and yield Web page at <http://growthandyield.com/main/index.htm>. These sections contain compilations of area, volume, and other estimates for Georgia and 13 other southern States and allow users to compare data from the last periodic inventory and the three annual panels. We used estimates published by the FIA as points of reference for our compilation procedures. We were not able to reproduce some of these estimates after several attempts using various programs and software.

In our analysis, we used identical compilation procedures for all 4 years. All procedures were downloaded from FIA Web sites (table 1) and were performed with the same program, with the only difference in the program between years being an identification of the year. Many different modifications of the algorithms (i.e., different combinations of variables used in the equations and different combinations of values of variables used in the requirement filters) were tested; the ones that produced the best results (compiled estimate closest to published estimate) for at least one of the 4 years’ data are included in this report.

The compiled condition-level estimate that we tested (timberland area) differed from published estimates only for the 2001 data when the suggested algorithm was used. When the suggested algorithms were used for the tree-level estimates, volumes of all live and of growing stock trees, the largest difference between compiled and published estimates was from the 2000 data. Results for the two tree-level estimates were very similar in that using the suggested algorithm resulted in compiled and published estimates matching for only the 1997 data, being much different for the 2000 data, and less different for the 2001 and 2002 data. They were also similar because modifying the suggested algorithms for the two tree-level estimates by including the variable for condition proportion as a product in the equation and summing over missed live trees as well as live trees resulted in the compiled estimates equaling the published estimates for the 2000 data.

Also, from the 2001 and 2002 data, modifying which conditions were included according to site productivity class when compiling the two tree-level estimates resulted in the algorithm giving compiled estimates most close to published estimates. Modifying the algorithms in this way created an estimate only 0.01 percent more precise.

All differences between compiled and published estimates when the suggested algorithms are used might be considered small (less than 1 percent) except for both tree-level estimates from the 2000 data, which were both just under 67 percent different. The causes of anomalies presented here are not known. Compilations made from FIA data, however, could be made more efficient and less error prone with the data revision history knowledge, access to the code the FIA used to compile the published estimates, and knowing when these compilations were made; this would allow the published estimates to be matched with the data revision history. Also, we recommend access to data sets used for compiling published estimates before any revisions or changes are made, so that users can compile the unrevised data to check their procedures against the published estimates. Such measures would help FIA data users to make their own compilations by providing benchmarks for their own routines and eliminating the possibility of their estimates not matching those published by the FIA.

Literature Cited

Hansen, M.H.; Frieswyk, T.; Glover, J.F.; Kelly, J.F. 1992. The eastwide forest inventory database: users manual. Gen. Tech. Rep. NC-151. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 48 p.

Miles, P.D.; Brand, G.J.; Alerich, C.L.; *et al.* 2001. The forest inventory and analysis database: database description and users manual, version 1.0. http://www.ncrs2.fs.fed.us/4801/fiadb/fiadb_documentation/FIADB_DOCUMENTATION.htm. [Date accessed unknown.]

Thompson, M.T. 1998. Forest statistics for Georgia, 1997. RB-SRS-036. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 92 p.



K-Nearest Neighbor Estimation of Forest Attributes: Improving Mapping Efficiency

Andrew O. Finley, Alan R. Ek, Yun Bai, and Marvin E. Bauer¹

Abstract.—This paper describes our efforts in refining k-nearest neighbor forest attributes classification using U.S. Department of Agriculture Forest Service Forest Inventory and Analysis plot data and Landsat 7 Enhanced Thematic Mapper Plus imagery. The analysis focuses on FIA-defined forest type classification across St. Louis County in northeastern Minnesota. We outline three steps in the classification process that highlight improvements in mapping efficiency: (1) using transformed divergence for spectral feature selection, (2) applying a mathematical rule for reducing the nearest neighbor search set, and (3) using a database to reduce redundant nearest neighbor searches. Our trials suggest that when combined, these approaches can reduce mapping time by half without significant loss of accuracy.

The k-nearest neighbor (kNN) multisource inventory has proved timely, cost-efficient, and accurate in the Nordic countries and initial U.S. trials. (Franco-Lopez *et al.* 2001, Haapanen *et al.* 2004, McRoberts *et al.* 2002). This approach for extending field point inventories is ideally suited to the estimation and monitoring needs of Federal agencies, such as the U.S. Department of Agriculture (USDA) Forest Service, that conduct natural and agricultural resource inventories. It provides wall-to-wall maps of forest attributes, retains the natural data variation found in the field inventory (unlike many parametric algorithms), and provides precise and localized estimates in common metrics across large areas and various ownerships.

At a pixel-level classification, the kNN algorithm assigns each unknown (target) pixel the field attributes of the most similar reference pixels for which field data exists. Similarity is defined in terms of the feature space, typically measured as Euclidean or Mahalanobis distance between spectral features. The kNN algorithm is not mathematically complex; however,

using multiple image dates and features from each date, along with several thousand field reference observations, makes kNN pixel-based mapping of large areas very inefficient. Specifically, the kNN classification approximates to $F \cdot N$ distance calculations, where F is the number of pixels to classify and N is the number of references. For example, standard kNN mapping of a 1.3×10^6 ha area, with a pixel resolution of 30 m², and approximately 1,500 FIA field reference observations requires about 22 billion distance calculations and around 16 hours to process on a Pentium 4, single-processor computer.

Our study examined using USDA Forest Service Forest Inventory and Analysis (FIA) plot data and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) imagery in kNN classification of FIA-defined forest types. Specific emphasis is placed on improving mapping efficiency by reducing classification feature space, decreasing the number of distance calculations in the nearest neighbor search, and eliminating redundancy in redundant nearest neighbor searches by building a database of feature patterns associated with different forest type classes.

Study Area and Data

Study Area

St. Louis County, in northeastern Minnesota, is located in the FIA aspen-birch unit. For a detailed description of the study area, see Bauer *et al.* (1994).

FIA Plot Data

The FIA program began fieldwork for the sixth Minnesota forest inventory in 1999. This effort also initiated a new annual inventory or monitoring system. In this new system, approximately one-fifth of the field plots in the State are measured each year. The new inventory protocol collected field data on the four-subplot cluster plot configuration (USDA Forest Service 2000). This plot design consists of four 1/60-ha, fixed-radius, circular subplots linked as a cluster, with each of the three outer subplots located

¹ Department of Forest Resources, College of Natural Resources, University of Minnesota, 1530 Cleveland Ave. N., St. Paul, MN 55108. E-mail: afinley@gis.umn.edu

36.6 m from the center subplot. FIA assigns each subplot to the land use class recorded at the subplot center. The 2001 inventory sampled 1,853 forested subplots in our study area. We removed 83 subplots because they fell under cloud-covered areas in the Landsat imagery. We removed an additional 58 subplots because the FIA field crew and FIA algorithm disagreed in the subplot forest type. The subsequent analysis used the remaining 1,712 subplots.

Satellite Imagery

We used Landsat 7 ETM+ satellite images for the analysis. The study area fell within two Landsat image scenes—path 27, rows 26 and 27. Bands 1 to 5 and 7 of three year 2000 dates were used including a late winter scene from March 12, a spring scene from April 29 and a late spring scene from May 31.

The images were geo-referenced to the Universal Transverse Mercator coordinate system using the following parameters: spheroid GRS80, datum NAD83, and zone 15. The resampling method was nearest neighbor using a 30-m by 30-m pixel size. The geo-referencing reference map was road vectors from the Minnesota Department of Transportation. For image portions with few roads, we used the U.S. Geological Survey digital orthophoto quads from the years 1991 to 1992 with 3-m resolution. The number of control points used in geo-referencing was 38 to 46 per date in path 27, row 27 images and 20 to 22 in 27/26 images. A second order polynomial regression model was used to fit the image. The root mean square error for all six images was less than 8 m. The clouds were digitized by hand and a cloud mask was created.

A forest/nonforest mask was generated using a kNN classification described in Haapanen *et al.* (2004). This mask was used to define the area extent of our forest type classification.

Methods

k-Nearest Neighbor Algorithm

For estimating with Euclidean distances, consider the spectral distance $d_{p_i,p}$, which is computed in the feature space from the target pixel p to each reference pixel p_i for which the forest type class is known. For each pixel p , sort the k -nearest field plot pixels (in the feature space) by $d_{p_1,p} \leq \dots \leq d_{p_k,p}$. The imputed value for the pixel p is then expressed as a function of

the closest units, each such unit weighted according to this distance decomposition function:

$$w_{p_i,p} = \frac{1}{d_{p_i,p}^t} \bigg/ \sum_{j=1}^k \frac{1}{d_{p_j,p}^t}, \quad (1)$$

where t is a distance decomposition factor set equal to 1 for all trials. To impute class variables such as forest type, the distance decomposition function calculates a weighted mode value.

For a class variable, the error rate (*Err*) indicates the disagreement between a predicted value \hat{y} and the actual response y in a dichotomous situation such that y does or does not belong to class i (Efron and Tibshirani 1993). Thus, we used the overall accuracy (*OA*) (Stehman 1997, Congalton 1991) defined as follows:

$$OA = 1 - Err, \quad (2)$$

where

$$Err = \sum_{i=1}^n (y_i - \hat{y}) / n \quad (3)$$

This is a special case of the mean square error for an indicator variable. These estimators were preferred over the usual Kappa estimator for reasons given by Franco-Lopez *et al.* (2001).

Errors were estimated by leave-one-out cross-validation. This technique omits training sample units one by one and mimics the use of independent data (Gong 1986). For each omission, we applied the kNN prediction rule to the remaining sample. Subsequently, the errors from these predictions were summarized. In total, we applied the prediction rule n times and predicted the outcome for n units. Such estimates of prediction error are nearly unbiased (Efron and Tibshirani 1993).

Spectral Feature Selection

As described by McRoberts *et al.* (2002), it is useful to select a parsimonious set of image features to use in the nearest neighbor search. Specifically, McRoberts *et al.* caution that including features unrelated to the attribute being estimated can reduce classification accuracy. When using a non-weighted Euclidian measure for the minimum distance criterion, the inclusion of these unrelated features directly reduces the class discriminating power of the entire feature set.

Instead of testing all combinations of spectral features in our analysis set, we used the statistical separability measure of transformed divergence to find feature subsets that adequately

discriminate among forest type classes. As described by Swain and Davis (1978), measures of divergence can be used to select a subset of feature axes that maximally separate class density functions. The degree to which class density functions diverge, or are separated in a multidimensional space, determines the classification accuracy of parametric classifiers. This approach to feature subsetting should also be effective with non-parametric classifiers such as kNN.

We used the transformed divergence measure implemented in ERDAS, Inc. Imagine® geospatial imaging software to derive feature subsets. Then kNN classification accuracy statistics were generated for each subset and judged against the classification accuracy of the full set of 18 features (i.e., 6 bands from each of 3 Landsat images). The smallest feature subset that performed at least as well as the full feature set was then moved forward in the analysis.

Stratification

In a similar study using the kNN classifier to characterize forested landscape in northeastern Minnesota, Franco-Lopez *et al.* (2001) found that simply stratifying by upland and lowland significantly improved forest type classification accuracy. Based on these findings, we divided the study area into upland and lowland strata as delineated by the U.S. Wildlife and Fish Service (USWFS) National Wetland Inventory. These strata were then classified separately using their respective subplots.

Nearest Neighbor Search Reduction

As previously noted, using minimum Euclidian distance as a nearest neighbor criterion is not mathematically complex; however, mapping large areas computing $F \cdot N$ distance calculations can take significant computer processing time.

Ra and Kim (1993) proposed the mean-distance-ordered partial codebook search (MPS) algorithm to reduce the number of Euclidian distance calculations required in a nearest neighbor search. The first component of the algorithm is the minimum distance criterion, defined by Ra and Kim as the squared Euclidian distance (SED):

$$d_{E,i} = \sum_{j=i}^m (x_j - c_{ij})^2, \quad (4)$$

where x_j and c_{ij} are the j^{th} component of the target and reference vector respectively, and m is the dimension of the vector (i.e., number of features). The next element in the algorithm is the squared mean distance (SMD), defined as follows:

$$d_{M,i} = \left(\sum_{j=i}^m x_j - \sum_{j=i}^m c_{ij} \right)^2. \quad (5)$$

The algorithm calculates and sorts the first k nearest neighbor distances in the reference set. Then, the SMD is calculated for the $k + 1$ vector in the reference set. This value is then tested with this inequality:

$$d_{M,i} \leq m d_{E,max}, \quad (6)$$

where $d_{E,max}$ is the largest distance in the sorted set of k nearest neighbors. If this inequality is true, the SED is calculated for the $d_{M,i}$ and the set of $k + 1$ nearest neighbors is resorted and the maximum value is discarded. If the inequality is false, the $d_{M,i}$ reference vector is discarded. This procedure is repeated for every subsequent vector in the reference set until each is either included in the k nearest neighbor set or rejected.

Depending on the amount of dispersion in the reference set, the MPS algorithm can significantly reduce the number of Euclidian distance calculations required to classify a given target pixel. Specifically, a reference set that contains observations that are highly dispersed in feature space will require fewer SED calculations to find the k nearest neighbor set when compared to a reference set that contains observations that are underdispersed. Further, as k increases, the number of observations that pass the inequality will also increase and need to be considered using a full Euclidian comparison. Therefore, analyses with lower values of k will be more time efficient than analyses with higher values of k . We evaluate the usefulness of the MPS algorithm by its ability to reduce the average number of Euclidean distance calculations for different levels of k in the classification of our study area.

Database-Assisted Mapping

The use of the kNN classifier, or any classifier, relies on a correlation between characteristics of the target pixel (e.g., spectral features) and the characteristics of observations in a reference set for which additional information is known. It is this correlation

that allows for meaningful assignment of class-specific reference pixel information to target pixels. In forest type classification, for example, we hope for low variability of spectral features in a forest type class and high variability of spectral features among forest type classes. This desirable “within” relationship versus “among” class variability can also help to increase mapping efficiency.

When implementing a typical kNN classification, the algorithm discards the target/reference similarity information and imputed value after each pixel is processed. If within class variability is low, the typically discarded information could be saved and reused successfully to classify pixels elsewhere in the image. Both the storage and subsequent retrieval of this information would be more efficient than computing the kNN for a given image pixel. Using this premise, we tested the marginal efficiency of incorporating a database system into a kNN image classification.

Using the MySQL database system and the MySQL++ API, we designed a program that would insert, search, and retrieve records that hold pixel spectral features and the kNN imputed values associated with each feature pattern. To allow the database to efficiently search for a feature pattern, it was necessary to discretize the 0–256 range of the Landsat bands into a smaller number of units, referred to in this report as bins.

Many methods are available to discretize continuous data ranging from arbitrary bin assignments across the variable’s distribution to using complex algorithms for deciding bin range and placement (Chmielewski and Grzymala-Busse 1996). For our study, we used a relatively simple approach based on the normal probability density function. Each band in the image was divided into the same number of bins. The range and placement of the bins was contingent on the band’s mean and standard deviation. This approach holds the area under the bands’ theoretical distribution equal for each bin. That is, the bins that occurred near the mean are narrow and the bins on the distribution’s tails are proportionally wider. Bin counts of 6, 8, and 10 were tested for database search and retrieval efficiency.

Repeated discretization of a variable’s distribution will ultimately result in information loss. A balance must be struck between the amount of reduction in classification accuracy and improved classification efficiency. For this reason, we compared the information loss and efficiency gain through reducing the bin counts.

Our database-assisted mapping program started by making a connection with a predefined MySQL database. The database contained one table with columns to hold the discretized image features and a column to hold the kNN estimated imputation value. For more efficient record insertion, search, and retrieval, the database existed as a hash table in the main memory.

For each image pixel, the feature pattern was extracted and compared to all records in the database. If a match was found, the pixel received the imputation from the matching database record; otherwise, the kNN algorithm was used to assign the value. Each time the kNN algorithm was implemented to assign a classification value, the associated feature pattern and resulting imputation value were inserted in the database.

As noted above, for this database-assisted mapping to be useful, the information insertion, search, and retrieval process must be more efficient than implementing a single kNN search of the reference data set. Further, the discretization process required to make pattern matching efficient must not significantly degrade classification accuracy. Therefore, we evaluated the utility of database-assisted mapping through a series of time tests and classification accuracy comparisons.

Results and Discussion

Spectral Feature Selection

Using the transformed divergence measure implemented in ERDAS Imagine, we derived optimal subsets of 16, 14, 12, 10, 8, and 6 spectral features. Subsets that contained fewer than 14 features produced suboptimal classification accuracy and degraded confusion matrices. Therefore, the remainder of our analysis was performed using the 14 spectral feature subset.

Stratification

Based on results from previous studies, we divided our study area by upland and lowland strata. These areas were delineated based on USWFS National Wetland Inventory classification maps. Approximately 10 percent of the forested landscape in the study area was designated as lowland and contained 149 FIA subplots. The upland portion of the study area contained the remaining 1,563 FIA subplots.

Table 1.—Forest type group classification confusion matrix for combined upland and lowland strata, using $k = 7$.

		FIA Forest Type Groups				R. tot.	P. Acc.
		100	120	800	900		
Observed	100	37	67	1	71	176	43.53
	120	28	562	0	93	683	81.21
	800	0	0	0	28	28	0.00
	900	20	63	0	742	825	79.44
C. tot.		85	692	1	934	1712	

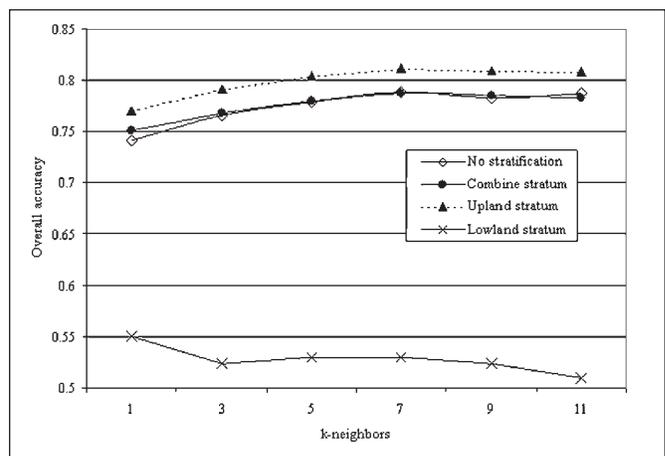
Recognizing that most classification errors resulted from within forest type group confusion, we collapsed the 12 forest types into their respective base groups. Figure 1 shows stratified and nonstratified classification accuracy for the four forest type groups sampled in our study area. Table 1 presents the combined classification confusion metrics for both strata. The spruce-fir group (forest type code 120) and aspen-birch group (forest type code 900) show satisfactory classification accuracy. Generalizing to the group level, however, does not address poor classification of the maple-beech-birch group (forest type code 800) or the aspen-birch group's overclassification.

Nearest Neighbor Search Reduction

Substituting the brute force nearest neighbor search, which considers the distance to all reference observations, with the MPS algorithm significantly reduced the number of distance calculations needed to classify each pixel. In leave-one-out cross-validation trials of 1,712 observations, each consisting of 14 features, we recorded the average number of observations in the $n - 1$ reference set that failed the inequality described in Equation 6. The trial results were $k = 1$ (74.4 percent failed), $k = 3$ (69.3 percent failed), $k = 5$ (66.4 percent failed), $k = 7$ (64.4 percent failed), $k = 9$ (62.6 percent failed), and $k = 11$ (61.2 percent failed).

As noted in the Methods section, as k increases, a higher probability exists that a reference observation will pass the inequality and require a full Euclidean distance comparison with the target. Our trials confirmed this relationship between increasing k and number of Euclidean distance measurements. Most

Figure 1.—Overall classification accuracy of four forest type groups at increasing values of k for no stratification, upland stratum, lowland stratum, and combined upland/lowland strata.



importantly, our research shows that using the MPS algorithm can significantly improve mapping efficiency by reducing the number of calculations needed to classify each target pixel.

Database-Assisted Mapping

Time trials using our database-assisted mapping program were conducted on a Pentium 4, 2 GHz, Linux OS-based- computer with 1 GB of memory. Our k NN program was written in C++, using the MySQL++ API to interact with a local MySQL version 4.0.14 server. Our program was compiled with g++ (GCC) 3.2.2.

The program was tested on the 3-date, 14-feature image of St. Louis County, which contains 14.72×10^6 pixels. The k NN reference set contained 1,712 subplot observations. Across all bin counts, the average time required for our program to search

the database for a given feature pattern and return a null or imputed value was 372 milliseconds. If a null value was returned, the kNN algorithm was initiated and took an average of 3,947 milliseconds to run. When a kNN instance was complete, the program used an additional average of 196 milliseconds to insert the feature pattern and imputed record in the database.

For individual bin counts, the insert, search, and retrieval time was contingent on the number of records in the table. Figure 2 shows that the total interaction time with the database increases with the database size. After processing completed, the table held 10-bin = 10.6×10^6 records; 8-bin = 7.29×10^6 records; and 6-bin = 3.2×10^6 records.

Figure 3 shows the frequency at which imputed values were retrieved from the database as the image is processed. The fewer bins in the image, the greater redundancy there was in feature patterns. The greater the redundancy in feature patterns, the greater dependence on the database was required to provide imputed values. Figure 3 also shows a rough plateau starting at about 6.5×10^5 processed pixels. This leveling off point describes the percent redundancy in the image for the given bin count (e.g., approximately 80-percent redundancy in the 6-bin image).

The brute force kNN procedure took 16.1 hours to process the sample image. Incorporating the database with a bin count of 10 decreased mapping time by 23.3 percent. Reducing the

bin count to 8 provided a 43.2 percent decrease in mapping time. Generalizing the image further to a bin count of 6 reduces mapping time by 67.1 percent.

This improvement in mapping efficiency must be balanced against accuracy loss from the discretization process. Figure 4 compares the forest type group classification overall accuracy of the binned images against the nonbinned image and shows that minor loss of information occurred in the 10- and 8-bin images. At the 6-bin image count of, accuracy declined more substantially.

The reduction in overall accuracy does not appear to be significant despite the severity of the image generalization. Deciding on the level of acceptable loss of classification accuracy in return for increased efficiency, however, is specific to the mapping project.

Our simple approach to imposing bin boundaries on each feature appears to maintain a significant portion of information; however, many other discretization procedures exist that may more effectively preserve class discriminating information contained in the images. Because of high spectral similarity, it is difficult to differentiate between forest types or forest type groups. Using our database-assisted mapping may enjoy experience success if classes were more spectrally distinct.

Figure 2.—Total database interaction time versus number of pixels processed for image bin counts of 10, 8, and 6.

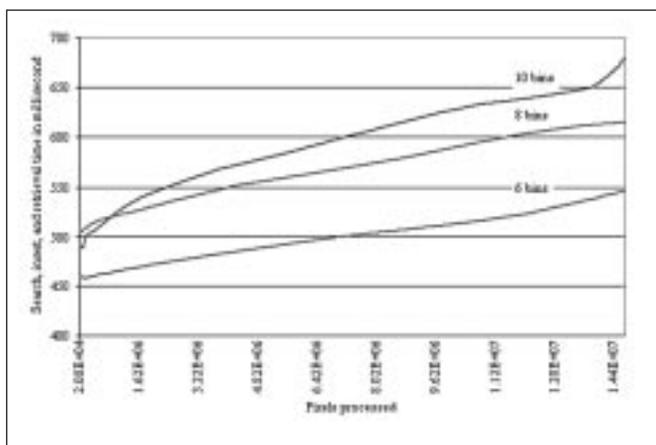


Figure 3.—Percent of imputed values retrieved from the database versus number of pixels processed for image bin counts of 10, 8, and 6.

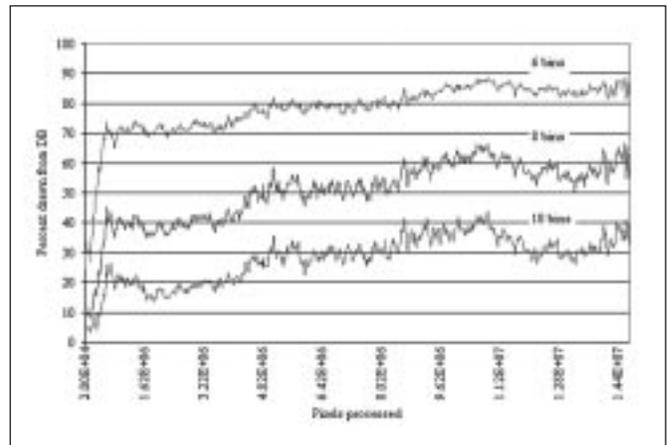
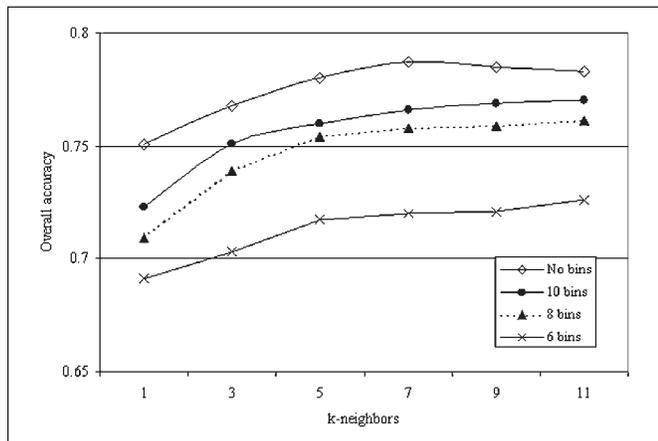


Figure 4.—Overall classification accuracy of four forest type groups at increasing values of k for varying degrees of image generalization.



Conclusions

Our study documented efforts to refine the process of kNN forest attributes classification using FIA plot data and Landsat 7 ETM+ imagery. We outlined three steps in the classification process that highlighted mapping efficiency improvements.

First, our analysis indicated that using transformed divergence may provide an objective way to reduce the dimensionality of the feature set without compromising classification accuracy. Second, the MPS algorithm proved to significantly reduce the number of distance calculations needed to classify each pixel. Third, the proposed database-assisted mapping provides a way to store and retrieve computationally expensive information.

Unlike the MPS algorithm, the discretization step needed for database-assisted mapping requires the analyst to compromise between increased mapping efficiency and loss of classification accuracy. Depending on the structure of the dataset and the degree of discretization, the loss of classification accuracy can be minimal.

Literature Cited

- Bauer, M.E.; Burk, T.E.; Ek, A.R.; Coppin, P.R.; *et al.* 1994. Satellite inventory of Minnesota's forest resources. *Photogrammetric Engineering and Remote Sensing*. 60(3): 287–298.
- Chmielewski, M.R.; Grzymala-Busse, J.W. 1996. Global discretization of continuous attributes as preprocessing for machine learning. *International Journal of Approximate Reasoning*. 15: 319–331.
- Congalton, R.G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*. 37: 35–46.
- Efron, B.; Tibshirani, R.J. 1993. *An introduction to the bootstrap*. New York: Chapman & Hall. 436 p.
- Franco-Lopez, H.; Ek, A.R.; Bauer, M.E. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sensing of Environment*. 77(3): 251–274.
- Gong, G. 1986. Cross-validation, the jackknife, and the bootstrap: excess error estimation in forward logistic regression. *Journal of the American Statistical Association*. 81: 108–113.
- Haapanen, R.; Ek, A.R.; Bauer, M.E.; Finley, A.O. 2004. Delineation of forest/nonforest and other land-use/cover classes using k nearest neighbor classification. *Remote Sensing of Environment*. 89(3): 265–271.
- McRoberts, R.E.; Nelson, M.D.; Wendt, D.G. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the k-nearest neighbors technique. *Remote Sensing of Environment*. 82: 2–3, 457–468.

Ra, S.W.; Kim, J.K. 1993. A fast mean-distance-ordered partial codebook search algorithm for image vector quantization. *IEEE Transactions on Circuit and System, Part II: Analog and Digital Signal Processing*. 40(9): 576–579.

Stehman, S.V. 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*. 62: 77–89.

Swain, P.H.; Davis, S.M., eds. 1978. *Remote sensing: the quantitative approach*. New York: McGraw-Hill. 396 p.

U.S. Department of Agriculture (USDA) Forest Service. 2000. *Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 1.4*. U.S. Department of Agriculture, Forest Service. Internal report. On file at U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, Washington Office.

Proof of Concept for an Approach to a Finer Resolution Inventory

Chris J. Cieszewski¹, Kim Iles², Roger C. Lowe³, and Michal Zasada⁴

Abstract.—This report presents a proof of concept for a statistical framework to develop a timely, accurate, and unbiased fiber supply assessment in the State of Georgia, U.S.A. The proposed approach is based on using various data sources and modeling techniques to calibrate satellite image-based statewide stand lists, which provide initial estimates for a State inventory on a common timeline. The system is based on using Georgia ground inventory data from the forest products industry, enhanced by various geographic information system and remote sensing data, and applied with the k-th “nearest neighbor” methods to time-series-stratified satellite imagery. The initial estimates are then scaled regionally to the Forest Inventory and Analysis (FIA) summary totals to eliminate potential bias in the initial estimates. The system enhances the FIA inventory data in four significant ways. First, it removes the need for the specific FIA plot coordinates; although, the coordinates, if available, would probably enhance the analysis. Second, it provides a current common timeline of inventory estimates based on the Landsat Thematic Mapper imagery for the given year and season. Third, it provides currently accurate high-resolution area estimates. Last, it uses various auxiliary data available from private and public sources in the State and can easily take advantage of other data as they become available.

The American Forest and Paper Association’s Second Blue Ribbon Panel (BRP) on the Forest Inventory and Analysis (FIA) program called for developing and implementing a plan

to conduct a national inventory to be coordinated with State foresters, Federal land management agencies, forest industry, nongovernmental organizations, and others. In response to the second BRP’s recommendations, the Agricultural Research, Extension, and Education Reform Act of 1998 (Section 253c) mandated that the U.S. Department of Agriculture (USDA) Forest Service conduct forest inventories in all States at the 20 percent annual rate of sample plots. The forest community is expected to obtain timelier and more accurate estimates of timber inventories and changes in fiber supply due to harvests, urbanization/sub-urbanization, natural disasters, and reforestation programs. Georgia was one of the first southern States to participate in the Southern Annual Forest Inventory System (SAFIS) in partnership with the USDA Forest Service FIA program. Currently, the Georgia Forestry Commission (GFC) provides 11 full-time equivalent positions, trucks, per diem travel, and other supplies to collect FIA data in Georgia.

Full implementation of an annual system by FIA requires reliable forested-area estimates, and standardized operating procedures to maximize benefits from informational resources such as satellite data.

Without locally accurate area estimates, the informational value of an annual sample is greatly reduced. One option for timely and accurate area estimation is to use remotely sensed spatial data such as Landsat Thematic Mapper (TM) satellite imagery, which has supplied reliable land cover area estimates in many parts of the country (e.g., Evans 1994, Rack 2001, Scrivani *et al.* 2001, Wynne *et al.* 2000).

Georgia is the third fastest growing State in the United States, although 72 percent of its land is forest cover with 9.55 million hectares in commercial forests, more than any other State (Smith *et al.* 2002). With more than two-thirds of that forest owned by approximately 630,000 private landowners and forestry contributing more than \$30 billion to Georgia’s economy each

¹ Associate Professor, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602 USA; Corresponding Author. Phone: 706-542-8169; fax: 706-542-8356; Web site: www.growthandyield.com/chris.

² Consultant on Forest Inventory, Kim Iles and Associates Inc., 412 Valley Place, Nanaimo BC, V9R 6A6, Canada.

³ GIS Analyst.

⁴ Postdoctoral Fellow, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602 USA; Assistant Professor, Department of Dendrometry and Forest Productivity, Faculty of Forestry, Warsaw Agricultural University, Nowoursynowska 159, 02-776 Warsaw, Poland.

year (Cieszewski *et al.* 2000), the results of this project will be useful to many forest managers and the State's economy.

Objective

Although Georgia government and forest industry support many forest activities and multiple sources of forestry data exist, the State does not have a timely and accurate high-resolution, spatially explicit forest inventory. The FIA-generated State inventory produced a low-resolution statewide survey with reliable estimates for very large areas, but does not provide accurate local, fine-scale values. In addition, the FIA inventory estimates are not derived for a common timeline and define a moving lagged average of the resource availability that is delayed by 3 to several years.

The objective of our project is to use the FIA and other data to derive timely and accurate high-resolution estimates for Georgia forests every year for the current year.

Proposed Approach

A large-scale inventory is very precise at the overall level but imprecise at the polygon level. More field plots will *not* solve this problem; they may exacerbate the problem by adding cost and delay. The objective of obtaining polygon-level precision, therefore, must be sought without the benefit of any additional fieldwork. The most promising approach to achieving this goal is to estimate every polygon volume or other characteristic and to ensure that these estimates add up to an appropriate total. How to determine this total is a separate topic and is beyond the scope of our study. The British Columbia Vegetation Inventory and a number of private forest companies have employed this approach.

The large-scale inventory maintained by FIA assumes that useful data and more precise results must come from statistical samples. Any inventory's design must, of course, be based on certain properties, such as unbiased data and correct measurements. The same problem arose during the design phase for the British Columbia Vegetation inventory; our project will apply the solution and approach used for that inventory.

An inventory that adds up to the same total as any unbiased estimate, regardless the source of that unbiased estimate, is itself an unbiased estimate. Therefore, we can make an estimate for every polygon on the land base, and then ensure that the sum of these estimates is constrained to add up to the total provided by an unbiased statistical process. The FIA is better equipped than any other organization to provide such an unbiased total for large areas in the United States. The FIA does not have the resources to provide the fine scale resolution of this total in individual polygon values. Other organizations, however, are prepared to make the individual polygon estimates, which can then be constrained by the FIA results.

Polygon estimates can be made using several methods, including "nearest neighbor" estimates, historical data, old inventory data, projected past values, aerial photos, and personal judgment, and any remote sensing technology. Inconsistent or partially available data is not a problem. Currently, the only advantage satellite methods offer is their ability to process large amounts of data, which increases the refresh cycle frequency. Satellite imagery, however, does not provide acceptable accuracy, because of a self-imposed attempt to produce the estimates automatically and because of insufficient resolution. One important advantage of satellite information is its ability to detect large-magnitude change.

Typically, the first objection made to using many estimates of polygon values is that they are biased. Adjusting the estimated parts to an unbiased overall total addresses this problem. A second objection, that such estimates are only available for a portion of the area or provide inconsistent precision, is not a serious constraint.

Table 1 presents a simple example of a small group of polygons that have been changed based on local knowledge of some sort to provide a more precise polygon-level estimate. The initial sample that the total of the polygon is based on was unbiased and provided a set of statistics describing that total. The absence of bias in the initial procedure and the value of any statistics regarding that total or average also apply for the revised polygon values. The difference is the improved polygon level resolution. The process's flexibility and inclusiveness are evident because other groups can contribute to the process; a "ground truth" visit, however, can verify any potential change.

Three significant changes in forest inventory data use and maintenance have occurred during the past few decades:

Table 1.—A simple example of a small group of polygons that have been changed based on local knowledge to provide a more precise polygon level estimate.

Polygon	Initial estimate	Final estimate	Further revised estimate	Further revision criteria
a	1,877	2,141.0	6,000	ecological guess
b	1,836	2,094.2	1,000	actual cruise
c	1,941	2,214.0	2,000	field visit
d	717	817.8	500	pure guess
e	1,584	1,806.8	2,200	10% more than c
f	996	1,136.1	600	similar neighbor
g	1,580	1,802.2	500	same as d
h	866	987.8	200	1/10 of c
	11,397	13,000	13,000	

Unbiased total = 13,000

Simple Correction Ratio = 1.141

- Aerial photography and other methods enable areas and forest value estimates to be made without fieldwork.
- Fast and high capacity databases allow individual polygons to have individual values; strata averages are not needed to store and report data only.
- Geographic information systems now function reliably, after a long and frustrating wait. Field information can be matched with information available from many sources.

The FIA contributed to these changes with the following actions:

- Developed a sample process to cover the entire land base, or at least make the grid extendable to all areas.
- Performed the fieldwork and made a continual effort to improve definitions, consistency, and data quality.

FIA data offers rigorous, statistically valid data with good quality control; other sources may offer only the ability to discriminate on a relative basis, and for only a portion of the overall land base.

The University of Georgia plans to combine these various types of data to create a fine-resolution inventory with location-specific information that is unbiased over some area to which it has been balanced. Because this information can be further refined, many specialist groups may be able to provide insight into improving the distribution of individual values that sum to a specific total.

How can this data be maintained and improved? When a newer or better estimate of the total is available, the individual polygons can be adjusted. Some polygons may be adjusted more than others, depending on how reliable the current estimate may be. Over time, the polygons should be grown or depleted according to the best information available. Although technically any inventory is biased as soon as the stands age, this detail is not expected to cause any serious errors.

The Forest Service has been working on several projects involving imputation and estimation that fit well with our project's approach. One closely associated method is the "most similar neighbor" work by Melinda Moeur, Al Stage, and others in the Forest Service (described in Moeur and Stage 1995).

Other Initial Estimation Aspects

The high-resolution inventory will be compiled in several steps. This section briefly describes the general framework for the unbiased, fine-resolution, spatially explicit estimation. First, to improve the analysis' accuracy, we will prestratify the Landsat TM images using multi-image change tracking.

Second, we will use various available inventory data provided by the forest industry and private forest land owners to develop models that stratify the satellite images to different species groups and volume/basal area classes in the prestratified classes. Although availability of the FIA exact coordinates would provide

more reference points to calibrate the k-th nearest neighbor models, it is not imperative because the FIA estimates are used to adjust all high-resolution estimates to the unbiased total or average.

Third, we will remove any bias in the high-resolution estimates by scaling them so that the sums of their volumes or basal areas in each satellite image will be equal to the corresponding sums in the FIA estimates for a corresponding timeframe. For example, the corresponding classification can be applied to scenes from the time of the inventory estimates and, after scaling, the corrected estimates can be forwarded to the current time. One challenge for us is to determine how to do the scaling. The FIA estimates are not for any given time but for an average in a 5-year period. At any time during this period, we can expect removals and growth that are intractable; ignoring the removals and growth, however, can create a bias.

Fourth, the adjusted estimates for the k-th nearest-neighbor-calibrated polygons will be used to compute the current inventory for the given year.

We expect satellite data to help quantify forested resource areas. In addition, using consecutive images over the last 30 years, we will be able to identify when specific areas were cleared and reestablished, so that we will be able to estimate their current ages. Optimal success in this effort requires reliable ground data on a large number of acres at different ages. Some acres will be used for training sites (that is, sites to develop classification algorithms); the remaining acres will be used to test modeled sites (that is, sites to evaluate the efficacy of the classification algorithms). We have industrial and private nonindustrial cooperators willing to provide these data.

Stand structure data from the Piedmont and Coastal Plain regions of Georgia will be obtained from our cooperators and supporters. Study location selections will be based on digital data availability from our large industrial cooperators. Data from neighboring nonindustrial private landowners who volunteer to be partners in this project (through cooperation with GFA) will fill in around these industrial land holdings.

We will generate an urban mask to disregard areas within city limits, such as parks, and the confusing satellite signatures from suburban areas. Using data from U.S. Geological Survey (USGS) paper maps and other available sources that show remote building locations will ensure that most dwellings and other structures are masked.

Using the field data in conjunction with the TM data, we will determine TM signatures for the forest types of interest. We will evaluate the consistency of these signatures in each satellite image. From the combination of the summer and winter TM data, we will generate a hardwood mask to help prevent us from confusing the signatures of pine at different ages with the signatures of hardwoods and pasturelands. Ancillary elevation and stream data will be used to help separate hardwoods in riparian zones from upland hardwoods and help define buffers along drainages where it may be difficult to distinguish between hardwoods and pines.

For each polygon (delineated area) of data provided by our cooperators and field crews, we will determine the “overriding” land cover class in the TM signatures for that same area. We will then evaluate the accuracy with which our list of the forest-types and stand ages can be classified. We will verify the accuracy of our forest/nonforest polygon classification and forest-type polygon classification based on the match between polygon field class and polygon satellite class. We will then reevaluate the TM signatures and recheck some field locations before we report which forest-types are most commonly confused in TM signatures and why.

The primary analyses will focus on investigating the images’ changes over time, which mark the harvested polygons. The change points will be examined with geostatistical methods, such as variograms and cross-semivariograms (Zawadzki *et al.*, 2005) that define the cross-sectional changes consistently over time, except for periods and locations of disturbances, which this approach will attempt to identify.

Data

In this analysis, we use FIA data, forest ground inventory data obtained from the local forest industry, geographic information system (GIS) data, and the Landsat TM imagery. The FIA data came from the plot FIA database (Hansen *et al.* 1992, Miles *et al.* 2001). The initial estimates of a high-resolution statewide forest inventory will be based on various spatial data available publicly and privately. Some examples of the publicly available data, other than the USDA FIA data, are described below.

-
- The Georgia GIS Data Clearinghouse (GGDC) (<http://www.gis.state.ga.us/Clearinghouse/clearinghouse.html>) provides access to numerous county-level GIS data for the entire State.
 - Hydrology data in vector format are available at the 1:24,000 scale. These data sets were captured from the USGS 1:24,000-scale topographic quadrangles and include linear features such as rivers and streams and polygonal features like lakes and ponds. Most features are attributed by class (e.g., perennial, intermittent) so that major and minor rivers and streams can be determined.
 - Road and highway data are available at the GGDC at the 1:12,000 scale. These data were captured from the 1993 digital orthophoto quarter quadrangles. They contain public roads including interstates, State highways, county roads, and city streets. These vector data are well suited to incorporate in various distance-related analyses in which the features are buffered to create polygons for further investigation.
 - The GGDC also serves raster data. Digital elevation models (DEMs) are available at the 1:24,000 scale and a 30-meter pixel size.
 - DEMs contain elevation information from which slope and aspect data sets can be derived.
 - Land cover data are available at the 1:100,000 scale and a 30-meter pixel size. These data, developed using satellite imagery from the late 1980s and the early 1990s, divide the landscape into different classes such as conifer, deciduous, agriculture, and urban. Though dated, they provide a source for stratifying the landscape into broad cover types.
 - Aerial photographs, historical and recent, are available from the GGDC in digital format and in paper format from the University of Georgia's Science Library and the GGDC. The GGDC sells two sets of digital aerial photographs:
 - The 1993 black and white digital orthophoto quarter quadrangles (DOQQs) have a 1-meter pixel and are available for the entire State.
 - The 1999 color-infrared photos (1-meter pixel) are available for select counties.
 - The University of Georgia's Science Library maintains a large set of historical paper aerial photographs from the early to mid to late 1990s.
 - USGS sells recent paper aerial photographs from the 1980s through the current decade (http://edcns17.cr.usgs.gov/finder/finder_main.pl?dataset_name=NAPP). These data provide a good model verification foundation.
 - Satellite imagery is available from the USGS EROS Data Center (<http://edc.usgs.gov>).
 - Landsat Multispectral Scanner (MSS), Landsat 5 TM, and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) satellite data are suitable for these types of landscape studies (Note: Recent malfunctioning equipment for ETM+ has yielded some suspect data in a scene).
 - Other satellite imagery available includes ASTER, MODIS, and AVHRR, each of which can be used to discriminate between land cover types.
- Our industrial partners supplied various GIS data, including boundaries and tabular data that may be the richest source of forest information. The final inventory of Georgia's forest resources will be scaled to be consistent with the FIA inventory regional and subregional statistical summaries (Thompson 1998).

Summary

This report describes a proof of concept to develop a high-resolution inventory based on pooling information from various types and sources of data. Because the data do not originate in a consistent statistical framework, they are likely to initially generate a biased inventory. Therefore, the initial inventory estimates are scaled to make the summary values equal to the summary values of the FIA inventory estimates, which will remove any existing bias in the final estimates of the high-resolution inventory. The proposed approach should allow timelier and more accurate inventory estimate compiling than either the initial remote-sensing-only based inventory or the moving average FIA survey estimates.

Literature Cited

- Cieszewski, C.J.; Borders, B.E.; Whiffen, H.; Harrison, W.M. 2000. Forest inventory in Georgia. In: Zawila-Niedzwiecki, T.; Brach M., eds. Proceedings, IUFRO conference on remote sensing and forest monitoring; 1999 June 1–3. Rogów, Poland: Warsaw Agricultural University, Faculty of Forestry.
- Evans, D.L. 1994. Forest cover from Landsat Thematic Mapper data for use in the Catahoula Ranger District geographic information system. Gen. Tech. Rep. SO-99. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 14 p.
- Hansen, M.H.; Frieswyk, T.; Glover, J.F.; Kelly, J.F. 1992. The eastwide forest inventory database: users manual. Gen. Tech. Rep. NC-151. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 48 p.
- Miles, P.D.; Brand, G.J.; Alerich, C.L.; *et al.* 2001. The forest inventory and analysis database: database description and user's manual, version 1.0, rev. 8 (1 June 2001). St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station (http://www.ncrs2.fs.fed.us/4801/fiadb/fiadb_documentation/fiadb_chapter1.htm).
- Moeur, M.; Stage, A.R. 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. *Forest Science*. 41(2): 337–359.
- Rack, J. 2001. Forest/nonforest classification of Landsat TM data for annual inventory phase one stratification. In: Reams, G.A.; McRoberts, R.E.; Van Deusen, P.C., eds. Proceedings, 2nd annual forest inventory and analysis symposium; 2000 October 17–18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 8–10.
- Scrivani, J.A.; Wynne, R.H.; Blinn, C.F.; Musy, R.F. 2001. Phase I forest area estimation using Landsat TM and iterative guided spectral class rejection: assessment of possible training data protocols. In: Reams, G.A.; McRoberts, R.E.; Van Deusen, P.C., eds. Proceedings, 2nd annual forest inventory analysis symposium; 2000 October 17–18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 11–14.
- Smith, B.W.; Vissage, J.S.; Darr, D.R.; Sheffield, R.M. 2002. Forest resources of the United States, 1997, METRIC UNITS. Gen. Tech. Rep. NC-222. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 127 p.
- Thompson, M.T. 1998. Forest statistics for Georgia, 1997. Resour. Bull. SRS-36. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 92 p.
- Wynne, R.H.; Oderwald, R.G.; Reams, G.A.; Scrivani, J.A. 2000. Optical remote sensing for forest area estimation. *Journal of Forestry*. 98(5): 31–36.
- Zawadzki, J.; Cieszewski, C.J.; Zasada, M. 2005. Use of semi-variances for studies of Landsat TM image textural properties of loblolly pine forests. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; McWilliams, W.H.; Cieszewski, C.J., eds. Proceedings, 4th annual forest inventory and analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 258 p.

Can a Forest/Nonforest Change Map Improve the Precision of Forest Area, Volume, Growth, Removals, and Mortality Estimates?

Dale D. Gormanson, Mark H. Hansen, and Ronald E. McRoberts¹

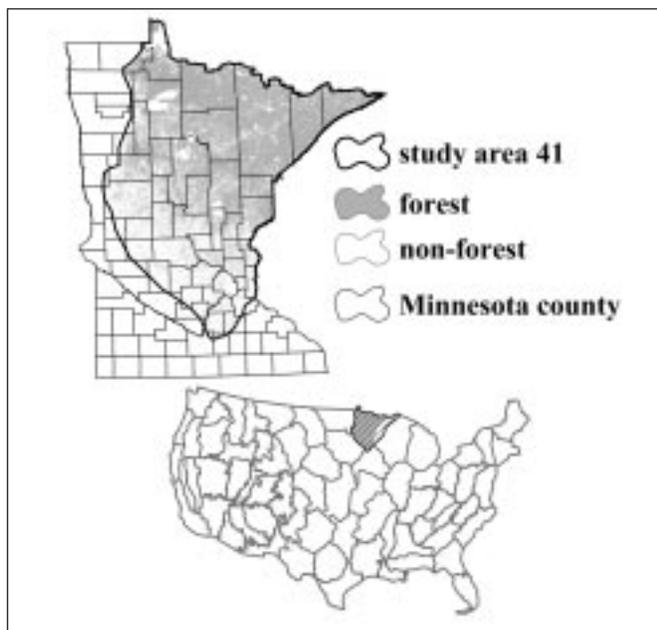
Abstract.—In an extensive forest inventory, stratifications that use dual-date forest/nonforest classifications of Landsat Thematic Mapper data approximately 10 years apart are tested against similar classifications that use data from only one date. Alternative stratifications that further define edge strata as pixels adjacent to a forest/nonforest boundary are included in the test. The variance of stratified estimates of total forest area, volume, growth, removals, and mortality are used to compare results. The two classifications tested are the 1992 and 2001 National Land Cover Data (NLCD) maps, referred to as NLCD-92 and NLCD-01, respectively. The study area is the Minnesota portion of Mapping Zone 41, an area that covers 60 percent of Minnesota and contains 90 percent of the State's forest. Permanent plot data from nearly 6,500 samples measured over the period 1999 to 2002 are used to compare the alternatives that include four different edge stratifications at the forest/nonforest boundary: one 2-pixel wide stratum, one 4-pixel wide stratum, two 1-pixel wide strata, and two 2-pixel wide strata. Estimates that use stratifications based only on the NLCD-92 were found to be more precise than estimates that used only the more recent NLCD-01 for stratification. Change-based stratifications (those that incorporated the NLCD-92 and NLCD-01) produced estimates with lower variances than estimates based on either single-date stratification alone, with the largest differences observed for the forest area estimates and smaller differences observed for estimates of volume, mortality, growth, and removals.

The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture (USDA) Forest Service has implemented an annual forest inventory system on a State-by-State basis in which a proportion of plots are measured in each State each year. The sampling design consists of rotating panels of permanent sample plots; the intent is to measure one complete panel each year. The time between inventories varies among States because of funding limitations and varying information needs. Population estimates and their sampling errors are calculated using stratified estimation with strata derived from classified satellite imagery. For the 11-State North Central region, stratifications derived from the 1992 National Land Cover Data data set (NLCD-92), developed by the Multi-Resolution Land Characterization Consortium (MRLC), have demonstrated that they can decrease forest area (FA) estimate variances by a factor exceeding 3.5 and volume (V) estimates by a smaller factor (Hansen and Wendt, 2000; McRoberts *et al.* 2002).

If the above stratifications derived from forest/nonforest classifications are effective in increasing the precision of estimates of forest attributes such as FA and V, analogy stratifications derived from forest change classifications may be effective in increasing the precision of estimates of forest change attributes such as growth (G), removals (R), and mortality (M). Completion of a final draft of a new National Land Cover Data data set (NLCD-01) by MRLC for most of Minnesota's forested area (fig. 1) provides the opportunity to construct such a land cover change classification. Our study's objective was to compare the precision of stratified FA, V, G, R, and M estimates using single-date stratifications derived from NLCD-92 and NLCD-01 and change-based classifications derived from NLCD-92 and NLCD-01 combined.

¹ Supervisory Forester, Research Forester, and Mathematical Statistician, respectively, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

Figure 1.—Mapping zones of the contiguous United States, with Minnesota study area.



Data

Study Area

Our study was conducted for the Minnesota portion of the MRLC Mapping Zone 41 (fig. 1) for which the final draft of NLCD-01 is completed. Mapping Zone 41 encompasses 33,025,274 acres, including approximately 60 percent of Minnesota's total land area and about 90 percent of the State's forested land area.

Land Cover Classifications

NLCD-92 is a 21-class land cover map of the conterminous United States at 30 m x 30 m resolution. The U.S. Geological Survey developed NLCD-92 under the auspices of MRLC using nominal 1992 Landsat 5 Thematic Mapper (TM) imagery and ancillary data (Vogelmann *et al.* 2001). NLCD-01 has similar characteristics but is based on nominal year 2001 Landsat 7 Enhanced Thematic Mapper Plus (ETM+) images (Homer *et al.* 2002).

FIA Plot Data

Inventory data were obtained from FIA plots in the study area that were measured between 1999 and 2002. Each FIA plot

consists of a cluster of four 1/24-acre, fixed radius, circular subplots distributed over approximately 1 acre. FA and V estimates were based on observations of all 6,492 plots in the study area; G, R, and M estimates were based on observations from the 3,535 plots measured for the 1990 periodic inventory and the first four panels (1999–2002) of the annual inventory. Fewer plots were available for estimating G, R, and M because estimates for these change variables require measurements for two inventories; FA and V estimates, however, require measurements for only one inventory. The sample size decreased for two reasons: some plots were removed from the inventory during the transition from periodic to annual inventories, and other plots could not be relocated during the annual inventory. Each FIA plot is linked to the NLCD-92 and NLCD-01 pixels in which the plot center is located. Most plot center locations were obtained using global positioning system (GPS) receivers, although locations of plots that digital orthophoto quad analysis determined had no accessible forest land were obtained instead with digitization methods. The accuracy standard for FIA plot locations is ± 140 feet (43.7 m) of the true location for 99 percent of the plots. A USDA Forest Service study reported the accuracy of GPS receivers of the kind used by FIA field crews to average approximately 7.9 m with maximum errors of approximately 20 m (Karsky *et al.* 2001). Therefore, we were certain that plot centers were linked to the 30 m x 30 m TM pixel containing the plot center or an adjacent pixel.

Methods

Stratifications

All stratifications were based on aggregations of NLCD classes into two initial classes, forest and nonforest. For both NLCD classifications, the forest class was constructed by aggregating the three NLCD forest classes (41—deciduous forest, 42—evergreen forest, and 43—mixed forest), the woody wetlands class (91—woody wetlands), and the shrub land class (51—shrub land for NLCD-92 and 52—short shrub land for NLCD-01). In addition, the NLCD-92 transitional class representing land in transition to forest was also grouped in the NLCD-92 forest class. No comparable class was identified in NLCD-01. The aggregated forest classes were constructed to be consistent with

FIA forest land definitions. To conform to the FIA requirement that forest land be at least 1 acre, a clump and eliminate algorithm (ERDAS 1997) was applied to the initial forest/nonforest classifications to eliminate 1-acre and smaller areas in each class.

Five stratifications were derived from each of the two NLCD aggregated forest/nonforest classifications. Two single-date forest/nonforest stratifications were constructed, one obtained directly from the NLCD-92 aggregated forest/nonforest classification and designated SDS2-92 and the other obtained directly from the NLCD-01 aggregated forest/nonforest classification and designated SDS2-01. In stratification designations, SDS is a single-date stratification, 2 means two strata [forest (F) and nonforest (NF)], and 92 and 01 refer to the NLCD classifications the stratifications were based on. Two edge strata were derived from each single-date forest/nonforest stratification based on the recommendations of Hansen and Wendt (2002) and McRoberts *et al.* (2002). The forest edge stratum (FE) consisted of pixels in the F stratum within one pixel of a forest/nonforest boundary, and the nonforest edge stratum (NFE) consisted of pixels in the NF stratum within one pixel of a forest/nonforest boundary. These stratifications are designated SDS4-92-1P and SDS4-01-1P, where 4 refers to four strata and 1P to the one-pixel edge strata width. Similar stratifications were constructed using a two-pixel edge stratum width and were designated SDS4-92-2P and SDS4-01-2P. The two edge strata, FE and NFE, were also collapsed into a single transition (T) stratum. For stratifications with one-pixel edge strata widths, the T stratum was two pixels wide; for the stratifications with two-pixel edge strata widths, the T stratum was four pixels wide. These stratifications were designated SDS3-92-2T, SDS3-01-2T, SDS3-92-4T, and SDS3-01-4T. The five SDS classifications based on NLCD-01 are depicted in figure 2, where the white dot indicates an FIA plot center.

Five change-based stratifications were constructed by creating two-way classifications of comparable single-date stratifications derived from the NLCD-92 and NLCD-01 (fig. 3). The n main diagonal cells of the two-way classifications are strata for which the assignments of pixels to NLCD 92 and NLCD 01 strata are identical; n^2-n off diagonal cells are strata for which assignments of pixels to NLCD 92 and NLCD 01 strata changed. Thus, NF-NF defines the change-based stratum for which the NLCD 92 and NLCD 01 strata pixel assignments were NF; F-NF defines the

change-based stratum for which the NLCD 92 and NLCD 01 strata pixels assignments changed from F to NF. However, we exercise caution in attributing actual ground change to pixels assigned to change strata, i.e., any strata off the diagonal in figure 3. Classification errors in NLCD-92, NLCD-01, or both will cause some erroneous pixel assignment to change strata even though no actual ground change occurs. In addition, error in image or NC-FIA plot registration will cause some erroneous pixel assignment to change strata. Although image misclassification and GPS plot location registration error reduces the effectiveness of the strata derived from classification by

Figure 2.—GPS monumented FIA plot (white dot) that straddles two land use conditions on a 1992 digital orthophotograph subset (a). Also shown is the stratification mapping progression of the 1992 digital orthophotograph compared to the (b) SDS2-01 (c), SDS3-01-2T (d), SDS4-01-1P (e) SDS3-01-4T, and (f) SDS4-01-2P stratifications.

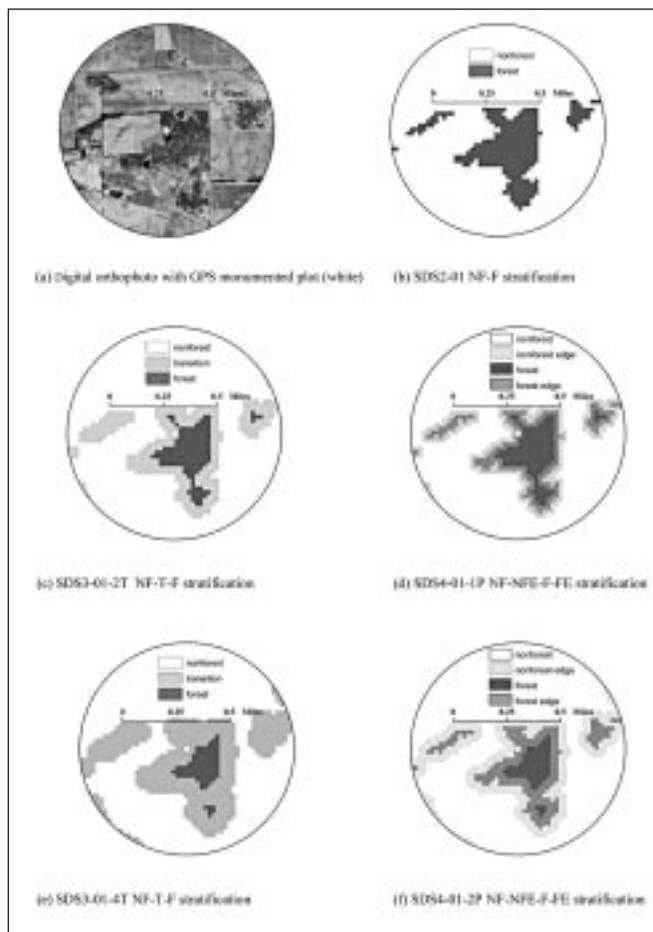
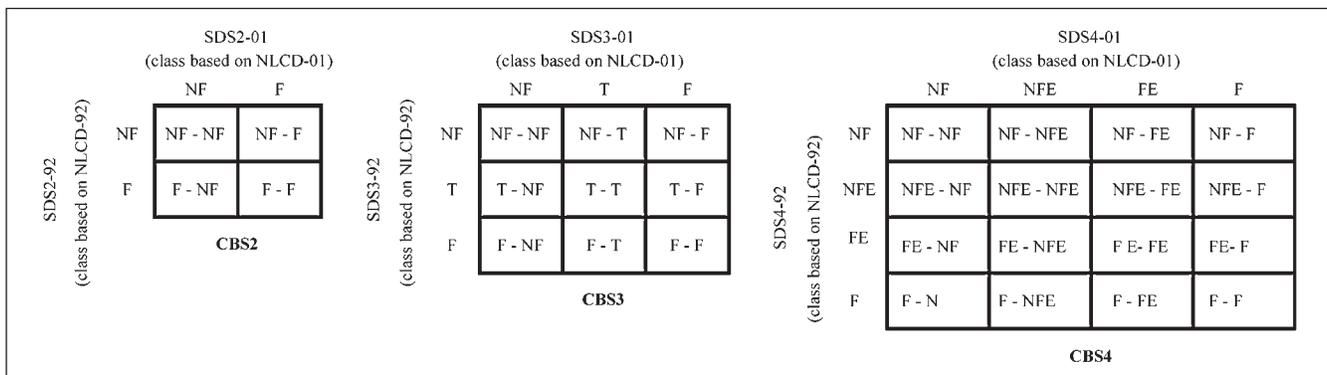


Figure 3.—CBS2, CBS3, and CBS4 change-based stratifications. For change-based stratifications, n^2 classes are possible, where n is the number of classes and $(n^2 - n)$ off-diagonal possible change classes.



decreasing precision, they will not produce bias in the estimates as long as the misclassifications are independent of ground sample plots. If using a consistent technique that does not incorporate the ground plot classification used later for estimation, the estimates will be unbiased.

The five change-based stratifications are designated using change-based stratification (CBS). CBS2 is based on SDS2-92 and SDS2-01, CBS3-2T on SDS3-92-2T and SDS3-01-2T, CBS3-4T on SDS3-92-4T and SDS3-01-4T, CBS4-1P on SDS4-92-1P and SDS4-01-1P, and CBS4-2P on SDS4-92-2P and SDS4-01-2P. Table 1 summarizes all stratifications.

Table 1.—Stratification alternatives.

Stratification	Description	Edge/transition strata	No. strata
<i>No STRATIFICATION</i>			
SRS	Simple random sample	None	1
<i>Single-date, NLCD-92</i>			
SDS2-92	NLCD 92 F/NF	None	2
SDS3-92-2T	NLCD 92 F/T/NF	2 pixel transition	3
SDS3-92-4T	NLCD 92 F/T/NF	4 pixel transition	3
SDS4-92-1P	NLCD 92 F/FE/NFE/NF	1 pixel edge	4
SDS4-92-2P	NLCD 92 F/FE/NFE/NF	2 pixel edge	4
<i>Single-date, NLCD-01</i>			
SDS2-01	NLCD 01 F/NF	None	2
SDS3-01-2T	NLCD 01 F/T/NF	2 pixel transition	3
SDS3-01-4T	NLCD 01 F/T/NF	4 pixel transition	3
SDS4-01-1P	NLCD 01 F/FE/NFE/NF	1 pixel edge	4
SDS4-01-2P	NLCD 01 F/FE/NFE/NF	2 pixel edge	4
<i>Change-based</i>			
CBS2	NLCD 92-01 F/NF change	None	4
CBS3-2T	NLCD 92-01 F/T/NF change	2 pixel transition	9
CBS3-4T	NLCD 92-01 F/T/NF change	4 pixel transition	9
CBS4-1P	NLCD 92-01 F/FE/NFE/NF change	1 pixel edge	16
CBS4-2P	NLCD 92-01 F/FE/NFE/NF change	2 pixel edge	16

Stratified Estimation

Stratified estimates and sampling errors were calculated using post-stratification estimators (Cochran 1977) with finite population correction ignored. Estimates of population totals were calculated with this formula:

$$\hat{Y}_{Str} = A \sum_{h=1}^L W_h \bar{y}_h \quad (1)$$

Estimates of sampling errors were calculated as follows:

$$S_{\hat{Y}_{Str}} = A \sqrt{\frac{1}{n} \sum_{h=1}^L W_h^2 s_h^2} \quad (2)$$

where h denotes stratum, L is the number of strata, n is the total number of plot observations, \bar{y}_h and s_h^2 are the observed sample mean and variance in stratum h , W_h is the weight for stratum h calculated as the proportion of pixels assigned the stratum, and A is the population area defined as the Minnesota portion of Mapping Zone 41.

Analyses

FA, V, G, R, and M estimates were calculated using [1] and [2] for the 15 stratifications and compared using two measures. The first measure, relative efficiency (RE), was calculated with this formula:

$$RE = S_{\hat{Y}_{SRS}}^2 / S_{\hat{Y}_{Str}}^2 \quad (3)$$

where $S_{\hat{Y}_{SRS}}^2$ is the variance of the estimate obtained under the assumption of simple random sampling (SRS) that uses no stratification, and $S_{\hat{Y}_{Str}}^2$ is the variance obtained using stratified estimation. RE values close to 1.0 indicate that the stratification has little use in increasing precision; RE values greater than 1.0 indicate increasing utility. RE is equivalent to the factor by which the sample size must be increased to obtain the same variance under SRS obtained using stratified estimation. The second measure was based on converting the observed sampling error to a percent per specified area or volume (Hansen 2001). This measure is reported in FIA publications as the sampling error per million acres for FA estimates and the sampling error per billion cubic feet for V, G, R, and M estimates. The percent sampling error per S , where S is 1 million acres or 1 billion cubic feet, is calculated as follows:

$$E_s = \frac{S_{\hat{Y}_{Str}}}{\sqrt{S}} 100\% \quad (4)$$

FIA guidelines are $E_s \leq 3.0$ for FA and $E_s \leq 5.0$ for V, G, and R.

Results

We noted several important results. First, the stratifications generally improved the precision of estimates (table 2). For FA, $RE > 2.0$ indicates that the stratifications were effective; however, for V, G, M, and R, RE only slightly greater than 1.0 indicates that the stratifications were less effective.

Second, FIA precision guidelines were satisfied or nearly satisfied for all variables except V. $E_s \approx 3.0$ indicates that the FIA precision guidelines were nearly satisfied for FA; $E_s < 5.0$ indicates that the guidelines were satisfied for G and R estimates; but E_s near 7.0 indicates that the precision guidelines were not satisfied for V.

Third, generally, the most precise estimates were obtained using one of the change-based stratifications, although RE with the change-based stratifications was often only slightly greater than with single-date stratifications. We attributed this result to the increased precision from more strata with the change-based stratifications.

Fourth, estimates obtained with single-date stratifications with three or four strata were more precise than estimates obtained with single-date stratifications with only two strata, although precision estimates for the stratifications with three and four strata were similar. Nevertheless, adding edge and/or transition strata increased precision.

Fifth, estimates were more precise when using single-date stratifications derived from NLCD-92 than those derived from NLCD-01 for all variables except V, for which the results were mixed. Although the differences were small, the trend is pronounced. The only apparent explanation for this phenomenon is that the aggregated forest and nonforest classes obtained from NLCD-92 conformed better to FIA definitions of forest land than did the classes obtained from NLCD-01, even with the nearly 10-year gap between the two sets of images. One possible reason is including the transition class in the NLCD-92 aggregated forest class.

Table 2.—Estimates by stratification alternative.

Stratification	Forest area (million ac)			Volume (billion ft ³)			Growth (million ft ³ /yr)			Mortality (million ft ³ /yr)			Removals (million ft ³ /yr)								
	L	\hat{Y}	$S_{\hat{Y}}$	E_S	RE	\hat{Y}	$S_{\hat{Y}}$	E_S	RE	\hat{Y}	$S_{\hat{Y}}$	E_S	RE	\hat{Y}	$S_{\hat{Y}}$	E_S	RE				
SRS	1	14.72	0.195	5.08	1.00	16.36	0.325	8.04	1.00	306.8	17.06	3.08	1.00	336.1	13.32	2.30	1.00	305.3	20.75	3.75	1.00
SDS2-92	2	15.19	0.128	3.28	2.33	16.89	0.287	6.98	1.28	284.3	15.47	2.90	1.22	309.9	11.72	2.10	1.29	282.4	19.02	3.58	1.19
SDS2-01	2	15.25	0.131	3.36	2.20	16.97	0.288	6.99	1.28	286.0	15.54	2.91	1.21	312.6	11.85	2.12	1.26	291.9	20.02	3.71	1.07
SDS3-92-2T	3	15.16	0.122	3.12	2.57	16.86	0.283	6.90	1.32	279.4	15.21	2.88	1.26	305.0	11.52	2.09	1.34	277.6	18.79	3.57	1.22
SDS3-01-2T	3	15.17	0.125	3.20	2.45	16.86	0.285	6.94	1.30	281.8	15.35	2.89	1.24	308.0	11.70	2.11	1.30	288.7	19.81	3.69	1.10
SDS3-92-4T	3	15.14	0.128	3.28	2.33	16.82	0.288	7.02	1.27	282.5	15.39	2.90	1.23	307.7	11.65	2.10	1.31	279.8	18.80	3.55	1.22
SDS3-01-4T	3	15.05	0.131	3.39	2.20	16.73	0.287	7.02	1.28	282.4	15.43	2.90	1.22	308.9	11.80	2.12	1.27	289.1	19.81	3.68	1.10
SDS4-92-1P	4	15.19	0.120	3.08	2.64	16.89	0.283	6.88	1.32	279.7	15.24	2.88	1.25	305.0	11.52	2.09	1.34	278.0	18.83	3.57	1.21
SDS4-01-1P	4	15.22	0.123	3.15	2.52	16.94	0.284	6.90	1.31	282.2	15.38	2.89	1.23	308.5	11.70	2.11	1.30	289.1	19.82	3.69	1.10
SDS4-92-2P	4	15.20	0.121	3.09	2.61	16.90	0.284	6.91	1.31	281.4	15.28	2.88	1.25	306.6	11.54	2.08	1.33	279.0	18.73	3.55	1.23
SDS4-01-2P	4	15.19	0.124	3.19	2.46	16.91	0.284	6.91	1.31	282.3	15.37	2.89	1.23	308.7	11.69	2.10	1.30	288.9	19.76	3.68	1.10
CBS2	4	15.28	0.121	3.09	2.60	16.99	0.283	6.86	1.32	282.3	15.34	2.89	1.24	308.4	11.60	2.09	1.32	281.9	18.97	3.57	1.20
CBS3-2T	9	15.16	0.114	2.94	2.91	16.87	0.280	6.81	1.35	278.9	15.14	2.87	1.27	302.3	11.40	2.07	1.37	271.9	18.61	3.57	1.24
CBS3-4T	9	15.11	0.119	3.05	2.70	16.81	0.283	6.90	1.32	279.4	15.13	2.86	1.27	303.1	11.41	2.07	1.36	274.3	18.92	3.61	1.20
CBS4-1P	16	15.20	0.114	2.92	2.92	16.94	0.280	6.80	1.35	279.7	15.23	2.88	1.26	303.6	11.47	2.08	1.35	271.7	18.45	3.54	1.26
CBS4-2P	16	15.20	0.114	2.92	2.93	16.95	0.281	6.83	1.34	280.3	15.17	2.87	1.26	304.1	11.40	2.07	1.37	272.9	18.50	3.54	1.26

L = number of strata.

 \hat{Y} = estimated total. $S_{\hat{Y}}$ = sampling error. E_S = sampling error per million acres (area) or billion cubic feet (volume).

RE = relative efficiency.

Finally, stratified estimates of the mean were similar for all variables, but SRS estimates differed from the stratified estimates. This result is attributed to forested strata undersampling for two reasons: first, some private land owners refused to allow FIA crews to measure plots on their lands, and second, some forested plots could not be relocated at the time of the second inventory. The undersampling generally did not occur for nonforested strata because plots determined to have no accessible forest land based on digital orthophoto quad analyses were not visited by field crews; hence, land owner permission was not required, and plots did not have to be relocated. Stratified estimation compensates for this phenomenon by producing independent, within-stratum estimates. Although the SRS estimates are biased, their precision estimates are still useful to compute RE for comparison.

The observed values of n_h , \bar{y}_h , s_h^2 , and W_h for the CBS3-2T stratification provide additional information on the effectiveness of the change-based stratifications (table 3). Overall, this nine-strata, change-based stratification produced estimates with the same or higher precision than other strata. Precision estimates for CBS4-2P were comparable but required nearly twice as many strata.

Three salient points are noted. First, more than 80 percent of the population did not change stratum assignment from NLCD-92 to NLCD-01. The changed portion of the population was distributed over four strata with weights ranging from 0.004 to 0.042. The few plots in some strata corresponding to change in strata assignments indicates that change-based stratifications may not be useful for subpopulations such as counties. Second, the general trend of smaller strata having large within-stratum variances indicates that the stratifications are correctly grouping high-variance plots into strata with small weights that contributes to stratification effectiveness. Third, the within-stratum means tend to conform to expectations that strata corresponding to F for NLCD-01 have greater means. One exception is the estimate of the FA mean in the F-NF stratum that, although expected to be relatively small, was actually large. This result could be attributed to the clump and eliminate procedure that reclassified predicted forested areas of less than 1 acre into the nonforest class. Thus, some forested plots in small, isolated patches of forest land that were actually larger than 1 acre may have been classified in error as nonforest and assigned to a nonforest stratum.

Table 3.—Observations and within-stratum estimates for CBS3-2T stratification.

NLCD-92 Attribute ² Stratum		NLCD-01 stratum														
		NF					T					F				
		p _f	v	g	m	r	p _f	v	g	m	r	p _f	v	g	m	r
NF	n _h	2261	2261	982	982	982	241	241	107	107	107	30	30	18	18	18
	W _h	0.328	0.328	0.328	0.328	0.328	0.035	0.035	0.035	0.035	0.035	0.006	0.006	0.006	0.006	0.006
	\bar{y}_h	0.007	5.1	0.21	0.04	0.29	0.099	99.4	2.88	1.42	3.93	0.3275	149.5	2.46	3.39	0.00
	s _h	0.072	82.4	3.03	0.79	7.10	0.2630	354.4	12.33	7.57	40.68	0.4472	280.3	10.82	8.13	0.00
T	n _h	236	236	117	117	117	656	656	366	366	366	268	268	177	177	177
	W _h	0.039	0.039	0.039	0.039	0.039	0.107	0.107	0.107	0.107	0.107	0.042	0.042	0.042	0.042	0.042
	\bar{y}_h	0.088	93.1	2.78	0.30	10.24	0.362	458.0	8.56	8.21	4.77	0.7139	697.5	10.15	12.63	8.81
	s _h	0.244	425.0	11.54	1.91	54.92	0.4071	740.0	30.92	22.94	20.75	0.3935	862.6	33.49	24.46	31.14
F	n _h	34	34	28	28	28	399	399	255	255	255	2367	2367	1485	1485	1485
	W _h	0.004	0.004	0.004	0.004	0.004	0.061	0.061	0.061	0.061	0.061	0.378	0.378	0.378	0.378	0.378
	\bar{y}_h	0.632	277.6	5.85	10.41	64.31	0.753	821.8	12.15	17.05	24.98	0.8752	983.7	16.00	17.38	13.00
	s _h	0.466	468.4	25.33	23.67	85.45	0.3746	903.7	36.12	29.34	54.92	0.3008	886.2	38.82	29.37	43.47

p_f=proportion of forest.
v=volume (ft³/acre).
g=growth (ft³/acre/yr).

m=mortality (ft³/acre/yr).
r=removals (ft³/acre/yr).

² n_h=sample size for stratum h.
W_h=weight for stratum h.

\bar{y}_h =sample mean for stratum h.
s_h=sample standard deviation for stratum h.

Conclusions

Our study confirms two previous findings. First, stratifications derived from NLCD aggregated forest/nonforest classifications are effective in increasing the precision of forest attribute estimates. Second, creating transition or edge strata at forest/nonforest boundaries enhances the effectiveness of the stratifications.

Three primary conclusions can be drawn from our study: (1) change-based stratifications improved the precision of forest attribute estimates, even though increases in precision over single-date stratifications were not great; (2) differences among the precision estimates for stratifications using three and four strata and two- and four-pixel widths for edge or transition strata were minimal; and (3) stratifications were more effective in increasing the precision of FA estimates than V, G, M, and R estimates. A possible reason for the last conclusion is that the classification of land as forest is based on the presence of forest canopy, which is not necessarily a good indicator of below canopy forest attributes such as volume, growth, removals, and mortality.

Acknowledgments

The authors gratefully acknowledge very helpful reviews by Dr. Thomas E. Burk, Department of Forest Resources, University of Minnesota, and Dr. James Westfall, USDA Forest Service Northeastern Research Station. We also appreciate and acknowledge the contribution of Dr. Limin Yang, EROS Data Center, U.S. Geological Survey, Sioux Falls, SD U.S.A., for providing the NLCD 2001, Zone 41 data.

Literature Cited

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

ERDAS, Inc. 1997. ERDAS field guide, 4th ed., rev. Atlanta, GA: ERDAS, Inc. 656 p.

Hansen, M.H. 2001. Remote sensing precision requirements for FIA estimation. In: Reams, G.A.; McRoberts, R.E.; Van Deusen, P.C., eds. Proceedings, 2nd annual forest inventory and analysis symposium; 2000 October 17–18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 43–51.

Hansen, M.H.; Wendt, D.G. 2000. Using classified Landsat Thematic Mapper data for stratification in a statewide forest inventory. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C., eds. Proceedings, 1st annual forest inventory and analysis symposium; 1999 November 2–3; San Antonio, TX. Gen. Tech. Rep. NC-213. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 20–27.

Homer, C.G.; Huang, C.; Yang, L.; Wylie, B. 2002. Development of a circa 2000 landcover database for the United States. In: Proceedings, American society of photogrammetry and remote sensing annual conference; 2002 April; Washington, DC. Bethesda, MD: American Society of Photogrammetry and Remote Sensing. CD-ROM.

Karsky, D.; Chamberlain, K.; Mancebo, S.; *et al.* 2001. Comparison of GPS receivers under a forest canopy: after selective availability has been turned off. Gen. Tech. Rep. 0171-2809-MTDC. Missoula, MT: U.S. Department of Agriculture, Forest Service, Missoula Technology and Development Center. 18 p.

McRoberts, R.E.; Wendt, D.G.; Nelson, M.D.; Hansen, M.H. 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment*. 81: 36–44.

Vogelmann, J.E.; Howard, S.M.; Yang, L.; *et al.* 2001. Completion of the 1990s National Land Cover Data data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*. 67: 650–662.

Stratum Weight Determination Using Shortest Path Algorithm

Susan L. King¹

Abstract.—Forest Inventory and Analysis uses post-stratification to calculate resource estimates. Each county has a different stratification, and the stratification may differ depending on the number of panels of data available. A “5 by 5 sum” filter was passed over the reclassified forest/nonforest Multi-Resolution Landscape Characterization image used in Phase 1, generating an image in which each pixel represents the count of forested pixels inside a 5 by 5 window. The forested pixel count ranges from 0 to 25 or 26 classes. In the next step, the ground plots are overlaid on the class map generated by the 5 by 5 window. The objective is to find the break points in the 26 classes that minimize the difference in the number of acres/plot between strata while simultaneously maximizing the number of strata. These are conflicting goals. More strata imply larger deviances between the strata. Also, the stratum must have contiguous classes with at least four plots. This is a nonlinear integer programming problem. Because software is not readily available to solve a nonlinear integer programming problem, the problem was reformulated to finding the shortest path through the network. For each county, the optimal one, two, three, four, five, six, and seven strata are found, and various heuristics for determining the final solution are investigated and compared.

Introduction

The annual Forest Inventory and Analysis (FIA) sampling design is composed of three phases. Phase 1 uses satellite imagery to classify the land area in a State as forest or nonforest. Phase 2 is the traditional ground sample. An interpenetrating hexagonal grid is placed across a State with one ground plot

per grid cell. Each hexagonal grid represents 5,937 acres.

One-fifth of the ground plots spread uniformly across the State are visited yearly. Each year’s plots/hexagonal grids are referred to as a panel. On a subset of the Phase 2 plots, additional variables are measured to determine forest health. This subset is the Phase 3 sample. This article focuses on finding an automated and efficient procedure for determining the optimal number of strata and, hence, the stratum weights for Phase 2 forest land estimates or “on the fly” resource estimates of user-defined polygons. The objective is to minimize the difference in the Phase 1-to-Phase 2 ratio between the strata (deviance) while simultaneously maximizing the number of strata. Each stratum must have a minimum of four ground plots. When the population is divided into as many homogenous strata as possible, the variance of the population estimates tends to be lower. As the number of strata increase, it becomes more difficult to find break points in which all the strata have approximately equal Phase 1-to-Phase 2 ratios.

Methods

Phase 1 and Phase 2 Cost Information

The satellite imagery used for the Phase 1 sample was a forest/nonforest map acquired from National Land Cover Data (formerly Multi-Resolution Land Characterization [MRLC]). This vegetation map was made by the U.S. Geological Survey Earth Resources Observation Systems (EROS) Data Center (Vogelmann *et al.* 2001) and is based on 1992 Landsat 7 Thematic Mapper data; other intermediate-scale spatial data were used as ancillary data. For the forest/nonforest call, the MRLC was reclassified so that the forest classes and woody wetland received a value of 1, and other pixels received a value of 0. A “5 by 5 sum” filter was passed over the reclassified forest/nonforest MRLC image, generating an image in which each pixel represents the count of forested pixels inside a 5 by 5 window. The forested pixel count ranges from 0 to 25, which creates 26 classes (bins). The Phase 2 plots were overlaid on the filtered image to obtain a

¹ Operations Research Analyst, U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA, 19073. Phone: 610-557-4048; fax: 610-557-4250; e-mail: sking01@fs.fed.us.

network has nodes joined by arcs, and, in this problem, the arcs are directed in only one direction. The objective is to traverse the network from the start node to the stop node with the least cost. Each arc has an associated cost, which could be time, money, distance, or another measure. In this case, cost is the deviance or smoothed deviance objective function. The stages in the network correspond to the strata. The network in figure 1 is a three-stage or three-strata network. The cost encountered from the start node to stage 1 is 0, and the cost encountered from stage 3 to the stop node is 0. The nodes in each stage correspond to the number of bins. In stage 1, all the nodes include the first bin, and, in stage 3, all the nodes contain bin 26. In stage 1, if bins 1 through 24 are selected, in the three-stratum case, bin 25 must be selected in stage 2, and bin 26 selected in stage 3. If only bin 1 is in the first stratum, many combinations exist for strata 2 and 3. If four plots are not in the bin combination at a node, the arc is assigned a large cost so that this path never will be selected. Paths containing these infeasible arcs are pruned from the network before solving for the shortest path through the network.

Many approaches and algorithms are available to solve a shortest path problem. Dijkstra's Algorithm (1959) is the classic method for computing the shortest path from a single source node to every other node in a weighted (stratum weight cost on an arc is a weight) network. This algorithm, a simple and consequently easily implemented algorithm for finding the shortest routes, is the most widely used in GIS software packages. Dijkstra's Algorithm is used for solving problems that require real-time solutions—for example, routing an ambulance to an accident site and from there to the nearest hospital. Its performance depends on the data structures (for example, heaps or priority queues) used to represent the network (Derekenaris *et al.* 2001). Improving the data structure efficiency of Dijkstra's Algorithm and other approaches to solving the shortest path problem are active areas of research. Nevertheless, shortest path algorithms are used routinely to solve large-scale problems and are available in most programming languages.

Almost any problem that can be formulated as a shortest path through a network also can be solved using dynamic programming, that is, the problem can be solved using recursive equations without special software. The drawback of dynamic programming is the "curse of dimensionality." As both the

number of stages and nodes increase, so do the number of recursions. Information for the recursions must be stored in a lookup table. The feasibility of using dynamic programming for solving the stratum weight problem was not investigated.

Another mathematical optimization approach to optimally allocating bins to the strata is nonlinear integer programming. If bin *i* is assigned to stratum *j*, the decision variable is 1; otherwise, the decision variable is 0. In addition to the constraints requiring at least four ground plots and consecutive bins, constraints are added to ensure that each bin is assigned to only one stratum, and that each stratum has at least one bin. The objective function is either equation (5) or (6). The objective function is nonlinear because the denominator term, number of ground plots in a stratum, is an integer variable. Commercial software is not readily available to solve nonlinear integer programming problems, but is readily available to solve a shortest path problem.

Data

Three panels of annual inventory data from Pennsylvania were used to evaluate the procedures for determining the stratum weights for each county. From this information, estimates are calculated for the number of acres of nonforest and forest for each county and the State. The complete range of forested conditions is found in Pennsylvania, from heavily nonforested to heavily forested counties. Heavily forested or nonforested counties may require only one stratum, whereas counties with a mixture of forest conditions may require as many as seven strata.

Table 1 shows the possible stratifications from the shortest path algorithm for Mifflin County using the smoothed deviance objective function and three annual panels. The shortest path cost increases as the number of strata increases. The one-stratum solution starts at the first bin and stops at the last bin. The cost is 9,763 acres/plot. For the two-stratum solution, the first stratum has bins 1 through 25, and the second stratum has bin 26. The shortest path is the difference between the cost of 10,194 acres/plot and 9,417 acres/plot, or 776.7 acres/plot. The shortest path cost increases as the number of strata increases. This precludes building one network and allowing the algorithm to select the strata combination with the lowest cost. From the table, dividing the 26 bins into two groups of equal cost is easier than three groups of equal cost. Larger numbers of strata should have

Table 1.—Possible stratifications for Mifflin County.

Strata number	Start1	Stop1	Cost1	Start2	Stop2	Cost2	Start3	Stop3	Cost3	Start4	Stop4	Cost4	Shortest path
1	1	26	9,763										
2	1	25	9,417	26	26	10,194							776.76
3	1	1	5,116	2	24	11,204	25	26	11,005				12,174.29
4	1	1	5,116	2	6	6,471	7	22	12,244	23	26	12,149	27,059.84

lower sampling errors, however. From a mathematical point of view, stopping at bin 25 for the first of two strata makes sense, but is it wise from a biological perspective? Bins may have no or a sparse number of ground plots, and they are grouped with the strata that best balances the cost.

Because the shortest path cost directly increases as the number of strata increases, a single network cannot be built because the smaller strata solutions would prevail. Therefore, several heuristics were investigated for selecting the “optimal” number of strata. (Optimal is in quotation marks because the procedures are rules of thumb and not mathematically based procedures that develop necessary and sufficient conditions for optimality). One heuristic was to divide the shortest path by the number of difference pairs and graph the new cost versus the number of strata. The hypothesis was that the curve would decrease, reach a minimum, and then increase. The solution would be the number of strata at which the curve reached its minimum. The curves for each county did not follow the expected pattern, however; the cost per difference pair tended to increase with an increasing number of strata. A second heuristic ties the last bin in the first stratum and the first bin in the last stratum to NLCD imagery classification break points developed by Hoppus *et al.* (2001). Not all the counties could be classified with NLCD imagery break points because the requirements are too stringent or the county is essentially all forest or nonforest and the only appropriate stratification is one stratum. The final procedure is a series of relaxations on the NLCD imagery requirement. The procedure is as follows.

Imagery-Based Heuristic for Selecting the “Optimal” Number of Strata

Step 1. Create table 1 for each county. For each strata combination, calculate:

Cost range = largest cost of a strata – smallest cost of a strata

Next, sort by county and descending cost range. This sorting guarantees that the solution with the largest number of strata that meets the remaining criteria will be selected first.

- Step 2.** Separate the counties that can be stratified only by one stratum (group A) from the remaining counties. From the remaining counties, select the counties with a cost range of less than 6,000 acres/plot, a break point between the first and second stratum at bin 7 or lower, and the break point between the highest stratum and its adjacent lower stratum at bin 24 or higher. (These break points are the NLCD imagery classification break points.) From the counties that meet these criteria, select the solution with the largest number of strata (group B). Remove group B counties from the remaining data.
- Step 3.** Relax the standards on the remaining data (original data set minus groups A and B). From the remaining counties, select the counties with a cost range of less than 6,000 acres/plot, fewer than 12 bins in the first stratum, and the last stratum in bin 19 or higher. From the counties that meet these criteria, select the solution with the largest number of strata (group C). Remove group C counties from the remaining data.
- Step 4.** Relax the standards on the remaining data (original data set minus groups A, B, and C). From the remaining counties, select those counties with a cost range of less than 6,000 acres/plot. From the counties that meet these criteria, select the solution with the largest number of strata (group D). Remove group D counties from the remaining data.
- Step 5.** Select any remaining counties based on the smallest cost range. Place these counties in group E.
- Step 6.** Add groups A, B, C, D, and E to form the final solution.

Ratio Heuristic

Another heuristic is the ratio heuristic:

$$\text{Ratio} = \frac{\text{Total Land Area (acres) in State}}{\text{Total Number of Plots Measured}} \quad (7)$$

For the first panel, the ratio is approximately 30,000 acres/plot. For the second and third panel, the ratio is approximately 15,000 and 10,000 acres/plot, respectively. Exact numbers for the ratio heuristic could easily be calculated, but the approximations are used in this study.

To implement the ratio heuristic, apply Step 1 in the imagery-based heuristic. For Step 2, find the solution with the highest number of strata so that the cost range is less than the ratio in equation (7).

Results

Currently, a human expert performs the stratification for the annual inventory. From the number of plots in the county and their distribution in the 26 classes, the expert can estimate the number of strata. Using this information and a spreadsheet macro, the expert can visually examine the impact of different

break points in the cost per stratum equation (4). The final solution is achieved when the expert believes that the cost cannot be further balanced among the strata. Table 2 shows the division of land between nonforest (0) and forest (1) for the three panels of annual inventory data in Pennsylvania.

The statistics in table 2 are calculated using FINSYS (Born and Barnard 1983), the computer program used by the Northeastern FIA unit to calculate sampling statistics. Because the data must be in a special format for FINSYS, “what if” questioning is difficult. As a result, the remaining statistics were calculated with a user-written SAS macro (SAS Institute 1999) and benchmarked against FINSYS. The results for the three-panel problem with a deviance objective function and a smoothed deviance objective function are shown in tables 3 and 4, respectively. According to the ratio rule, for three panels, the maximum cost range should be 10,000 acres/plot. For evaluation purposes, 6,000 acres/plot also was considered. The sampling

Table 2.—Human expert's result for three-panel stratification in Pennsylvania.

Forest land	Area (acres)	Mean area (%)	Sampling error (%)
0	12,030,500	41.9	1.2
1	16,652,100	58.1	0.9

Table 3.—Three-panel stratification using optimization and deviance objective function.

Procedure	Forest land	Area (acres)	Mean area (%)	Sampling error (%)
Imagery-based	0	11,767,610	41.03	1.179
	1	16,915,021	58.97	0.820
Cost range < 6,000 acres/plot	0	11,794,962	41.12	1.164
	1	16,887,670	58.88	0.813
Cost range < 10,000 acres/plot	0	11,769,360	41.03	1.172
	1	16,913,272	58.97	0.816

Table 4.—Three-panel stratification using optimization and smoothed deviance objective function.

Procedure	Forest land	Area (acres)	Mean area (%)	Sampling error (%)
Imagery-based	0	11,760,770	41.00	1.175
	1	16,921,862	59.00	0.817
Cost range < 6,000 acres/plot	0	11,794,956	41.12	1.169
	1	16,887,675	58.88	0.817
Cost range < 10,000 acres/plot	0	11,736,291	40.92	1.165

errors are lower for the optimization procedures. The mean area percentages differ by approximately 1 percent between the expert and the optimization. In the optimization groups, the imagery-based procedure had slightly higher sampling errors. For the optimization procedure, the lowest sampling errors were for the smoothed deviance objective function and the ratio decision rule using a cost range of less than 10,000 acres/plot. Consequently, in the remaining statistics, the imagery-based decision rule is not further investigated, and only the smoothed deviance objective function is investigated.

Table 5 shows the results for two-panel combinations. Panels 1 and 2 are shaded light gray, panels 1 and 3 are shaded dark gray, and panels 2 and 3 have a white background. For two panels,

each plot is worth approximately 15,000 acres. A decision rule of cost range of less than 10,000 acres/plot is for comparison. The sampling errors are close. The cost range of less than 10,000 acres/plot decision rule has slightly lower sampling errors in two of the three cases. In the first and second panel combination, the lowest sampling error for the forest land is for the cost range of less than 10,000 acres/plot decision rule, and the lowest sampling error for nonforest is the cost range of less than 15,000 acres/plot decision rule.

Table 6 presents the results for the single panel. Panels 1 and 2 are shaded light gray, panels 1 and 3 are shaded dark gray, and panels 2 and 3 have a white background. The ratio decision rule would be to accept the stratification with the largest number

Table 5.—Two-panel stratification using optimization and smoothed deviance objective function.

Procedure	Forest land	Area (acres)	Mean area (%)	Sampling error (%)
Cost range < 10,000 acres/plot	0	11,824,972	41.23	1.457
	1	16,857,659	58.77	1.022
Cost range < 15,000 acres/plot	0	11,916,395	41.55	1.443
	1	16,766,236	58.45	1.025
Cost range < 10,000 acres/plot	0	11,824,972	41.23	1.457
	1	16,857,659	58.77	1.022
Cost range < 15,000 acres/plot	0	11,810,641	41.18	1.498
	1	16,871,991	58.82	1.048
Cost range < 10,000 acres/plot	0	11,633,254	40.56	1.457
	1	17,049,378	59.44	0.994
Cost range < 15,000 acres/plot	0	11,660,065	40.65	1.495
	1	17,024,023	59.35	1.024

Table 6.—One-panel stratification using optimization and smoothed deviance objective function.

Procedure	Forest land	Area (acres)	Mean area (%)	Sampling error (%)
Difference < 15,000 acres/plot	0	12,105,840	42.21	2.367
	1	16,576,792	57.79	1.728
Difference < 30,000 acres/plot	0	12,215,132	42.59	2.289
	1	16,467,500	57.79	1.698
Difference < 15,000 acres/plot	0	11,841,438	41.28	2.396
	1	16,841,194	58.72	1.685
Difference < 30,000 acres/plot	0	11,632,008	40.55	2.344
	1	17,050,624	49.45	1.599
Difference < 15,000 acres/plot	0	11,872,889	41.39	2.363
	1	16,809,742	58.61	1.669
Difference < 30,000 acres/plot	0	11,871,364	41.39	2.338
	1	16,811,268	58.61	1.651

of strata so that the cost range between the stratum with the largest and smallest cost is less than 30,000 acres/plot. These results are contrasted with those obtained from using 15,000 acres/plot. The sampling errors are larger with fewer panels. The sampling errors are lower for the difference of less than 30,000 acres/plot for all three panels.

Conclusions

The procedure described in this article is an automated approach to determining the stratification for a county, State, or other polygon. Using this procedure achieved lower sampling errors than with the current human expert procedure. By formulating the problem as a shortest path through the network, fast and efficient computational procedures are available to provide a real-time solution. Actual solution time depends on the number of arcs in the network. Arcs increase with the number of polygons to be simultaneously processed and the number of strata. Pruning of infeasible paths before optimization and a fast computer processor reduce the solution time. Different shortest path algorithms affect the solution speed.

From the shortest path formulation, the optimal one-, two-, three-, four-, five-, six-, and seven-strata solutions are found. To find the “optimal” solution, several heuristics were investigated. The ratio heuristic is easily implemented and provided the smallest sampling errors.

Literature Cited

- Born, J.D.; Barnard, J.E. 1983. FINSYS-2: subsystem table-2 and output-2. Gen. Tech. Rep. NE-84. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experimental Station. 68 p.
- Derekenaris, G.; Garofalakis, J.; Makris, A.; *et al.* 2001. Integrating GIS, GPS and GSM technologies for the effective management of ambulances. *Computers, Environment, and Urban Systems*. 25: 267–278.
- Dijkstra, E. 1959. A note on two problems in connection with graphs. *Numerische Mathematik*. 1: 269–271.
- Hoppus, M.; Arner, S.; Lister, A. 2001. Stratifying FIA ground plots using a 3-year old MRLC forest cover map and current TM delivered variables selected by “decision tree” classification. In: Reams, G.A.; Mc Roberts, R.E.; Van Deusen, P.C., eds. *Proceedings, 2nd annual forest inventory and analysis symposium; 2000 October 17–18; Salt Lake City, UT*. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 19–24.
- SAS Institute. 1999. SAS macro language: reference, version 8. Cary, NC: SAS Institute. 324 p.
- Vogelmann, J.E.; Howard, S.M.; Yang, L.; *et al.* 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering & Remote Sensing*. 67(3): 650–661.



A Three-Step Approach To Model Tree Mortality in the State of Georgia

Qingmin Meng¹, Chris J. Cieszewski², Roger C. Lowe³, and Michal Zasada⁴

Abstract.—Tree mortality is one of the most complex phenomena of forest growth and yield. Many types of factors affect tree mortality, which is considered difficult to predict. This study presents a new systematic approach to simulate tree mortality based on the integration of statistical models and geographical information systems. This method begins with variable preselection using multiple linear regression models and logistic models and employs spatial autocorrelation detection and random sampling. Three random sampling methods are applied and compared to reduce the effects of spatial autocorrelation, and systematic random sampling significantly reduces the spatial autocorrelation among the observations and is used for the final variable selection and model fitting. Using Forest Inventory and Analysis (FIA) data for the State of Georgia, this systematic approach provides significant implications for future tree mortality studies and other spatial analysis in forestry or geography.

Forest tree mortality is an important factor in nutrient cycling as well as global climate warming because mortality and net primary production are two critical processes of forest carbon budgets (Brown and Schroeder 1999). In addition, a large portion of the threatened forest species lives in dead wood (Rouvinen *et al.* 2002). At the same time, forest tree mortality may reduce the productivity of forests and increase the risk of wildfires. Tree mortality, however, is considered difficult to predict.

The literature contains many reports from different studies on tree mortality. For example, Greene *et al.* (1992) and Pedersen and McCune (2002) conducted research on mortality rates, using

wind as the primary disturbance contributing to tree mortality, and found that the total biomass declined from 1979 to 1989 because the new biomass production was less than mortality. Pedersen and McCune (2002) modeled tree mortality rates as a function of diameter at breast height (d.b.h.), species, decades, and site index. In their study, they reconstructed tree mortality rates for the years 1968 to 1977 and 1978 and 1987 in oak-hickory forest. Rouvinen *et al.* (2002) researched tree mortality, its causes, and spatial pattern in four transects with a total area of 48.8 ha of Vienansalo wilderness in eastern Fennoscandia, Finland. They divided mortality into three categories: current, recent, and predicted mortality. They concluded that tree mortality was continuous at the landscape scale, although some spatial aggregations occurred. Osawa *et al.* (1986) conducted systematic research on forest tree mortality using Baxter State Park as the study area and compared tree mortality among various onsite topographical conditions. Basal area, d.b.h., or stand age is supposed to significantly contribute to tree mortality (Fridman and Stahl 2001, Yang *et al.* 2003, Monserud and Sterba 1999, Avila and Burkhart 1992, and Zhang *et al.* 1997).

In the 1960s and 1970s, linear and polynomial models were commonly used (e.g., Lee 1971). Osawa *et al.* (1986) concluded, however, that multiple regression analysis was unsuccessful in relating tree mortality to forest structural characteristics and topographical properties. Nonlinear models, especially the logistic functions, have been the most widely used functions for mortality modeling from Walker and Duncan (1967) and Neter and Maynes (1970). Guan and Gertner (1991) pointed out that the best function to model individual tree mortality may be the logistic function based on statistical tests.

We conducted a systematic study on tree mortality in the State of Georgia. In our research, we performed three steps: variable preselection based on original data, sampling and spatial autocorrelation comparison, and model fitting and selection.

¹ Graduate Student Assistant.

² Associate Professor, Corresponding Author, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602 USA. Phone: 706-542-8169; fax: 706-542-8356; home page: www.growthandyield.com/chris.

³ GIS Analyst, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602 USA.

⁴ Postdoctoral Fellow, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602 USA; Assistant Professor, Department of Dendrometry and Forest Productivity, Faculty of Forestry, Warsaw Agricultural University, Nowoursynowska 159, 02-776 Warsaw, Poland.

Study Area and Data

Our study area was the entire State of Georgia, the largest State east of the Mississippi River, with an area of 152,576 square kilometers. We based our research on available, online, forest inventory data provided by the Forest Inventory and Analysis (FIA) program.

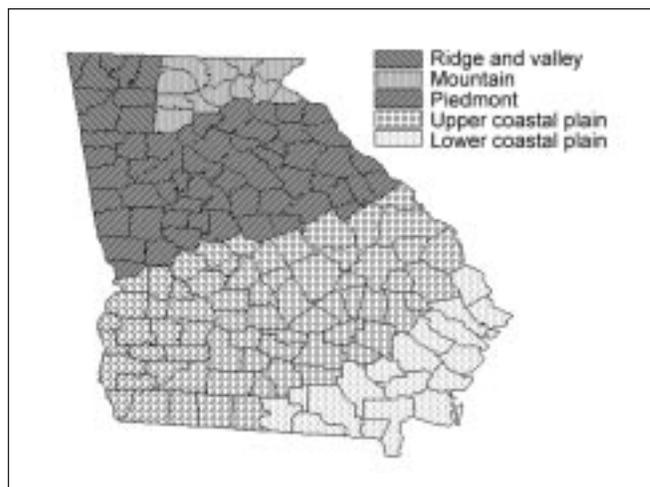
We analyzed plot data for the periodic 2001 inventories and considered every single FIA plot as the basic unit for our study. We processed the original data in this manner:

1. All forest type data, i.e. the data of all trees, was segregated into two broad forest categories, hardwoods and softwoods. All tree sizes inventoried by the FIA were included to account for any potential mortality of small trees.
2. The number of mortality trees per acre per year (TPAMORT) was used as the variable to calculate tree mortality on every plot and also as the response variable.
3. Latitude (LAT), longitude (LON), elevation (ELEV), condition proportion (CONDPROP), stand age (STDAGE), stand size code (STDSZCD), site productivity code (SITECLCD), slope (SLOP), aspect (ASPECT), physiographical code (PHYCLCD), growing-stock stocking code (GSSTKCD), stand treatment 1 code (TRTCD1), basal area of all live trees (BALIVE), current diameter (DIA), and trees per acre (TPACURR) are used as independent variables. STDSZCD, SITECLCD, PHYCLCD, GSSTKCD, and TRTCD1 are five categorical variables; the remaining 10 variables are numerical.
4. These data are aggregated into three groups: all trees, hardwood, and softwood.
5. In the sampling process, the State of Georgia is divided into five regions (fig 1): the northwest corner is a ridge and valley region; the northeast corner is a mountain region; north central Georgia is the piedmont region; south central Georgia is an upper coastal plain region; and the southeast corner is a lower coastal plain region. In each subregion, few differences exist in natural environment, landscape, or forest species.

Methodology

We used four types of methods in our study. Multiple linear regression and logistic regression were applied for variable selection and model fitting. Sampling methods were used to

Figure 1.—Georgia's five study subregions.



make samples from the original data. For geographic information systems (GIS), Environmental Systems Research Institute, Inc. (ESRI) ArcInfoSM and Arcview[®] products are used to process location data and related attributed data and calculate the coefficients of spatial autocorrelation.

Multiple Linear Regression and Logistic Regression

Equation 1 is a multiple linear regression function; equation 2 is a logistic regression function. The two functions, used for all trees, hardwood and softwood, have categorical variables. We applied a stepwise method to select variables in fitting logistic models:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_{15} X_{15} \quad (1)$$

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_{15} X_{15} \quad (2)$$

Where: Y is the numbers of tree mortality, p is the probability of tree mortality, and x_1, x_2, x_3, \dots and x_{15} are the variables of LAT, LON, ELEV, CONDPROP, STDAGE, STDSZCD, SITECLCD, SLOP, ASPECT, PHYCLCD, GSSTKCD, TRTCD1, BALIVE, DIA, and TPACURR.

Sampling Methods

We used simple random sampling (SRS), the simplest form of random sampling. It is easy to perform and explain to others, a fair way to select a sample, and reasonable to generalize the

results from the sample back to the population. Second, we employed systematic random sampling (SYS), also fairly easy to perform. Third, we used a stratified random sampling (STS) method. For STS, the population of all trees, hardwood and softwood, were divided into five groups based to the five sub-regions (fig. 1) noted above.

Geographic Information Systems

We used the following two types of coefficients of spatial autocorrelation this study. Equation 3 is Geary's coefficient C, and equation 4 is Moran's I coefficient.

$$C = \frac{\sum \sum W_{ij} C_{ij} / 2 \sum \sum W_{ij} \sigma^2}{\sum \sum W_{ij} C_{ij} / S^2 \sum \sum W_{ij}} \quad (3)$$

$$I = \frac{\sum \sum W_{ij} C_{ij} / S^2 \sum \sum W_{ij}}{\sum \sum W_{ij} \sigma^2} \quad (4)$$

where C_{ij} is the similarity of attributes, W_{ij} is the similarity of distance, and s^2 and δ^2 is the variance of attributes.

Results

Variable Selection Based on Original Data

First, a multiple linear regression function is applied for all trees to select significant variables based on original data. In equation 1, the tree mortality numbers is the response, and the other 15 variables are independent variables. The residual plot indicates the absence of constant variance among the residuals. Then, a straightforward log transformation for the data of all trees, hardwood and softwood, is used. Next, the multiple linear regression models are fitted again for these three groups of data. The residual plots are good, and the linear models are acceptable. At the level of $\alpha = 0.05$, some differing variables exist, but some are also the same (table 1). For all trees, hardwood and

softwood, TPAGROW and DIA are significant variables, which means that tree density may be a critical factor for tree mortality, and tree size may be also a better variable for mortality prediction. For all trees and hardwood, ELEV is the other common variable that affects tree mortality. In addition, at the 0.05 level, the physiographic class variable is significant for hardwood mortality, and the treatment class variable is significant for softwood mortality.

The logistic models provided some of the same significant variables and some different ones, too, compared with multiple linear models. For all trees, only one variable, DIA, was still significant; the other four variables—BALIVE, CONPROP, SITECLC, and STDSZCD—are added, which means that basal area, condition proportion, site productivity class, and stand size class were important for mortality of all trees. For hardwood, the four variables DIA, TPAGROW, PHYSCLCD, and ELEV were still significant compared with the above multiple linear regression analysis, and two other variables, BALIVE and STDAGE, were added. Basal area and stand age were also important for hardwood mortality based on logistic regression analysis. For softwood, only one variable, TPAGROW, is still significant compared with the multiple linear regression, and three other variables that are important for softwood mortality were added: stand age, site productivity class, and stand size class.

Sampling and Spatial Autocorrelation Calculation

Spatial autocorrelation is typically over looked in most tree mortality research. In our study, we calculated spatial autocorrelation for different kinds of sampling methods (table 2), and selected a better sampling method, SYS, for model fitting again.

Variable Reselection and Model Fitting

For multiple regression functions, variables are reselected after log transformation of tree mortality data. Table 3 lists the significant variables.

Table 1.—Multiple linear regression analysis based on original data.

Variable	All trees		Variable	Hardwood		Variable	Softwood	
	F-value	P-value		F-value	P-value		F-value	P-value
DIA	15.57	0.0001	DIA	30.74	0.0001	DIA	10.21	0.0015
TPAGROW	60.34	<.0001	TPAGROW	20.44	0.0001	TPAGROW	42.15	0.0001
ELEV	3.85	0.05	ELEV	9.97	0.0016	TRTCD	2.91	0.0338
			PHYSCLCD	1.87	0.0397			

Table 2.—Spatial autocorrelation comparison of different sampling methods.

	All trees		Hardwood		Softwood	
	Moran's I	Geary's C	Moran's I	Geary's C	Moran's I	Geary's C
No sampling	0.103	0.100	0.018	0.122	0.097	0.049
SRS	0.071	0.042	0.003	0.040	0.042	0.060
SYS	0.056	0.034	0.004	0.048	0.087	0.029
STRAT	0.138	0.059	0.006	0.036	-0.0009	0.039

Table 3.—Multiple regression analysis after SRS.

Variable	All trees		Variable	Hardwood		Variable	Softwood	
	Wald-Chi	P		Wald-Chi	P		Wald-Chi	P
DIA	5.65	.0179	DIA	16.43	<.0001	TPAGROW	11.46	.0009
TPAGROW	22.12	<.0001	TPAGROW	3.91	.0488			
			ELEV	5.57	.0189			

Table 4.—Logistic regression analysis after systematical random sampling

Variable	All trees		Variable	Hardwood		Variable	Softwood	
	Wald-Chi	P		Wald-Chi	P		Wald-Chi	P
BALIVE	9.02	0.0027	STDAGE	9.66	0.0019	STDSZCD	22.49	<0.0001
TPAGROW	6.42	0.0113	DIA	8.29	0.0040	STDAGE	7.87	0.0050
SITECLCD	12.35	0.0149	TPAGROW	5.32	0.0211	SITECLCD	11.36	0.0288
STDSZCD	12.74	0.0052						

Compared with outcomes of logistic regression without random sampling, the numbers of significant variables decreased, and some significant variables were no longer significant (table 4). For mortality of all trees, BALIVE, SITECLCD, and STDCLCD were still significant after SYS, and TPAGROW became a significant variable. For hardwood, STDAGE, DIA, and TPAGROW were still significant, but BALIVE, PHYSCLCD, and ELEV were no longer significant. For softwood, TPAGROW was no longer significant, but STDSZCD, STDAGE, and SITECLCD were still significant after SYS.

Conclusions

Based on our analyses, the logistic mortality functions for all trees, hardwood and softwood, are equations 5, 6, and 7. \hat{Y} is the probability of tree mortality; ba is basal area; g is TPAGROW;

s is the stand size class variable with from one to four classes; si is the site productivity class variable, which has from one to five classes; a is stand age; and d is DIA.

Tree mortality has several common characteristics. For all trees and softwood, STDSZCD classes 1 and 2 (large diameter and medium diameter classes) have the highest mortality probability; STDSZCD 4 (chaparral class) has a medium mortality probability; class 3 (small diameter class) has the lowest mortality probability. SITECLCD classes 2 (site productivity between 165 and about 224 cubic feet/acre/year) and 3 (site productivity between 124 and approximately 165 cubic feet/acre/year) have higher mortality probability than classes 4 (site productivity between 85 around 119 cubic feet/acre/year) and 5 (site productivity between 50 and about 84 cubic feet/acre/year). Class 1 (site productivity more than 225 cubic feet/acre/year) has a medium mortality probability. For all trees and hardwood, the probability of tree mortality slightly decreases as TPAGROW increases.

$$\hat{Y} = \frac{\ell \cdot \frac{-0.4436+0.0037ba-0.0039g+0.5336I_{s=1}+0.5704I_{s=2}-0.0122I_{s=3}+0.0235I_{si=2}+0.3099I_{si=3}-0.3029I_{si=4}-0.2806I_{si=5}}{-0.4436+0.0037ba-0.0039g+0.5336I_{s=1}+0.5704I_{s=2}-0.0122I_{s=3}+0.0235I_{si=2}+0.3099I_{si=3}-0.3029I_{si=4}-0.2806I_{si=5}}}{1 + \ell \cdot \frac{-0.4436+0.0037ba-0.0039g+0.5336I_{s=1}+0.5704I_{s=2}-0.0122I_{s=3}+0.0235I_{si=2}+0.3099I_{si=3}-0.3029I_{si=4}-0.2806I_{si=5}}{-0.4436+0.0037ba-0.0039g+0.5336I_{s=1}+0.5704I_{s=2}-0.0122I_{s=3}+0.0235I_{si=2}+0.3099I_{si=3}-0.3029I_{si=4}-0.2806I_{si=5}}}} \quad (5)$$

$$\hat{Y} = \frac{\ell \cdot \frac{-1.1575+0.0096a+0.0891d-0.0032g}{-1.1575+0.0096a+0.0891d-0.0032g}}{1 + \ell \cdot \frac{-1.1575+0.0096a+0.0891d-0.0032g}{-1.1575+0.0096a+0.0891d-0.0032g}} \quad (6)$$

$$\hat{Y} = \frac{\ell \cdot \frac{-0.7276-0.0164a+0.6130I_{s=1}+0.4721I_{s=2}-0.5783I_{s=3}+0.5153I_{si=2}+0.0379I_{si=3}+0.0344I_{si=4}-0.4810I_{si=5}}{-0.7276-0.0164a+0.6130I_{s=1}+0.4721I_{s=2}-0.5783I_{s=3}+0.5153I_{si=2}+0.0379I_{si=3}+0.0344I_{si=4}-0.4810I_{si=5}}}{1 + \ell \cdot \frac{-0.7276-0.0164a+0.6130I_{s=1}+0.4721I_{s=2}-0.5783I_{s=3}+0.5153I_{si=2}+0.0379I_{si=3}+0.0344I_{si=4}-0.4810I_{si=5}}{-0.7276-0.0164a+0.6130I_{s=1}+0.4721I_{s=2}-0.5783I_{s=3}+0.5153I_{si=2}+0.0379I_{si=3}+0.0344I_{si=4}-0.4810I_{si=5}}} \quad (7)$$

Much research indicates that logistic regressions appear to be the best method for individual tree mortality modeling and have been widely applied (Monserud, 1976; Monserud and Sterba 1999, Fridman and Stahl 2001, Woolons 1998, Yang *et al.* 2003). In our study, logistic models are selected, and tree mortality analysis is summarized based on these logistic models.

Literature Cited

Avila, O.B.; Burkhart, H.E. 1992. Modeling survival of loblolly pine trees in thinned and unthinned plantations. *Canada Journal of Forest Research*. 22: 1878–1882.

Brown, S.L.; Schroeder, P.E. 1999. Spatial patterns of above-ground production and mortality of woody biomass for Eastern U.S. forests. *Ecological Application*. 9(3): 968–980.

Fridman, J.; Stahl, G. 2001. A three-step approach for modeling tree mortality in Swedish forests. *Scandinavian Journal of Forest Research*. 16: 455–466.

Greene, S.E.; Harcombe, P.A.; Harmon, M.E.; Spycher, G. 1992. Patterns of growth, mortality and biomass change in a coastal picea-sitchensis-tsuga-heterophylla forest. *Journal of Vegetation Science*. 3(5): 697–706.

Guan, B.T.; Gertner, G. 1991. Using a parallel distributed-processing system to model individual tree mortality. *Forest Science*. 37(3): 871–885.

Hosmer, D.W.; Lemeshow, S. 1989. *Applied Logistic Regression*. New York: John Wiley & Son. 309 p.

Lee, Y.L. 1971. Predicting mortality for even-aged stands of lodge-pole pine. *The Forestry Chronicle*. 47: 29–32.

Monserud, R.A. 1976. Simulation of tree mortality. *Forest Science*. 22: 428–444.

Monserud, R.A.; Sterba, H. 1999. Modeling individual tree mortality for Austrian forest species. *Forest Ecology and Management*. 113: 109–123.

Neter, J.; Maynes, E.S. 1970. On the appropriateness of the correlation coefficient with a 0, 1 dependent variable. *Journal of the American Statistical Association*. 65: 501–509.

Osawa, A.; Spies, C.J.; Dimond, J.B. 1986. Patterns of tree mortality during an uncontrolled spruce budworm outbreak in Baxter State Park, 1983. *Maine Agricultural Experiment Station Technical Bulletin*. 121: 1–69.

Pedersen, B.S.; McCune, B. 2002. A non-invasive method for reconstructing the relative mortality rates of trees in mixed-age, mixed-species forests. *Forest Ecology and Management*. 155(1–3): 303–314.

Rouvinen, S.; Kuuluvainen, T.; Siitonen, J. 2002. Tree mortality in a *Pinus sylvestris* dominated boreal forest landscape in Vienansalo wilderness, eastern Fennoscandia. *Silva Fennica*. 36(1): 127–145.

Walker, R.H.; Duncan, D.B. 1967. Estimation of the probability of an event as a function of several independent variables. *Biometrika*. 54: 167–179.

Woollons, R.C. 1998. Even-aged stand mortality estimation through a two-step regression process. *Forest Ecology and Management*. 105: 189–195.

Yang, Y.; Titus, S.J.; Huang, S. 2003. Modeling individual tree mortality for white spruce in Alberta. *Ecological Modeling*. 163: 209–222.

Zhang, S.; Amateis, R.L.; Burket, H.E. 1997. Constraining individual tree diameter increment and survival models for loblolly pine plantations. *Forest Science*. 43(3): 414–423.

Optimal Tree Increment Models for the Northeastern United States

Don C. Bragg¹

Abstract.—I used the potential relative increment (PRI) methodology to develop optimal tree diameter growth models for the Northeastern United States. Thirty species from the Eastwide Forest Inventory Database yielded 69,676 individuals, which were then reduced to fast-growing subsets for PRI analysis. For instance, only 14 individuals from the greater than 6,300-tree eastern white pine sample were used to fit its PRI model. The Northeastern northern red oak model predicted faster small tree growth than those derived for the Lake States or Midsouth, but it soon fell behind the other regional models and never again matched their performance. Predicted maximum increment differences between regions rarely exceeded 0.25 cm, however. The PRI methodology also can help identify possibly erroneous individual tree records.

Introduction

For tree growth modelers, increment “optimality” is often defined as an idealized or maximal rate of increase in a specified dimension (usually diameter or height). This concept assumes that all environmental conditions are at their most favorable, and, thus, anything suboptimal decreases growth accordingly. The primary advantage to an optimized approach is that the modifiers influencing increment can be separated from the model used to predict growth, allowing for many different constructs to be applied (Bragg 2003a). Not surprisingly, potential increment models have become the cornerstone of many ecological simulators (e.g., Botkin *et al.* 1972).

Potential growth formulations have their critics. Purely theoretical designs, while often intellectually appealing, are problematic because they rarely incorporate real-world meas-

urements and sometimes contain biological flaws. For example, the gap model’s potential increment design includes a number of unsupportable assumptions about diameter accumulation and maximum tree dimensions (Bragg 2001). Lessard *et al.* (2001) dismissed potential growth constructs because they cannot be directly observed and may be difficult to estimate. Finally, some have argued that empirical models predicting average (realized) growth are more precise, even if they lack mechanism (Fleming 1996).

Biologically meaningful optimal growth curves can be empirically derived, however. The potential relative increment (PRI) methodology (Bragg 2001) uses the Eastwide Forest Inventory Database (EFIDB) (Hansen *et al.* 1992) to estimate optimal growth based on actual inventories. A set of simple post-processors (Bragg 2002a), when properly applied to data on rapidly growing individuals fit to a nonlinear model, produce response patterns identified as crucial by Shvets and Zeide (1996) and Zeide (1993). PRI models have been developed for the Lake States (Michigan, Minnesota, and Wisconsin) and Midsouth (Arkansas, Louisiana, Missouri, Oklahoma, and Texas) (Bragg 2001, Bragg 2002b, Bragg 2003b). This article presents PRI models for the common tree species of the Northeastern States of Connecticut, Maine, Massachusetts, New Hampshire, New York, Rhode Island, and Vermont.

Methods

A detailed description of the PRI methodology is beyond the scope of this article (rather, see Bragg [2001] and Bragg [2002a]). The PRI approach is a type of boundary line analysis (Webb 1972). Boundary line analysis has shown promise for identifying the role of maximal growth in ecological and mensurational applications (for example, Black and Abrams 2003). Briefly, PRI calculates actual relative increment (ARI) from:

$$ARI = \frac{d.b.h._c - d.b.h._o}{d.b.h._o} \quad (1)$$

¹ Research Forester, Southern Research Station, U.S. Department of Agriculture, Forest Service, P.O. Box 3516 UAM, Monticello, AR 71656. Phone: 870-367-3464; e-mail: dbragg@fs.fed.us.

where the initial (d.b.h.₀) and final (d.b.h._c) inventory diameters are in centimeters. ARI values were then annualized by dividing by the remeasurement period. A record was considered eligible if the tree was alive during both inventories, was of the species of interest, and showed a positive increment (d.b.h._c > d.b.h.₀).

The PRI methodology does not consider every ARI value. Because virtually all trees are negatively affected by local environmental conditions, e.g., competition or poor site quality, their diameter growth decreases markedly. Easily identified by their slower growth rates, these individuals were eliminated from further consideration, leaving only a handful of the fastest growing trees in a specified diameter class. Maximally performing individuals that fail to reach the levels of adjacent diameter classes are also removed from further consideration. The final subset represents only a fraction (usually 6 to 12 trees) of the original data, for which the following model was fit:

$$E(PRI) = \hat{b}_1 d.b.h. \frac{\hat{b}_2}{\hat{b}_3^{d.b.h. - \hat{b}_1}} \quad (2)$$

where d.b.h._{MAX} is the d.b.h. of an individual tree growing at the highest rate in its respective diameter class, and \hat{b}_1 , \hat{b}_2 , and \hat{b}_3 are nonlinear ordinary least squares regression parameter estimates.

As an example, ARI values were calculated for 6,348 eastern white pines (*Pinus strobus*) from the Northeastern EFIDB (fig. 1a). Selecting only the pines (by 2-cm d.b.h. classes) with maximal ARI reduced this number to 51 individuals (fig. 1b). Because most of this subset of d.b.h. class maximal ARI points fell appreciably below the “optimal” frontier, they were removed before the final curve fitting. Hence, a PRI model for eastern white pine in the Northeastern United States was generated with only 14 trees (fig. 1c). Eastern white pine displayed a characteristic curve (fig. 1d), with the highest predicted PRI in the smallest pines. Multiplying the result of equation (2) by the tree’s current diameter yielded an increment curve (fig. 2), with the greatest optimal annual growth of approximately 2 cm occurring at 20- to 40-cm d.b.h.

Figure 1.—Step-by-step PRI methodology for eastern white pine taken from the Northeastern United States EFIDB. After the original 6,348 eligible pines were identified (a), the 51 individuals growing at the highest rate per 2-cm diameter class (b) were retained and further reduced to the final subset (c) of 14 data points, to which the actual PRI equation was fit (d).

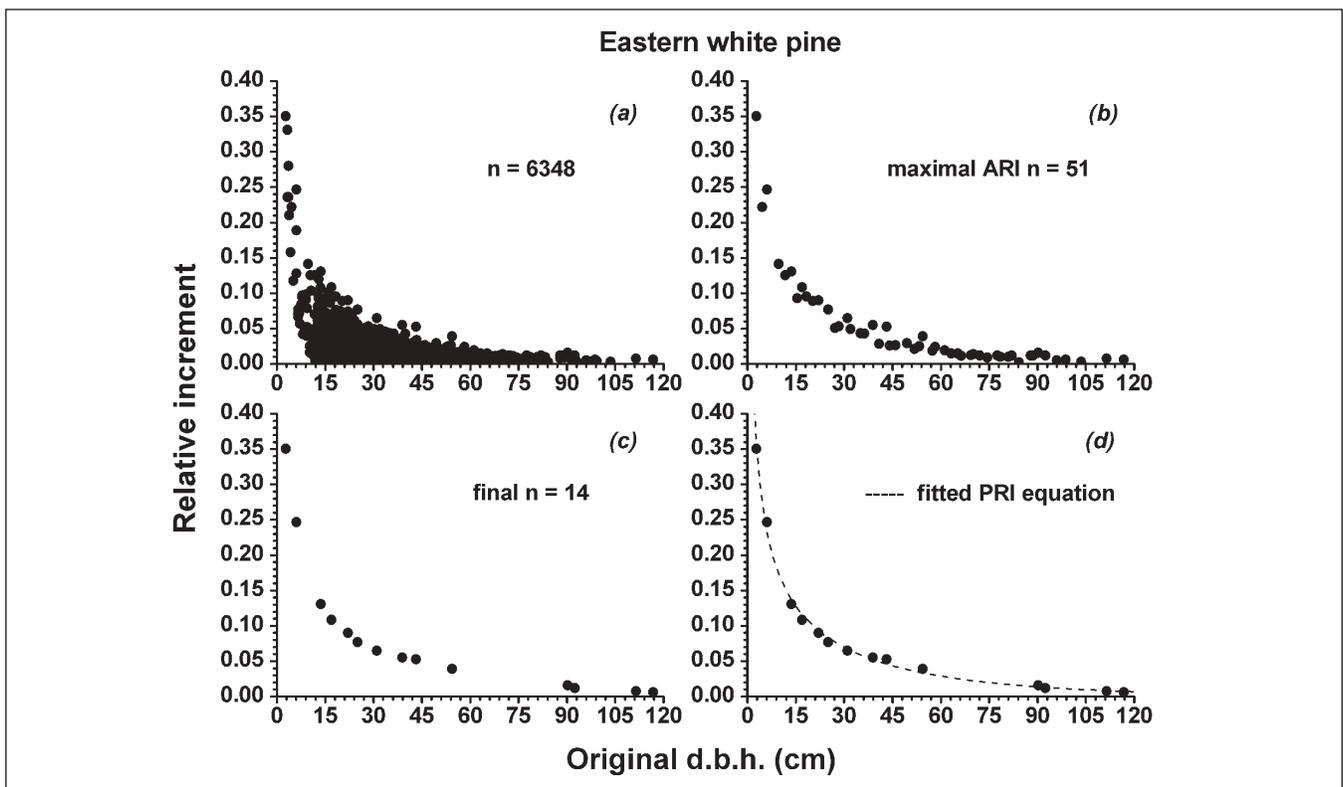
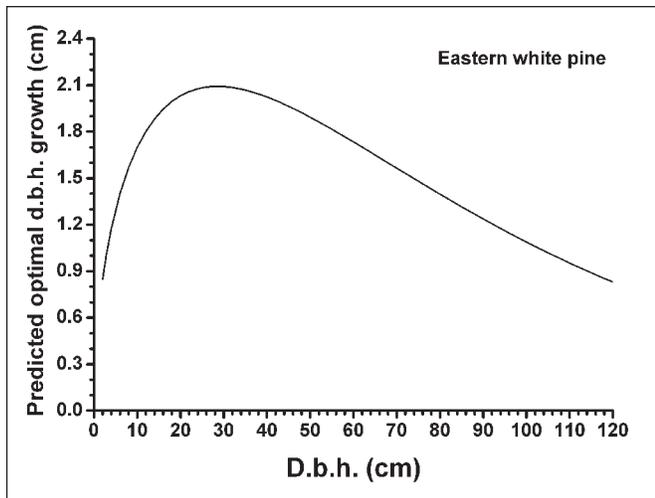


Figure 2.—Predicted optimal d.b.h. annual increment for eastern white pine in the Northeastern United States.



Similar steps were used to produce PRI models for the most common species of the Northeastern EFIDB. Additionally, I compared northern red oak PRI curves for three areas (Northeast, Lake States, and Midsouth) to highlight regional differences in predictions of optimal diameter increment. Finally, two examples of extremely fast growing individuals were used to demonstrate the potential of PRI to identify inventory outliers.

Results and Discussion

Northeastern PRI Results

I found 30 species sufficiently abundant ($n \geq 100$) in the Northeastern EFIDB for the PRI methodology (table 1). Most (23 of 30) species produced at least 450 individuals, and 1 in 6 had more than 5,000 trees. Combined, these taxa yielded 69,676 individuals for preliminary analysis. A sample size of this magnitude, even if most are rejected for growing too slowly, is far more comprehensive than typical growth modeling efforts.

Although individuals greater than 10 cm and less than 50 cm in d.b.h. (averaging 20- to 30-cm d.b.h.) predominated, this sample contained very small and very large trees (table 1). For example, 10 species had individuals greater than 100-cm d.b.h., including a 165.9-cm d.b.h. northern red oak (*Quercus rubra*) and a 185.4-cm d.b.h. black willow (*Salix nigra*).

Depending on the species, only 6 to 15 individuals were needed to develop the PRI curves. All parameter estimates were significant at a significance level of $\alpha = 0.05$ (table 2).

Regional Comparison Using Northern Red Oak

Very few obvious differences arose between the regional PRI models (fig. 3a). After converting the PRI values to potential increments, relative growth performance primarily differed by absolute tree diameter. Up to about 10-cm d.b.h., the Northeastern northern red oak model predicted the highest optimal increment. It was then replaced by the Lake States model (to 76-cm d.b.h.), after which the Midsouth version produced the highest predicted optimal northern red oak increment (fig. 3b). Estimated optimal increments peaked at approximately 1.25 cm for the Northeastern and Midsouth models, and at just over 1.4 cm for the Lake States model. These maxima were reached at about 15-cm d.b.h. for the Northeastern model, roughly 25-cm d.b.h. for the Midsouth model, and approximately 30-cm d.b.h. for the Lake States model.

Overall, potential diameter increment differences among the regions were minor, with residual differences rarely exceeding 0.2 cm annually at any given diameter (fig. 3c). This difference is not trivial when accumulated over years of growth, however, especially because PRI-based growth projection systems are nonlinear functions of current tree diameter.

Identifying Potential Inventory Errors With PRI

As a conservative estimate of optimal growth, PRI curves can identify individuals growing dramatically faster than expected. For instance, two individuals from the New York data set were obvious outliers when maximal ARI points were plotted. An 80.3-cm d.b.h. black cherry grew to 107.4-cm d.b.h. in just 12 years (fig. 4a), while an 80.5-cm d.b.h. white oak increased to 106.7-cm, also in 12 years (fig. 4b). Although this growth is possible for vigorous young individuals of either species, this level of productivity was highly suspect in trees of 80-cm d.b.h.

These extremely fast-growing outliers came from plots of low stand density (basal areas of 6.9 m²/ha for the black cherry and 3.7 m²/ha for the white oak), and thus could reflect the pronounced release of previously suppressed individuals. More likely, they probably reflect measurement or transcription errors. Given their large girth, these outliers could prove highly influential in any extrapolations based on their size.

Table 1.—Species, preliminary counts, and diameter at breast height (d.b.h.) ranges of species used in the Northeastern United States PRI analysis.

Species ^a	EFIDB code ^a	Initial number	Minimum d.b.h. (cm)	Mean d.b.h. (cm)	Maximum d.b.h. (cm)
Balsam fir (<i>Abies balsamea</i>)	12	3,387	3.3	17.9	45.7
Tamarack (<i>Larix laricina</i>)	71	357	11.7	20.5	52.1
White spruce (<i>Picea glauca</i>)	94	763	4.6	21.9	55.9
Black spruce (<i>Picea mariana</i>)	95	499	12.7	18.6	39.1
Red spruce (<i>Picea rubens</i>)	97	5,384	7.1	22.2	67.1
Red pine (<i>Pinus resinosa</i>)	125	501	3.8	24.7	72.4
Pitch pine (<i>Pinus rigida</i>)	126	165	9.9	24.7	53.6
Eastern white pine (<i>Pinus strobus</i>)	129	6,348	2.8	29.0	116.8
Northern white-cedar (<i>Thuja occidentalis</i>)	241	4,222	6.6	23.1	73.4
Eastern hemlock (<i>Tsuga canadensis</i>)	261	6,424	5.8	25.5	105.4
Red maple (<i>Acer rubrum</i>)	316	11,283	3.0	23.6	114.0
Silver maple (<i>Acer saccharinum</i>)	317	157	4.6	29.7	74.2
Sugar maple (<i>Acer saccharum</i>)	318	7,540	3.0	27.3	125.2
Yellow birch (<i>Betula alleghaniensis</i>)	371	2,905	4.8	25.9	109.5
Sweet birch (<i>Betula lenta</i>)	372	810	3.6	23.4	75.4
Paper birch (<i>Betula papyrifera</i>)	375	2,538	9.4	20.9	59.7
American beech (<i>Fagus grandifolia</i>)	531	3,430	4.6	24.2	76.5
White ash (<i>Fraxinus americana</i>)	541	2,402	3.6	24.1	115.6
Black ash (<i>Fraxinus nigra</i>)	543	257	8.9	19.0	43.9
Bigtooth aspen (<i>Populus grandidentata</i>)	743	757	9.1	24.3	59.9
Quaking aspen (<i>Populus tremuloides</i>)	746	1,572	3.0	21.3	97.3
Black cherry (<i>Prunus serotina</i>)	762	1,449	4.1	27.1	85.9
White oak (<i>Quercus alba</i>)	802	744	9.9	28.8	93.0
Scarlet oak (<i>Quercus coccinea</i>)	806	351	6.9	25.1	65.0
Chestnut oak (<i>Quercus prinus</i>)	832	484	8.1	25.5	83.6
Northern red oak (<i>Quercus rubra</i>)	833	3,384	5.3	28.6	165.9
Black oak (<i>Quercus velutina</i>)	837	651	3.8	28.7	100.1
Black willow (<i>Salix nigra</i>)	922	104	6.1	46.9	185.4
American basswood (<i>Tilia americana</i>)	951	568	4.1	29.3	119.1
American elm (<i>Ulmus americana</i>)	972	240	3.6	18.6	46.2
TOTAL = 69,676					

^a Species nomenclature consistent with the EFIDB as reported by Hansen *et al.* (1992).

Table 2.—Parameter estimates by species for Northeastern United States PRI models.

Species	Final number	Estimated parameters		
		\hat{b}_1	\hat{b}_2	\hat{b}_3
Balsam fir	14	0.42	-0.21	0.937051
Tamarack	11	1.63	-1.13	0.991027
White spruce	15	1.69	-1.11	0.989269
Black spruce	7	0.87	-0.59	0.930458
Red spruce	7	1.06	-0.56	0.959002
Red pine	6	0.45	-0.31	0.955000
Pitch pine	7	0.67	-0.56	0.942390
Eastern white pine	14	0.61	-0.47	0.981844
Northern white-cedar	7	0.82	-0.36	0.945433
Eastern hemlock	11	1.02	-0.83	0.984948
Red maple	15	0.96	-0.66	0.986284
Silver maple	7	0.51	-0.27	0.968776
Sugar maple	10	0.40	-0.46	0.988446
Yellow birch	13	0.66	-0.67	0.976746
Sweet birch	6	0.46	-0.29	0.946431
Paper birch	8	0.48	-0.36	0.945841
American beech	14	0.33	-0.51	0.977587
White ash	12	0.98	-0.86	0.996465
Black ash	7	0.58	-0.58	0.949105
Bigtooth aspen	7	1.97	-1.06	0.995803
Quaking aspen	12	1.14	-0.64	0.955174
Black cherry	14	1.04	-0.80	0.988091
White oak	8	0.41	-0.23	0.957937
Scarlet oak	7	3.05	-1.42	0.999900
Chestnut oak	11	0.12	-0.38	0.979383
Northern red oak	12	0.88	-0.80	0.988439
Black oak	12	0.52	-0.75	0.993657
Black willow	7	0.95	-0.66	0.989004
American basswood	9	0.58	-0.74	0.985406
American elm	9	0.78	-0.30	0.935231

Figure 3.—PRI comparison for northern red oak between the Northeastern (NE), Midsouth (MS), and Lake States (LS) regions. PRI curves differed slightly for all three regions (a), which translated into noticeable increment differences (b and c).

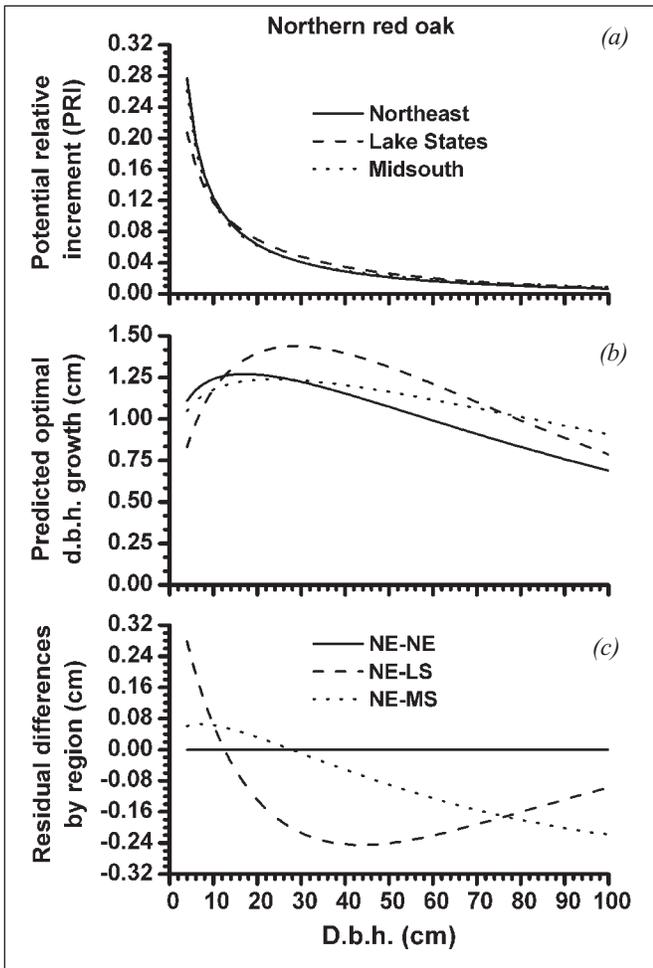


Figure 4.—Very prominent outliers (large open symbols) identified by the PRI methodology. Both the 80-cm d.b.h. black cherry (a) and white oak trees (b) grew at a very high rate, given their large size, identifying them as individuals of concern.

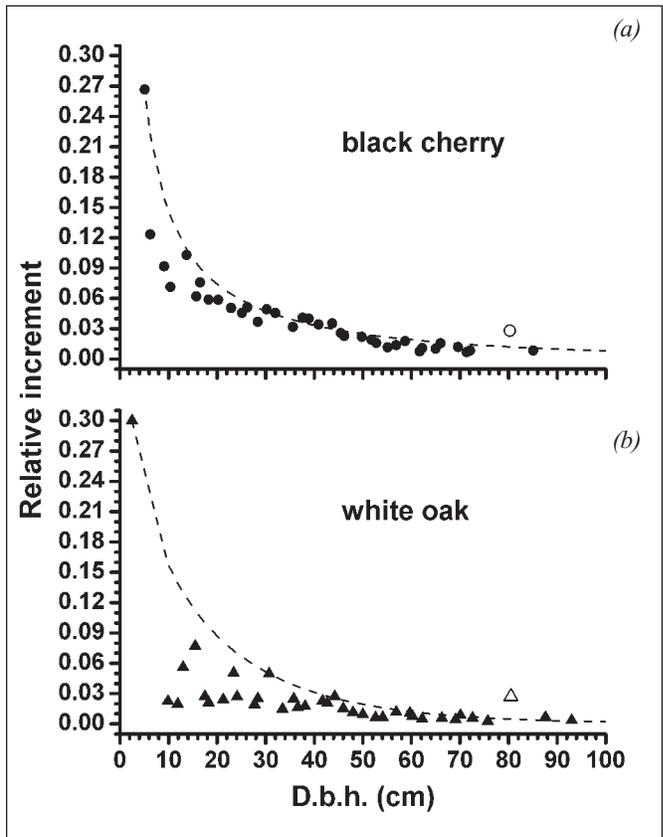


Table 3.—Annualized parameters for northern red oak models developed from the Northeastern United States (this article), the Lake States (Bragg 2001), and the Midsouth (Bragg 2003b).

Parameter	Northeast	Lake States	Midsouth
\hat{b}_1	0.880309	2.241167	0.843941
\hat{b}_2	-0.800826	-0.506656	-0.823358
\hat{b}_3	0.988439	0.983046	0.992617

Conclusions

The EFIDB for the Northeastern United States contained enough data to construct PRI growth models for 30 tree species. A comparison of northern red oak models among several regions, including the Northeast, produced noticeable differences in the magnitude and timing of the predicted maximal increment (table 3). Because diameter growth is cumulative, even subtle differences over time would lead to substantial variation in tree size, assuming all other environmental conditions are held constant. Key to any effort, however, is ensuring that the inventory records accurately reflect tree dimensions before they are incorporated into any type of predictive environment.

Acknowledgments

I thankfully recognize the following people for their helpful review comments: Paul Doruska (University of Arkansas at Monticello), Eric Heitzman (University of Arkansas at Monticello), Ron McRoberts (U.S. Department of Agriculture, Forest Service, North Central Research Station), and Paul Smith (U.S. Department of Agriculture, Forest Service, Southern Research Station).

Literature Cited

Black, B.A.; Abrams, M.D. 2003. Use of boundary-line growth patterns as a basis for dendroecological release criteria. *Ecological Applications*. 13: 1733–1749.

Botkin, D.B.; Janak, J.F.; Wallis, J.W. 1972. Some ecological consequences of a computer model of forest growth. *Journal of Ecology*. 60: 849–873.

Bragg, D.C. 2001. Potential relative increment (PRI): a new method to empirically derive optimal tree diameter growth. *Ecological Modelling*. 137: 77–92.

Bragg, D.C. 2002a. A system to derive optimal tree diameter increment models from the Eastwide Forest Inventory Data Base (EFIDB). *Southern Journal of Applied Forestry*. 26(4): 214–221.

Bragg, D.C. 2002b. Empirically derived optimal growth equations for hardwoods and softwoods in Arkansas. In: Outcalt, K.W., ed. *Proceedings, the 11th biennial southern silvicultural research conference*. Gen. Tech. Rep. SRS-48. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 421–424.

Bragg, D.C. 2003a. Efficacy of using inventory data to develop optimal diameter increment models. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; Moser, J.W., eds. *Proceedings, 3rd annual forest inventory and analysis symposium*. Gen. Tech. Rep. NC-230. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 121–129.

Bragg, D.C. 2003b. Optimal diameter growth equations for major tree species of the Midsouth. *Southern Journal of Applied Forestry*. 27(1): 5–10.

Fleming, R.A. 1996. Better derivations and an improvement for the specific increment equations of tree growth. *Canadian Journal of Forest Research*. 26: 624–626.

Hansen, M.H.; Frieswyk, T.; Glover, J.F.; Kelly, J.F. 1992. The Eastwide Forest Inventory Data Base: users manual. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 48 p.

Lessard, V.C.; McRoberts, R.E.; Holdaway, M.R. 2001. Diameter growth models using Minnesota Forest Inventory and Analysis data. *Forest Science*. 47(3): 301–310.

Shvets, V.; Zeide, B. 1996. Investigating parameters of growth equations. *Canadian Journal of Forest Research*. 26: 1980–1990.

Webb, R.A. 1972. Use of the Boundary Line in the analysis of biological data. *Journal of Horticultural Science*. 47: 309–319.

Zeide, B. 1993. Analysis of growth equations. *Forest Science*. 39: 594–616.

Northeastern FIA Tree Taper Study: Current Status and Future Work

James A. Westfall and Charles T. Scott¹

Abstract.—The northeastern unit of the Forest Inventory and Analysis program (NE-FIA) is engaged in an ongoing project to develop regionwide tree taper equations. Sampling intensity is based on NE-FIA plot data and is stratified by species, diameter class, and height class. To date, modeling research has been aimed largely at evaluating existing model forms (and hybrids thereof) and incorporating mixed-effects parameters to account for correlations among measurements. In conjunction with the taper study, bark thickness estimates are being developed from wood utilization studies. When fully implemented, the bark thickness/taper equation system will provide a wide range of analytical flexibility for tree species in northeastern forests, and may reduce or eliminate the costs of collecting data on merchantable heights.

Introduction

To compute merchantable volume of standing trees, the northeastern unit of the U.S. Department of Agriculture (USDA) Forest Service Forest Inventory and Analysis program (NE-FIA) has traditionally taken field measurements of tree height at certain diameter limits. Often, these measurements lack repeatability due to difficulties in observing the bole in upper portions of the tree and determining the point at which the diameter limit occurs. To improve data quality and increase fieldwork efficiency, NE-FIA is developing regionwide taper models. A parallel effort is underway for estimating the bark thickness of northeastern tree species. Prediction of bark thickness will increase analytical flexibility by enabling computation of inside-bark diameters. These models will be applicable across the 13 States in which NE-FIA collects resource inventory data.

Tree Taper Sample Development

The sampling list was developed from inventory data obtained on NE-FIA sample plots to represent the range of geography and tree size. The primary goal was to sample 150 trees in each of 18 species groups for a total sample of 2,700 trees. The species groups arise from species assignments used by NE-FIA for tree volume estimates.

Information on frequency of occurrence was tabulated and stratified by species group, tree species, diameter class, and height class (table 1). This stratification indicated that the sampling intensity for a specific species/diameter class/height class combination (S/D/H) could not be based on frequency alone due to the dominance of certain S/D/H combinations within some species groups. To spread the sample more evenly among species and tree sizes, a limit of six sample trees was imposed for any S/D/H arrangement. Conversely, it would be undesirable to devote the necessary resources to sample relatively rare S/D/H combinations. Thus, to be included in the sample, S/D/H combinations must have at least five observed trees across all NE-FIA sample plots. S/D/H combinations with at least five observations but comprising less than 0.1 percent of a species group are limited to a sample size of one.

Table 1.—*Tree diameter (inches) and height (feet) classes used to stratify tree sample.*

Tree d.b.h.	d.b.h. class	Tree height	Height class
3.0"–4.9"	1	0.0'–29.9'	1
5.0"–8.9"	2	30.0'–49.9'	2
9.0"–12.9"	3	50.0'–69.9'	3
13.0"–16.9"	4	70.0'–89.9'	4
17.0"–23.9"	5	90.0' +	5
24.0" +	6		

¹ Research Forester and Program Manager, respectively, U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, Northeastern Research Station, 11 Campus Blvd., Suite 200, Newtown Square, PA 19073. Phone: 610-557-4043; fax: 610-557-4250; e-mail: jameswestfall@fs.fed.us.

An example of the development of the sample list for red pine is given in table 2. The Sample₀ column indicates the S/D/H combinations found in the data and the original sample size based on frequency information alone. The numbers in this column are determined from the values in the percent of Group column, which includes frequency percentages for each S/D/H combination. Using the first row as an example, this S/D/H combination comprised 4.82 percent of the trees in the species group (i.e., 4.82 of every 100 trees in the group were in this category). However, we wanted to determine the number of trees in this category that should be sampled among the 150 sample trees for the entire species group. This is indicated in the percent of 150 column, which shows that the sample size should be 7.23 percent of the 150 sample trees. The result is 10.8 (11 in the Sample₀ column). The remaining values in Sample₀ column also were computed in this manner.

The limit of six sample trees per S/D/H combination is imposed in Sample₁ column (note that the total number of trees is reduced from 41 to 30). The Sample₂ column shows the reallocation of trees into cells with fewer than six sample trees and comprising more than 0.1 percent of the species group (i.e., an increase of two sample trees for each eligible S/D/H combination). The remaining sample tree to be accounted for is placed into the most common S/D/H category with fewer than six sample trees (column Sample₃), and the overall total is reconciled at 41 sample trees. This approach to reallocation of sample trees maintains the original sample size while permitting sampling of most

S/D/H combinations found on NE-FIA inventory plots. Sample lists for each State were created from the overall sample list.

Data Collection

During the 2002-03 leaf-off season, tree taper data were collected in Ohio, Maryland, Pennsylvania, and West Virginia. Tree form was measured with a Barr & Stroud dendrometer. Paired height/diameter data were obtained at 1, 2, 3, 4.5, 6 feet, and at taper intervals of about 1 inch thereafter. A measurement also was taken at the base of the live crown. Additional data were collected for each sample tree (d.b.h., crown ratio, crown class, etc.) and plot-level characteristics (slope, aspect, etc.) were noted. In all, 267 sample trees were measured; yellow poplar was the most common species.

Efforts to collect taper data have been greatly expanded during the 2003-04 leaf-off season, with collection occurring in all 13 States under NE-FIA auspices. Cooperators in this effort include Ohio State University, State University of New York College of Environmental Science and Forestry, and Maine Forest Service. Additional data were obtained from studies by the USDA Forest Service Eastern Region (Region 9).

Data on bark thickness were obtained primarily from wood-utilization studies conducted by NE-FIA; Region 9 also contributed information. All of these data are from studies of felled trees. Protocols have differed over time and between studies, but there are measures of bark thickness for most trees from 1-foot stump height to a 4-inch top diameter limit.

Table 2.—Development of sample list for red pine from frequency information and abundance limitations.

D class	H class	Percent of group	Percent of 150	Sample ₀	Sample ₁	Sample ₂	Sample ₃
3	4	4.82	7.23	11	6	6	6
3	3	4.35	6.52	10	6	6	6
2	2	3.31	4.97	8	6	6	6
2	3	1.21	1.81	3	3	5	6
4	4	0.72	1.08	2	2	4	4
4	3	0.61	0.92	2	2	4	4
1	1	0.19	0.29	1	1	3	3
5	4	0.13	0.19	1	1	3	3
1	2	0.07	0.11	1	1	1	1
3	2	0.06	0.09	1	1	1	1
2	1	0.04	0.06	1	1	1	1
Count =				41	30	40	41

Taper Modeling

Data on yellow poplar collected during 2002-03 were used to initiate the taper modeling process. Initial analyses consisted of comparing existing taper equations. Max and Burkhart (1976) presented a segmented polynomial model with estimated join points. Each segment was specified as representing the neiloid, parabolic, or conic sections of a tree. This is consistent with the approach taken by many other researchers, i.e., the lower portion of the bole is similar to a neiloid; the middle section is parabolic in shape; and the top section generally is conic:

$$d^2/DBH^2 = \beta_1(h/H - 1) + \beta_2(h^2/H^2 - 1) + \beta_3(\alpha_1 - h/H)^2 I_1 + \beta_4(\alpha_2 - h/H)^2 I_2 + \varepsilon \quad (1)$$

where: d = diameter outside bark (in)
 DBH = diameter at breast height (in)
 h = height (ft) at diameter d
 H = total tree height
 I_1 = indicator (= 1 if $\alpha_1 \geq h/H$; = 0 if $\alpha_1 < h/H$)
 I_2 = indicator (= 1 if $\alpha_2 \geq h/H$; = 0 if $\alpha_2 < h/H$)
 α_1, α_2 = segment join points (estimated from data)
 β_{1-4} = parameters to be estimated from data
 ε = random deviation

Kozak (1988) eliminated the necessity for specifying different functions for various parts of the stem by developing a variable-exponent taper equation. This approach allows the exponent to change with relative tree height, which allows a single function to describe neiloid, paraboloid, and conic forms:

$$d = \beta_5 DBH^{\beta_6} \times \beta_7^{DBH} \times X^{\beta_8 Z^2 + \beta_9 \ln(Z + 0.001) + \beta_{10} \sqrt{Z} + \beta_{11} e^Z + \beta_{12} (DBH/H)} + \varepsilon \quad (2)$$

where: $X = (1 - \sqrt{h/H}) \times (1 - \sqrt{p})$
 p = percentage of total height where change from neiloid to paraboloid occurs
 $Z = h/H$
 \ln = natural logarithm
 e = base of natural logarithm
 β_{5-12} = parameters to be estimated from data
 other variables as previously defined

Valentine and Gregoire (2001) described a taper model in which numerical switching functions are used to smooth the transition between the neiloid, parabolic, and conic forms. The model is similar to that Max and Burkhart (1976) in that the three classic shape descriptors provide the basis for the model. Rather than being estimated from the data, their join points were fixed at 4.5 feet and height to live crown. To account for repeated measures on individual trees, 2 random-effects parameters were specified in one of the switching functions:

$$A_d = A_{DBH} \left(\frac{H-h}{H-4.5} \right)^{\alpha_1 + S_1} \times \left(\frac{H-h}{H-C} \right)^{\alpha_2 + S_2} + \varepsilon \quad (3)$$

where: A_d = cross-sectional area (ft²) at diameter d
 A_{DBH} = cross-sectional area (ft²) at diameter at breast height
 C = height to base of live crown (ft)
 α_1 = estimated shape parameter of the middle segment
 α_2 = estimated shape parameter of the top segment
 S_1 = numerical switch exhibiting switch-off behavior
 S_2 = numerical switch exhibiting switch-on behavior
 other variables as previously defined

When the Max and Burkhart (hereafter MB) and Kozak models were developed, there was no practical means by which correlations among measurements on individual trees could be accounted for. However, advances in statistical theory and computing capabilities now allow researchers to account for this lack of independence when fitting equations. With respect to correlated observations, one approach is specifying a mixed-effects model, as was done by Valentine and Gregoire (hereafter VG). To make valid comparisons among models, the MB and Kozak equations were modified by incorporating random-effects parameters. Random components were added to the estimated join points in the MB equation, allowing the join points to vary among trees. The Kozak model also was altered to incorporate random effects into parameters associated with tree size.

The ability of these three models to describe bole shape was evaluated by fitting each of the equations to taper data from 34 yellow poplar trees. For comparisons among models, each was modified to produce diameter outside bark squared (d^2) as the dependent variable. The models were fitted using the SAS NL MIXED (Version 8.01) procedure. The efficacy of

each model was measured by Akaike's Information Criteria (AIC) (see Gregoire *et al.* 1995). A smaller AIC value indicates a better model fit:

Model	AIC
Max and Burkhardt	4,153.0
Valentine and Gregoire	4,254.5
Kozak	4,324.3

The results indicate that the MB model outperforms the other models in predicting tree taper for yellow poplar. On the basis of this limited analysis, two observations can be made. First, both the MB and VG models have smaller AIC values than the Kozak model. This implies that specifying the neiloid, paraboloid, and conic terms in the model provides better predictions of tree taper than the variable-exponent approach. The primary difference between the MB and VG models is that the former utilizes estimated join points while the latter has fixed values. The smaller AIC for the MB specification suggests that estimating join points provides a better fit to the data.

To test this assumption, the VG model was generalized to have estimated join points. This was accomplished by recasting the model to use relative rather than actual tree height and replacing fixed join points at 4.5 feet and crown height with parameters. The fitted regression produced an AIC of 4,041.2, which was a notable improvement over 4,254.5 obtained from the original model. This specification also surpassed the MB model in minimizing AIC. The primary gain in predictive accu-

racy is found in the lower section of the bole (fig. 1). This is particularly important if the taper equation is used to derive tree volume because a relatively large percentage of the volume occurs in this area. Both the estimated and fixed join point models performed similarly above 0.10 relative tree height.

An additional investigation was undertaken to determine whether moving the random-effects parameters to another location in the VG model would improve the fit. The original specification by VG placed the random effects in the S_1 switch. The improvement in fit statistics for the VG model obtained by estimating join points led to the supposition that moving the random effects into the estimated join points could result in further improvements in AIC. Fitting of this specification produced an AIC statistic of 3,962.2, a reduction of 2.0 percent from the previous formulation and 6.9 percent from the original model.

The work thus far provides evidence that a segmented model with estimated join points provides the best description of the shape of the bole. Also, it appears that specification of random effects in the join points produces better fit to the data than other formulations, and it is thought that the use of switching functions improves model fit, though additional evaluation is needed. These findings are based on limited analyses of a single tree species. Research on the applicability of these results to other tree species and species groups is warranted.

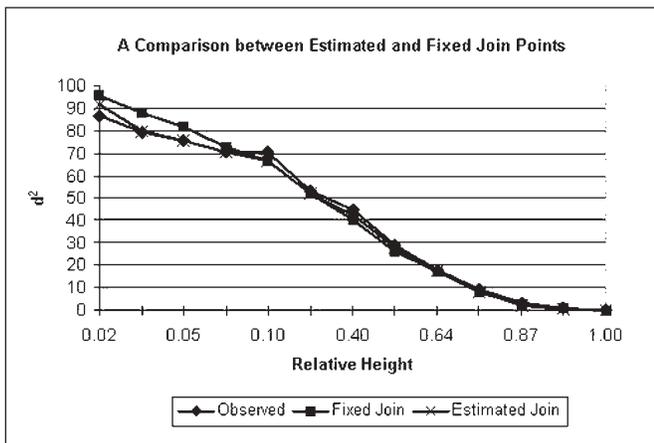
Bark Thickness Estimation

Estimates of bark thickness are needed to obtain diameter inside bark (dib) and diameter outside bark (dob) for volume estimation. In most previous work on bark thickness, an average dib/dob ratio has been applied (Martin 1981) or the ratio was predicted as a function of tree size (Hilt 1985).

To date, data on bark thickness indicate that for many species, the dib/dob ratio depends on d.b.h. and height along the bole. Further, the dependence of dib/dob ratio on d.b.h. and height along bole can be described adequately by a linear model. The following model was fitted to tree species (as opposed to groups) for which there were a minimum of 30 observations of bark thickness:

$$\frac{\text{dib}}{\text{dob}} = \beta_{13} + \beta_{14}\text{DBH} + \beta_{15}h + \varepsilon \quad (4)$$

Figure 1.—Comparison between estimated and fixed join points using the Valentine and Gregoire model.



where: dib = inside-bark diameter (in.) at height h
 dob = outside-bark diameter (in.) at height h
 β_{13-15} = parameters to be estimated from data
 other variables as previously defined

However, because the ratios for certain species show considerable variability (e.g., white ash), it is difficult to justify a modeling approach (table 3). For these species, application of an average ratio may be sufficient. Predicted dib/dob ratios for 3 d.b.h. sizes of slippery elm are shown in figure 2. As expected, the ratio increases as tree size increases due to bark thickness occupying a relatively smaller portion of the overall diameter.

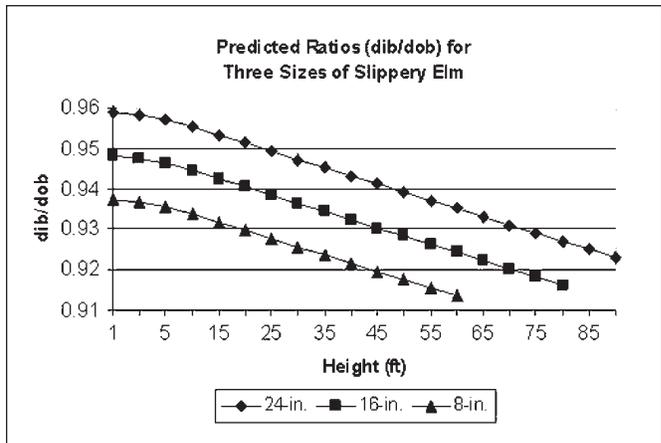
Future Work

Taper and bark thickness data still are being collected. As additional data become available, expanded analyses will be possible. The application of taper modeling results presented in this article to other species needs to be addressed, and the need to model bark thickness vs. applying an average value requires additional study. If bark thickness models are developed, mixed-effects parameters should be used to account for correlations among observations. These analyses should allow for determination of the best approaches for modeling tree taper and estimating bark thickness by comparisons of fit statistics and validation using independent data. When data collection is completed, the estimates of model parameters and other necessary statistics (e.g., average dib/dob ratios) can be finalized. When fully implemented, the bark thickness/taper equation system will provide a wide range of analytical flexibility for tree species in northeastern forests, and may reduce or eliminate the costs of collecting data on merchantable heights.

Table 3.—Nonsignificant variables and adjusted R^2 for equation [4] fit to various species (minimum of 30 observations).

Species	Nonsignificant	Adj. R^2
American basswood		0.399
Bigtooth aspen	h	0.086
Bitternut hickory	d.b.h.	0.179
Black cherry		0.192
Chestnut oak		0.544
Cucumbertree	d.b.h.	0.331
Eastern white pine	d.b.h.	0.111
Northern red oak	d.b.h.	0.119
Pignut hickory		0.474
Red maple		0.315
Scarlet oak		0.442
Slippery elm		0.315
Sugar maple		0.237
Sweet birch		0.200
White ash	d.b.h., h	0.024
White oak	d.b.h.	0.097
Yellow-poplar		0.199

Figure 2.—Predicted dib/dob ratios for three different sizes of slippery elm.



Literature Cited

- Gregoire, T.G.; Schabenberger, O.; Barrett, J.P. 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Canadian Journal of Forest Research*. 25: 137–156.
- Hilt, D.E. 1985. OAKSIM: an individual-tree growth and yield simulator for managed, even-aged, upland oak stands. U.S. Department of Agriculture Forest Service Research Paper NE-562. 21 p.
- Kozak, A. 1988. A variable-exponent taper equation. *Canadian Journal of Forest Research*. 18: 1363–1368.
- Martin, A.J. 1981. Taper and volume equations for selected Appalachian hardwood species. U.S. Department of Agriculture Forest Service Research Paper NE-490. 22 p.
- Max, T.A.; Burkhart, H.E. 1976. Segmented polynomial regression applied to taper equations. *Forest Science*. 22(3): 283–289.
- Valentine, H.T.; Gregoire, T.G. 2001. A switching model of bole taper. *Canadian Journal of Forest Research*. 31: 1400–1409.

Mapped Plot Patch Size Estimates

Paul C. Van Deusen

Abstract.—This paper demonstrates that the mapped plot design is relatively easy to analyze and describes existing formulas for mean and variance estimators. New methods are developed for using mapped plots to estimate average patch size of condition classes. The patch size estimators require assumptions about the shape of the condition class, limiting their utility. They may have some value as landscape metrics.

The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture Forest Service demarcates various forest conditions that occur on field plots. This relatively new concept results in a map for each plot showing where the conditions occur. A new condition is defined by changes in characteristics such as land use, tree species, tree size, and tree density. In the past, some FIA regions forced plots to contain a single condition by rotating them into a pure condition class. Plot rotation led to a small bias in the estimates (Birdsey 1995), and was replaced with plot mapping as recommended by Hahn *et al.* (1995).

A recent paper (Van Deusen 2004) provides a simple derivation of mapped plot estimators for the mean and variance for particular conditions. This paper briefly reviews those results and develops a new method for estimating condition patch size from mapped plots. The new method requires assumptions about the shape of the condition, i.e., circular or square. Such shape assumptions are crude approximations to the actual shape. Regardless, the resulting patch size estimates may have value as a landscape metric.

Review of Theory

The simple forest inventory model used in the original derivation (Van Deusen 2004) is used here. Assume two conditions, C and B, where C is a circular condition surrounded by condition B (fig. 1). We want estimates of the mean and variance of type C,

and the other types that surround it are denoted as B. The spatial shape of type C could be anything in practice. Sampling is systematic or simple random and uses fixed area circular plots with radius d . The edge of type C is shown by a dash line and a perimeter band overlaps the outer edge of area C shown by solid lines. The plot contains both conditions when the plot center falls within the perimeter band. The condition boundary is mapped when it crosses a plot, but there is no need to map the perimeter band.

The following notation is used (Van Deusen 2004):

a_i = the proportion of the area of plot i that is within condition C.

A = area of condition C.

\tilde{A} = area in the perimeter band that is outside of condition C plus the area of C.

$r = \frac{A}{\tilde{A}}$, the ratio of area C to the area of C plus the outer perimeter band.

y_i = a variable that can be measured on each randomly located plot that completely or partially overlaps condition C. For plots that do not overlap C, $y_i=0$.

μ = the per unit area mean of variable y for condition C, e.g., cubic meter per hectare pine volume.

$\tilde{\mu}$ = the per unit area mean of variable y in condition C inclusive of the outer perimeter.

area; the outer perimeter area contains no variable y by definition, and $\mu = \frac{\tilde{\mu}}{r}$.

n = the number of plots that contain some condition C.

The FIA sampling design involves randomly locating plot centers in a forest area. The FIA plot consists of a fixed configuration of four circular 1/24-acre subplots. The development here is simplified by using a single circular plot, but the results apply directly to the FIA plot design. After a plot is located, the amount of variable y is recorded and expanded to a per acre value. With the FIA subplot configuration, a total of 1/6 acre is sampled, so y is multiplied by 6 to expand it to a per acre value.

Consider randomly locating plot centers within the model forest area (fig. 1) and recording the amount of variable y in the plot. When plot i contains no condition C, $y_i=0$ and $a_i=0$. It follows immediately that an estimator for the ratio, r , is

¹ National Council for Air and Stream Improvement, 600 Suffolk Street, Fifth Floor, Lowell, MA 01854. E-mail: Pvandeusen@ncasi.org.

$$\hat{r} = \sum_{i=1}^n \frac{a_i}{n} \quad (1)$$

Likewise, an estimator of $\tilde{\mu}$ is

$$\hat{\tilde{\mu}} = \sum_{i=1}^n \frac{y_i}{n} \quad (2)$$

Putting equations (1) and (2) together provides an estimator for the average amount of y within condition C,

$$\hat{\mu} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n a_i} \quad (3)$$

Equation (3) is not unbiased because it involves the ratio of two random variables (Thompson 2002), but it is consistent.

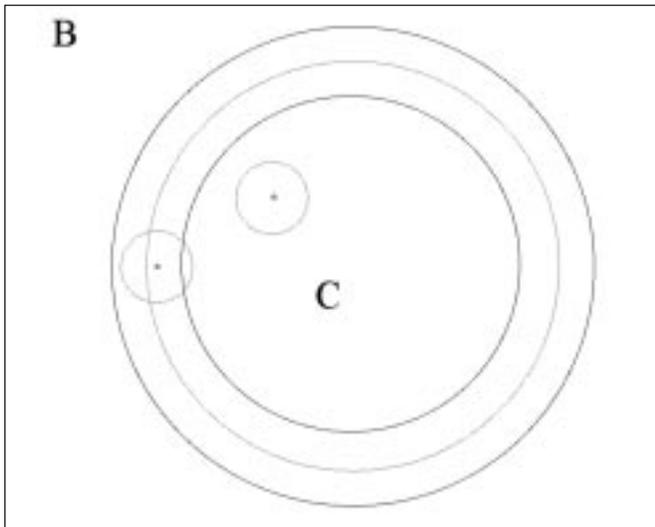
A model-based derivation of estimator (3) is provided in Van Deusen (2004), which yields the following variance estimator:

$$\text{Var}(\hat{\mu}) = \frac{\sigma^2}{\sum_{i=1}^n a_i} \quad (4)$$

where

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{e}_i^2}{\left(\sum_{i=1}^n a_i\right) - 1} \quad (5)$$

Figure 1.—An area of condition C surrounded by condition B. Condition C is bounded by the dashed line that is contained in a perimeter band of width equal to the diameter of the fixed area circular plots. Plots with centers that fall within the band will contain some of both conditions. All other plots contain only one condition. One plot that is fully in condition C is shown along with a plot that overlaps the boundary. Plots that contain no C are not of interest.



and $\hat{e}_i = y_i - a_i \hat{\mu}$.

Note that $\sum_{i=1}^n a_i$ when all plots are completely in the condition of interest. Therefore, equations (3), (4), and (5) reduce to the usual equations for simple random sampling when all plots are fully in the condition of interest or plot mapping is not implemented.

Area Estimation

Within the FIA design, three layers are referred to as P1, P2, and P3. The basic FIA plots are called the P2 layer, whereas forest health plots comprise the P3 layer. The P1 layer traditionally consisted of photo points that were classified as being forest or nonforest. Thus, the P1 layer is a higher resolution sample used to estimate forest area. The P2 layer has traditionally been a subset of the P1 layer and could potentially be used to produce a double sampling variance estimate of area. FIA is currently in the process of moving toward a P1 layer that provides wall-to-wall forest/nonforest coverage using remotely sensed data such as TM. Some uncertainty exists in forest area estimates with either approach, although FIA does not typically show the variance in forest area estimates. Confidence intervals for total volume estimates for a county or State include only the between plot sampling variance.

FIA data users have typically relied on plot expansion factors to determine the number of acres in their area of interest. These expansion factors are included in the standard FIA database (FIADB) and indicate the number of acres that a plot represents. It is difficult to justify the concept of a plot expansion factor because it implies that plots are selected with variable probability when they are actually established on a grid. Recognizing this, FIA is phasing out the plot expansion factor concept. Area expansion in this paper is based on the assumption that each plot represents 6,000 acres, adjusted depending on how many panels are being used.

Suppose we want to estimate the area or volume in condition C from mapped plots in a State. For any approach, one first needs to obtain a list of all the plots in the State that contain the desired condition. The approach used here is to compute the average volume in condition C with equation (3) and then multiply by the summed and adjusted expansion factors as follows:

$$\hat{V}_C = \hat{\mu} \sum_{i \in C} a_i E_i \quad (6)$$

where the expansion factor is reduced in proportion to the amount of C on the plot. For the example application, $E_i = 6000 * 5/3 = 10000$ for all plots, because three of five panels are used.

Equation (6) provides a way to estimate total volume in a condition class from mapped plots. The total area part of the estimate, however, could be improved with a remote sensing derived P1 layer. In any event, the focus of this paper is the patch area estimator, which does not require knowledge of the total forest area.

Patch Area

The average patch area of condition classes can be estimated from mapped plot data, if one is willing to make an assumption about the shape of a condition. A feasible shape assumption would be that the condition of interest is approximately circular. If the average radius of circular condition C (fig. 1) is R, the average area of condition C is $Area(C) = \pi R^2$. The radius of the plots used (fig. 4) is d, so the average area of condition C plus the perimeter band is $Area(C+band) = \pi (R+d)^2$. The ratio of this area is

$$r = \frac{R^2}{(R+d)^2} \quad (7)$$

the same ratio estimated by equation (1). Given the known value for d and an estimate of r, the unknown condition average-radius is estimated from

$$\hat{R} = \frac{d}{\frac{1}{\sqrt{\hat{r}}} - 1} \quad (8)$$

A circular patch area estimator follows immediately from equation (9):

$$\hat{P}_C = k\pi\hat{R}^2 \quad (9)$$

where k is a constant to convert the units used for R and d into appropriate units for area. This estimator requires only the mapped plot information that is already available in the FIA database.

A patch-area estimator can also be derived under the alternative assumption that the condition shape is square. Let the length of one side of the condition be S. The analogous procedure as used with circular shapes gives the following estimate:

$$\hat{S} = \frac{d}{\frac{1}{\sqrt{\hat{r}}} - 1} \quad (10)$$

The square patch-area estimator is

$$\hat{P}_S = k\hat{S}^2 \quad (11)$$

Therefore, the square patch assumption results in smaller patches by a factor of π . This shows that patch area estimates are sensitive to the shape assumption.

Patch-Area Variance Estimates

The patch-area estimators involve the denominator, $1/\sqrt{\hat{r}} - 1$, which presents the possibility of dividing by zero. This occurs when the estimate of r is 0 or 1. A suggested solution to this potential problem is to use an ad hoc, but more robust, estimator for r:

$$\tilde{r} = \frac{\sum a_i}{\frac{ieC}{n(n-1)}} \quad (12)$$

Estimator (11) prevents the estimate of r from reaching 1.0. An r estimate of 0.0 implies that no plots were in the condition, so this contingency is not a problem.

The only random component of the patch-size estimators is \tilde{r} , and a suggested standard error estimator for it is

$$SE(\tilde{r}) = \frac{\sum (a_i - \hat{r})^2}{\frac{ieC}{n(n+1)}} \quad (13)$$

which is the usual standard error estimator for \hat{r} . Since \tilde{r} is divided by n+1 instead of n, this should provide a conservative estimate. Now treat the patch-size estimator as a function of \tilde{r} , i.e., $P_C(\tilde{r})$. Establish approximate 95 percent upper and lower bounds on the patch-size estimates from the bounds on \tilde{r} as $P_C(\tilde{r} \pm 2\sqrt{SE(\tilde{r})})$. In practice, this involves computing the upper and lower bounds on \tilde{r} , and then using these values to compute upper and lower patch-size estimates.

Example Application

Data from the first three annual panels in Maine were used to demonstrate the patch size estimators. The analysis consists of computing the results broken down by forest type for the entire State. The estimated number of acres, number of sample plots, and mean condition volume estimates are displayed (table 1) for forest types that have at least 10 sample plots. A sample plot

can be counted in more than one row, because it can contain more than one forest type. The per acre estimate of net cubic foot volume is computed with equation (3). These figures may not correspond to official FIA reports, because different methods are being used and this is an out-of-date data set.

Patch size means and confidence bounds (table 2) are computed for each forest type assuming circular patches. The plot

radius, d , is set to the radius of a circular 1/6-acre plot. The mean patch size varies from less than an acre for nonstocked areas to 107 acres for the sugar maple/beech/yellow birch type.

It is also possible to estimate the number of patches by type (table 3) by dividing the patch sizes (table 2) into the total area in the forest type (table 1). It isn't clear how these statistics would be used, but they might be useful in conjunction with the average patch size estimates (table 2) as another landscape metric.

Table 1.—Number of acres, sample size, and net cubic feet per acre by forest type based on the first three panels for Maine. A constant expansion factor of 10,000 acres per plot is assumed. Results are shown for sample sizes of 10 or more.

Forest type	Total acres	Sample size	Cubic feet/acre
Non-stocked	51,471.00	10.00	67.57
Tamarack	80,283.00	12.00	1,122.99
Balsam poplar	121,199.00	18.00	947.07
Red maple/lowland	200,563.00	29.00	943.21
White spruce	149,275.00	21.00	1,449.91
Northern red oak	178,840.00	24.00	1,923.19
Black ash/American elm/red maple	144,677.00	19.00	967.62
Eastern white pine	535,345.00	73.00	2,736.53
Red maple/upland	640,588.00	85.00	1,204.09
White oak/red oak/hickory	127,683.00	16.00	1,246.48
Cherry/ash/yellow-poplar	168,696.00	21.00	1,090.47
Red spruce	1,021,258.00	126.00	2,166.10
White pine/hemlock	122,117.00	14.00	2,801.73
White pine/red oak/white ash	377,068.00	45.00	1,864.22
Aspen	926,498.00	112.00	1,252.95
Northern white-cedar	1,020,298.00	123	1,907.68
Black spruce	436,904.00	52.00	911.06
Eastern hemlock	700,215.00	83.00	2,049.81
Paper birch	1,386,688.00	162.00	937.76
Balsam fir	1,780,338.00	206.00	949.10
Red spruce/balsam fir	1,044,594.00	118.00	1,117.62
Sugar maple/beech/yellow birch	6,626,599.00	715.00	1,427.02

Table 2.—Mean patch size estimates (acres) for Maine by forest type, with approximate 95 percent upper and lower confidence bounds (assuming circular patches and sorted by mean patch size).

Forest type	Lower bound	Mean	Upper bound
Non-stocked	0.47	0.78	1.30
Tamarack	1.64	2.24	3.12
Balsam poplar	1.93	2.62	3.63
Red maple/lowland	2.82	3.35	4.01
White spruce	2.79	3.64	4.84
Northern red oak	3.99	5.01	6.39
Black ash/American elm/red maple	4.14	5.40	7.18
Eastern white pine	4.99	5.40	5.85
Red maple/upland	6.17	6.62	7.11
White oak/red oak/hickory	4.99	7.04	10.34
Cherry/ash/yellow-poplar	5.97	8.27	11.89
Red spruce	11.89	12.57	13.30
White pine/hemlock	9.28	14.21	23.48
White pine/red oak/white ash	13.53	15.26	17.30
Aspen	14.36	15.30	16.32
Northern white-cedar	15.00	15.89	16.85
Black spruce	14.06	16.21	18.84
Eastern hemlock	16.89	18.36	20.02
Paper birch	22.42	23.52	24.69
Balsam fir	26.15	27.19	28.29
Red spruce/balsam fir	34.41	36.76	39.35
Sugar maple/beechn/yellow birch	105.45	106.99	108.57
Means	14.70	16.03	17.85

Conclusions

FIA is installing mapped plots nationwide. Some estimators for basic statistics, such as per acre volume and variance, have been reviewed. New estimators for average patch size are also presented. The possibility of making patch size estimates is unique to the mapped plot design. Patch size estimates depend on

assumptions about patch shape and are nonrobust to the shape assumption. This limits their application and interpretation. They may be useful, however, as a basic landscape or ecological metric.

All the estimators presented in this study depend only on standard FIA measurements. They can therefore be implemented with little cost when deemed appropriate. They are a fortuitous byproduct of the mapped plot design that may prove useful.

Table 3.—Number of patches in Maine by forest type, with approximate 95 percent upper and lower confidence bounds (assuming circular patches).

Forest type	Lower bound	Mean	Upper bound
Non-stocked	39,545.96	65,886.00	109,046.95
Tamarack	25,693.23	35,770.69	48,887.27
Balsam poplar	33,367.62	46,203.98	62,727.81
Red maple/lowland	50,054.23	59,856.03	71,080.45
White spruce	30,839.63	41,016.24	53,579.08
Northern red oak	27,989.92	35,671.28	44,790.60
Black ash/American elm/red maple	20,148.13	26,812.97	34,944.24
Eastern white pine	91,525.70	99,165.71	107,257.23
Red maple/upland	90,042.63	96,765.52	103,834.91
White oak/red oak/hickory	12,350.25	18,138.77	25,600.04
Cherry/ash/yellow-poplar	14,187.58	20,404.11	28,271.76
Red spruce	76,762.60	81,251.39	85,908.70
White pine/hemlock	5,201.45	8,593.88	13,160.27
White pine/red oak/white ash	21,790.56	24,710.10	27,869.38
Aspen	56,775.40	60,561.61	64,506.11
Northern white-cedar	60,554.92	64,217.45	68,018.59
Black spruce	23,190.67	26,953.18	31,083.90
Eastern hemlock	34,981.36	38,138.02	41,468.53
Paper birch	56,173.46	58,969.30	61,849.32
Balsam fir	62,926.01	65,466.82	68,069.13
Red spruce/balsam fir	26,545.79	28,415.00	30,360.33
Sugar maple/beechn/yellow birch	61,032.87	61,934.75	62,844.00
Means	41,894.54	48,404.67	56,598.12

Literature Cited

- Birdsey, R.A. 1995. A brief history of the “Straddler Plot” debates. *Forest Science Monograph*. 31: 7–11.
- Cochran, W.G. 1977. *Sampling techniques*, 3rd ed. New York: John Wiley & Sons.
- Hahn, J.T.; MacClean, 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.
- Scott, C.T.; Bechtold, W.A. 1995. Techniques and computations for mapping plot clusters that straddle stand boundaries. *Forest Science Monograph*. 31: 46–61.
- Thompson, S.K. 2002. *Sampling*. New York: John Wiley & Sons.
- Van Deusen, P.C. 2004. Forest inventory estimation with mapped plots. *Canadian Journal of Forest Research*. 34: 493–497.

FIA Quality Assurance Program: Evaluation of a Tree Matching Algorithm for Paired Forest Inventory Data

James E. Pollard¹, James A. Westfall², Paul A. Patterson³, and David L. Gartner⁴

Abstract.—The quality of Forest Inventory and Analysis inventory data can be documented by having quality assurance crews remeasure plots originally measured by field crews within 2 to 3 weeks of the initial measurement, and assessing the difference between the original and remeasured data. Estimates of measurement uncertainty for the data are generated using paired data statistical analyses. Because plot remeasurements are taken at different, but similar, times by different crews, it can be difficult to match the remeasured trees with the original tree measurements. In the past, this process required a laborious exercise of manual review and assignment of matching codes for the paired tree measurements. An automated process for matching tree data was developed and tested using a previously hand-matched data set. Results of the two matching processes were compared. More than 95 percent of the individual trees could be reliably matched using the automated matching program. The effects of unmatched data being excluded from the uncertainty analysis was minimal.

Introduction

The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture (USDA) Forest Service provides information needed to assess the status and trends of environmental quality in the Nation's forests. The FIA program works to continually improve monitoring and assessment activities by controlling, identifying, and documenting errors and sources of variability that could be detrimental to the quality of FIA

inventory results. The quality assurance (QA) program within FIA involves the overall system of management activities designed to assure that quality data are collected. This program can be further divided into quality control and quality evaluation activities. Quality control within the program encompasses the operational techniques and activities that control the data acquisition process such as use of standardized field protocols. Quality evaluation activities involve application of statistical tools to determine if the uncertainty in the data will support programmatic decisions.

A large portion of the QA effort in the FIA program is focused on error control during the field measurement and data collection processes. One key element is provided through crew training and certification with specific national standards. Another key element of quality control in the program is development and annual updating of standardized field protocols that are documented in National Field Manuals (USDA 2003). In addition, the possibility of data entry error is reduced through use of portable field data recorders by inventory crew members. This onsite data recording reduces the chances of transcription-type data entry errors that are common problems in paper transfers. Finally, a variety of field check protocols provide immediate feedback to the crews and provide data to score crew performance.

In addition to extensive quality control activities discussed above, data quality is assessed and documented using performance measurements and post-survey assessments. These assessments identify areas of the data collection process that need improvements or refinements to meet the quality objectives of the program. Specific measurement quality objectives (MQOs) have been developed for the program and are presented in detail in the field methods guides. These quality standards were developed from extensive knowledge of measurement processes in forestry and to meet the program needs of FIA. Evaluation of data quality is accomplished by analysis of plot remeasurement data and comparison of the results to the MQO.

¹ FIA Quality Assurance Advisor, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154.

² Research Forester, U.S. Department of Agriculture (USDA) Forest Service, Northeast Research Station, 11 Campus Blvd. Suite 200, Newtown Square, PA 19073.

³ Mathematical Statistician, USDA Forest Service, Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401-2394.

⁴ Mathematical Statistician, USDA Forest Service, Southern Research Station, 4700 Old Kingston Pike, Knoxville, TN 37919.

Methods

Description of the Problem

An ongoing problem encountered when analyzing QA data is assuring that observations of individual trees are matched for paired statistical analysis. When plots are measured by two independent crews, it is not unusual for the crews to number or identify the trees slightly differently. This creates two data sets that may not be matched by tree number for a variety of reasons. For example, crews began numbering trees at different places on the plot, or crews missed a tree on the plot, setting the numbering sequence off. In addition, crews can number trees using a different spatial rule that can alter the numbering sequence for trees in a data file. Assuring that data are properly matched, and evaluating the consequences of mismatched trees in an inventory data set, is the subject of the current article. This study evaluates two different methods of assuring tree matching prior to data analysis.

The remeasurement process used to generate QA data sets in the FIA program is known as a blind check. This process involves a full reinstallation of a production inventory plot, performed by a qualified inspection crew, without access to the crew data. This results in two data sets that are independent of one another and can be subjected to paired data statistical analyses to obtain an unbiased estimate of the measurement uncertainty associated with crew performance. To analyze the quality of the two independent crews' data it is essential to have the data paired tree-to-tree so any error in the measurement process can be attributed to crew measurement error rather than data management or other nonmeasurement process errors.

The quality of FIA data has been evaluated in the past using blind check data (Pollard and Smith 1999, Pollard and Smith 2000). These data have been incorporated into a national forest health inventory report to document the basic data quality associated with these inventories (Conklin *et al.*, in press). However, to produce these assessments, it was necessary to obtain unbiased remeasurement data that was representative of the FIA program both operationally, temporally, and regionally. Once regional data sets were obtained we began a laborious process of preparing the data for analysis. This included normalizing regional differences in naming conventions and variables measured, as well as matching paired observations to the greatest extent possible.

The most time-consuming aspect of data preparation was assuring that paired observations of tree level variables were correctly matched. As increasing amounts of QA data are generated in the FIA program, and additional States are added to the national inventory, it becomes highly desirable to automate this tree matching process to the fullest extent possible.

Development of the Matching Process

Experience gained in analysis of QA data from inventories from 1998 through 2001 led to development of a hand-matching process for pairing tree-level data. The following steps were involved in this process:

- Two independently measured data files from a given inventory plot were obtained and identified as the crew data and the QA data. Each plot file was composed of four subplots of tree-level data that needed to be tree matched by subplot.
- Each tree in a given file was assigned a number within a subplot, which may or may not match the corresponding tree in the paired file depending on how the sequence was assigned (see discussion above).
- Tree-level variables were renamed in the QA data file and both files were sorted by subplot number, species of tree, horizontal distance of the tree from plot center, azimuth of tree measured at plot center, and diameter of tree at breast height.
- The data from both plots were printed with crew and QA data side by side and the data were visually compared for closeness of all matching parameters including the assigned tree number.
- If the tree numbers were not identical for sorted crew and QA observations within a subplot, then the numbering sequence was adjusted in the QA data file to match the crew data file.
- The decision to adjust the tree number was based on visual inspection for closeness of all matching parameters for a given tree as well as the total number of trees within a given subplot. For example, if crew data tree numbering started at 1 and the QA data had an extra tree in the subplot, then the numbering sequence would be off by one. In this case the tree numbers in the QA data file were adjusted to match the tree numbers in the crew data file. Then an extra number was assigned for the extra tree in the QA data file.

Once the tree numbering sequences in both crew and QA data files were matched, then differences between crew and QA crew observations could be calculated using the subplot and tree number as the identification key for a given tree.

This matching process can be very labor intensive, depending on the type of numbering discrepancy in the data files. For example, it was simple to identify an extra tree in a file and adjust the tree numbering sequences accordingly. However, if the files contained the same number of trees of the same species and the numbering sequence for more than two trees of the same species were transposed, it was much more difficult to identify which tree was the corresponding tree for a given number in the sequence. Occasionally the data for matching parameters for a number of trees in a given file were so close that it was necessary to align tree numbering sequences as a “best judgment” call. This hand-matching process was applied to a large set of Phase 3 FIA blind check data collected between 1998 and 2001 and required person months of effort totaling more than 3 years.

Refinement and Automation of the Matching Process

Refinement of the hand- matching process was initiated as a cooperative effort of three regional statisticians and the FIA Quality Assurance Coordinator. Automation of the process was developed in the SAS programming language and involved the following steps:

- QA variables were renamed in the QA data file and crew and QA data files were merged by State, county, plot number, and subplot number.
- A “distance” was computed for each QA tree to each crew tree using a function based on horizontal distance, azimuth, and diameter of the trees.
- Each QA tree was matched to the crew tree with the smallest distance. Pairs of trees were removed from the matched list because either multiple QA trees were matched to the same crew tree and only the QA with the shortest distance was matched or the distance was too great, or other technical reasons.
- A decision rule was incorporated in the matching algorithm that rejected potential matches having relatively large computed distances. This distance criteria was established to provide a conservative tree matching basis for this exercise. This distance matching criteria can be adjusted in the program if desired.

- The first iteration of matches was saved in a list file.
- Unsuitable matches were removed using similar standards as were used after the first iteration.
- A second iteration of distance functions were computed for those trees not matched in the first iteration.
- The two iterations of matched trees were combined and outputted into a matched tree list.
- The unmatched trees and/or extra/missed trees were separated into subfiles for manual examination to determine any remaining matches and to determine any missed and extra trees.

Description of the Data Files

The data were composed of inventory measurements from approximately 100 Phase 3 inventory plots measured between 1998 and 2001. Data were aggregated from the five FIA regions, for all years, which resulted in a national data set with reasonable representation from all FIA regions. The combined data set contained a total of 4,269 tree records in the QA file and 4,138 tree records in the crew file. The records in one file included trees that had corresponding matches in the other file, as well as additional trees that were unique to one or the other file. These “missed” or “extra” trees were screened from the combined data set resulting in 3,981 pairs of matched tree data that were assigned tree numbers based on best judgment of the analyst.

Results

Automated Matching Process

Application of the automated matching process to the national QA data set produced 3,576 pairs of matched trees after two iterations. Additional matched pairs of data could have been added to this data set by examination of the unmatched tree file and performing a hand-matching process. However, for the purposes of this exercise, it was decided to only use the trees matched by the fully automated process. Once the programming was complete, the actual matching process required less than one day’s effort that included multiple runs of the program to verify comparability of the results of the two matching processes.

Uncertainty Analysis

The two data sets (hand matched and automated with two passes) were subjected to an analysis of mean differences between crews and estimates of MQO compliance. Simple MQO values were used to evaluate the robustness of the data sets. The tree level variables chosen for analysis represented characteristics of tree diameter, height, and crowns. The variables analyzed were diameter at breast height (DBH), diameter at root collar (DRC), total length of the tree (Total Length), actual length of the tree (Actual Length), foliar transparency (Transparency), foliar dieback (Dieback), and foliar density (Density) of the crown, as well as the crown class.

Mean Differences Between Crews

One estimate of measurement uncertainty that can be easily calculated is the average or mean difference between crew and QA measurements. Ideally we would expect the mean differences between the two crews to be zero, which would indicate that the two estimates for a given variable were not biased.

In addition to the central tendency of the differences the dispersion of these differences is an indicator of the overall reproducibility of the data set. The Means Procedure in SAS calculates the mean, standard error of the mean, and the minimum and maximum differences. This procedure also allows the mean differences to be tested to determine if they were significantly different from zero (biased) using a Student's t test (Probability | t | Value).

The results of these calculations for both matching processes showed that the hand-matched and automated matched data sets provided very similar estimates of data uncertainty (table 1). The mean differences between investigators were very similar with some variables having slightly larger differences for the QA crews and some having slightly smaller differences for the QA crews. The pattern of probability that the mean differences were not zero was also very similar. There was a tendency for the range of differences to be somewhat larger for the hand-matched data set than for the automated matching process. This would make sense because the automated matching process set

Table 1.—Mean differences between investigators for diameter, crown, and length variables computed from the hand-matched data set (A) and automated matching data set (B).

A. Hand-matching process						
Variable	N	Mean	Standard error	Probability t value	Minimum	Maximum
DBH	3,573	-0.03	0.02	0.0684	-15.9	22.7
DRC	408	-0.25	0.08	0.0019	-26.2	5.5
Transparency	2,884	0.03	0.14	0.8236	-60	79
Crown Class	1,410	0.10	0.02	<.0001	-4	4
Die Back	2,884	0.08	0.11	0.4536	-80	94
Density	2,884	-1.21	0.21	<.0001	-60	50
Total Length	1,529	0.47	0.22	0.0356	-66	63
Actual Length	1,590	-0.17	0.24	0.4912	-116	92
B. Automated matching process						
Variable	N	Mean	Standard error	Probability t value	Minimum	Maximum
DBH	3,250	-0.03	0.00	<.0001	-6.3	4.8
DRC	326	-0.17	0.09	0.044	-25.9	4.0
Transparency	2,594	0.09	0.14	0.5416	-60	79
Crown Class	1,341	0.10	0.02	<.0001	-4	4
Die Back	2,594	0.06	0.12	0.5912	-80	94
Density	2,594	-1.37	0.22	<.0001	-60	45
Total Length	1,443	0.3	0.19	0.1162	-30	57
Actual Length	1,498	-0.05	0.17	0.7749	-67	57

aside 405 sets of measurements for manual inspection based on the matching criteria provided in the program.

It is of interest to note that, with one exception (total length), the variables that had significant bias at < 10 percent probability were the same in both data sets. However, with the exception of density, the mean differences were very small, which would make the significance of the biases somewhat irrelevant.

Measurement Quality Objective Achievement

Analyzing the field crews' performance against the program assigned MQOs can be complex. For example, the MQO for DBH is ± 0.1 inch for every 20 inches of diameter. During this exercise, simplified MQO were assigned to variables as follows to allow easy interpretation of the efficacy of the matching processes (table 2).

To compare MQO compliance between hand-matching and the automated processes, cumulative frequency distributions were computed and the percentage of the differences noted for four levels of differences: zero differences; differences within the

MQO; differences within two times the MQO; and differences within three times the MQO (table 3).

As with the results for mean differences, MQO compliance was very similar in both data sets. There was a slight tendency for the automated process to produce slightly improved MQO compliance although the improvement was rarely greater than a 2 percent improvement. It is likely that addition of the hand matched trees at the end of the automated process would result in virtually identical results.

Table 2.—Simplified measurement quality objectives.

Variable	Measurement quality objective
DHH	± 0.2 feet 95% of the time
DRC	± 0.4 feet 95% of the time
Transparency	$\pm 10\%$ Class 90% of the time
Crown Class	no errors 85% of the time
Crown Die Back	$\pm 10\%$ Class 90% of the time
Crown Density	$\pm 10\%$ Class 90% of the time
Total Length	± 5 feet 90% of the time
Actual Length	± 5 feet 90% of the time

Table 3.—Cumulative percentage of the data set with zero differences between crews and one times, two times, and three times the simplified MQO for the hand-matching process (A) and the automated process (B).

A. Hand-matching process					
Variable	N	Percent zero differences	Percent 1X MQO	Percent 2X MQO	Percent 3X MQO
DBH	3,573	50	90	94	95
DRC	408	25	74	85	91
Transparency	2,884	38	94	99	100
Crown Class	1,410	68	97	100	100
Die Back	2,884	61	98	99	99
Density	2,884	22	76	95	99
Total Length	1,529	22	74	88	95
Actual Length	1,590	23	75	89	95
B. Automated matching process					
Variable	N	Percent zero differences	Percent 1X MQO	Percent 2X MQO	Percent 3X MQO
DBH	3,250	54	94	97	98
DRC	326	29	70	82	86
Transparency	2,594	38	94	99	100
Crown Class	1,341	69	97	100	100
Die Back	2,594	61	98	99	99
Density	2,594	23	77	96	99
Total Length	1,443	22	76	90	96
Actual Length	1,498	23	77	91	97

Summary and Conclusions

Development of an automated tree matching shows much promise for time saving and simplification of data base manipulations within the FIA program for the following reasons:

- Hand-matching trees in an inventory data set produced more tree matches but required much more office labor.
- The mean differences between crews (bias) were similar for both matching methods.
- MQO compliance was similar for the two tree-matching procedures although the automated procedure tended to provide slightly better MQO compliance. It is likely that addition of hand matched trees from the list generated by the automated process would have generated very similar MQO compliance.

One needs to consider the size of the data set used in this study, however. With a sample size of thousands of trees, an automated tree-matching algorithm provided estimates of uncertainty and MQO compliance comparable to the laborious hand-matching data screening. However, if regional data sets or data sets for a given State are analyzed, the exclusion of unmatched trees from the data set may have a significant impact on the uncertainty analysis. Additional analyses are needed to evaluate this technique with smaller, regionally representative data sets. In addition, the matching program provides a list of unmatched trees. Using this much-reduced data set, hand screening of the unmatched trees becomes feasible, which should allow application of this process to much smaller data sets than were used in this study.

Acknowledgments

The authors would like to thank Susan Wright from the Northeastern Research Station for providing excellent technical editing of the manuscript. Her attention to detail and timely response is greatly appreciated. In addition, we wish to thank Mark Hansen from the North Central Research Station who provided many helpful suggestions and observations during the development of the automated tree-matching program.

Literature Cited

- Conkling, B.L.; Coulston, J.W.; Ambrose, M.J., eds. [In press.] Forest health monitoring national technical report for 2001. Asheville, NC: U.S. Department of Agriculture, Forest Service.
- Pollard, J.E.; Smith, W.D. 1999. Forest health monitoring 1998 plot component quality assurance report, vol. I. Research Triangle Park, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.
- Pollard, J.E.; Smith, W.D. 2001. Forest health monitoring 1999 plot component quality assurance report. Research Triangle Park, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.
- SAS/STAT. 1999. Users guide, ver. 8, vol. 2. Cary, NC: SAS Institute Inc. 2552 p.
- Steel, R.G.D.; Torrie, J.H.; Dickey, D.A. 1997. Principles and procedures of statistics: a biometrical approach, 3rd edition. New York: McGraw-Hill. 666 p.

Allocating Fire Mitigation Funds on the Basis of the Predicted Probabilities of Forest Wildfire

Ronald E. McRoberts, Greg C. Liknes, Mark D. Nelson¹, Krista M. Gebert², R. James Barbour³, Susan L. Odell⁴, and Steven C. Yaddof⁴

Abstract.—A logistic regression model was used with map-based information to predict the probability of forest fire for forested areas of the United States. Model parameters were estimated using a digital layer depicting the locations of wildfires and satellite imagery depicting thermal hotspots. The area of the United States in the upper 50th percentile with respect to predicted probability of forest wildfire was intersected with areas within 25 miles of rural communities needing economic assistance using a geographic information system. The proportion of total forest wildfire mitigation funds to be allocated to each national forest region was calculated as the ratio of intersected area in the region to all intersected areas nationwide.

Among the environmental issues confronting the United States in recent years, none has been more visible or compelling than the frequency and severity of forest wildfires in the Western States. This phenomenon is generally attributed to two causes: several years of widespread and intense drought and a century of aggressive fire suppression practices. These practices have resulted in a substantial accumulation of highly combustible woody material throughout much of the Nation's forested regions, particularly in the Western States.

Numerous agencies of the Federal Government participate in efforts to mitigate forest wildfire risks. Among them, Cooperative Forestry (CF), State and Private Forestry, and U.S. Department of Agriculture (USDA) Forest Service allocate Federal funds to regional, State, and community entities for mitigating wildfire risk and stimulating local economies. In

particular, the Economic Action Programs (EAP) of CF seek rural communities to which funds may be allocated, directly or indirectly, to treat forested areas as a means of mitigating wildfire risk and building industrial infrastructure. Because sufficient funds are not available to satisfy all funding needs, EAP needs defensible methods for allocating funds.

The objective of the study was to develop a defensible procedure for determining the proportion of available funds to be allocated to the national forest regions of the USDA Forest Service. The allocation to each region was to be in proportion to the area of forested lands at risk of wildfire that were in close proximity to rural communities that need economic assistance.

Methods

Data

A set of nationally consistent maps in the form of digital data layers was assembled and aggregated into three categories: Community, Ecosystem, and Fire. The Community category consisted of two layers: Populated Places and Economic Need. The Populated Places layer includes the locations and selected demographic attributes of populated places in the United States identified by the U.S. Census Bureau (ESRI 2002). Communities were selected based on CF's definition of rural communities as populated places with populations between 100 and 50,000. Selecting areas of high forest wildfire risk in close proximity to these communities, defined as a distance of 25 miles or less, simultaneously accomplishes two objectives: first, it identifies communities that are at risk of loss due to forest wildfires, and second, it identifies communities with labor forces that are sufficiently close to high wildfire risk to implement treatment prescriptions. The Economic Need layer, based on county income information from the 1980, 1990, and 2000 decennial

¹ Group Leader for Research and Analysis, Computer Specialist, and Computer Specialist, respectively, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108. Phone: 651-649-5174; fax: 651-649-5285; e-mail: rmcroberts@fs.fed.us.

² Economist, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Missoula, MT 59807.

³ Program Manager, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR 97208.

⁴ Associate Leader, U.S. Department of Agriculture, Forest Service, State and Private Forestry, Washington, DC 20250.

censuses, classifies counties into five categories. Using Geographic Information System (GIS) functions, an overall Communities layer was created that depicts areas of the contiguous 48 States within 25 miles of communities characterized as rural and in the three classes of the Economic Need layer corresponding to greatest economic need.

The Ecosystem category consisted of five layers: Total Biomass (TB), Removable Biomass (RB), Palmer Drought Index (PDI), Historical Natural Fire Regime (HNFR), and Fire Regime Current Condition (FRCC). TB is estimated using individual tree measurements on plots measured by the Forest Inventory and Analysis (FIA) program of the USDA Forest Service. The FIA sampling design is based on a nationwide array of approximately 6,000-ac hexagons, each of which includes at least one FIA plot. These hexagons are derived from the former Forest Health Monitoring Program (FHM) array of approximately 160,000-ac hexagons, which in turn are adapted from the U.S. Environmental Protection Agency's EMAP hexagon array that tessellates the contiguous 48 States (White *et al.* 1992). TB is calculated for each FIA plot, and the mean over all plots in each FHM hexagon is attributed to the hexagon as a whole. RB is an estimate of the biomass per unit area that could be removed from a forest stand to create more optimal forest conditions and partially addresses CF's desire to reduce forest fuels. RB is based on the concept of stand density index (SDI) (Reineke 1933, Avery and Burkhart 1994), a measure of forest stocking, and is estimated as the difference between observed SDI and 30 percent of maximum empirical SDI (USDA Forest Service 2003). Maximum stand density is based on the self-thinning rule (Yoda *et al.* 1963), which describes the maximum ecologically sustainable biomass on a per unit area basis. Greater overstocking is assumed to contribute to greater wildfire risk. As with TB, RB is estimated for individual FIA plots, and the mean over all plots in each FHM hexagon is attributed to the hexagon as a whole.

PDI indicates prolonged and abnormal moisture deficiencies or excesses for 350 climatic divisions in the United States (Heim 2000). On the PDI scale, 0 is normal, - 2 is moderate drought, - 3 is severe drought, and - 4 is extreme drought. PDI is an important tool for evaluating the scope, severity, and frequency of prolonged periods of abnormally dry or wet weather and has been used to indicate the potential intensity of forest fires. Mean PDI over June, July, August, and September was calculated for

2000, 2001, and 2002 for each climate division. From the climate division means, mean PDI was calculated for each FHM hexagon and attributed to the hexagon as a whole.

HNFR and FRCC are coarse-scale characterizations of pre-settlement natural fire return intervals and current vegetation conditions (Schmidt *et al.* 2002). The concept of risk is defined in terms of losing key components that define a system as a result of either wildfire or prescribed fire. Current conditions are characterized in terms of departures from historical natural conditions. These measures integrate biophysical information, remotely sensed products, and disturbance and successional processes including combinations of HNFR and potential natural vegetation (Hann and Bunnell 2001). HNFR describes the frequency and severity of pre-settlement fire processes in three categories of fire return intervals: less than 35 years, 35–100 years, and greater than 100 years. FRCC describes the relative risk of losing one or more key components that define an ecosystem in three categories of increasing wildfire risk (Schmidt *et al.* 2002). HNFR and FRCC are mapped at the resolution of 1-km² pixels, and classifications are assumed to be unchanging over periods of several years. The proportions of all 1-km² pixels in each FHM hexagon were determined for each category of both HNFR and FRCC.

TB and RB values were aggregated to the resolution of FHM hexagons to obtain enough plot observations to produce sufficient precision. HNFR and FRCC values were aggregated to the same scale for two reasons. First, aggregation reduced the size of the data set from approximately 180,000 records, each representing 1 km², to a more manageable data set of approximately 7,500 records, each representing approximately 64,800 ha. Second, the assumed low accuracy of the HNFR and FRCC classifications for the 1-km² pixels was expected to introduce an unacceptable level of measurement error into the predictor variable set. Aggregation at a coarser spatial scale alleviated some of this problem. Thus, because aggregation was considered necessary, and because TB and RB were already aggregated to the resolution of FHM hexagons, PDI, HNFR, and FRCC were also aggregated to the same resolution.

The Fire category consisted of two layers: fire perimeter data obtained from the Geospatial Multi-Agency Coordination (GeoMAC) Wildland Fire Support site (DOI and USDA 2003) and thermal data obtained from satellite imagery. The scarcity of

appropriate wildfire location data makes it difficult to calibrate national models for predicting the probability of forest wildfire. One of the few appropriate sources, GeoMAC, depicts the locations and perimeter boundaries of 2000, 2001, and 2002 forest fires on Federal lands that were sufficiently large to be recorded by geographic information specialists working on the fires. Although the layer provides excellent coverage for Western States, it includes only three fires in the Eastern United States, all of which were in close proximity to each other. Therefore, a second layer not specific to Federal lands was obtained from the Remote Sensing Applications Center, USDA Forest Service, and was used as a surrogate for fire locations and sizes. This layer, based on the thermal band of the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite sensor, identifies locations of summer 2001 and 2002 thermal hotspots. Variables related to presence or absence of fires for the GeoMAC and MODIS hotspots layers were also aggregated for FHM hexagons. If a hexagon included any portion of the perimeter of a fire recorded by GeoMAC, a variable was coded 1; otherwise, the variable was coded 0. Although the MODIS hotspots layer depicts the locations of forest wildfires, it also depicts the locations of prescribed burns and prairie, agricultural, and other fires. Thus, four MODIS hotspots fire variables were created, one corresponding to each of the threshold values of 6, 8, 10, and 12 hotspots per FHM hexagon. For each variable, if the number of hotspots equaled or exceeded the threshold value, the variable was coded 1; otherwise, the variable was coded 0.

Models

Predictions of the probability of forest wildfire for each FHM hexagon were obtained by combining the Ecosystem and Fire layers using a logistic model,

$$E(P) = \left[1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_m X_m) \right]^{-1}, \quad (1)$$

where:

$E(\cdot)$ denotes statistical expectation,

P is the probability of a forest wildfire,

β is a vector of parameters to be estimated, and

X is a vector of predictor variables consisting of values of TB,

RB, and PDI and proportions of pixels in FHM hexagons for each category of HNFRC and FRCC.

The Statistical Analysis System (SAS) CATMOD procedure with the maximum likelihood option was used to estimate the model parameters. The model was calibrated twice: once for the two eastern national forest regions collectively and once for the six western national forest regions collectively. The model was separately calibrated for three reasons: first, the model was calibrated using the GeoMAC layer only for the western regions because of the inadequate number of observations for the eastern regions; second, the model was calibrated separately for the western regions using the MODIS hotspots layers to facilitate comparisons of results with those obtained for calibration with the GeoMAC layer; and third, because of differences in species composition, topography, and forest management practices, relationships between the probability of forest wildfire and the predictor variables were expected to differ between the eastern and western regions. Thus, nine sets of model parameters were estimated: one set for the western regions using the GeoMAC layer; four sets for the western regions, one for each of the four MODIS hotspots threshold levels; and four sets for the eastern regions—one for each of the four MODIS hotspots threshold levels. Predictions for the five model calibrations for the western regions were compared by evaluating the similarity in the rankings of individual hexagons with respect to their predicted probabilities of forest wildfire.

Estimation

Using model [1] with estimates of the model parameters, the predicted probability of forest wildfire was calculated for July 2002 for each FHM hexagon, and a map depicting the 50 percent of hexagons with the greatest predicted probabilities of forest wildfire was constructed. Using a GIS, the selected areas from these maps were intersected with the Community layer depicting areas within 25 miles of rural communities in need of economic assistance. The proportion of EAP funds to be allocated to each national forest region was calculated as the ratio of the area selected for each region to the total area selected for all regions.

Results and Discussion

For the Western States, the maps depicting the predicted probabilities of forest wildfire using models calibrated with the four

MODIS hotspots variables were similar to each other and to the map obtained using the model calibrated with the GeoMAC variable. For each of the five maps, each hexagon was classified with respect to whether its predicted probability of forest wildfire exceeded probability percentiles ranging from 0.05 to 0.95 in steps of 0.05. Comparisons of the classifications of individual hexagons for each of the four MODIS-based maps to the GeoMAC-based map revealed that proportions of hexagons with the same classification always exceeded 95 percent. These results indicate that although the models calibrated using different fire location layers produced different predictions of the probability of forest wildfire, the relative rankings of the hexagons with respect to percentiles of the probability predictions were very similar. Thus, for the western regions, the MODIS hotspots data layers were concluded to be acceptable surrogates for forest fire locations for ranking the hexagons.

The greatest similarity between rankings with the GeoMAC variable and a MODIS variable was obtained for the MODIS variable corresponding to a threshold value of eight hotspots per hexagon. Thus, this MODIS variable was used as a surrogate for the presence or absence of a forest wildfire for calibration of the model for the two eastern regions also. A map depicting areas of the country in the upper 50th percentile with respect to the probability of forest wildfire was constructed by selecting

the hexagons at or above the median predicted probability separately for the Eastern and Western United States (fig. 1). The map indicated much more area with high relative probabilities in the Southeastern United States than the Northeastern United States, but the area with high relative probabilities was more concentrated in the western regions. Because separate models were calibrated for the eastern and western regions, the relatively greater amount of area selected in the Eastern United States should not necessarily be construed to mean that more area is at greater risk of wildfire in the East. This phenomenon may possibly be attributed to different calibration data sets, responses to predictor variables, species compositions, forest management practices, and climate.

This digital layer corresponding to the upper 50th percentile of the country relative to the probability of forest wildfire was intersected with the Community layer (fig. 2). Proportions of funds to be allocated to national forest regions were calculated as the ratios of areas in the intersections for a particular national forest region to the collective area of the intersection for all national forest regions. Based on the intersected areas, the proportional allocations were estimated for the East as 0.739 to Region 8 and 0.261 to Region 9, and for the West as 0.216 to Region 1, 0.211 to Region 2, 0.220 to Region 3, 0.084 to Region 4, 0.129 to Region 5, and 0.141 to Region 6 (fig. 2).

Figure 1.—Percentile identity of hexagons with respect to relative probability of forest fire for July 2002 (light gray = nonforest; dark gray = 50% of forested area with smallest predicted probabilities; black = 50% of forested area with greatest predicted probabilities; white = hexagons with eight or more MODIS hotspots in 2002).

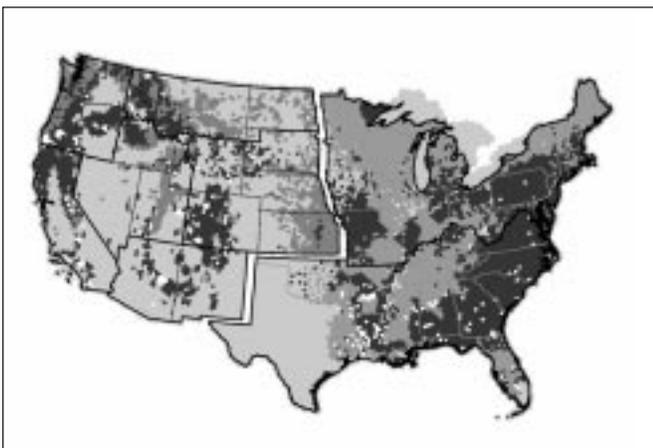
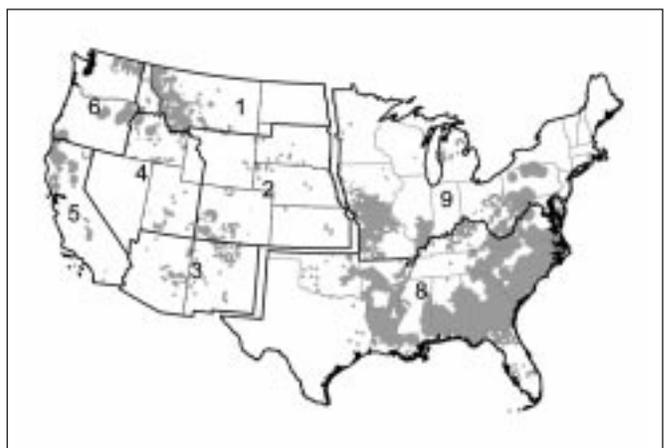


Figure 2.—Areas with the 50% largest predicted probabilities of forest wildfire for 2002 within 25 miles of a rural community needing economic assistance (light gray); numerals refer to national forest regions.



Acknowledgments

The authors thank Pat Miles and John Vissage, North Central Research Station, St. Paul, MN, for providing the TB and RB layers, and Mark Finco and Tom Bobbe, Remote Sensing Applications Center, Salt Lake City, UT, for providing the MODIS hotspots layer. All four people are employees of the USDA Forest Service.

Literature Cited

- Avery, T.E.; Burkhart, H.E. 1994. Forest measurements. New York: McGraw-Hill. 408 p.
- ESRI. 2002. ESRI data & maps 2002. An ESRI white paper. Redlands, CA.: ESRI. 24 p. <http://www.esri.com/library/whitepapers/pdfs/datamaps2002.pdf>. (25 April 2005).
- Hann, W.J.; Bunnell, D.L. 2001. Fire and land management planning and implementation across multiple scales. *International Journal of Wildland Fire*. 10: 389–404.
- Heim, Jr., R.R. 2000. Drought indices: a review. In: Whilwhite, D.A., ed. *Drought: a global assessment*. Routledge Hazards and Disasters Series, vol. 1. London: Routledge: 159–167.
- Reineke, L.H. 1933. Perfecting a stand-density index for even-aged stands. *Journal of Agricultural Research*. 46: 627–638.
- Schmidt, K.M.; Menakis, J.P.; Hardy, C.C.; *et al.* 2002. Development of coarse-scale spatial data for wildland fire and fuel management. Gen. Tech. Rep. RMS-87. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 41 p.
- U.S. Department of Agriculture (USDA) Forest Service, Research and Development. 2003. A strategic assessment of forest biomass and fuel reduction treatments in western states. http://www.fs.fed.us/research/pdf/Western_final.pdf. (25 April 2005).
- U.S. Department of the Interior and U.S. Department of Agriculture (DOI and USDA). 2003. Geospatial multi-agency coordination wildland fire support. http://geomac.usgs.gov/homepage/usersguide/geomac_ug.pdf. (25 April 2005).
- 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.
- Yoda, K.; Kira, T.; Ogawa, H.; Hozumi, K. 1963. Self thinning in overcrowded pure stands under cultivated and natural conditions (intraspecific competition among higher plants XI). *Journal of Biology (Osaka City University)*. 14: 107–129.



Comparison of U.S. Forest Land Area Estimates From Forest Inventory and Analysis, National Resources Inventory, and Four Satellite Image-Derived Land Cover Data Sets

Mark D. Nelson¹, Ronald E. McRoberts², and Veronica C. Lessard³

Abstract.—Our objective was to test one application of remote sensing technology for complementing forest resource assessments by comparing a variety of existing satellite image-derived land cover maps with national inventory-derived estimates of United States forest land area. National Resources Inventory (NRI) 1997 estimates of non-Federal forest land area differed by 7.5 percent from estimates based primarily on Forest Inventory and Analysis data reported in the Forest and Rangeland Renewable Resources Planning Act of 1974 (RPA) draft 2002 forest resource assessment. The NRI estimates differed only 2.2 percent from non-Federal land area, with the NRI estimate slightly smaller than the RPA estimate. Comparisons of statewide forest land area estimates derived from these two inventories with four satellite image-derived maps reveal area-weighted root mean square deviations ranging from 2.5 to 41.0 percent across the conterminous United States. In general, estimates of non-Federal forest land area from RPA and NRI were more closely related to each other than to image-derived estimates. The Forest Cover Types map and the National Land Cover Data set produced image-derived estimates that were most similar to the RPA estimate of forest land area across all land ownerships.

For more than half a century, global forest resource assessments (FRAs) have been conducted by the Forest Resources Assessment Programme of the Food and Agriculture Organization (FAO) of the United Nations to “provide information on the state of forest

resources worldwide on a continuing basis.” These FRAs are based primarily on national forest inventory information provided by countries, supplemented by state-of-the-art technology. The global FRA of 2000 (Food and Agriculture Organization of the United Nations 2001) identified a need to complement future inventories of forest parameters through remote sensing technology. Zawila-Niedziecki (2000) edited a compilation of works on this effort, presented at an International Union of Forest Research Organizations conference on remote sensing and forest monitoring. Our study sought to test one application of remote sensing technology for complementing FRAs by comparing estimates of forest land area from a variety of existing satellite image-derived land cover maps with national inventory-derived estimates of U.S. forest land area.

The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture (USDA) Forest Service (<http://fia.fs.fed.us>) conducts detailed surveys of the Nation’s forests across all ownerships. The USDA Natural Resources Conservation Service (NRCS) monitors land use, status, condition, and trends of the Nation’s soils, water, and related natural resources on non-Federal lands through its National Resources Inventory (NRI) (<http://www.nrcs.usda.gov/technical/NRI>). Differences in sampling designs and definitions of land cover/use categories contribute to differences in estimates of forest land and other common land cover/use categories between these two inventories (Lessard *et al.* 2003). Czaplowski *et al.* (2002) reported that NRI statewide estimates of forest area can differ by more than 30 percent from FIA estimates, although these large relative differences occur only in a few sparsely forested states where forest land area is small.

Satellite image-derived land cover data and related geospatial data layers provide an alternative source of information from which forest land area estimates can be calculated and compared

¹ GIS/RS Specialist, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

² Mathematical Statistician, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

³ Statistician, USDA Natural Resources Conservation Service, Natural Resources Inventory and Analysis Institute, 1992 Folwell Avenue, St. Paul, MN 55108.

with inventory estimates. Conversely, field-based inventory data provide a reference for assessing the accuracy of satellite image-derived data. For example, Owens (2001) reported that lowland conifer, pines, and nonforest groups had the largest differences with respect to area when comparing FIA plot-based and Landsat Thematic Mapper (TM) image-based estimates in Michigan's Upper Peninsula. She reported difficulty in creating a common legend between FIA forest types and forest type classes in a TM image-based map, and reported that differences in spatial resolution between FIA and TM maps led to differences in area estimates (Owens 2001).

Häme *et al.* (2001), Päivinen *et al.* (2001), Kennedy and Bertolo (2002), and Schuck *et al.* (2003) compared pan-European forest area estimates derived from forest inventory and satellite image-derived sources. For some European countries, forest land area estimates derived from Advanced Very High Resolution Radiometer (AVHRR) satellite imagery and official statistics were within ± 5 percent. For other European countries, area-weighted root mean square errors (RMSE) of estimates derived from AVHRR imagery and forest inventory statistics were tens of percent (Päivinen *et al.* 2001, Schuck *et al.* 2003). Thus, satellite image-derived estimates of forest land area appear relatively comparable across large geographic areas such as the European Union, but differences in these estimates vary among regions and tend to increase within smaller geographic regions.

In our study we explored the efficacy of satellite image-derived maps for estimating forest land area in the United States by comparing estimates obtained from FIA, NRI, and four satellite image-derived data sets: 1991 Forest Cover Types, 1992–93 Land Cover Characteristics, 2001 Vegetation Continuous Fields, and the 1992 National Land Cover Data set. We address differences among FIA and NRI estimates by incorporating ancillary geospatial data. Comparisons are made for the entire United States, the conterminous United States (CONUS), and for individual States.

Data and Methods

Land Ownership

Polygons in the Conservation Biology Institute's Protected Areas Database (PAD) 2001 (DellaSala *et al.* 2001) that delineate boundaries of Federal ownership were recoded into a single

Federal lands class. Areas within detailed State boundaries not delineated in PAD as Federal lands or surface water were assumed to have non-Federal ownership. PAD Federal lands were used as a geospatial filter when comparing satellite image-derived estimates with NRI and FIA estimates of non-Federal forest land area.

Inventory Estimates

Forest and Rangeland Renewable Resources Planning Act of 1974

Estimates used in this study come from the Forest and Rangeland Renewable Resources Planning Act of 1974 (RPA), P.L. 93-378, 99 Stat. 4765 (USDA Forest Service) FRA 2002 Draft Tables on U.S. forest resources, with source dates ranging between 1983–2000 and an average of 1994 (http://ncrs2.fs.fed.us/4801/fiadb/rpa_table/Draft_RPA_2002_Forest_Resource_Tables.pdf). RPA data were derived from FIA data, except for portions of some western States where National Forest System (NFS) lands were inventoried independently (Smith *et al.* 2001, USDA Forest Service 2003). Each of the five regions in the national FIA program report estimates of forest land area for their respective States. These estimates are obtained by multiplying total area inventoried by the mean proportion forest land estimated from forest inventory plot observations. National FIA precision standards "are designed to meet statistical guidelines for accuracy within one standard deviation at the 67 percent level for each State: ± 3 –5 percent per million acres of timberland, ± 5 –10 percent per million acres of all other forest land" (Smith *et al.* 2001). Because natural variability among plots and budgetary constraints limit the sufficiency of sample sizes, national FIA precision standards may not be achieved using estimation techniques based on simple random sampling. A technique known as stratified estimation (post-sampling stratification) is used to reduce uncertainty of FIA estimates (Cochran 1977, Hansen 2001). Sampling errors used in this study were obtained from a compilation of published statewide FIA reports (Hansen unpublished report) or by updating published data from previous inventories using formula 3 in Hansen (2001). FIA sampling errors for Alaska and Hawaii were estimated based on a conservative assumption that their forest land area estimates meet the FIA national precision standard (likely an underestimate of sampling error) because no FIA sampling errors were available for these two States.

FIA defines forest land as “timberland,” “reserved forest land,” or “other forest land,” including some pastured land with trees, forest plantations, and unproductive forest land. This definition of forest land also requires 10-percent minimum stocking level or, for several western woodland types where stocking cannot be determined, 5-percent canopy cover; minimum area of 0.405 ha (1 acre); and a minimum continuous canopy width of 36.58 m (120 feet) (USDA Forest Service 2003).

National Resources Inventory

NRI is a statistical survey designed to help gauge natural resource status, conditions, and trends on non-Federal land in the United States and is carried out under the authority of a number of legislative acts including the Rural Development Act of 1972, the Soil and Water Resources Conservation Act of 1977, the Federal Agriculture Improvement and Reform Act of 1996, and the Farm Security and Rural Investment Act of 2002. Although NRI data are currently collected on an annual basis, NRI inventories were conducted every 5 years from 1977 through 1997. For this project, statewide NRI estimates of non-Federal forest land were obtained from the NRI 1997 inventory. The 1997 NRI database was chosen because of the temporal similarity to the 2002 RPA’s mean data source date of 1994. Although future NRI inventories will also include Alaska, no 1997 NRI data were collected for that State; Alaska is therefore excluded from NRI statewide estimates. The NRI is a longitudinal sample survey based on scientific statistical principles and procedures. The NRI is designed as a stratified cluster sample. Estimates and standard errors of the estimates are calculated using standard statistical procedures (Cochran 1977, Fuller *et al.* 1986, Särndal *et al.* 1992).

The NRI land cover/use definition of forest land is similar to that of FIA in minimum size (0.405 ha or 1 acre) and stocking (10 percent) requirements. Although both require a minimum area of 1 acre, NRI specifies a minimum width of 100 feet, while FIA specifies a minimum width of 120 feet. In some areas of the west, FIA interprets the 10-percent stocking requirement to be equivalent to 5-percent canopy cover, but the stocking definition used by FIA most often is calculated from field measurements of basal area and number of trees per unit area. NRI interprets 10-percent stocking to be equivalent to 25-percent canopy cover when viewed from a vertical direction. Also

included in both FIA and NRI forest land definitions are lands not currently developed for nonforest use that bear evidence of natural regeneration of tree cover (for example, cutover forest or abandoned farmland).

Satellite Image-Derived Estimates

This study used four satellite image-derived maps to estimate forest land area. Statewide estimates of forest land area for each of these image sources was obtained by overlaying a detailed State boundary geospatial dataset (ESRI® Data & Maps 2002) using ArcGIS® software (ESRI).

Forest Cover Types

Forest Cover types (FC) data were produced by the Forest Service and the United States Geological Survey (USGS) and are distributed on the National Atlas Web site (<http://www.nationalatlas.gov/fortypem.html>). Sometimes referred to as the “RPA map” because of its inclusion in the 1997 RPA report (Smith *et al.* 2001), FC is a thematic classification of 25 forest cover types derived from 1991 AVHRR imagery at 1-km spatial resolution (Zhu and Evans 1994). When estimating forest land area, we included all 25 forest types and excluded four nonforest classes (ocean fill, non-U.S. land, U.S. nonforest, and lakes) from the 29 available classes.

Land Cover Characteristics

Land Cover characteristics (LC) data were produced by USGS and are distributed on the National Atlas website (<http://www.nationalatlas.gov/landcv.html>) as a map of 25 land cover classes at 1-km spatial resolution (Loveland *et al.* 2000). This data set was created using AVHRR imagery from 1992–93. When estimating forest land area, we excluded 20 nonforest classes and included 5 forest classes: Deciduous Broadleaf Forest, Deciduous Needleleaf Forest, Evergreen Broadleaf Forest, Evergreen Needleleaf Forest, and Mixed Forest.

Vegetation Continuous Fields

Vegetation Continuous Fields (VCF) data provide per-pixel tree cover estimates as percent tree canopy cover data and are derived from 2001 Moderate Resolution Imaging Spectroradiometer (MODIS) imagery at 500-m spatial resolution. VCF data are produced and distributed by the Global Land Cover Facility at

the University of Maryland (<http://modis.uniaccs.umd.edu/vcf.htm>). Hansen *et al.* (2002) reported that a VCF minimum percent tree canopy cover threshold of 35 percent produced a map of CONUS forest land similar in forest land area to a 1992 Forest Service estimate (Powell *et al.* 1993). We calculated two independent VCF percent tree cover thresholds that produce national estimates of forest land area equivalent to NRI 1997 estimates of non-Federal forest land in the United States (excluding Alaska) and RPA draft 2002 estimates for all 50 states, across all land ownerships. Statewide estimates of forest land area were obtained from image pixels having VCF percent tree canopy values greater than or equal to national thresholds corresponding to NRI and RPA estimates.

National Land Cover Data Set

The circa 1992 National Land Cover Data set (NLCD) is a 30-m spatial resolution national land cover data set produced and distributed by the USGS EROS Data Center (EDC), available at <http://landcover.usgs.gov/natl/landcover.asp>, using early 1990s Landsat Thematic Mapper imagery and other sources of digital data. The classification system used for NLCD provides a consistent hierarchical approach to defining 21 classes of land cover across CONUS (Vogelmann *et al.* 2001). For estimating forest land area we examined eight combinations of up to six NLCD classes: transitional (33)⁴, deciduous forest (41), evergreen forest (42), mixed forest (43), shrubland (51), and woody wetland (91). Table 1 provides definitions for each combination of NLCD

Table 1.—CONUS estimates of forest land area (thousand acres) and proportion of CONUS in forest land for non-Federal lands (both NRI and RPA) and for lands of all ownerships (RPA only) derived from NRI, RPA, lower (95low) and upper (95up) limits of NRI and RPA 95-percent confidence intervals, FC, LC, VCF25, VCF36, and combinations of NLCD 1992 classes: 41, 42, and 43 (NLCD3); 33, 41, 42, and 43 (NLCD4a); 41, 42, 43, and 51 (NLCD4b); 41, 42, 43, and 91 (NLCD4c); 33, 41, 42, 43, and 51 (NLCD5a); 33, 41, 42, 43, and 91 (NLCD5b); 41, 42, 43, 51, and 91 (NLCD5c); 33, 41, 42, 43, 51, and 91 (NLCD6); and all 21 classes combined (NLCD21).*

Estimate	Non-Federal lands		All land ownerships	
	Thousand acres	Proportion	Thousand acres	Proportion
NRI	404,680	0.271	—	—
NRI95low	391,655	0.262	—	—
NRI95up	417,704	0.279	—	—
RPA	437,315	0.292	620,306	0.321
RPA95low	432,205	0.289	613,819	0.318
RPA95up	442,425	0.296	626,793	0.325
FC	451,091	0.302	631,897	0.329
LC	473,881	0.317	620,503	0.323
VCF36	407,699	0.273	—	—
VCF25	—	—	623,413	0.325
NLCD3	—	—	562,986	0.292
NLCD4a	—	—	575,248	0.298
NLCD4b	—	—	912,741	0.473
NLCD4c	—	—	617,539	0.320
NLCD5a	—	—	925,003	0.479
NLCD5b	—	—	629,801	0.326
NLCD5c	—	—	967,294	0.501
NLCD6	—	—	979,556	0.507
NLCD21	—	—	1,930,619	1.000

*Values in **bold** type indicate that for pair-wise comparisons, one estimate of CONUS forest land area falls within the 95 percent Confidence Interval of the NRI estimate (for non-Federal land) or the RPA estimate (for all land ownerships).

⁴ The correct numerical designation for the transitional class is 33; its designation as 31 in Vogelmann *et al.* (2001) is attributed to a manuscript error (Vogelmann, EROS Data Center, U.S. Geological Survey, personal communication, 10 October 2001).

classes. NLCD forest land area estimates are compared with other estimates across all ownerships, but not with estimates of non-Federal forest land as with the other three satellite image-derived estimates.

Comparisons

RPA/NRI Comparisons

Using statewide sampling errors described above for the two inventory estimates, three separate sets of 95-percent confidence intervals were computed for RPA estimates of forest land area across all ownerships, RPA estimates of non-Federal forest land area, and NRI estimates of non-Federal forest land area. Nonoverlapping confidence intervals for RPA and NRI estimates of non-Federal forest land were interpreted as indicating the plot-based estimates were significantly different.

No estimate of uncertainty (e.g., confidence interval) was available for the image-based estimates. Differences in estimates of forest land area between image- and plot-based estimates are reported as significantly different if image-based estimates fell outside the 95-percent confidence intervals for RPA (across all land ownerships) and NRI (non-Federal lands). Using PAD, Federal lands were excluded from satellite image-based maps when comparing image and NRI estimates of non-Federal forest land area.

Root Mean Square Deviation

In this paper we use area-weighted root mean square deviation (RMSD) rather than RMSE (as was cited from previous studies) for comparing differences in pairs of statewide forest land area estimates derived from plot- or image-based sources:

$$RMSD_{rs} = \sqrt{\sum_r \frac{a_i}{A} (\hat{p}_{ir} - \hat{p}_{is})^2} \quad (1)$$

where a_i is the area of the i^{th} state, A is the total area (sum of a_i s for all states), and \hat{p}_{ir} and \hat{p}_{is} denote the estimated proportion of forest land area in the i^{th} state obtained from two (r, s) of the six sources compared in this study (Häme *et al.* 2001). Values for a_i 's, \hat{p}_{ir} 's, and \hat{p}_{is} 's pertain either to non-Federal lands only or to all land ownerships (depending on estimate pairs), but are consistent in each pair.

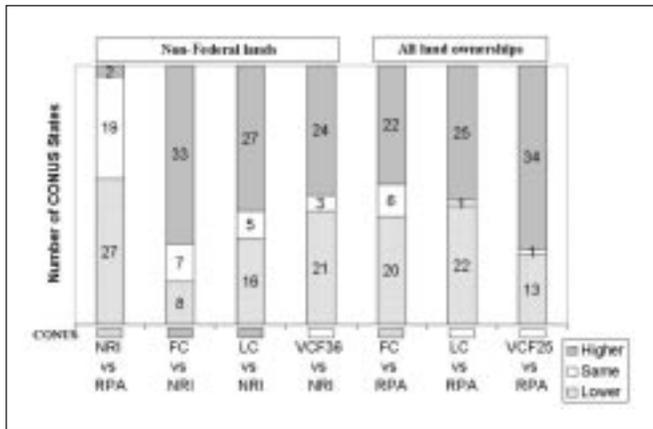
Results

We observed a 2.8-percent difference between 1997 U.S. Census Bureau statistics and 2001 PAD-derived non-Federal land area for the Nation (1.8-percent difference when excluding Alaska). Likewise, statewide 2001 PAD estimates of non-Federal land area generally were within a few percent of 1997 Census Bureau statistics, with notable exceptions for Wyoming (24 percent), Alaska (15 percent), and Idaho (11 percent). (Appendix A; appendixes are not included in this manuscript due to space constraints but are available from the senior author.) An RMSD of 2.5 percent (2.3 percent when excluding Alaska) was observed when comparing PAD to Census Bureau estimates of U.S. non-Federal land.

RPA and NRI estimates of non-Federal forest land area differ by 7.5 percent (NRI-RPA/RPA) across CONUS and as much as 54 percent for individual statewide comparisons (Appendix B). Relative to non-Federal land area (NRI-RPA/non-Federal), the CONUS difference is only 2.2 percent and the maximum statewide difference is 12 percent. The NRI estimate (± 95 percent confidence interval) of CONUS non-Federal forest land area (404.7 ± 13.0 million acres) was significantly less than the RPA estimate (437.3 ± 5.1 million acres) (Appendix B). Statewide NRI estimates were significantly less than RPA estimates in 27 States, similar in 19 States, and significantly greater in 2 of 48 CONUS States (Appendix B, fig. 1). The RMSD for NRI versus RPA forest land area estimates was 3.9 percent (table 2).

Minimum VCF tree canopy cover thresholds of 36 percent (VCF36) and 25 percent (VCF25) resulted in national estimates of forest land area within 95-percent confidence intervals of NRI estimates for CONUS non-Federal lands (404.7 ± 13.0 million acres) (Appendix B) and RPA estimates for the entire United States across all ownerships (748.9 ± 7.2 million acres) (Appendix C). VCF thresholds resulting in statewide estimates of forest land area equivalent to RPA statewide estimates (all ownerships) ranged from 2 percent in Arizona, Nevada, New Mexico, and Utah to more than 55 percent in Connecticut, Massachusetts, and Rhode Island, with six states (Alaska, Illinois, Indiana, Iowa, Ohio, and Oklahoma) having thresholds within 5 percent of RPA's nationwide VCF threshold (25 percent) (Appendix C).

Figure 1.—Comparison of statewide estimates of forest land area within the conterminous United States (CONUS), excluding Alaska and Hawaii. Bar areas and numerals portray the number of States in each pair-wise comparison where the first estimate is higher, the same as, or lower than the 95-percent confidence interval of the second estimate. Rectangular bars below the x-axis illustrate CONUS-wide comparisons, using the same scheme as described for statewide comparisons.



Of the three estimates of CONUS non-Federal forest land derived from image products, both FC and LC were significantly higher than NRI while VCF36 was not significantly different than NRI (Appendix B, fig. 1). Based on RMSD, FC appears more similar to both NRI (5.0 percent RMSD) and RPA (2.9 percent RMSD) than do three other image-derived estimates of non-Federal forest land percent (table 2, Appendix B). Slightly larger RMSD values were observed when comparing non-Federal forest land percent estimates based on VCF36 to NRI (6.0 percent) and RPA (7.3 percent) (table 2, Appendix B).

For CONUS estimates across all land ownerships, NLCD3 and NLCD4a (table 1) were significantly lower; FC, NLCD4b, NLCD5a, NLCD5b, NLCD5c, and NLCD6 were significantly higher; and LC, VCF25, and NLCD4c did not differ significantly from the RPA estimate of forest land area (table 1, fig. 1, Appendix C, Appendix D). Of these three estimates, NLCD4c had the smallest RMSD (5.8 percent) (table 1) for CONUS while VCF25 resulted in the smallest RMSD for the entire United States (9.9 percent), relative to RPA estimates. Compared to the RPA estimate across all ownerships, however, FC had the lowest RMSD of all image-derived estimates (2.5 percent) (table 1, table 2).

Table 2.—Area-weighted RMSD (percent) for pair-wise comparisons of forest land area estimates for CONUS non-Federal lands (above diagonal, *italics font*) and for CONUS lands across all ownership (below diagonal, *regular font*).*

Comparison	RPA	NRI	FC	LC	VCF25	VCF36
RPA	3.9	2.9	13.2	—	7.3
NRI	—	5.0	14.2	—	6.0
FC	2.5	—	14.0	—	7.6
LC	12.2	—	13.1	—	13.1
VCF25	10.7	—	10.6	12.0	—
NLCD3	7.0	—	8.3	12.9	13.0
NLCD4a	6.4	—	7.7	12.4	12.1	—
NLCD4b	32.2	—	32.4	38.6	41.0	—
NLCD4c	5.8	—	6.3	12.4	8.3	—
NLCD5a	32.2	—	32.4	38.5	40.8	—
NLCD5b	6.2	—	6.5	12.6	7.3	—
NLCD5c	31.4	—	31.4	37.7	39.3	—
NLCD6	31.5	—	31.5	37.7	39.2	—

*Values in **bold** type indicate that for pair-wise comparisons, one estimate of CONUS forest land area falls within the 95-percent Confidence Interval of the other estimate in that pair.

Discussion

Because of its similarity with the 1997 Census Federal land area statistics, PAD was deemed suitable for use as a geospatial filter to exclude Federal land in image-derived analyses for national estimates and for most statewide estimates of non-Federal forest land area. Differences in shoreline definitions and census vintage between ESRI® Data & Maps (2002) State boundary delineations; census statistics; and NRI statistics may result in slight differences in statewide total land area estimates. This study does not address possible effects of these differences on estimates of statewide forest land area.

In general, estimates of non-Federal forest land area from the two inventory sources were more closely related to each other than to the image-derived estimates. The notable exception was the closer agreement between FC and RPA than between NRI and RPA. The relatively small differences between NRI and FIA may be the result of differences in sampling intensity, year of photography, definitions of forest land cover/use, or inventory conventions on range and pasturelands. Greatest differences occurred in the arid western states, which is also where the most rangeland occurs, often with substantial patches of woody canopy).

VCF Estimates

The VCF36 threshold that produced estimates comparable with the NRI 1997 CONUS estimate of non-Federal forest land is nearly identical to the 35 percent-threshold reported by Hansen *et al.* (2002) as being equivalent to a 1992 CONUS Forest Service estimate of forest land area across all ownerships. It is different, however, from the 25-percent threshold we observed for the RPA 2002 nationwide estimate of forest land across all ownerships. No single VCF threshold appears suitable for all national inventories.

By identifying VCF percent tree canopy thresholds specific to inventory, geographic area, land ownership, and date, we obtained VCF-derived estimates of forest land area that fell within 95-percent confidence intervals of NRI and RPA forest land area estimates. For most States, it is inappropriate to extract statewide estimates of forest land area using maps created with a nationwide VCF threshold. The vast majority of VCF-derived statewide estimates were significantly larger or smaller than their

corresponding inventory-based estimates, and RMSD values were approximately 6 and 11 percent for comparisons of NRI:VCF36 (non-Federal lands) and RPA:VCF25 (all land ownerships), respectively.

The VCF data set includes continuous estimates of percent tree, percent herbaceous, and percent bare cover, allowing for user-defined thresholds, varying by inventory, land ownership, and geographic extent. Although not analyzed in this study, a similar product (forest density, derived from 1-km AVHRR) also is available as a companion to the FC data (Zhu 1994). Improvements to VCF-derived estimates of forest land area would be expected by determining thresholds in smaller geographic areas; stratifying areas by forest types, type groups, or life forms; or by calibrating per-pixel continuous estimates of land cover to match inventory-based estimates.

NLCD Estimates

All eight NLCD combinations used in this study included three “pure” forest classes (deciduous forest, evergreen forest, and mixed forest) and seven of the eight combinations included one, two, or three additional NLCD classes (transitional, shrubland, and woody wetland). Area estimates derived from NLCD varied widely with class combinations. Estimates from NLCD combinations that included shrubland were about 50 percent larger than RPA estimates of CONUS forest land area across all ownerships and had RMSD values of 31–32 percent. In contrast, NLCD combinations that excluded shrubland resulted in estimates similar to RPA estimates, with RMSD values of about 6–7 percent. In particular, the NLCD4c four-class combination (deciduous forest, evergreen forest, mixed forest, and woody wetland) had the lowest RMSD (5.8 percent) of any NLCD-derived estimate of CONUS forest land area across all ownerships and was the only NLCD-derived estimate to fall in the 95-percent confidence interval surrounding the RPA estimate. NLCD5b included one additional class (transitional), had an RMSD value (6.2 percent) slightly larger than NLCD4c, and produced an estimate of forest land area only slightly larger than the RPA 95 percent upper confidence interval. The apparent superiority of NLCD over some of the other satellite image-derived estimates may result from its finer spatial resolution (30 m versus 500–1000 m) combined with its temporal similarity (~1992) to the mean date of RPA data collection (~1994). Although NLCD estimates

were not calculated for non-Federal lands, such a calculation might be useful for comparison with NRI estimates of non-Federal forest land and other land cover classes.

Forest Cover Types Estimates

Compared with RPA estimates across all land ownerships, the FC-derived estimates had the lowest RMSD of any image-derived estimate, even though the CONUS estimate from FC was significantly larger than the RPA estimate of CONUS forest land area. It is not surprising that FC performed well, despite its coarser spatial resolution (1-km pixel size), because the “Forest Cover Types” used in the FC classification were defined to be consistent with RPA definitions.

Conclusions

Five conclusions may be drawn from the results of this study. First, inventory-derived estimates of non-Federal forest land area from draft 2002 RPA data and 1997 NRI data were closer to each other than to image-derived estimates, with the exception of the Forest Cover Types map. Second, compared to both RPA and NRI, estimates of forest land area derived from FC resulted in the smallest image-derived RMSD, both for non-Federal lands and for all land ownerships. Third, a combination of four land cover classes (deciduous forest, evergreen forest, mixed forest, and woody wetland) from the 1992 National Land Cover Data set resulted in an estimate of CONUS forest land that was similar to the RPA estimate and had an acceptably low RMSD. Addition of “transitional” class resulted in a slightly larger estimate, and also may account for differences between forest land use (e.g., RPA) and forest land cover (e.g., NLCD) by including forest clearcuts and other areas of forest regeneration not usually recognized by satellite imagery as forest cover. Fourth, thresholds of VCF percent tree canopy data can be selected that produce estimates comparable with plot-based estimates at either State or national levels. A single threshold based on nationwide or CONUS-wide plot-based estimates, however, is inappropriate for obtaining estimates within smaller geographic areas, e.g., States. Finally, multiple satellite image-derived land cover maps with a variety of characteristics (date of imagery, classification scheme, spatial resolution, etc.) show potential for complementing U.S. forest resource assessments.

Literature Cited

- Cochran, W. 1977. Sampling techniques, 3rd edition. New York: John Wiley & Sons. 428 p.
- Czaplewski, R.L.; Rack, J.; Lessard, V.C.; *et al.* 2002. Coordination, cooperation, and collaboration between FIA and NRI. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; McWilliams, W.H.; Cieszewski, C.J., eds. Proceedings, 4th annual forest inventory and analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC–252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 141–148.
- DellaSala, D.A.; Staus, N.L.; Strittholt, J.R.; Hackman, A.; Iacobelli, A. 2001. An updated protected areas database for the United States and Canada. *Natural Areas Journal*. 21(2): 124–135.
- Environmental Systems Research Institute, Inc. (ESRI). ArcGIS, ver. 8.1. Redlands, CA: ESRI.
- Environmental Systems Research Institute, Inc. (ESRI). 2002. ESRI data & maps 2002. CD-ROM. Redlands, CA: ESRI.
- Food and Agriculture Organization (FAO) of the United Nations. 2001. Global forest resources assessment 2000: main report. FAO Forestry Paper; 0258-6150; 140. Rome, Italy: FAO. 479 p.
- Fuller, W.A.; Kennedy W.; Schnell, D.G.S.; Park, H.J. 1986. PC CARP®. Ames, IA: Iowa State University, Statistical Laboratory.
- Häme, T.; Stenberg, P.; Andersson, K.; *et al.* 2001. AVHRR-based forest proportion map of the Pan-European area. *Remote Sensing of Environment*. 77(1): 76–91.
- Hansen, M.C.; DeFries, R.S.; Townshend, J.R.G.; *et al.* 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment*. 83: 303–319.

- Hansen, M.H. Unpublished report. State total estimates from FIADB with reported and estimated sampling errors from FIA state reports. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station, Forest Inventory and Analysis.
- Hansen, M.H. 2001. Remote sensing precision requirements for FIA estimation. In: Reams, G.A.; McRoberts, R.E.; Van Deusen, P.C., eds. Proceedings, 2nd annual forest inventory and analysis symposium; 2000 October 17–18 ; Salt Lake City, UT. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 43–51.
- Kennedy, P.; Bertolo, F. 2002. Mapping sub-pixel forest cover in Europe using AVHRR data and national and regional statistics. *Canadian Journal of Remote Sensing*. 28(2): 302–331.
- Lessard, V.C.; Hansen, M.H.; Nelson, M.D. 2003. Comparing Minnesota land cover/use area estimates using NRI and FIA data. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; Moser, J.W., eds. Proceedings, 3rd annual forest inventory and analysis symposium; 2001 October 17–19; Traverse City, MI. Gen. Tech. Rep. NC-230. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 36–43.
- Loveland, T.R.; Reed, B.C.; Brown, J.F.; Ohlen, D.O.; *et al.* 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*. 21(6/7): 1303–1330.
- Owens, K.E. 2001. Development and analysis of current and presettlement forest cover mapping methods for the eastern Upper Peninsula of Michigan, Luce District. Houghton, MI: Michigan Technological University. 215 p.
- Päivinen, R.; Lehtikoinen, M.; Schuck, A.; *et al.* 2001. Combining earth observation data and forest statistics. Joensuu, Finland: Research Report 14, European Forest Institute, Joint Research Center-European Commission. 101 p.
- Powell, D.S.; Faulkner, J.L.; Darr, D.R.; Zhu, Z.; MacCleery, D.W. 1993. Forest resources of the United States, 1992. Gen. Tech. Rep. RM-234. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 132 p.
- Särndal, C.E.; Swensson, B.; Wretman, J. 1992. Model assisted survey sampling. New York: Springer-Verlag. 694 p.
- Schuck, A.; Päivinen, R.; Häme, T.; *et al.* 2003. Compilation of a European forest map from Portugal to the Ural Mountains based on earth observation data and forest statistics. *Forest Policy and Economics*. 5(2): 187–202.
- Smith, B.W.; Vissage, J.S.; Darr, D.R.; Sheffield, R.M. 2001. Forest resources of the United States, 1997. Gen. Tech. Rep. NC-219. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 190 p.
- U.S. Department of Agriculture (USDA) Forest Service. 2003. Forest inventory and analysis national core field guide, vol. 1: field data collection procedures for phase 2 plots, ver. 2.0. USDA Forest Service, Washington Office, internal report. On file with: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, 201 14th St., Washington, DC 20250. 281 p.
- Vogelmann, J.E.; Howard, S.M.; Yang, L.; *et al.* 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering & Remote Sensing*. 67(6): 650–662.
- Zawila-Niedzwiecki, T., ed. 2000. Remote sensing and forest monitoring. Warsaw, Poland: Institute of Geodesy and Cartography. 101 p.
- Zhu, Z. 1994. Forest density mapping in the lower 48 States: a regression procedure. Res. Pap. SO-280. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 11 p.
- Zhu, Z.; Evans, D.L. 1994. U.S. forest types and predicted percent forest cover from AVHRR data. *Photogrammetric Engineering & Remote Sensing*. 60(5): 525–531.



Accuracy Assessment of FIA's Nationwide Biomass Mapping Products: Results From the North Central FIA Region

Geoffrey R. Holden, Mark D. Nelson, and Ronald E. McRoberts¹

Abstract.—The Remote Sensing Band of the Forest Inventory and Analysis (FIA) program has developed a nationwide map of forest biomass to be distributed as a geospatial raster data set with 250-meter spatial resolution. The accuracy of the forest biomass map depends on both an intermediate forest/nonforest classification and the biomass estimation. For the North Central FIA region, we assessed the accuracy of the forest classification and biomass estimation using three approaches: (1) per pixel percent correctly classified; (2) comparisons of pixel- and FIA plot-based estimations within delineated areas; and (3) utility for stratified estimation of FIA attributes. Results showed the forest/nonforest map to be accurate based on all three assessments, while the forest biomass map performed well only for area estimates.

Introduction

The mission of the Forest Inventory and Analysis (FIA) program is to provide statistical information about America's forests. In the past, much of this information was summarized in statistical reports covering large areas, which offered little insight into describing where forest attributes occur on the ground. With improvements in geospatial technologies and the increased availability of remotely sensed imagery, developing maps of predicted forest attributes across landscapes is now feasible. During the summer of 2003, the Remote Sensing Band (RSB)—an interregional working group of the FIA program—developed a nationwide forest biomass map. The purpose of this mapping effort was to predict forest biomass across the Nation by combining FIA measurements with remotely sensed imagery and

other raster data sets. This effort produced two mapping products: a forest/nonforest (FNF) map and the forest biomass map itself. Although this mapping effort is notable, its utility depends on the accuracy of the resulting products. In this article, we report and discuss results related to our validation of the portions of the FNF and biomass maps in the North Central FIA (NCFIA) region. We assessed accuracy using the following approaches: (1) per pixel correctly classified, (2) comparisons of pixel-based and FIA plot-based estimates within delineated areas, and (3) utility for stratified estimation of FIA attributes.

Biomass Mapping Effort

Forest/nonforest classes and continuous estimates of biomass were mapped at a 250-meter spatial resolution across the continental United States, Alaska, and Puerto Rico. FIA attributes were integrated with spectral information from a variety of data layers collectively referred to as the national geospatial predictor data set. These data layers included Moderate Resolution Imaging Spectroradiometer (MODIS) reflectance data; MODIS derivatives such as Enhanced Vegetation Index, Normalized Difference Vegetation Index, and Vegetation Continuous Fields (Hansen *et al.* 2002); derivatives of the National Land Cover Data of 1992 (NLCD92), a 30-meter spatial resolution land cover product of the Multi-Resolution Land Characterization Consortium derived from nominal 1992 Landsat Thematic Mapper (TM) imagery (Vogelmann *et al.* 2001); and elevation, precipitation, and temperature data.

Software from a suite of data mining packages developed by RuleQuest Research was used for creating models specific to mapping zones developed for the National Land Cover Data of 2000 (Homer *et al.* 2002). See5, a data mining software package, was used for modeling forest land by combining FNF data observed on FIA plots with corresponding information from the predictor data set to produce decision trees or rule sets (RuleQuest Research 1997). Using a custom software interface tool developed by the Forest Service's Remote Sensing

¹ Geographic Information System (GIS) Analyst, GIS Analyst, and Mathematical Statistician, respectively, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108. Phone: 651-649-5149; fax: 651-649-5285; e-mail: gholden@fs.fed.us.

Applications Center (RSAC), we incorporated these decision trees or rule sets with ERDAS IMAGINE image processing software to assign each 250-meter pixel to either a forest or nonforest class.

Modeling and mapping of biomass was performed in a similar manner, but relied on Cubist, RuleQuest Research's data mining software package for deriving predictive models of continuous response variables (RuleQuest Research 1997). The Cubist rule-based linear models were used for predicting gross biomass of all live trees (oven-dry tons/acre). Again, the models were brought into ERDAS IMAGINE using the custom-built RSAC tool to map the biomass predictions for each pixel. Because the model was predicting forest biomass values, biomass was mapped only on forest land, and the FNF map was used to mask appropriate nonforest areas.

Accuracy Assessment

Per Pixel

FNF Map

Because the FNF map included discrete classes, the accuracy assessment for the FNF map used only FIA plots that were both single condition (either 100-percent forested or 100-percent nonforested) and independent of the plots used to train the model (10 percent were randomly selected from the total plots available). For pixels containing these "test" plot locations, predicted forest or nonforest classes were extracted and compared to observed forest or nonforest assignments reported for those specific plots. Results were tabulated in a confusion matrix. Originally, assessments were to be performed for each State. Based on a rule of thumb proposed by Congalton and Green (1999), which suggests at least 50 sample points per class for accuracy assessment, however, some States were combined into larger areas. One of these aggregations included Iowa (IA) and the Plain States (PS): North Dakota (ND), South Dakota (SD), Nebraska (NE), and Kansas (KS). Indiana (IN) and Illinois (IL) were also combined into a single area. Per State assessments were performed for the remaining four States in NCFIA: Michigan (MI), Minnesota (MN), Missouri (MO), and Wisconsin (WI). In addition, a nationwide assessment was conducted.

Four accuracy measures were calculated from the confusion matrix: (1) overall accuracy, (2) producer's accuracy, (3) user's accuracy, and (4) Kappa. Overall accuracy, a measure of the accuracy across all classes, is calculated as the ratio of the total number of correctly classified reference points to the total number of reference points. Producer's accuracy, which describes how well the classification matched the reference data for each class individually, is calculated as the ratio of the number of correctly classified reference points for a class to the total number of reference points for that class. User's accuracy, also referred to as mapping accuracy, measures the likelihood that an area assigned to a class will actually be that class when visited in the field. This number is calculated as the ratio of the number of correctly classified reference points for a class to the total of reference points classified as that class. The Kappa, or KHAT statistic, measures the percent improvement the classification has over a classification based purely on a random assignment of pixels to classes (Congalton and Green 1999, Jensen 1996, Lillesand and Kiefer 2000).

Forest Biomass

Cubist, the data mining software used to model biomass, produces three measures for assessing model accuracy. These measures—average error, relative error, and correlation coefficient—are based on all model predictions. Average error is calculated as the average absolute difference between observed and predicted values. Relative error is the ratio of the average error to the average absolute difference between observed values and the mean value of the training observations. A relative error value near to or greater than 1 indicates little or no improvement by the model over the assignment of the mean of the observed values to each case. The correlation coefficient (r) describes the linear relationship between observed and predicted values. These three measures were also recalculated for the independent validation data set, using only plots with 100-percent forest condition and located in pixels classified as "forest" in the FNF map. Assessment was performed on the nationwide map.

Plot- Versus Pixel-Based Area Estimates

Although per pixel accuracy assessments are important for validating maps at the local level, area estimates can provide an indication of correctness for larger areas. Reese *et al.* (2002)

reported that using a k-Nearest Neighbor (kNN) algorithm with SPOT 3 and Landsat TM/Enhanced Thematic Mapper Plus data performed poorly at the pixel level but produced acceptable results for larger area estimates of wood volume and biomass in Sweden's forests. For our study, we assessed accuracy of area estimates by comparing FIA plot-based estimates to RSB map-based estimates. This comparison, relative difference (RD), was calculated as:

$$RD = \frac{(\text{predicted} - \text{observed})}{\text{observed}}$$

Because FIA plot-based estimates are based on a sample, 95-percent confidence limits were calculated for each estimate and compared to the RSB map-based estimates.

FNF Map

Total and forested pixel counts were determined for each State in the 11-State NCFIA region. Proportion forest land area for a State was computed as the ratio of the number of forested pixels in a State to the total number of pixels in that State. A consistent intensity of FIA plots in a State was used to calculate estimates of plot-based proportion forest area based on the assumption of simple random sampling (SRS). For comparison to another map product, estimates of statewide proportion forest land were calculated using the NLCD92. The NLCD92 data set consists of 21 land cover classes. We aggregated these classes into forest and nonforest, as recommended by McRoberts *et al.* (2002), and computed statewide proportion forest land, as was done for the RSB FNF map. In addition, RD was calculated to compare plot-based estimates to NLCD92-based estimates.

Forest Biomass

Pixel-based area estimates for forest biomass were calculated in much the same way as forest area estimates. The mean biomass in tons/acre was calculated for an area using all pixels classed as forest in the RSB FNF map in the area. Plot-based estimates included only plots that had a forested condition.

Stratified Estimation

NCFIA currently uses a stratified estimation (post-sampling stratification) approach that incorporates satellite-imagery-derived land cover classification data for improving the precision of FIA estimates (McRoberts *et al.* 2002). According to Hansen

(2001), "The sampling errors of the area estimates are very dependent on the quality of the stratification" (Hansen 2001, 45). We assessed the accuracies of the mapping products by examining their utility for stratified estimation. This was accomplished by comparing the precision of a stratified estimate using map-based strata to the precision of estimates based on SRS.

For SRS, plot-based estimates calculated for the area estimate comparisons discussed above were used. The variance estimates were then calculated to determine the precision of each SRS estimate. For stratified estimation, the same plots used in the SRS technique were assigned to strata based on overlays of the stratification layers. Again, the estimate, $\bar{\rho}$, and associated variance estimate, $\text{Var}(\bar{\rho})$, were calculated, but this time Cochran's (1977) formulas for stratified analyses were used:

$$\bar{\rho} = A \sum_{j=1}^J W_j \hat{P}_j$$

and

$$\text{Var}(\bar{\rho}) = A^2 \sum_{j=1}^J W_j^2 \hat{\sigma}_j^2 / n_j$$

where:

A is total area,

$j=1, \dots, J$ denotes stratum,

W_j is the weight for the j th stratum calculated as the ratio of the number of pixels assigned to the j th stratum to the total number of pixels for all strata,

\hat{P}_j denotes the mean proportion forest land for plots assigned to the j th stratum,

$\hat{\sigma}_j^2$ is the within-stratum variance for the j th stratum calculated as:

$$\hat{\sigma}_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (P_{ij} - \hat{P}_j)^2$$

where:

P_{ij} is the proportion forest land observed by the field crew for the i th plot in the j th stratum, and

n_j is the number of plots assigned to the j th stratum.

FIA precision estimates are scaled to a constant area to report precisions for estimation units of varying sizes. These precisions, as a percent estimate per million acres of forest land or a percent per billion tons of biomass, were calculated as follows:

$$PREC = \frac{(\text{sampling error}) \sqrt{\text{estimated attribute}}}{\sqrt{\text{standard}}}$$

Relative efficiency (RE) of the stratified estimation precision to the SRS precision is calculated as:

$$RE = \left(\frac{SE(SRS)}{SE(\text{stratified estimation})} \right)^2$$

FNF Map

Using a consistent intensity of FIA plots from the first annual inventory, a SRS estimate of forest land area was calculated for each State. The FNF map was used as a stratification layer to compute a stratified estimate of forest land area. Two stratification layers were derived from this map. The first was a two-stratum layer created using the forest and nonforest classes directly. The second was a four-stratum layer that included two interior classes, forest and nonforest, and two edge classes, one located adjacent to forest interior and the other adjacent to nonforest interior. The four-strata layer required manipulation of the FNF map including (1) a division of each 250-meter pixel into 25 50-meter pixels, and (2) a search and recode operation to reassign the first tier of 50-meter pixels adjacent to each interior class located along the transition zone to its respective edge class. This same procedure was repeated for the entire 11-State FIA region, using FIA plots from fiscal year (FY) 2001, the only year for which annual plot data are available across all 11 States.

Forest Biomass

Per area estimates of forest biomass were computed for the entire NCFIA region. Plots from a single year of the annual inventory FY 2001 were used. To create the stratification layer, the continuous biomass values were aggregated into three biomass

classes representing low, medium, and high biomass. Class breaks were based on those proposed for distribution in the RSB national map: less than 30 tons/acre (low), 30–50 tons/acre (medium), and greater than 50 tons/acre (high). These class breaks are comparable to one standard deviation of the predicted data in the NCFIA region. A per pixel accuracy assessment of these biomass classes was performed using observed values from FIA plot data grouped in the same classes as test data.

Results

FNF Map

Per Pixel

Table 1 shows a summary of the results for the per pixel correctly classified assessment of the RSB FNF map. Overall accuracies for all areas were good, exceeding the 85-percent minimum described as acceptable by Anderson (1971) for land cover classifications. For many areas, overall accuracy exceeded 90 percent. When compared to the reference data (producer's accuracy), the nonforest class had relatively high values (88 percent or higher) in all analysis areas. The accuracy of the forest class tended to reflect the abundance of forest land in that area, because classifications of sparsely forested areas (PS and IA) performed poorly, while classifications of more heavily forested areas (MI, MN, MO, and WI) performed better. The mapping accuracy (user's accuracy) exhibited a similar trend. Except for the PS and IA area (Kappa = 43 percent), the classifications performed better than a random assignment of pixels to classes (Kappa > 70 percent).

Table 1.—*Confusion matrix summary for the FNF map based on a per pixel accuracy assessment.*

Analysis area	Overall accuracy (%)	Producer's accuracy (%)		User's accuracy (%)		Kappa (%)
		Forest	Nonforest	Forest	Nonforest	
MI	91	92	89	90	91	81
MN	89	91	88	84	93	78
MO	88	77	94	88	88	73
IL and IN	95	82	97	77	98	77
PS and IA	98	40	99	48	99	43
Region	94	84	96	86	96	81

Table 2.—Comparison of proportion forest area estimates.

Analysis area	Proportion forest land			Relative difference (%) to plot-based	
	RSB map-based	NLCD92 map-based	Annual plot-based	RSB map-based	NLCD92 map-based
IL	0.09	0.15	0.11	-25	31
IN	0.14	0.20	0.19	-25	7
IA	0.04	0.08	0.07	-35	19
KS	0.02	0.04	0.04	-49	-13
MI	0.51	0.56	0.52	-3	7
MN	0.29	0.36	0.29	0	22
MO	0.28	0.39	0.32	-11	21
NE	0.01	0.02	0.03	-76	-23
ND	0.01	0.04	0.02	-67	144
SD	0.03	0.06	0.03	-12	62
WI	0.41	0.47	0.44	-7	6
Region	0.16	0.20	0.18	-11	11

Table 3.—Comparison of RSB map-based forest proportion to plot-based 95-percent confidence limits (C. I.)

Analysis area	RSB map-based	Plot-based 95-percent C.I.	
		Lower	Upper
IL	0.09	0.10	0.13
IN	0.14	0.18	0.20
IA	0.04	0.06	0.08
KS	0.02	0.04	0.05
MI	0.51	0.51	0.54
MN	0.29	0.28	0.30
MO	0.28	0.31	0.33
NE	0.01	0.02	0.03
ND	0.01	0.01	0.02
SD	0.03	0.03	0.04
WI	0.41	0.43	0.46
Region	0.16	0.17	0.19

Plot- Versus Pixel-based Per Area Estimates

Table 2 shows the RSB map-based, NLCD92 map-based, and plot-based proportion forest area estimates. Compared to the plot-based estimates, the RSB map-based forest land estimates underestimate all areas except MN. For MN, the estimates were within 1 percent. The relative difference between the plot- and RSB map-based estimates tended to reflect the abundance of forest land calculated for an area, so that more heavily forested areas had closer estimates, and more sparsely forested areas

had greater differences between estimates. Conversely, the NLCD92 map-based estimates tended to overestimate forest area when compared to the plot-based estimates except for two sparsely forest areas, KS and NE. For most States/areas, the RSB map-based estimates were below the range of the 95-percent confidence limits for the plot-based estimates (table 3).

Stratified Estimation

Table 4 reports the forest area estimates in thousands of acres for the SRS, RSB map two- and four-class stratifications, and NCFIA's reported estimates for FY 2002 that use an NLCD92-derived classification to assign plots to strata. As was expected, the estimates were all comparable with stratified estimates and within a few percent of the SRS estimates. The RSB FNF map improved the precision of the estimates over the SRS for all States and the entire NCFIA region. Table 5 lists these precisions and the relative efficiencies of the stratified estimates. Because an improvement in precision for most States was obtained using the four-class stratification over the two-class stratification, the relative efficiency was calculated only for the four-class stratification. Sparsely forested States such as NE, ND, IA, and KS showed a slight improvement in precision over the SRS (relative efficiency was less than 1.5), while heavily forested States, such as MI, MN, MO, and WI, showed a greater improvement, which is consistent with the poorer accuracies

seen with the other two assessment methods described above. Also, increased sampling intensity in these four heavily forested States most likely contributed to this improvement. Stratified estimation using the NLCD92-based stratification layer

achieved higher precisions for all analysis areas except one, SD (relative efficiency of 2.14 compared to 3.72 when the RSB FNF map was used as a stratification layer).

Forest Biomass

Per Pixel (Continuous)

Average error of the Cubist biomass model predictions was about 20 tons/acre, meaning that, on average, the predicted value differed from the observed value by plus or minus 20 tons/acre. The relative error was 0.88. Because values nearing 1.0 indicate little improvement of the predicted values over the assignment of the observed mean to all cases, this result might indicate that the modeling tended to predict values close to the mean. The correlation coefficient was calculated as 0.4. Figure 1 is a scatter diagram of the predicted biomass values plotted against the observed biomass values with a “least squares” trendline and r^2 statistic. Initial observations show a positive correlation between the observed and predicted values, but the range of the predicted values (10–60 tons/acre) did not match that of the observed value range (0–286 tons/acre). Correlation was poor with $r^2 = 0.16$.

Table 4.—Estimated forest area (1,000 acres) based on an assumption of simple random sampling (SRS) and stratified estimation.

Analysis area	SRS	Stratified estimation		
		RSB map two-class	RSB map four-class	NCFIA four-class
IL	4,097	4,180	4,275	4,260
IN	4,359	4,216	4,223	4,545
IA	2,492	2,524	2,567	2,699
KS	2,172	2,171	2,175	2,229
MI	19,618	19,752	19,710	19,349
MN	15,922	16,394	16,416	16,353
MO	14,231	14,544	14,570	14,464
NE	1,355	1,443	1,449	1,364
ND	744	731	716	823
SD	1,700	1,702	1,705	1,714
WI	15,894	15,982	15,941	16,016
Region	82,390	81,755	81,924	83,815

Table 5.—Precisions of forest area estimates as a percent per million acres and relative efficiency of map-based precisions over precisions based on the assumption of simple random sampling (SRS).

Analysis area	FIA precision ^a				Relative efficiency	
	Simple random sampling	RSB map two-class	RSB map four-class	NLCD92 four-class	RSB map four-class	NLCD92 four-class
IL	6.74	4.96	4.82	3.98	1.95	2.88
IN	6.48	4.60	4.38	3.66	2.19	3.13
IA	6.92	5.88	5.87	4.39	1.39	2.48
KS	6.83	6.50	6.44	5.47	1.13	1.56
MI	2.40	1.45	1.43	1.34	2.82	3.21
MN	4.75	2.89	2.84	2.57	2.80	3.41
MO	5.56	4.11	3.96	3.16	1.97	3.08
NE	6.95	6.28	6.31	5.97	1.21	1.36
ND	7.02	6.42	6.42	6.12	1.19	1.32
SD	7.37	3.88	3.82	5.04	3.72	2.14
WI	3.87	2.64	2.58	2.19	2.25	3.14
Region	5.83	3.60	3.56	N/A	2.69	N/A

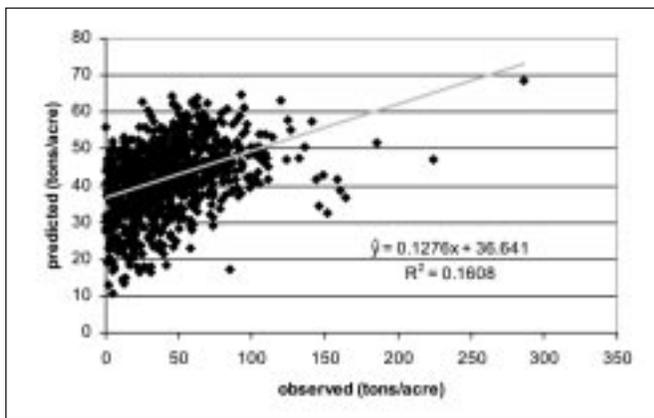
^a Calculated as a percent per million acres of forest land.
N/A = not available.

Per Pixel (Classed)

Overall accuracy for the classed biomass map was poor with only about 40 percent of the reference points correctly classified. The error matrix in figure 2 shows that the majority of the error occurred in the low and high biomass classes because of

a bias in the map toward the medium biomass class. Although the reference data included an almost equal number of points observed in each class, the classification showed the majority of the reference data (about 69 percent) predicted in the medium class. This bias is evident in the producer's accuracies of 73 percent for the medium class and 24 and 32 percent, respectively, for the low and high classes. The Kappa statistic of 14 percent indicates that the classification was not much of an improvement over a random assignment of classes to pixels.

Figure 1.—Scatter diagram of observed biomass values versus predicted biomass values.



Plot- Versus Pixel-based Per Area Estimates

Table 6 provides average biomass estimates by State and region based on the RSB biomass map and the FIA plot data. For many States and regionwide, the predicted average biomass per acre was within 10 percent of the observed average biomass, with estimates for two States, MO and NE, within less than 1 percent. Many of the pixel-based estimates were within the 95-percent confidence limits of plot-based estimates. Exceptions included areas with RD greater than 0.04.

Figure 2.—Confusion matrix for classed biomass map.

Classification	Reference			Total	Producer's Accuracy	User's Accuracy
	Low	Medium	High			
Low	87	32	19	138	23.7%	63.0%
Medium	259	212	215	686	72.9%	30.9%
High	21	47	110	178	32.0%	61.8%
Total	367	291	344	1002		

Table 6.—Pixel- versus plot-based estimates of average biomass per acre by analysis area.

Analysis area	Mean biomass (tons/acre)		Relative difference (%)	Plot-based 95 percent C.I.	
	RSB map	Plot-based		Lower	Upper
IL	50.66	52.34	-3	48.17	56.51
IN	53.37	60.76	-12	57.68	63.84
IA	45.01	46.06	-2	42.09	50.03
KS	39.84	41.04	-3	35.89	46.19
MI	45.66	50.16	-9	48.21	52.11
MO	45.07	45.29	0	44.06	46.53
NE	36.92	36.77	0	30.23	43.31
ND	29.24	28.62	2	21.52	35.72
SD	25.94	21.79	19	18.91	24.68
WI	43.81	45.18	-3	43.60	46.77
Region	42.43	44.87	-5	43.64	46.10

Stratified Estimation

Estimates of total biomass for the entire region were comparable between the SRS (3.73 billion tons) and stratified estimation (3.35 billion tons). Using biomass classes for stratification did little to improve the precision of total biomass stratified estimation with an RE of 1.08.

Conclusions

The 11-State NCFIA portion of the RSB FNF map exhibited favorable results for all three measures of accuracy assessment used in this study. The map appears to have a consistent negative bias, however, relative to plot-based estimates. Bias was smaller in more heavily forested States and larger in more sparsely forested States. In contrast, the 30-meter NLCD92 FNF map was positively biased relative to plot-based estimates, with magnitude of bias being less dependent on the amount of forest land within a State. The coarser spatial resolution of the RSB map (250-meter) versus the NLCD92 map (30-meter) may partially account for the different bias trends in these two image-based maps. Mayaux and Lambin (1995) related bias in image-based estimates to “effects of spatial aggregation.”

Although sums of biomass pixel values produced unbiased estimates of statewide forest biomass for most NCFIA States, individual pixel predictions tended to underestimate high biomass and overestimate low biomass with high correlations occurring in the middle values, similar to observations made by Reese *et al.* (2002). Given this observation, the minimal gain in precision over SRS estimates achieved when using biomass classes as strata for stratified estimation is not surprising. Additional work is needed to determine spatial scales at which the RSB FNF and biomass maps are biased or unbiased and compare precision of estimates derived from these maps to estimates derived from the FIA inventory.

Literature Cited

- Anderson, J.R. 1971. Land use classification schemes used in selected recent geographic applications of remote sensing. *Photogrammetric Engineering & Remote Sensing*. 37(4): 379–387.
- Cochran, W.G. 1977. *Sampling techniques*, 3rd edition. New York: John Wiley & Sons. 480 p.
- Congalton, R.G.; Green, K. 1999. *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton, FL: CRC Press. 137 p.
- Hansen, M.C.; DeFries, R.; Townshend, J.R.G.; *et al.* 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS. *Remote Sensing of Environment*. 83: 303–319.
- Hansen, M.H. 2001. Remote sensing precision requirements for FIA estimation. In: Reams, G.A.; McRoberts, R.E.; Van Deusen, P.C., eds. *Proceedings, the second annual forest inventory and analysis symposium; 2000 October 17–18; Salt Lake City, UT*. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 43–51.
- Homer, C.G.; Huang, C.; Yang, L.; Wylie, B. 2002. Development of a circa 2000 landcover database for the United States. In: *American Society for Photogrammetry and Remote Sensing Proceedings; April 2002; Washington, DC*.
- Jensen, J.R. 1996. *Introductory digital image processing: a remote sensing perspective*, 2nd edition. Upper Saddle River, NJ: Prentice Hall. 318 p.
- Lillesand, T.M.; Kiefer, R.W. 2000. *Remote sensing and image interpretation*, 4th edition. New York: John Wiley & Sons. 724 p.
- Mayaux, P.; Lambin, E.F. 1995. Estimation of tropical forest area from coarse spatial resolution data: a two-step correction function for proportional errors due to spatial aggregation. *Remote Sensing of Environment*. 53: 1–15.

McRoberts, R.E.; Wendt, D.G.; Nelson, M.D.; Hansen, M.H. 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment*. 81: 36–44.

Reese, H.; Nilsson, M.; Sandstrom, P.; Olsson, H. 2002. Applications using estimates of forest parameters derived from satellite and forest inventory data. *Computers and Electronics in Agriculture*. 37: 37–55.

RuleQuest Research. 1997. RuleQuest research data mining tools. <http://www.rulequest.com>. [Date accessed unknown.]

Vogelmann, J.E.; Howard, S.M.; Yang, L.; *et al.* 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering & Remote Sensing*. 67: 650–662.



A Proposal for Phase 4 of the Forest Inventory and Analysis Program

Ronald E. McRoberts¹

Abstract.—Maps of forest cover were constructed using observations from forest inventory plots, Landsat Thematic Mapper satellite imagery, and a logistic regression model. Estimates of mean proportion forest area and the variance of the mean were calculated for circular study areas with radii ranging from 1 km to 15 km. The spatial correlation among pixel predictions was incorporated into the variance calculations. The map-based estimates were compared to estimates obtained using only plot data for the same circular areas. For three circular study areas in Minnesota, the map-based estimates were similar to the plot-based estimates and more precise.

The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture (USDA) Forest Service surveys the Nation's forested lands using a combination of field plot measurements and remotely sensed data. Traditionally, the FIA program has used data from these inventories to respond to the "How Much?" question by reporting plot-based estimates of forest attributes by county or combinations of counties. The estimates are obtained using a three-phase approach: Phase 1 consists of using remotely sensed data, often satellite imagery, to stratify the land area of interest for increasing the precision of estimates, and, optionally, to determine if a plot has accessible forest land; Phase 2 consists of measuring a suite of mensurational variables for plots with accessible forest land; and Phase 3 consists of measuring a suite of variables related to forest health for a 1:16 subset of the Phase 2 plots.

Increasingly, users are also asking "Where?" and are requesting access to plot data for estimating for their own areas of interest (AOI), frequently for use as training or accuracy assessment data for spatial products. Many of these applications require the coordinates of plot locations to determine if a plot is located in the AOI. Although forest inventory programs generally release plot information to the public, many resist releasing

plot locations. First, release of plot locations may entice users to visit plots for additional data, which could artificially disturb the ecology of the sites and contribute to bias in inventory estimates. Second, many forest inventory programs rely on the goodwill of private forest landowners for permission to observe plots on their land. Landowners generally do not welcome unwarranted or frequent intrusions and often only permit visits by inventory crews contingent on assurances that plot locations and proprietary information will not be released.

In response to the "Where?" question, the FIA program has initiated local, regional, and national mapping efforts. Although the objectives of these efforts have been to map the spatial distribution of forest attributes, generally they have not included investigations of whether the maps may be used to produce unbiased and precise estimates of forest attributes. For the latter objective, plot-based estimation using data for plots located in the AOI has been necessary. However, if unbiased and sufficiently precise estimates of forest attributes could be obtained from maps, then several advantages would accrue. First, release of plot locations would be unnecessary for estimation for user AOIs. Second, because mapped values for individual mapping units would be based on aggregated data from multiple plots, proprietary information would not be released. Third, estimation would be possible for small areas for which the number of plots is insufficient for plot-based estimation. Fourth, efficiencies would be gained by simultaneously addressing both the "How Much?" and "Where?" questions.

The objective of the study was to investigate estimation of forest attributes from maps constructed using inventory plot data and satellite imagery. In this context, FIA Phase 4 is defined as the construction of forest attribute maps that satisfy two criteria: (1) estimates at all spatial scales are within tolerance limits relative to plot-based estimates, and (2) estimates of the precision of the map-based estimates are comparable to the precision of plot-based estimates. For this study, the terms plot-based and map-based estimates are considered equivalent to the terms design-based and model-based estimates, respectively. Although the second set of terms may be regarded as more correct from a

¹ Mathematical Statistician, U.S. Department of Agriculture, Forest Service, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108.

statistical perspective, the terms of the first set are generally used for this study because they are considered more descriptive. The study focused on constructing maps depicting pixel-level probabilities of forest ground cover using inventory plot data, Landsat Thematic Mapper (TM) satellite imagery, and a logistic regression model, and then comparing the map-based and plot-based estimates of forest land area.

Data

FIA Plot Data

Under the FIA program's annual inventory system (McRoberts 1999), field plot centers are established in permanent locations using a systematic sampling design that is assumed to produce a random, equal probability sample. In each State, a fixed proportion of plots is measured annually. For the FIA program of the North Central Research Station, plots measured in a single federal fiscal year (e.g., FY 2004: 1 October 2003 to 30 September 2004) comprise a single panel of plots with panels selected for annual measurement on a 5-year rotating basis. In aggregate, over a complete 5-year measurement cycle, a plot represents approximately 2,400 ha (slightly less than 6,000 ac). In general, locations of forested or previously forested plots are determined using global position system receivers, while locations of non-forested plots are determined using digitization methods.

Each field plot consists of four 7.31-m (24-foot) radius circular subplots. The subplots are configured as a central subplot and three peripheral subplots with centers located at 36.58 m (120 ft) and azimuths of 0°, 120°, and 240° from the center of the central subplot. Among the observations field crews obtain are the proportions of subplot areas that satisfy specific ground land use conditions. Subplot estimates of forest land proportion are obtained by aggregating ground land use conditions into forest and non-forest classes consistent with the FIA definition of forest land, and plot-level estimates are obtained as means over the four subplots. For this study, all plots were observed or measured between the beginning of FY1999 and the end of FY2002. Although the locations of the central subplots are considered a random, equal probability sample, the fixed spatial configuration of the peripheral subplots with respect to the central subplots requires accommodation of spatial correlation among subplot observations for subplot-level analyses.

Satellite Imagery

Landsat TM imagery for scenes Row 27 and Row 28 of Path 27 was obtained from the Multi-Resolution Characterization 2001 land cover mapping project (Homer *et al.*, in press) of the U.S. Geological Survey. The imagery was characterized by several salient features: (1) a combination of Landsat 5 and Landsat 7 Enhanced TM+ data, (2) three dates including early, peak, and late vegetation green-up (Yang *et al.*, 2001), (3) geometrically and radiometrically corrected, (4) cubic convolution resampling to 30 m x 30 m spatial resolution, (5) visible and infrared bands (1-5, 7), and (6) conversion to at-satellite reflectance.

For predicting the probability of forest cover, the satellite image spectral data were used in two forms: the 18 raw spectral band data and 12 transformations including normalized difference vegetation index (NDVI) and the tree tassel cap transformations (brightness, greenness, and wetness) for each image date (Kauth and Thomas 1976). The raw band data and the transformations were evaluated separately.

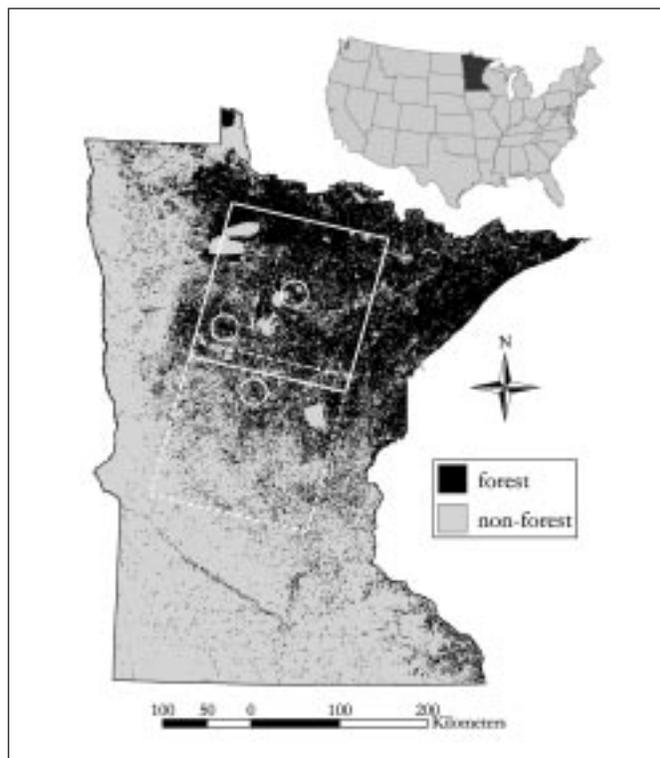
Combining FIA Data and Satellite Imagery

The spatial configuration of the FIA subplots with centers separated by 36.58 m and the 30 m x 30 m spatial resolution of the TM imagery permits individual subplots to be associated with individual image pixels. Further, the subplot area of 167.87 m² is approximately 19 percent of the 900 m² pixel area and is generally deemed an adequate sample of the ground characteristics of the pixel area. However, when describing relationships between forest attributes and satellite image spectral values, two phenomena must be considered. First, because a subplot is a single, 19-percent, contiguous sample of the pixel area, the proportion forest for a partially forested subplot may not accurately represent the proportion forest for the entire pixel. Second, Global Positioning System and image registration errors may cause a subplot to be associated with an incorrect pixel, resulting in the forest attribute of a subplot being erroneously associated with the spectral signature of a non-forested pixel, and vice versa. Both phenomena obscure relationships between observed forest attributes and spectral values, cause bias in estimates of parameters of models of the relationships, increase both model residual variability and the uncertainty in model parameter covariance estimates, and contribute to increasing the variance of map-based estimates of forest attributes. To avoid these phe-

nomena, when fitting models data were excluded for plots with a mixture of forest and non-forest cover. An important result of the exclusion of these data was that proportion forest area, Y , was a binary variable. This variable was quantified according to the convention that $Y = 1$ denoted a forested subplot and $Y = 0$ denoted a non-forested subplot.

Within the two TM scenes, three 15-km radius circular study areas were arbitrarily selected to represent a spectrum of forest/non-forest ground cover conditions (fig. 1). Because the relationship between spectral values and the forest/non-forest composition of ground cover was expected to be approximately constant over the scenes, data for calibrating models were combined for the three study areas. In addition, spatial variability was evaluated for the three study areas collectively under the assumptions that spatial correlation is stationary (i.e., it does not change within the scene) and that spatial correlation is isotropic (i.e., it is the same in all directions).

Figure 1.—Study areas (white circles represent study areas; solid white rectangle denotes TM Path 28 Row 27; dashed white rectangle denotes TM Path 28 Row 28).



Methods

Model Calibration

The relationship between the binary forest/non-forest dependent variable, Y , and the continuous spectral value independent variables, X , may be expressed as,

$$p_i = f(X_i; \beta) + \varepsilon_i, \quad (1a)$$

or

$$E(p_i) = f(X_i; \beta), \quad (1b)$$

where i denotes subplot, p_i is the probability that $Y_i = 1$, β is a vector of parameters to be estimated, $f(X_i; \beta)$ is a function expressing the relationship among the independent variables and the parameters, ε_i is unexplained residual uncertainty, and $E(\cdot)$ denotes statistical expectation. For binary data, $f(X_i; \beta)$ is often expressed as a logistic model of which one form is,

$$f(X_i; \beta) = \frac{1}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{im})}, \quad (2a)$$

and another is,

$$f(X_i; \beta) = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{im})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{im})}. \quad (2b)$$

Each of the two forms, (2a) and (2b), may be expressed as one minus the other, and parameter values for the two forms are the same, except that the signs are reversed.

The parameters of (2) are often estimated by maximizing the likelihood, L , expressed as,

$$L = \prod_{i=1}^n (p_i)^{Y_i} (1 - p_i)^{1 - Y_i}, \quad (3)$$

where n is the number of subplot observations. Maximum likelihood parameter estimates using (3) may be obtained using SAS PROCs LOGISTIC, CATMOD, or GLIM (SAS 1988), although care must be exercised to assure the model form, (2a) or (2b), to which the parameter estimates apply. However, because (3) cannot accommodate spatial correlation among subplot observations, another approach to parameter estimation must be considered if correct measures of uncertainty are required. Nevertheless, because (3) yields unbiased parameter estimates, it may be used to obtain parameter estimates if no measures of uncertainty are required or to obtain initial parameter

estimates for iterative approaches that accommodate spatial correlation.

An iterative approach based on generalized estimating equations (GEEs) as described by Albert and McShane (1995) and Gompertz *et al.* (2000) was used to accommodate spatial correlation. For the first iteration, ordinary logistic regression using maximum likelihood was used to fit (2) to the data. The GEE approach consisted of solving,

$$\sum_{i=1}^n Z_i' V_e^{-1} (Y_i - p_i) = 0, \quad (4)$$

where the elements, z_{ij} , of Z_i are,

$$z_{ij} = \frac{\partial f(X_i; \hat{\beta})}{\partial \beta_j}, \quad (5)$$

and $f(X_i; \hat{\beta})$ is the logistic function, (2), evaluated using the parameter estimates, $\hat{\beta}$. The subplot residual covariance matrix, V_e , was recalculated following each iteration. The standardized residuals,

$$\varepsilon_i = \frac{Y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}}, \quad (6)$$

were calculated and used to define an empirical semivariogram,

$$\hat{\gamma}(h) = \frac{1}{2n_h} \sum_{i,j}^{n_h} (\varepsilon_i - \varepsilon_j)^2, \quad (7)$$

where h is the distance between subplots, n_h is number of subplots h km apart, and the subscripts i and j indicate the i^{th} and j^{th} subplots, respectively. The exponential semivariogram,

$$\hat{\gamma}(h) = \alpha_0 + \alpha_1 [1 - \exp(-\alpha_2 h)], \quad (8)$$

was fit to the empirical semivariogram using weighted nonlinear least squares techniques. The elements, v_{ij} , of the subplot residual covariance matrix, V_e , were estimated as,

$$v_{ij} = \sqrt{\hat{p}_i(1 - \hat{p}_i)} \sqrt{\hat{p}_j(1 - \hat{p}_j)} \frac{\hat{\alpha}_1}{\hat{\alpha}_0 + \hat{\alpha}_1} \exp(-\hat{\alpha}_2 h). \quad (9)$$

The $k+1^{\text{st}}$ iterative updated estimate, $\hat{\beta}^{k+1}$, was obtained as,

$$\hat{\beta}^{k+1} = \hat{\beta}^k + \left(\sum_{i=1}^n Z_i' V_e^{-1} Z_i \right)^{-1} \sum_{i=1}^n Z_i' V_e^{-1} (Y_i - \hat{p}_i). \quad (10)$$

The iterative procedure of updating $\hat{\beta}$ and recalculating V_e continued until convergence. The covariance matrix, $V_{\hat{\beta}}$, for the parameter estimates was approximated as,

$$\hat{V}_{\hat{\beta}} \cong \left(\sum_{i=1}^n Z_i' V_e^{-1} Z_i \right)^{-1} \quad (11)$$

Individual variables from among the 18 raw spectral bands and the 12 transformations were selected for inclusion in the model by repeatedly fitting the model using the GEE technique after eliminating the least significant variable among those that did not contribute significantly to improving the quality of fit.

Map-Based Estimation

Because the estimate of forest area for an AOI may be expressed as the product of the AOI's total area and an estimate of its expected proportion forest land, the remaining discussion focuses on the estimation of the expected proportion forest land and the precision of the estimate. For the i^{th} pixel in the AOI, the probability, p_i , that $Y_i = 1$ was estimated as,

$$\hat{p}_i = \frac{1}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_{ij} + \dots + \hat{\beta}_m X_{im})} \quad (12)$$

where $\hat{\beta}$ is the vector of parameter estimates obtained from (10). The estimate, \hat{P} , of the expected proportion forest, P , for the entire AOI is the mean of the probability estimates over all pixels,

$$\hat{P} = \frac{1}{N} \sum_{i=1}^N \hat{p}_i, \quad (13)$$

where N is the number of image pixels in the AOI. The variance of \hat{P} was estimated as,

$$\text{Var}(\hat{P}) = \frac{1}{N^2} \left\{ \left[\sum_{i=1}^N \sum_{j=1}^N Z_i V_{\hat{\beta}} Z_j \right] + \left[\sum_{i=1}^N \sum_{j=1}^N \text{Cov}(\varepsilon_i, \varepsilon_j) \right] \right\} \quad (14)$$

where $V_{\hat{\beta}}$ is the covariance matrix for the parameter estimates from (11) and ε is residual uncertainty. The first component within the brackets of (14) quantifies the uncertainty in the estimate because predictions for each pixel are obtained from a model with parameter estimates obtained from the same sample. The second component within the brackets of (14) quantifies the effects on uncertainty of residual variation around the model predictions. The derivation of (14) is provided in the appendix.

An estimate of correlation among residuals at a distance h , $\hat{\rho}(h)$, was obtained from the relationship,

$$\hat{\rho}_h = 1 - \frac{\hat{\gamma}(h)}{\hat{\gamma}_{\text{tot}}}$$

where $\hat{\gamma}(h)$ is obtained from (8) and $\hat{\gamma}_{\text{tot}} = \hat{\alpha}_0 + \hat{\alpha}_1$. Thus, in application, $\text{Cov}(\epsilon_i, \epsilon_j)$ was estimated as,

$$\text{Cov}(\epsilon_i, \epsilon_j) = \sqrt{\hat{p}_i(1-\hat{p}_i)}\sqrt{\hat{p}_j(1-\hat{p}_j)}\frac{\hat{\alpha}_1}{\hat{\alpha}_0 + \hat{\alpha}_1}\exp(\hat{\alpha}_2 h) \quad (15)$$

where h is the distance between the centers of the i^{th} and j^{th} pixels.

Plot-Based Estimation

Plot-based estimates of the mean and variance of proportion forest land were calculated as a standard of comparison for map-based estimates. When calculating these estimates, data were included for all four subplots of each plot having any portion of a subplot in the AOI. In addition, all plots in the AOI were selected for calculation of the mean and variance, regardless of the homogeneity of the ground cover. Inclusion of subplots with a mixture of forest and non-forest ground cover meant that proportion forest area could not be considered a binomial variable when calculating the variance of the estimate of the mean. Because the sampling design was considered to produce an equal probability, random sample, central subplots of all plots were considered to be randomly located, and spatial correlation among observations from different plots was not considered. However, the fixed spatial configuration of peripheral subplots with respect to their corresponding central subplots meant that spatial correlation, ρ , among subplot observations of the same plot could not be ignored. Large area analyses indicated that the correlation between central and peripheral subplot observations for the same plot was approximately 0.9, while the correlation between peripheral subplot observations of the same plot was approximately 0.8. Estimates of the variances of the plot-based mean proportion forest land area estimates were calculated as,

$$\begin{aligned} \text{Var}(\bar{Y}) &= \frac{1}{n^2} \left[\sum_{i=1}^n \text{Var}(Y_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^n \text{Cov}(Y_i, Y_j) \right] \\ &= \frac{1}{n^2} \left[n \text{Var}(Y) + 2 \text{Var}(Y) \sum_{i=1}^n \sum_{j=i+1}^n \rho_{ij} \right] \\ &= \frac{\text{Var}(Y)}{n^2} \left\{ n + 2 \left[\frac{n}{4} 3(0.9) + \frac{3n}{4} (0.8) \right] \right\} \\ &= \frac{\text{Var}(Y)}{n} (3.55) \end{aligned} \quad (16)$$

where i denotes subplots, n is the number of subplots in the AOI, $\frac{n}{4}$ is the number of central subplots, and $\frac{3n}{4}$ is the number of peripheral subplots.

Analyses

The inventory plot data and the satellite image spectral data were pooled for the three study areas, and a single set of parameters for model (2a) was estimated for all three study areas. For each circular area, the prediction of the probability of forest land was calculated for each pixel, and a map of the predicted probabilities was created. Within each 15-km radius circular study area, the map-based estimates \hat{P} and $\text{Var}(\hat{P})$ were calculated using (13) and (14), respectively, for circular areas with radii ranging from 1 km to 15 km centered at the center of the study area. In addition, the plot-based estimates \bar{Y} and $\text{Var}(\bar{Y})$ were calculated using (15) and (16), respectively, for the same circular areas.

Results and Discussion

The NDVI and tassell cap transformations were better predictors of the probability of forest ground cover than were the raw spectral band data. However, these results may vary for different ground covers and for satellite imagery of different dates. Because of the relatively small numbers of plots in each study area and the common set of parameter estimates for all three study areas, a single semivariogram was fit to residual data collectively for all three study areas. The fitted semivariogram indicated that spatial correlation did not extend beyond 350 m.

The inventory data, satellite imagery, and logistic model produced considerable detail in the maps depicting the probability of forest land for each pixel (fig. 2). Comparisons of the map-based and plot-based estimates of mean proportion forest land and estimates of the standard errors of the estimates yielded three primary results (table 1). First, the map-based estimates of mean proportion forest land were very similar to the plot-based estimates. Of the 37 circular areas for which the plot-based standard error was greater than zero, all except four map-based estimates of proportion forest land were within two plot-based standard errors of the plot-based estimates. For these exceptional four cases, t-tests using plot-based variances in the denominator of the t-statistic yielded $P = 0.04$, $P = 0.04$, $P = 0.02$, and $P < 0.001$.

Table 1.—Mean proportion forest estimates by study area.

Radius (km)	Plot-based estimates			Map-based estimates	
	No. plots	Mean	SE	Mean	SE
Study Area 1					
1	0	-----	-----	0.4803	0.1038
2	0	-----	-----	0.6585	0.0563
3	1	1.0000	0.0000	0.7370	0.0468
4	2	0.5000	0.3561	0.7578	0.0449
5	3	0.6667	0.2678	0.7184	0.0418
6	6	0.8333	0.1464	0.7002	0.0384
7	6	0.8333	0.1464	0.6691	0.0359
8	7	0.8571	0.1269	0.6467	0.0299
9	10	0.7000	0.1383	0.5965	0.0281
10	13	0.6154	0.1284	0.5965	0.0281
11	16	0.6094	0.1158	0.6002	0.0309
12	18	0.6300	0.1072	0.6111	0.0314
13	23	0.5490	0.0975	0.6223	0.0317
14	26	0.5914	0.0906	0.6271	0.0321
15	29	0.5992	0.0855	0.6279	0.0332
Study Area 2					
1	0	-----	-----	0.8390	0.0069
2	1	1.0000	0.0000	0.8362	0.0400
3	1	1.0000	0.0000	0.8199	0.0378
4	1	1.0000	0.0000	0.7587	0.0402
5	3	0.6667	0.2678	0.7684	0.0381
6	7	0.8571	0.1269	0.7768	0.0359
7	7	0.8571	0.1269	0.7756	0.0398
8	7	0.8571	0.1269	0.7758	0.0359
9	12	0.9167	0.0760	0.7599	0.0390
10	12	0.9167	0.0760	0.7372	0.0361
11	14	0.9286	0.0654	0.7309	0.0351
12	17	0.8676	0.0780	0.7216	0.0351
13	22	0.8010	0.0796	0.7133	0.0344
14	24	0.8170	0.0737	0.7099	0.0333
15	32	0.7065	0.0755	0.7032	0.0332
Study Area 3					
1	0	-----	-----	0.7563	0.0388
2	1	0.7500	0.4710	0.7506	0.0570
3	1	0.7500	0.4710	0.7430	0.0714
4	2	0.8750	0.2355	0.7016	0.0608
5	4	0.9375	0.1178	0.6846	0.0675
6	5	0.9500	0.0942	0.6563	0.0705
7	6	0.7917	0.1596	0.6378	0.0719
8	8	0.8438	0.1229	0.6341	0.0726
9	10	0.6750	0.1413	0.6358	0.0717
10	13	0.7500	0.1142	0.6347	0.0694
11	16	0.6692	0.1111	0.6316	0.0713
12	18	0.5949	0.1093	0.6245	0.0726
13	23	0.5525	0.0979	0.6119	0.0734
14	27	0.5447	0.0905	0.5971	0.0742
15	30	0.5236	0.0861	0.5760	0.0735

Figure 2a.—Probability of forest for Study Area 1 (white=non-forest, black=forest).

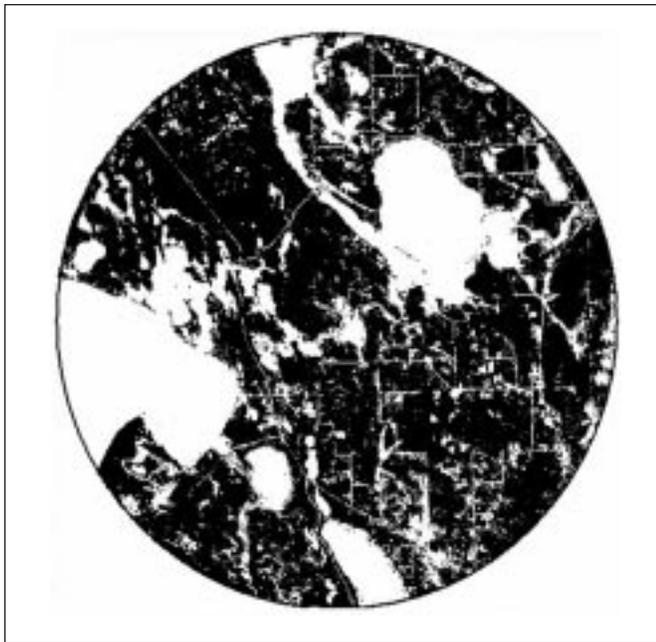


Figure 2b.—Probability of forest for Study Area 2 (white=non-forest, black=forest).

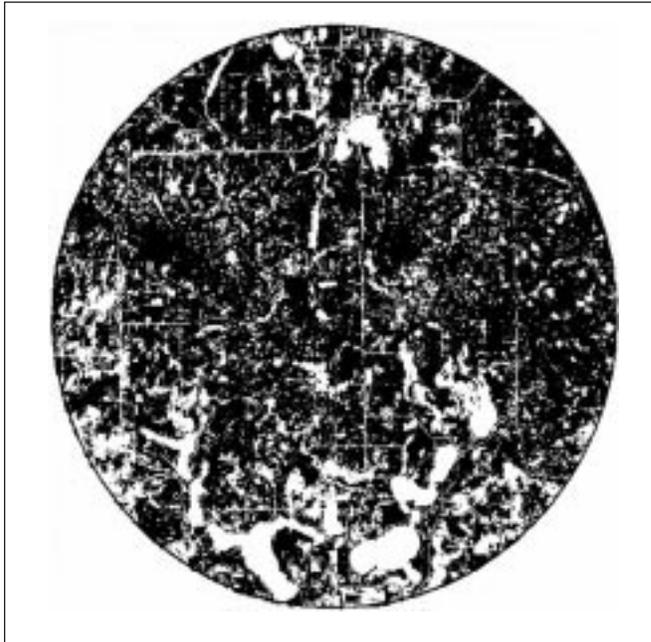
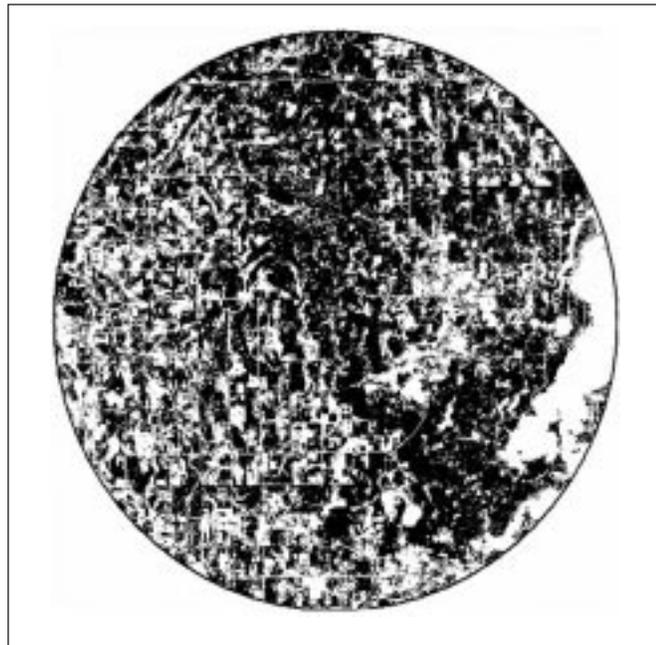


Figure 2c.—Probability of forest for Study Area 3 (white=non-forest, black=forest).



However, it must be noted that, on the one hand, these P-values are conservative because the denominator of the t-statistic did not account for uncertainty in the map-based estimates. On the other hand, P-values for circular areas of different radii in the same study area are not independent, because the plot data used to obtain an estimate for any particular circular area uses plot data for all circular areas of smaller radii in the same study area. Nevertheless, these results indicate that the map-based and plot-based estimates are quite similar. Second, the map-based estimates revealed a smooth transition as the radii of the circular areas increased from 1 km to 15 km. This result suggests that even when the number of plots in a small circular area was insufficient to obtain a reliable plot-based estimate, a reliable map-based estimate was possible. Third, the map-based standard errors were consistently smaller than the plot-based standard errors.

Two conclusions are warranted from this study. First, for forest land area, map-based estimation is not only feasible but also may produce estimates that are comparable, if not superior, to plot-based estimates. Second, although map-based estimation for other forest attributes is expected to be more difficult than for forest land area, the results of this study are sufficient to encourage the FIA program to initiate a Phase 4 focusing on construction of maps for estimation.

Literature Cited

Albert, P.; McShane, L. 1995. A generalized estimating equations approach for spatially correlated binary data: applications to the analysis of neuroimaging data. *Biometrics*. 51: 627–638.

Gompertz, M.L.; Wu, C.; Pye, J.M. 2000. Logistic regression for southern pine beetle outbreaks with spatial and temporal autocorrelation. *Forest Science*. 46(1): 95–107.

Homer, C.; Huang, C.; Yang, L.; Wylie, B.; Coan, M. [In press]. Development of a 2001 national landcover database for the United States. *Photogrammetric Engineering and Remote Sensing*.

Kauth, R.J.; Thomas, G.S. 1976. The tasseled cap—a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. *Proceedings of the symposium on machine processing of remotely sensed data*. West Lafayette, IN: Purdue University: 41–51.

McRoberts, R.E. 1999. Joint annual forest inventory and monitoring system: the North Central perspective. *Journal of Forestry*. 97: 21–26.

SAS Institute Inc. 1988. *SAS/STAT user's guide*, release 6.03. Cary, NC: SAS Institute. 1028 p.

Yang, L.; Homer, C.G.; Hegge, K.; Huang, C.; Wylie, B.; Reed, B. 2001. A Landsat 7 scene selection strategy for a National Land Cover Database. *Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium*; Sydney, Australia. CD-ROM, 1 disc.

Appendix

The estimate, \hat{P} , of the expected proportion forest, P , for an entire area of interest is simply the mean of the probability estimates, \hat{p}_i , over all pixels,

$$\hat{P} = \frac{1}{N} \sum_{i=1}^N \hat{p}_i,$$

where N is the number of image pixels in the AOI. The objective of the appendix is to provide a derivation of the estimate for the variance of \hat{P} . First, however, two intermediate results are demonstrated.

The first intermediate result is that for $Z_i = (z_{i1}, z_{i2}, \dots, z_{in})$, a vector of constants, and $\Delta' = (\delta_1, \delta_2, \dots, \delta_n)$, a random vector distributed $N(0, V_\Delta)$,

$$E[(Z_i \Delta)(Z_j \Delta)] = Z_i V_\Delta Z_j'. \quad (A1)$$

Although (A1) is not proven for the general case, a demonstration for the case of $n=2$ is provided.

$$\begin{aligned} E[(Z_i \Delta)(Z_j \Delta)] &= E[(z_{i1}\delta_1 + z_{i2}\delta_2)(z_{j1}\delta_1 + z_{j2}\delta_2)] \\ &= E(z_{i1}\delta_1 z_{j1}\delta_1 + z_{i1}\delta_1 z_{j2}\delta_2 + z_{i2}\delta_2 z_{j1}\delta_1 + z_{i2}\delta_2 z_{j2}\delta_2) \\ &= E(z_{i1}^2 \delta_1^2 z_{j1} + z_{i1}\delta_1 \delta_2 z_{j2} + z_{i2}\delta_2 \delta_1 z_{j1} + z_{i2}^2 \delta_2^2 z_{j2}) \\ &= z_{i1} E(\delta_1^2) z_{j1} + z_{i1} E(\delta_1 \delta_2) z_{j2} + z_{i2} E(\delta_2 \delta_1) z_{j1} + z_{i2} E(\delta_2^2) z_{j2} \\ &= z_{i1} \text{Var}(\delta_1) z_{j1} + z_{i1} \text{Cov}(\delta_1 \delta_2) z_{j2} + z_{i2} \text{Cov}(\delta_1 \delta_2) z_{j1} + z_{i2} \text{Var}(\delta_2) z_{j2} \\ &= Z_i V_\Delta Z_j'. \end{aligned}$$

The general case follows by analogy.

The second intermediate result is that for $Z_i = (z_{i1}, z_{i2}, \dots, z_{in})$, a vector of constants, and $\Delta' = (\delta_1, \delta_2, \dots, \delta_n)$, a random vector distributed $N(0, V_\Delta)$,

$$E[(Z_1 \Delta + Z_2 \Delta + \dots + Z_n \Delta)^2] = \sum_{i=1}^n \sum_{j=1}^n Z_i V_\Delta Z_j' \quad (A2)$$

Although (A2) is not proven for the general case, a demonstration for $n=2$ is provided.

$$\begin{aligned} E[(Z_1 \Delta + Z_2 \Delta)^2] &= E[(Z_1 \Delta)(Z_1 \Delta) + (Z_1 \Delta)(Z_2 \Delta) + (Z_2 \Delta)(Z_1 \Delta) + (Z_2 \Delta)(Z_2 \Delta)] \\ &= E[(Z_1 \Delta)(Z_1 \Delta)] + E[(Z_1 \Delta)(Z_2 \Delta)] + E[(Z_2 \Delta)(Z_1 \Delta)] + E[(Z_2 \Delta)(Z_2 \Delta)] \\ &= Z_1 V_\Delta Z_1' + Z_1 V_\Delta Z_2' + Z_2 V_\Delta Z_1' + Z_2 V_\Delta Z_2' \\ &= \sum_{i=1}^2 \sum_{j=1}^2 Z_i V_\Delta Z_j'. \end{aligned} \quad \text{by (A1)}$$

The extension of this result to the general case also follows by analogy.

The primary result is that for p_i equal to the probability that $Y_i = 1$ (i.e., that the i^{th} pixel has forested ground cover), $\hat{p}_i = f(X_i; \hat{\beta})$, and $\hat{p} = \frac{1}{N} \sum_{i=1}^N \hat{p}_i$, the variance of \hat{p} may be estimated as,

$$\text{Var}(\hat{p}) = \frac{1}{N^2} \left\{ \left[\sum_{i=1}^N \sum_{k=1}^N Z_i V_{\beta} Z_k' \right] + \left[\sum_{i=1}^N \sum_{k=1}^N \text{Cov}(\epsilon_i, \epsilon_k) \right] \right\} \quad (\text{A3})$$

where V_{β} is the estimated covariance matrix for the estimates of β , Z_i is a vector with elements,

$$Z_{ij} = \frac{\partial f(X_i; \hat{\beta})}{\partial \beta_j};$$

and ϵ_i is a residual. The proof follows.

$$\begin{aligned} \text{Var}(\hat{p}) &= E(\hat{p} - p)^2 \\ &= E \left\{ \left[\frac{1}{N} \sum_{i=1}^N \hat{p}_i \right] - \left[\frac{1}{N} \sum_{i=1}^N Y_i \right] \right\}^2 \\ &= \frac{1}{N^2} E \left\{ \left[\sum_{i=1}^N Z_i \hat{\beta} \right] - \left[\sum_{i=1}^N (Z_i \beta + \epsilon_i) \right] \right\}^2 \\ &= \frac{1}{N^2} E \left\{ \left[\sum_{i=1}^N Z_i (\hat{\beta} - \beta) \right] + \left[\sum_{i=1}^N \epsilon_i \right] \right\}^2 \\ &= \frac{1}{N^2} E \left\{ \left[\sum_{i=1}^N Z_i (\hat{\beta} - \beta) \right]^2 + 2 \left[\sum_{i=1}^N Z_i (\hat{\beta} - \beta) \right] \left[\sum_{i=1}^N \epsilon_i \right] + \left[\sum_{i=1}^N \epsilon_i \right]^2 \right\} \\ &= \frac{1}{N^2} E \left\{ \left[\sum_{i=1}^N Z_i (\hat{\beta} - \beta) \right]^2 + \left[\sum_{i=1}^N \epsilon_i \right]^2 \right\} \\ &= \frac{1}{N^2} E \left\{ \left[\sum_{i=1}^N \sum_{k=1}^N Z_i (\hat{\beta} - \beta) Z_k (\hat{\beta} - \beta) \right] + \left[\sum_{i=1}^N \sum_{k=1}^N \epsilon_i \epsilon_k \right] \right\} \\ &= \frac{1}{N^2} \left\{ \sum_{i=1}^N \sum_{k=1}^N E \left[Z_i (\hat{\beta} - \beta) Z_k (\hat{\beta} - \beta) \right] + \sum_{i=1}^N \sum_{k=1}^N E(\epsilon_i \epsilon_k) \right\} \\ &= \frac{1}{N^2} \left\{ \sum_{i=1}^N \sum_{k=1}^N E \left[Z_i (\hat{\beta} - \beta) Z_k (\hat{\beta} - \beta) \right] + \sum_{i=1}^N \sum_{k=1}^N \text{Cov}(\epsilon_i, \epsilon_k) \right\} \\ &= \frac{1}{N^2} \left\{ \left[\sum_{i=1}^N \sum_{k=1}^N Z_i V_{\beta} Z_k' \right] + \left[\sum_{i=1}^N \sum_{k=1}^N \text{Cov}(\epsilon_i, \epsilon_k) \right] \right\} \end{aligned} \quad \text{by (A2).}$$



Extending and Intensifying the FIA Inventory of Down Forest Fuels: Boundary Waters Canoe Area and Pictured Rocks National Lakeshore

Christopher W. Woodall¹ and Bruce Leutscher²

Abstract.—The sampling design for the Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture Forest Service allows intensification of fuel inventory sampling in areas of “special interest” and implementation of fuel sampling protocol by non-FIA personnel. The objective of this study is to evaluate the contribution of sampling intensification/extension toward furthering multiscale fire science investigations in two case study areas. In the Pictured Rocks National Lakeshore in Michigan, adoption of FIA’s fuel sampling protocols increased inventory efficiencies while linking local fuel estimates to the system of FIA plots. In Minnesota’s Boundary Waters Canoe Area, results indicate that 100- and 1,000-hr fuel loadings in wilderness blowdown areas may be twice those of the surrounding forest ecosystem. Both case studies illustrate the potential for the FIA program to provide estimates of fuels and facilitate local fire science initiatives and fuel inventories.

Introduction

The relative severity of recent fire seasons has highlighted the need for a more comprehensive forest fuels inventory across the forest ecosystems of the United States. The dispersal of fire risk mitigation efforts/funding is partially based on the amounts and condition of fuels at various locations. Unfortunately, the disparate efforts and sampling designs used to quantify down forest fuels across the United States has resulted in an inability to compare fuel loadings between forest regions. The down fuel sampling protocols of the U.S. Department of Agriculture

(USDA) Forest Service’s Forest Inventory and Analysis (FIA) program are applied in a systematic manner across forested regions of the United States, providing for the first standard national inventory (Woodall and Williams 2005). Beyond providing a “strategic-scale” fuels inventory, the sampling intensity of the FIA program’s fuels inventory may be augmented to increase fuel estimation precision in areas of “local interest.” Additionally, the fuels sampling protocol may be adopted by non-FIA entities so that their resulting fuels estimates may be directly comparable to those from the FIA program. The fuels inventory of the FIA program offers the ability to explore multiscale forest fuels issues through two initiatives described in this study: extension and intensification.

For this study, intensification is defined as increasing the sample intensity from a base Phase 3 intensity of one plot per 96,000 acres (Bechtold and Patterson, 2005). Besides the negative aspects of additional costs and logistics for field sampling and data management, numerous benefits of intensification exist. Intensification reduces the variance of fuel estimates by increasing sample sizes in areas of interest. Because intensification often increases the sample size in localized areas, estimates and analyses can now range in scale from regional to local, therefore increasing the number of analytical opportunities. With locally pertinent fuels data available, FIA can expand its user group through intensification. Because FIA’s down woody materials (DWM) inventory was intensified in the Boundary Waters Canoe Area (BWCA) Wilderness of northern Minnesota in 2001, this area can serve as a case study to examine intensification.

We define extension as the process by which non-FIA agencies or individuals use FIA sample protocols for their own inventories. Because these agencies/individuals have no FIA affiliation, sampling usually occurs off the FIA plot system, subsequent estimation procedures (i.e., population estimates) may vary from FIA’s. Non-FIA entities that adopt FIA’s sample

¹ Research Forester, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108. Phone: 651-649-5141; fax: 651-649-5140; e-mail: cwoodall@fs.fed.us.

² Biologist, Pictured Rocks National Lakeshore, Munising, MI 49862.

protocols may gain numerous benefits. First, field sampling protocols and associated quality assurance procedures are already developed, allowing non-FIA entities to allocate more resources to field sampling and less to sample protocol development. Secondly, because FIA provides base estimators and data management guidelines, non-FIA entities have peer-reviewed estimation procedures on which to base their own analyses. Finally, because the inventories of non-FIA entities use FIA sampling protocols, they may seamlessly link with the larger scale data from FIA and couch their inventory estimates within FIA's. The Pictured Rocks National Lakeshore (PRNL) began using FIA's DWM sample protocol to inventory fuels in 2001; subsequently, this area serves as a case study to examine extension.

Our goal was to evaluate the contribution of intensification (BWCA) and extension (PRNL) to DWM inventory and analysis objectives. Objectives included (1) estimating and comparing fuel loadings for blowdown and nonblowdown areas of the BWCA, the North Central State region, the Laurentian Mixed Forest ecosystem of the Lake States (Ecological Province 212) (Bailey 1995), and PRNL, (2) exploring multiscale mapping possibilities gained from intensification and extension, and (3) determining what novel, multiscale fuel dynamics information may be obtained from intensification and extension. For this study, local is defined as a distinct area of ownership, management, or ecological uniqueness (i.e., BWCA), and regional is defined as encompassing a political boundary (i.e., State or USDA Forest Service region) or an area of ecological similarity (i.e., ecological province).

Study Sites

The PRNL is located along the south shore of Lake Superior near the town of Munising in Michigan's Upper Peninsula (National Park Service 2003) (fig. 1). PRNL, designated by Congress as the first national lakeshore in 1966, includes 200-foot-high sandstone cliffs and unspoiled beaches/dunes. PRNL consists of 71,397 acres of mixed hardwood and boreal forest types on soils ranging from well-drained sand to hydric. Upland northern hardwoods, including American beech (*Fagus grandifolia*), sugar maple (*Acer saccharum*), red maple (*Acer rubrum*), and yellow birch (*Betula allegheniensis*), dominate about 80

percent of PRNL's forest area. Remaining forests are occupied by red, white, and jack pines (*Pinus resinosa*, *P. strobus*, and *P. banksiana*, respectively) on coarse outwash and coastal sands, paper birch (*Betula papyrifera*) and trembling aspen (*Populus tremuloides*) in successional areas, and black spruce (*Picea mariana*), white spruce (*Picea glauca*), and northern white cedar (*Thuja occidentalis*) in poorly drained lowlands. A primary management goal for PRNL is to develop and implement a comprehensive natural resource inventory and monitoring program (National Park Service 2003). Given that nearly 20 percent of PRNL's vegetation are fire-dependent forest communities, the National Park Service has mandated that PRNL preserve and perpetuate fundamental physical and biological processes to the fullest extent possible (National Park Service 2001). Therefore, PRNL intends to inventory forest and fuels conditions and develop a program to reintroduce fire in appropriate forested areas. Additionally, PRNL will compare fuel loads between Federal and non-Federal land within the lakeshore's boundary to assess effects of land management strategies on fuel loads and the potential for cooperative mitigation of hazardous conditions.

The BWCA Wilderness, located in northeast Minnesota, was established as a wilderness area in 1978 and contains 1.084 million acres (Heinselman 1996) (fig. 1). Nearly 18 percent of the BWCA is water with more than 1,000 portage-linked lakes and streams drawing more than 200,000 visitors a year

Figure 1.—Map of North Central States (white fill), Ecological Province 212 (grey), and BWCA and PRNL.



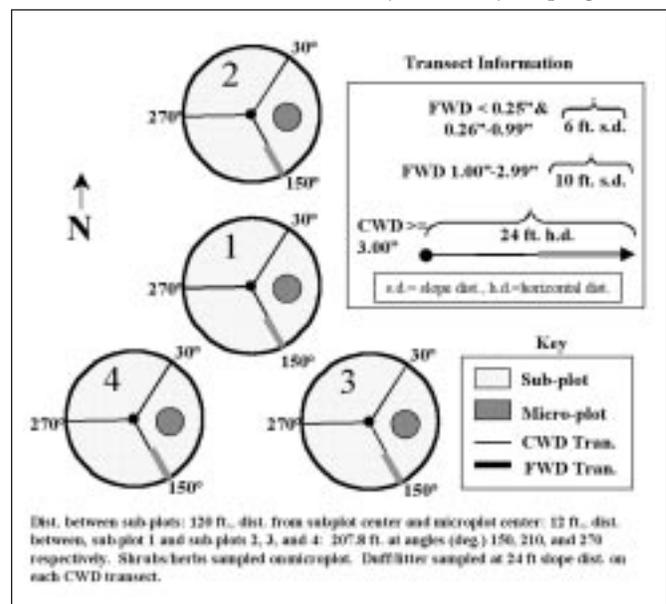
(Heinselman 1996, USDA Forest Service 2001). The forests of the BWCA, an intermix of north-central hardwoods and boreal forests, contain some of the largest tracts of virgin forest in the Eastern United States. (Heinselman 1996, USDA Forest Service 2001). Eighty-one percent of BWCA forests are occupied by upland species of black spruce, northern pines (jack, red, and eastern white), maple, aspen, and paper birch (Heinselman 1996). Black spruce bogs and alder (*Alnus rugosa*)/willow (*Justicia americana*) wetlands occupy the remaining forested areas (19 percent) (Heinselman 1996). Although mostly formed from the parent material of glacial till, soils in the BWCA range from moderately acidic granitic soils to slightly alkaline calcareous clay deposits (Heinselman 1996). The forests of the BWCA have historically been fire-dominated ecosystems with nearly all stands initiated by catastrophic wildfires (Carlson 2001). The historic fire regime that dominated the BWCA was large-scale running crown fires or high-intensity surface fires (Heinselman 1996). Since European settlement began more than a century ago, however, fire suppression activities have increased fire rotation length from 122 years to more than 2,000 years (Heinselman 1996). On July 4, 1999, an unheralded wind storm with winds in excess of 90 miles per hour affected approximately 367,000 acres of the BWCA. (USDA Forest Service 2001). The resulting blowdown substantially increased the volatility and amount of fuel in the BWCA, increasing the probability of wildfire escaping the wilderness (Leuschen *et al.* 2000). Fuel loadings were estimated to have increased from 5–20 tons per acre to 50–100 tons per acre; normal fuel loading in a typical BWCA stand has been proposed as 10 tons per acre (Leuschen *et al.* 2000). Since this wind event, the Superior National Forest has sought to monitor fuel loadings, assess blowdown effects, and mitigate fuel hazards through prescribed burns (USDA Forest Service 2001).

Methods

Varying sample intensities and field crews were used to sample DWM in the North Central States, Bailey's Ecological Province 212 (Bailey 1995), the BWCA (blowdown and nonblowdown), and PRNL. FIA field crews sampled 429 plots in the North Central States with 84 plots in the BWCA (9 in blowdown and

75 in nonblowdown), 131 plots in Ecological Province 212 (outside the BWCA), and 214 plots in the rest of the North Central States (outside Ecological Province 212). PRNL personnel sampled 121 plots in the lakeshore. In this study, we used 1-, 10-, 100-, and 1,000-hr fuels data, along with duff and litter depths, for analysis. Detailed description of the sampling design for DWM is provided by the North Central Research Station (USDA Forest Service 2003) and Woodall and Williams (2005). Briefly, 1,000-hr fuels were sampled on each of three 24-foot horizontal distance transects radiating from each FIA subplot center at 30, 150, and 270 degrees (fig. 2). Down woody pieces with a intersecting transect diameter of at least 3 inches and a length of at least 3 feet were considered 1,000-hr fuels (coarse woody debris [CWD]). Data collected for every 1,000-hr piece were transect diameter, length, small-end diameter, large-end diameter, decay class, species, evidence of fire, and presence of cavities. Fine woody debris (FWD) (1-, 10-, and 100-hr fuels) were sampled on the 150-degree transect on each subplot. FWD with transect diameters of 0.01 to 0.24 and 0.25 to 0.99 inches (1- and 10-hr, respectively) were tallied separately on a 6-foot slope distance transect (14 to 20 feet on the 150-degree transect). FWD with transect diameters of 1.00 to 2.99 inches (100-hr) were tallied on a 10-foot slope-distance transect (14 to

Figure 2.—Sampling design for the DWM indicator of the USDA Forest Service Forest Inventory and Analysis program.



24 feet on the 150-degree transect). The nonwoody surface fuels of duff and litter were sampled using an estimate of their respective depths at a 24-foot slope distance along each 1,000-hr transect (for a total of 12 sample points across all four subplots). Slight deviations exist between the 2001 and 2002 DWM sample protocols; the descriptions of these protocols are beyond the scope of this article but are detailed in Woodall and Williams (2005).

Per unit area estimates (tons per acre) for the fuel hour classes followed Brown's (1974) estimation procedures, while per unit area estimates (tons per acre) for litter and duff were based on average depth among all 12 sample points expanded to tons per acre units. CWD volume estimates were based on DeVries' line-intercept estimators (DeVries 1986). The total basal area (BA) for each sample plot was estimated from the inventory of standing trees conducted by field crews for the Phase 2 of each inventory. The means and associated standard errors were determined for all variables among plots stratified by PRNL, BWCA (blowdown and nonblowdown), Ecological Province 212 (excluding plots in the BWCA and PRNL), and North Central States outside Ecological Province 212. Fuel maps were based on interpolation of plot-level DWM estimates using ordinary kriging with an exponential model. To limit estimation to forested regions, fuel estimates were masked by applying a forest/nonforest map based on the Phase 1 National Land Cover Database (NLCD) imagery (Vogelmann *et al.* 2001).

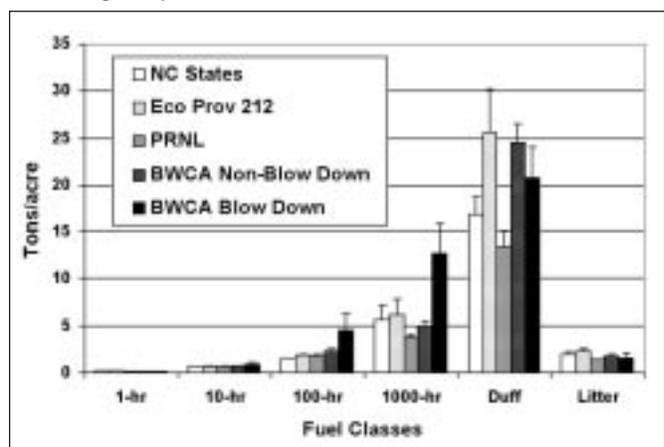
Results/Discussion

The national inventory of DWM provides context for smaller scale fuel inventories regardless of whether FIA conducted those inventories. Mean plot estimates of fuel classes show that blowdown areas have more than twice the 100- and 1,000-hr fuels than the nonblowdown BWCA area (fig. 3). Before FIA began its national inventory of fuels, any inventory of fuels in the BWCA would have been conducted in isolation with no regional or national context. Most local fuel inventories use unique sample designs, estimation procedures, and data management systems that restrict the direct comparison of local estimates to those regional estimates or estimates from adjoining lands. Through the process of increasing sample intensity in an area of interest (BWCA), however, the FIA program can provide a

multiscale assessment of fuels in local areas of interest. For the BWCA, fuel estimates may be compared to estimates from the greater North Central States region and inventories in Ecological Province 212 (fig. 3). Based on interpretation of sample mean standard errors, the means of 100- and 1,000-hr fuel loadings in BWCA blowdown areas are significantly different from those of BWCA nonblowdown areas (fig. 3). Additionally, 100- and 1,000-hr fuels in the BWCA are also greater than those in the rest of the North Central States and Ecological Province 212 (fig. 3). The same benefit of ecological context of local estimates is witnessed with PRNL fuel estimates (fig. 3). Before a consistent regional inventory of fuels was available, PRNL had only an estimate of approximately 4 tons per acre of 1,000-hr fuels (fig. 3). Management decisions and fuel mitigation efforts were based solely on that information with little relation or context of the surrounding forests/ecosystem/biome. Because PRNL used the same sample protocol and estimation procedures as the FIA program (extension), the PRNL data are essentially the same as FIA's and may be linked seamlessly with FIA regional estimates. PRNL management may know that its estimates of 4 tons per acre of 1,000-hr fuels are significantly below those for not only Ecological Province 212, but also for the remainder of the North Central States (fig. 3).

In addition to fuel inventory estimates, intensification/extension provides multiscale mapping capability that enable users to zoom in and out of areas of concern (fig. 4). The base

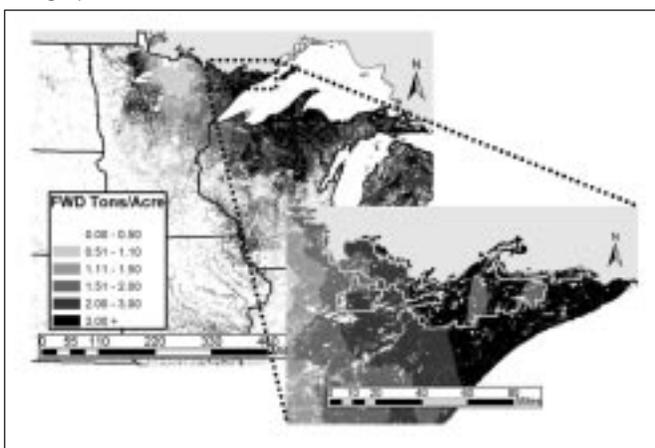
Figure 3.—Means and associated standard errors of fuel classes for forested inventory plots in various areas of the North Central States region of the United States.



sampling intensity of DWM provided by the FIA program allows construction of a regional map of FWD (fig. 4). Because sample intensity was increased in the local area of the BWCA, a smaller scale fuels map may be made for the area and framed within the regional FWD map (fig. 4). Intensification and extension of fuels sampling in local areas may allow local users to spatially appraise fuel loadings while zooming out to larger scales to couch local fuel inventories in regional assessments. Although numerous map-making methodologies are available for creating fuel maps, this study used a geostatistical approach. Regardless of which methodology is selected, the ability to provide context for fuel inventories at various scales (local to regional) may empower decisionmakers with an empirical basis for fire hazard mitigation planning.

Increasing sample intensity at local levels may refine the understanding of the dynamics between fuel and stand attributes. If a fuels inventory is conducted solely at regional scales, an absence of fuel inventories at finer (local) scales will prohibit assessments of local areas of “ecological interest.” Because the BWCA experienced a rare wind event, increasing DWM sampling intensity in the wilderness area allowed an examination of the fuel dynamics of this storm in the context of the region. In the BWCA, the volume per acre of 1,000-hr fuels actually decreases with increasing standing live stand basal area (fig. 5). In the

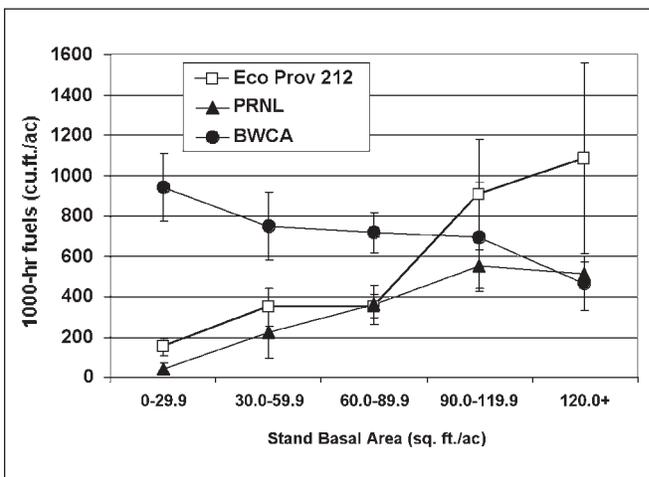
Figure 4.—Fine woody debris for selected region of the North Central States with “zoom-in” of Boundary Waters Canoe Area Wilderness based on ordinary kriging of DWM inventory plots masked using a nonforest map based on classified NLCD imagery.



PRNL, the volume per acre of 1,000-hr fuels appears to increase with increasing standing live basal area (fig. 5). These trends in 1,000-hr fuels and basal area may be examined in the context of the larger ecosystem. The remainder of the DWM plots in Ecological Province 212 has mean 1,000-hr per acre volumes that increase with increasing stand basal area (BA) in contrast to the trends found in the BWCA (fig. 5). Because the BWCA experienced a massive blowdown, stands that have little standing BA also have large amounts of 1000-hr fuels because most of their standing live BA was converted to 1,000-hr fuels in 1999 (fig. 5). Because this trend in 1,000-hr per acre volumes and BA was not found in the greater forest ecosystem or the PRNL, blowdowns of the severity found in the BWCA are likely to change forest ecosystem stand/fuel dynamics. Furthermore, because of an increasing linear relationship between standing live BA and 1,000-hr fuels in Ecological Province 212, it may be hypothesized that this ecosystem experiences density-induced individual tree mortality or sub acre-scale wind disturbances that increase 1,000-hr fuel volumes in older/greater density stands.

Although intensification and extension data may be seamlessly included with base DWM inventory data for fuel assessments and research, two caveats should be noted. First, extension data should not be used in standard reports or assessments in the FIA program; non-FIA personnel collected the data, and the subsequent data quality cannot be verified (through quality assurance/quality control procedures). Although this caveat does

Figure 5.—Means and associated standard errors of 1,000-hr fuels by total stand basal area for Ecological Province 212, BWCA, and PRNL.



not preclude use of extension data in FIA research projects, extension data should be explicitly defined and identified. Second, reduced sample and population estimate variances of means in intensification areas are possible. Hypothetically, if 60 plots exist statewide in Minnesota while 84 plots exist in the BWCA, less variance is likely with estimates in the limited area and forest types of the BWCA compared to the rest of the state. Unequal variances across fuel estimates by strata (i.e., forest types) may be acceptable provided variances are explicitly stated and discussed during hypothesis testing and data delivery.

Conclusions

Extension/intensification benefits not only helps meet the objectives of the FIA program, but they also better service FIA's customers. FIA's customer base is expanded by providing data at both regional and local scales that enable individuals/agencies with localized interests to use inventory data. Additionally, research opportunities are expanded because sampling may occur in relatively small areas of ecological interest. Quantification of dynamics between fuel and stand attributes may be refined because intensification may allow investigation of small-scale ecological phenomena such as wind events. For the BWCA, this study's preliminary estimates indicate that 100- and 1,000-hr fuel loadings in blowdown areas may be more than twice those of nonblowdown areas and the rest of the Lake States region. Also, relationships between fuel loadings and stand BA differed between the BWCA and surrounding forest ecosystem, a trend most likely caused by BWCA's recent wind event. For the PRNL, estimates of the lakeshore's fuel loadings appear less than those of surrounding regional forests. Given the current national efforts to assess and mitigate fire hazards at local and regional scales, FIA's DWM inventory may aid fuel management efforts as evidenced by intensification and extension cases in this study.

Literature Cited

- Bailey, R.G. 1995. Description of the ecoregions of the United States, 2nd ed. Misc. Pub. 1391. Washington, DC: U.S. Department of Agriculture, Forest Service. 108 p.
- Bechtold, W.A.; Patterson, P.L., eds. 2005. The enhanced Forest Inventory and Analysis program national sample design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.
- Brown, J.K. 1974. Handbook for inventorying downed woody material. GTR-INT-16. Ogden, UT: U.S. Department of Agriculture, Forest Service. 24 p.
- Carlson, D. 2001. Heterogeneity of fire severity in the Boundary Waters Canoe Area Wilderness: causes, consequences and management implications. M.S. thesis. St. Paul, MN: University of Minnesota, St. Paul. 100 p.
- DeVries, P.G. 1986. Sampling theory for forest inventory: a teach-yourself course. Berlin: Springer-Verlag: 242-279.
- Heinselmann, M.L. 1996. The Boundary Waters Wilderness ecosystem. Minneapolis, MN: University of Minnesota Press. 334 p.
- Leuschen, T.; Wordell, T.; Finney, M.; *et al.* 2000. Fuel risk assessment of the blowdown in the Boundary Waters Canoe Area Wilderness and adjacent lands. Duluth, MN: U.S. Department of Agriculture, Forest Service, Superior National Forest.
- National Park Service. 2001. National Park Service management policies 2001. Washington, DC.

National Park Service. 2003. Pictured Rocks National Lakeshore resource management plan, 2003. Munising, MI: Pictured Rocks Science Center. 68 p.

U.S. Department of Agriculture (USDA) Forest Service. 2001. Final environmental impact statement for the Boundary Waters Canoe Area Wilderness fuel treatments, vol. 1. Milwaukee, WI: U.S. Department of Agriculture, Forest Service, Eastern Region, Superior National Forest.

USDA Forest Service. 2003. Forest inventory and analysis national core field guide: field data collection procedures for phase 3 plots, vol. II. Ogden, UT: U.S. Department of Agriculture, Forest Service, North Central Research Station: 265–298.

Vogelmann, J.E.; Howard, S.M.; Yang, L.; *et al.* 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering & Remote Sensing*. 67(6): 650–661.

Woodall, C.W.; Williams, M.S. 2005. Sample protocol, estimation, and analysis procedures for the down woody materials indicator of the FIA program. Gen. Tech. Rep. NC–256. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 47 p.



Detection Monitoring of Crown Condition in South Carolina: A Case Study

William A. Bechtold¹ and John W. Coulston²

Abstract.—This article presents a case study of how indicators of forest health can be adjusted for natural factors, standardized to a common basis, and subjected to spatial analysis for the purpose of detecting potential problems related to forest health. Two of five Forest Inventory and Analysis inventory panels in South Carolina and surrounding States were completed in 2000 and 2001. The crown volume of each sampled live tree at least 5.0 inches in diameter was estimated from field measurements associated with the Phase 3 Crown Indicator. Regression models were then used to adjust each crown volume for differences in stem diameter by species. Model residuals were subsequently rescaled to a mean of 0 and standard deviation of 1, thereby enabling direct comparisons of deviations from expected crown volumes across species and tree sizes. The occurrence of trees below the 25th percentile on these adjusted statistical distributions was then examined for spatial cohesion. A statistically significant cluster of plots containing trees with below-threshold values was identified on the South Carolina-Georgia border. Additional spatial analyses in which thresholds were lowered to the 10th and 5th percentiles yielded similar results.

When the U.S. Department of Agriculture Forest Service Forest Inventory and Analysis (FIA) indicator advisors and analysts met in October 2002, the decision was made to showcase Phase 3 forest health indicators in the upcoming Maine and South Carolina analytical reports. The analyses included in this article began as a demonstration for the South Carolina report (Conner *et al.* 2004) to show how data for the crown indicator could be used in detection monitoring to check for unusual trends in tree crown health. Detection monitoring is the first of three analytical

processes applied by the Forest Health Monitoring (FHM) program to evaluate forest health (Riitters and Tkacz 2004). When a potential problem surfaces through detection monitoring, an evaluation monitoring project is initiated to increase the signal-to-noise ratio. If evaluation monitoring fails to yield a satisfactory explanation, and the potential problem is not a false signal, intensive site monitoring may then be used to more fully understand the situation.

With only the 2000 and 2001 data panels available for analysis, we were surprised to discover the demonstration in detection monitoring leading to significant results. This article describes the current status of the South Carolina crown analyses and discusses followup studies planned for evaluation monitoring.

Methods

Composite Crown Indicator

Field crews record the following tree-level variables as part of the crown indicator on all FIA Phase 3 plots: uncompact live crown ratio, crown density, crown dieback, foliage transparency, crown light exposure, and crown position. Complete descriptions of these variables are available in the FIA Phase 3 field guide (USDA Forest Service 2001). The crown variables can be analyzed singly or be combined to formulate composite indicators of crown condition. We decided that crown volume, a composite approximation of crown size that combines estimates of crown length, width, and density into a single value, was the most appropriate variable of interest for the purpose of detection monitoring. Net primary production originates at the tree crown; therefore, it logically follows that trees with small or sparsely foliated crowns might indicate a state of decline.

Field measurements of uncompact crown ratio, crown density, and tree length were thus combined with modeled crown diameter to estimate a composite crown volume (*CCV*) for each sampled tree:

¹ U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC 28802. Phone: 828-257-4357; fax: 828-257-4894; e-mail: wabehtold@fs.fed.us.

² North Carolina State University, P.O. Box 12254, Research Triangle Park, NC 27709. Phone: 919-549-4071; fax: 919-549-4047; e-mail: jcoulston@fs.fed.us.

$$CCV = 0.5(\pi R^2(CL)(DEN)) \quad (1)$$

where:

$$R = CD / 2,$$

$$CL = H (UCR) = \text{crown length (ft)},$$

$$UCR = \text{uncompacted crown ratio (percent)},$$

$$H = \text{total tree length (ft)},$$

$$DEN = \text{crown density (percent)}, \text{ and}$$

$$CD = \text{crown diameter (ft)}, \text{ which was estimated from the model:}$$

$$\hat{CD} = b_0 + b_1(D) + b_2(D^2) + b_3(UCR) \quad (2)$$

where:

$$D = \text{d.b.h. (in)}, \text{ and}$$

$$b_0 \dots b_3 = \text{regression coefficients unique to each species.}$$

The crown-diameter models were derived from trees on 1,740 FHM plots measured in 24 Eastern States between 1991 and 1999 (Bechtold 2003). Note that crown diameter had to be modeled because of the decision to drop the direct measurement of crown diameter when the FIA and FHM programs were integrated in 2000.

Standardized-Residualized Indicator

The next step was to adjust the computed crown volumes for natural factors known to influence crown size. Two of the most obvious and easily available factors in the data set were species and diameter at breast height (d.b.h.). Adjustments for species and d.b.h. were accomplished by solving the linear model specified in equation (3) for each species. Note that the model could have taken any form or been expanded to include any tree, stand, plot, or exogenous attributes for which adjustment is wanted.

$$CCV = b_0 + b_1(D) \quad (3)$$

The residuals from the least-squares solution of equation (3) serve to quantify deviations of individual trees from their expected crown volumes for a given species and tree size:

$$R_{ts} = CCV_{ts} - \hat{CCV}_{ts} \quad (4)$$

where:

$$R_{ts} = \text{the residualized indicator for tree } t \text{ of species } s.$$

Because the model residuals are scaled differently by species, one additional adjustment was made to standardize the residuals across species. The residualized indicators (R_{ts}) from equation (4) were rescaled to a standard deviation of 1 by dividing the model residuals by the standard deviation of the residuals for each species:

$$R'_{ts} = \frac{R_{ts}}{d_s} \quad (5)$$

where:

$$R'_{ts} = \text{the standardized-residual indicator for tree } t \text{ of species } s, \text{ and}$$

$$d_s = \text{the standard deviation of the model residuals for species } s.$$

At this point, we have a tree-level indicator of CCV (R'_{ts}) that has been adjusted for d.b.h. and standardized (by species) to a mean of 0 (i.e., the mean of the model residuals is 0) and a standard deviation of 1. Standardization in this manner allows trees to be combined across species for analysis. Trees can thus be averaged or otherwise grouped for comparison by tree-level attributes (e.g., overstory versus understory trees), condition-level attributes (e.g., public versus private ownership), or plot-level attributes (piedmont versus coastal plain). More details on standardization and residualization techniques are provided by Zarnoch *et al.* (2004).

Note that a regression model is not required to standardize indicators by species. Had the adjustment for d.b.h. not been wanted, a standardized indicator could have been produced by replacing the predicted (\hat{CCV}) in equation (4) with the mean (\bar{CCV}) from the data. Using the mean of the indicator allows standardization to proceed when adjustment is not necessary or possible.

Spatial Analysis

The spatial scan statistic developed by Kulldorff (1997) was used to search for potential clusters of plots with below-average crown conditions. This statistic was developed to test for randomness of disease occurrence in the spatial and spatiotemporal domains and has been applied to indicators of forest health by Coulston and Riitters (2003). The scanning proceeds by visiting every location (i.e., plot) in the study area. A series of circular windows of increasing size (up to 50 percent of the study area) is then superimposed over each location. The test statistic, Ψ_w , is then

calculated using the total number of “events” inside and outside each window. ψ_w is the likelihood ratio, based on the Bernoulli distribution, that the occurrence of events is the same everywhere after adjusting for differences in the total number of observations (events and nonevents) inside and outside the window:

$$\psi_w = \binom{E_c}{N_c} \binom{E_c'}{N_c'} \binom{E_c - E_c'}{N_c - E_c'} \binom{E_c'}{N_c'} \binom{N_c - E_c - E_c'}{N_c - E_c'} (I) \quad (6)$$

where:

- E_c = the number of events within the window,
- N_c = the number of nonevents within the window,
- E_c' = the number of events outside the window,
- N_c' = the number of nonevents outside the window, and
- $I = 1$ if $E_c/N_c > E_c'/N_c'$, or 0 otherwise.

The indicator function (I) in equation (6) sets up a one-sided test of the null hypothesis ($H_0: E_c/N_c = E_c'/N_c'$) against the alternative that the rate of events is higher inside the window.

The distribution of (ψ) across the study area and p -values associated with (ψ_w) were obtained by a Monte Carlo simulation that repeated the analysis for 9,999 random replications of the full data set under the null hypothesis of complete spatial randomness. The significance test for the cluster of observations within the window compared (ψ_w) for the window to the distribution of (ψ) from the Monte Carlo simulation. If the value of (ψ_w) exceeded 95 percent of the values from the Monte Carlo simulation, the cluster was considered significant at the 5-percent level.

We defined an event (E) as a plot with a mean adjusted crown volume (\bar{R}_p) below the 25th percentile of the frequency distribution of all plot-level means in the study area and non-events (N) as the complement:

$$E = 1 \text{ if } (\bar{R}_p) \leq T_p, \text{ or } 0 \text{ otherwise} \quad (7)$$

$$N = 1 \text{ if } (\bar{R}_p) > T_p, \text{ or } 0 \text{ otherwise} \quad (8)$$

where:

$$(\bar{R}_p) = \frac{\sum_1^{n_p} (R_{ip}')}{n_p},$$

n_p = the number of trees on plot p , and

T_p = the 25th percentile of the distribution of (\bar{R}_p) across all plots in the study area.

Plots up to 40 miles outside South Carolina were included in the spatial analysis to avoid any edge effect caused by truncating the analysis at the border. A total of 43 systematically distributed forest plots were available for analysis within the State, with an additional 33 plots contained in the band of border plots.

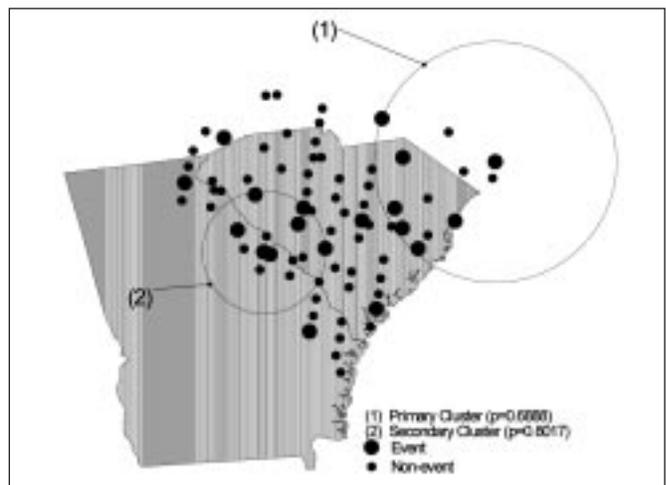
Results and Discussion

Detection Monitoring Analyses

Figure 1 shows the spatial distribution of forest plots in the study area classified as events and nonevents based on the 25th percentile of the (\bar{R}_p) frequency distribution. Two spatial clusters of plots with relatively small mean crown volumes (\bar{R}_p) were detected, but neither was statistically significant, and the observed clustering could have occurred by random chance.

The analysis was subsequently refined to increase the signal-to-noise ratio. Because the spatial clusters identified by the circles in figure 1 extended beyond the South Carolina border, the buffer area was expanded from 40 to 80 miles. This expansion increased the number of plots in the buffer from 33 to 80, yielding a total of 143 plots in the study area. We also revised the definition of an event. Recall that an event was previously defined as a plot in the lower 25th percentile of the distribution of (\bar{R}_p) across plots; thus, the number of events assigned to

Figure 1.—The distribution of FIA plots measured within 40 miles of South Carolina (2000–01), showing two clusters with a relatively high rate of events (mean plot-level standardized-residualized crown volumes below the 25th percentile). Neither cluster is statistically significant.



each plot location was either 0 or 1. In our refined analysis, trees in the lower 25th percentile of the distribution of the residuals for their species (R'_s) were identified as events, and the number of tree-level events and nonevents was then summed for each plot (equations 9 and 10) before calculating the test statistic (equation 6).

$$E = E_p = \sum_1^{n_p} E_{tp} \quad (9)$$

where:

E_p = the sum of events on plot p ,

E_{tp} = 1 if tree t is on plot p and $R'_s \leq T_s$, or 0 otherwise

T_s = the 25th percentile of the distribution of the residuals for species s (R'_s) across the study area.

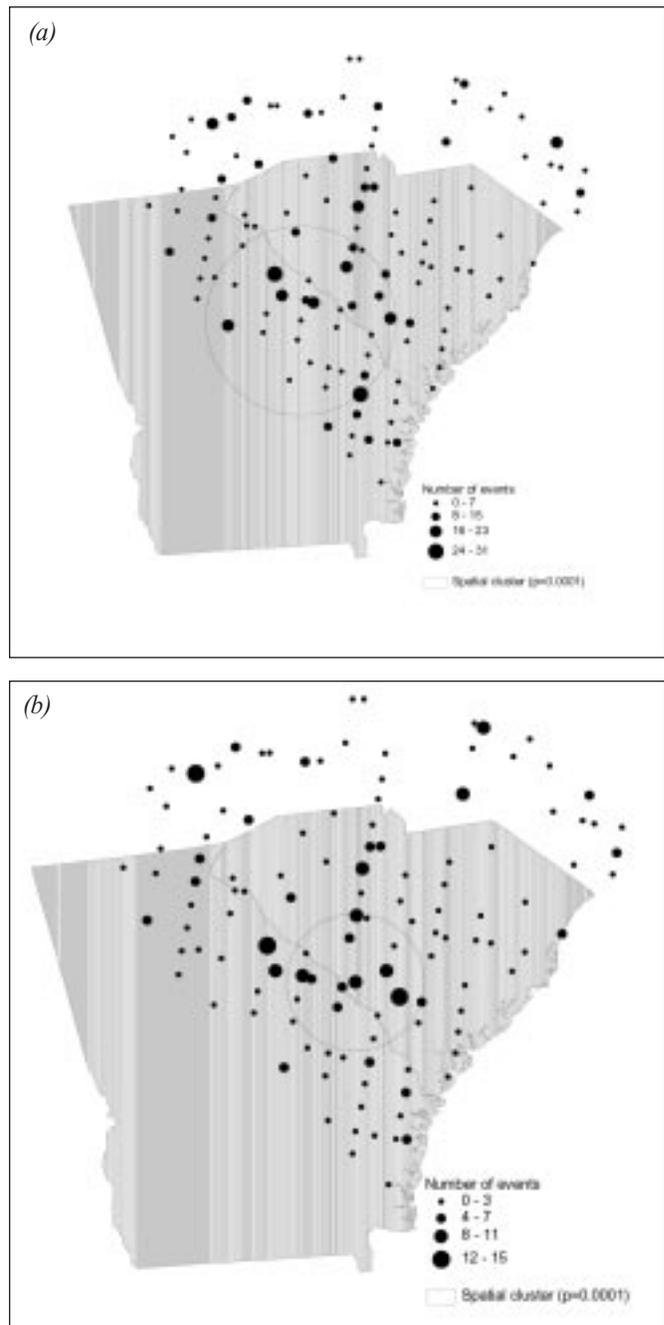
The number of nonevents on each plot (N) was then

$$N = N_p = n_p - E_p \quad (10)$$

Each plot was thus characterized by the number of events and nonevents observed, as opposed to the binary 0-1 classification used in the initial analysis. This adaptation was more consistent with Kulldorff's (1997) original technique and gave more precision and power to the analysis.

The revised analysis again detected a cluster of plots with small crowns (fig. 2a) in the same approximate location as the secondary cluster from the initial analysis (fig. 1), but this time the spatial cluster was statistically significant ($p = 0.0001$). In the cluster, 288 events were recorded when the expected number was 221. The threshold used to define an event was then progressively reduced to check the sensitivity of the cluster to the somewhat arbitrary threshold. Similar results were obtained when the threshold was lowered to the 10th percentile (fig. 2b). Again, the cluster was statistically significant ($p = 0.0001$), and 87 events were recorded when the expected number was 49. Although the 10th-percentile cluster was smaller and shifted slightly to the east, the cluster was mostly contained within the larger cluster associated with the 25th-percentile threshold. Further reducing the event threshold to the 5th percentile resulted in a significant spatial cluster ($p = 0.0002$) contained by the cluster from the 25th percentile, with 41 events recorded when only 18 were expected (fig. 2c). Substantial overlap was observed in the location of the spatial cluster across thresholds (fig. 2d).

Figure 2.—The distribution of FIA plots measured within 80 miles of South Carolina (2000–01), showing significant ($p < 0.01$) spatial clusters of FIA plots containing sample trees with standardized-residualized crown volumes below the (a) 25th, (b) 10th, (c) 5th percentile, and (d) overlay of significant spatial clusters.



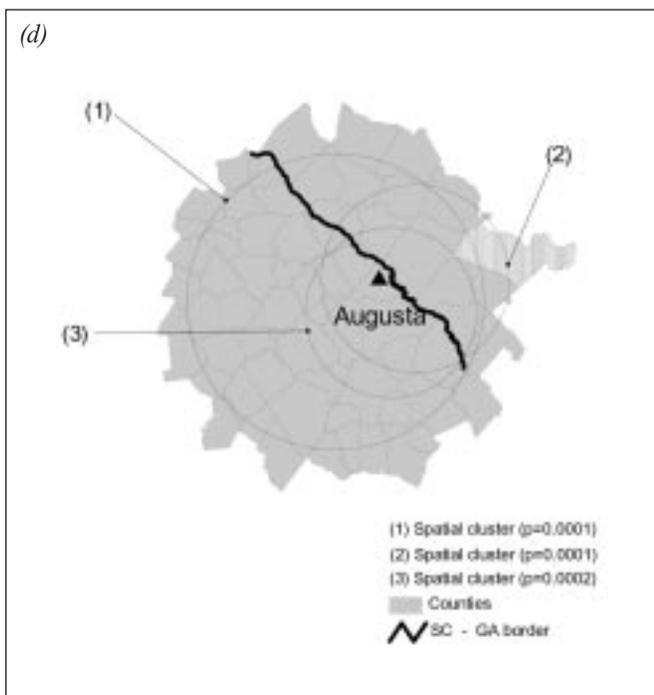
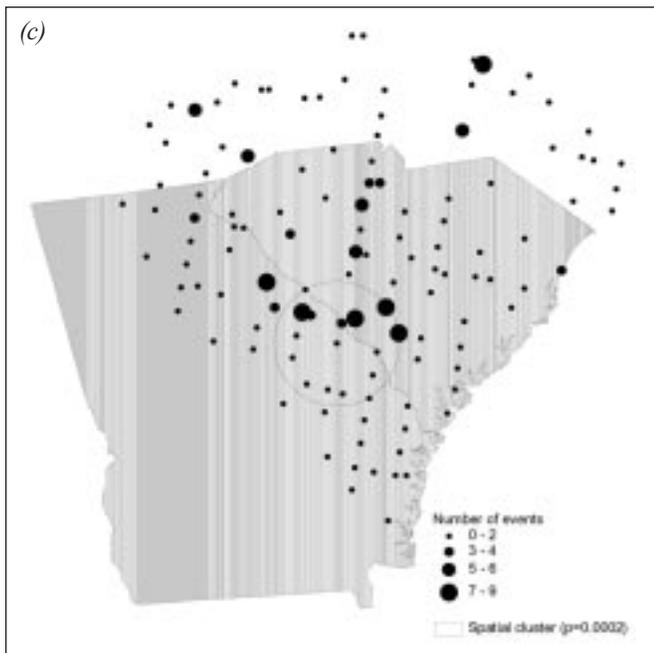
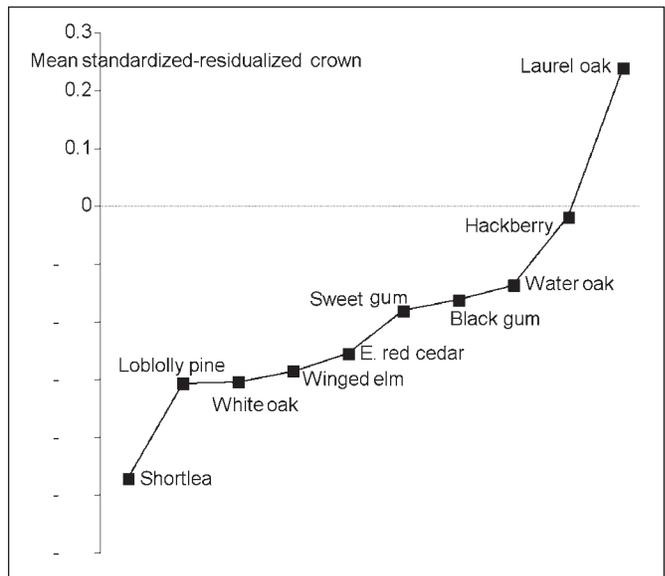


Figure 3 lists the 10 most common species encountered in the 25th-percentile cluster. By species, the mean standardized residuals of trees inside this cluster (\bar{R}'_{sc}) were below zero for all species except laurel oak, indicating that the spatial anomaly seems to cross species boundaries. (\bar{R}'_{sc}) was calculated as follows:

Figure 3.—Mean standardized-residualized crown volumes of FIA sample trees within a significant ($p < 0.01$) spatial cluster of trees below the 25th percentile on the South Carolina-Georgia border (2000–01), by species.



$$(\bar{R}'_{sc}) = \frac{\sum_{i=1}^{n_{sc}} (R'_{is})i}{n_{sc}} \quad (11)$$

where:

n_{sc} = the number of trees of species s in cluster c , and
 $i = 1$ if tree t is located in the cluster c and of species s , or 0 otherwise.

At -0.47 , shortleaf pine had the lowest (\bar{R}'_{sc}) of all species in the cluster. Loblolly pine, with a mean standardized residual of -0.30 , did not fare much better. Loblolly is by far the most common species in the region, accounting for 60 percent of the trees sampled within the cluster.

Evaluation Monitoring Proposal

Given these results, we conclude that the applied detection monitoring techniques have exposed a cluster of below-average crown volumes worthy of further investigation under evaluation monitoring. We have consequently proposed an evaluation monitoring study designed to probe deeper into the unusual cluster of trees with small crowns straddling the South Carolina/Georgia border. The objectives of the proposed study are as follows:

-
1. To examine the influence of specific crown dimensions and tree species on the location and significance of the cluster. This will be accomplished by using the standardization procedures and spatial scan statistics described above on the individual crown components (transparency, dieback, density, crown length, and crown width) and species separately. Pursuing this objective is important because the presence of a single crown dimension or tree species responsible for the geographic cluster will guide the selection of potential explanatory variables.
 2. To run the analysis again separately on individual panels and add a third panel with 2002 data (when available) to determine if a particular panel is driving the results, which may indicate a training issue.
 3. To validate the data. Field plots inside and outside the cluster will be visited to check the field measurements. Crown diameters will also be measured and checked against the crown diameters estimated with regression models.
 4. To identify potential explanatory variables. A tree pathologist will be included on the revisit team to examine tree and stand characteristics that might explain the cluster.
 5. To develop cause-effect hypotheses and test with statistical models. We will use multivariate statistical techniques to test potential explanatory variables for differences between plots inside and outside the cluster. Explanatory variables will include potential causal agents identified during the field visits, as well as environmental differences available from other data sets such as drought occurrence, ozone exposure, insect and pathogen activity, and soil characteristics. Based on results from the multivariate analysis, we intend to build models to evaluate and test cause-effect relationships.

Conclusions

Besides the potential problem with crown condition, some additional observations during the detection monitoring exercise are worthy of note.

Analysis of the crown indicator was severely handicapped by the absence of crown-diameter data. Because crown diameters had to be estimated using regression models, we essentially had to guess at one of the three variables needed to estimate the CCVs featured in the detection monitoring exercise, and are now faced with obtaining the missing data in the evaluation phase. We also had to delete species for which crown-diameter models were not available, such as palmetto and hawthorn. FIA should reconsider the decision to drop crown-diameter measurements. If not measured, crown diameters should at least be estimated in the field, which only requires 15–20 seconds per tree (Bechtold *et al.* 2002). We also experienced difficulty with the way uncompact crown ratios are measured on leaning and down trees, ultimately resulting in their deletion from the analysis. We have recently submitted a change proposal to correct that problem.

The detection monitoring techniques applied to the crown indicator can be used as a template for almost any indicator. One major advantage of the standardization approach is that it does not require biological thresholds, which involve difficult and time-consuming, process-level research, usually on a species-by-species basis. Statistical thresholds are quite useful and available for immediate use. In addition, the standardization approach easily lends itself to adjustment for the influence of natural factors through modeling. The spatial scan statistic also can be easily applied to other indicators, and it seems very efficient at identifying nonrandom spatial patterns with relatively few observations.

Finally, the spatial clustering detected in this analysis was surprisingly persistent (fig. 2d). Adjustments were made in the way crown volumes were computed for down trees and trees with broken tops, and the event thresholds were changed—all with essentially the same result. Whether a real problem exists in this area remains to be seen. Evaluation monitoring certainly is appropriate given the information at hand. Even if evaluation monitoring discovers a problem with the data or the applied analytical techniques, valuable experience will be gained in the effort to monitor forest health.

Literature Cited

- Bechtold, W.A. 2003. Crown-diameter prediction models for 87 species of stand-grown trees in the Eastern United States. *Southern Journal of Applied Forestry*. 27(4): 269–278.
- Bechtold, W.A.; Mielke, M.E.; Zarnoch, S.J. 2002. Comparison of field methods and models to estimate mean crown diameter. *Northern Journal of Applied Forestry*. 19(4): 177–182.
- Conner, R.C.; Adams, T.; Butler, B.; *et al.* 2004. South Carolina's forest resources, 2001. Resource Bulletin SRS-96. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 67 p.
- Coulston, J.W.; Riitters, K.H. 2003. Geographic analysis of forest health indicators using spatial scan statistics. *Environmental Management*. 31: 764–773.
- Kulldorff, M. 1997. A spatial scan statistic. *Communications in Statistics: Theory and Methods*. 26: 1481–1496.
- Riitters, K.; Tkacz, B. 2004. Forest health monitoring. In: Wierma, B., ed. *Environmental Monitoring*. Boca Raton, FL: CRC Press: 669–683.
- U.S. Department of Agriculture (USDA) Forest Service. 2001. Crowns: measurements and sampling. In: *Forest inventory and analysis national field guide (phase 3)*. <http://socrates.lv-hrc.nevada.edu/fia/dab/databandindex.html#2>.
- Zarnoch, S.J.; Bechtold, W.A.; Stolte, K.W. 2004. Using crown condition variables as indicators of forest health. *Canadian Journal of Forest Research*. 34: 1057–1070.



Forest Inventory and Analysis and Forest Health Monitoring: Piecing the Quilt

Joseph M. McCollum and Jamie K. Cochran¹

Abstract.—Against the backdrop of a discussion about patchwork quilt assembly, the authors present background information on global grids. They show how to compose hexagons, an important task in systematically developing a subset of Forest Health Monitoring (FHM) Program plots from Forest Inventory and Analysis (FIA) plots. Finally, they outline the FHM and FIA grids, along with their current problems and future issues.

Introduction

The curious title of this paper is meant to evoke an image of an old-fashioned patchwork quilt. One popular patchwork design, called *Grandmother's Flower Garden* (Brackman 1993), is featured in figure 1. One pattern represents the pistils of a flower, another pattern represents the petals, and a third pattern represents the actual garden. The flowers are reminiscent of the Steinman heptahex (Steinman 2001). A patchwork quilt with no fixed shape or size patch—for instance, one with triangles here and squares there—is called a crazy quilt.

Global grid developers have proposed lists of criteria for evaluating their grids. Clarke (2000) compared the proposed list of criteria for global grid development of Goodchild (1994) to that of Kimerling *et al.* (1999). Kimerling's list resembles Goodchild's, although with some obvious additions and an apparent subtraction. Carr (1998) cites the Goodchild criteria but reorders them as follows:

1. The domain is the globe.
2. Areas exhaustively cover the domain.
3. Areas are equal in size.
4. Areas are compact.
5. Areas are equal in shape.
6. Areas have the same number of edges.

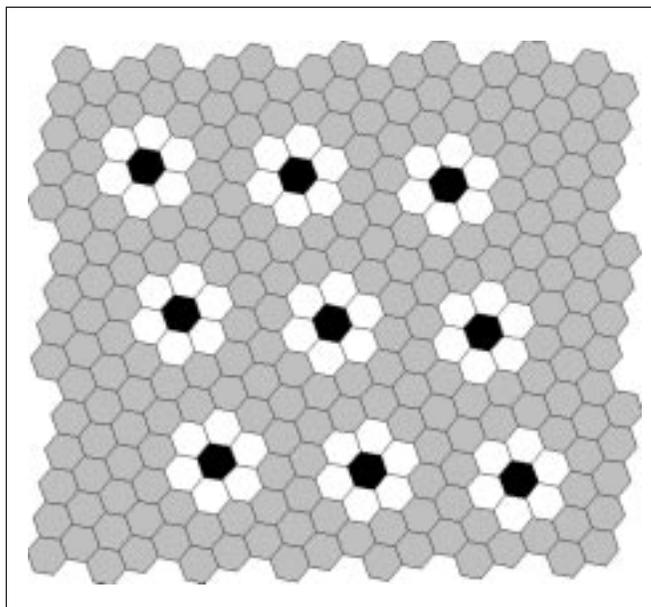
7. Edges of areas are of equal length.
8. Edges of areas are straight on some projection.
9. Areas form a hierarchy preserving some properties for $m < n$ areas.
10. Each area is associated with only one point.
11. Points are maximally central within areas.
12. Points are equidistant.
13. Points form a hierarchy preserving some properties for $m < n$ points.
14. Addresses of points and areas are regular and reflect other properties.

The authors offer the following points of clarification:

First, areas are grid cells. For Forest Inventory and Analysis (FIA), a plot (a piece of sampled landscape, currently 672 m², while cells are 2.4 x 10⁷ m²) assigned to a particular cell must fall inside that cell.

Second, many measures of compactness exist, but one possible measure is maximum area per unit perimeter. Because the first eight criteria limit the choice of grid cell to an equilateral

Figure 1.—*Grandmother's flower garden*.



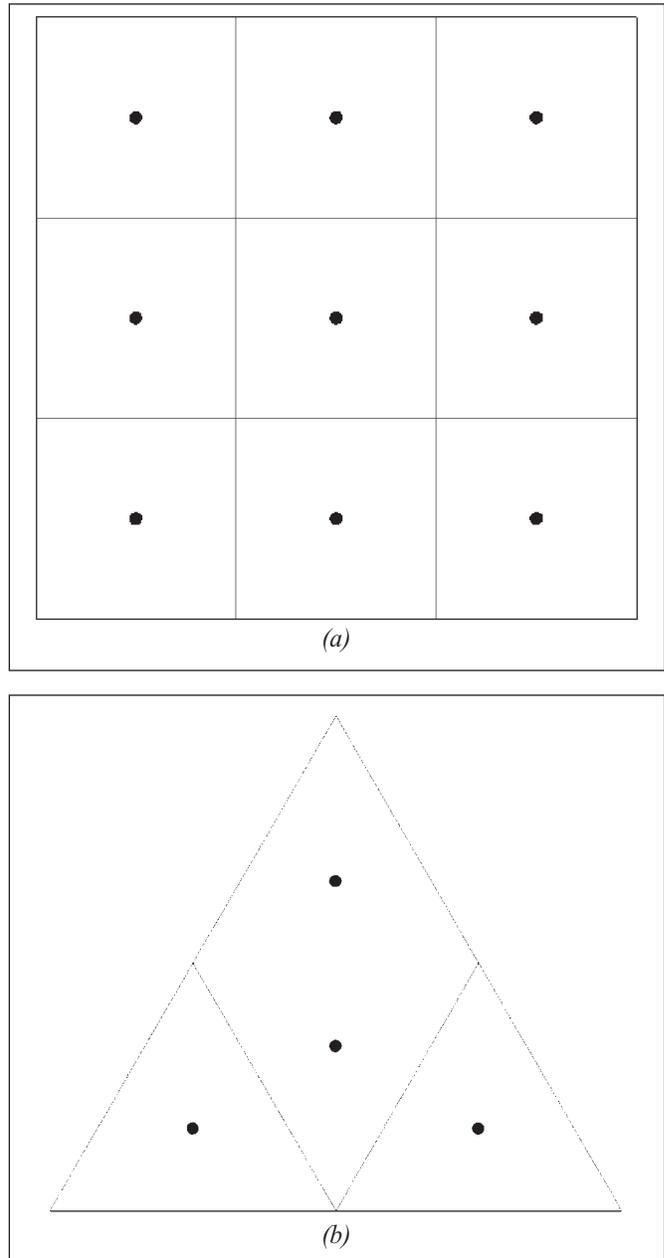
¹ Information Technology Specialist and Supervisory Forester, respectively, U.S. Department of Agriculture, Forest Service, Southern Research Station, Forest Inventory and Analysis Unit, 4700 Old Kingston Pike, Knoxville, TN 37919. E-mail: jmcollum@fs.fed.us.

triangle, rhombus, and regular hexagon, to show that the hexagon is the most compact cell is easy. Another possible measure of compactness is to minimize the maximum distance between any point and the grid point or any point and a potential plot. For FIA hexagonal cells of 2,400 hectares, no point is more than 3,040 meters from a grid point and no more than twice that distance from an actual plot. With the same size squares (which was the cell shape used in much of the United States before 1998), no point is more than 4,900 meters from a grid point and no more than twice that distance from an actual plot.

Third, what the hierarchy criteria mean is that the cell sizes should be somewhat flexible, as figure 2 shows. According to the criteria, the domain is divided into n cells, but the domain should be able to be divided into m cells, where m is some (but not necessarily any) number less than n , and m and n are integers. For instance, a network consisting of squares (which are rhombi with equal angles) is hierarchical because a square may be decomposed into h^2 smaller squares, where h is an integer. If h is odd, the centroid of the larger cell will coincide with a centroid of a smaller cell. An equilateral triangle, as figure 2(b) illustrates, may be decomposed into h^2 smaller equilateral triangles. A hexagon may be divided into $T = h^2 + hk + k^2$ smaller hexagons, where T stands for *triangulation*, and h and k are integers. The first several triangulation numbers are 1, 3, 4, 7, 9, 12, 13, 16, 19, 21, 25, and 27. Any product, rather than the multiple, of triangulation numbers is also a triangulation number. The decomposition of the hexagon is not quite as elegant as that of the square or the triangle because some of the smaller hexagons are cut in pieces around the perimeter of the larger hexagon. Reorientation of axes permits that $h \geq k \geq 0$. When $h = k$ (e.g., when $h = 3$, $k = 3$, and $T = 27$), shown in figure 3(a), or when $k = 0$ (e.g., when $h = 4$, $k = 0$, and $T = 16$), shown in figure 3(b), the decompositions are a bit more elegant. In the former case, partial hexagons are either half-hexagons or third-hexagons; in the latter case, partial hexagons are half-hexagons. Further details on partial hexagons may be found in McCollum (2001).

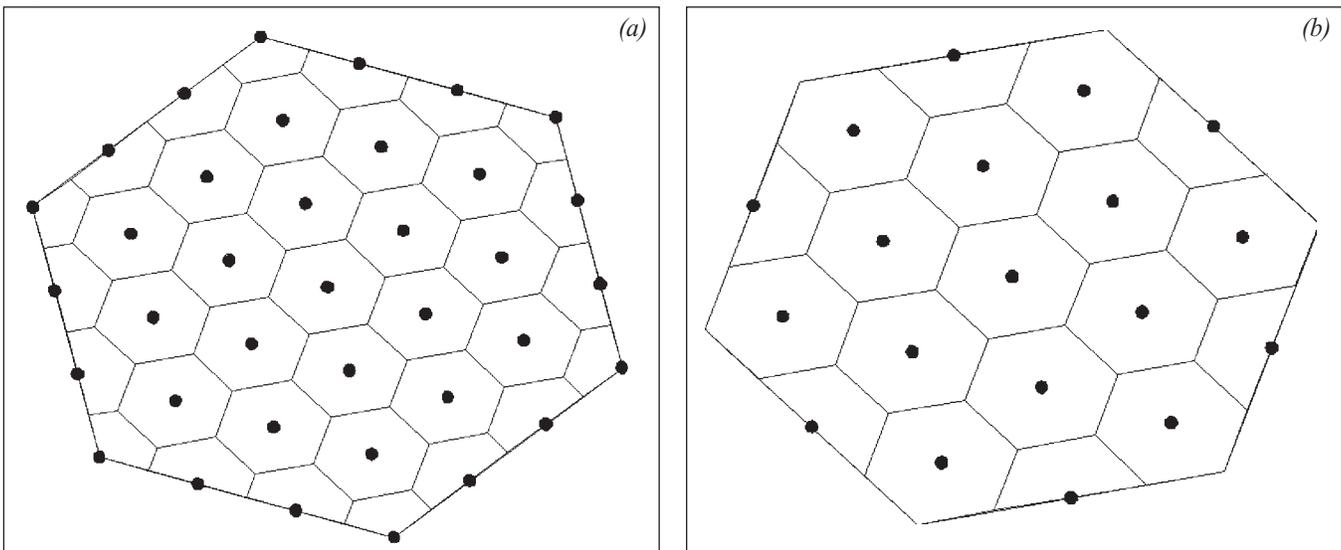
Fourth, points are grid points, and by “the maximally central point,” global grid developers mean the centroid of the cell. Such points are equidistant from each other. Many global grid developers believe the sample should be taken at the centroid, but FIA and Forest Health Monitoring (FHM) programs have avoided this rule. Details of how plots were assigned to hexagons

Figure 2.—Hierarchy of (a) squares and (b) triangles.



are available in Brand *et al.* (2000). Basically, legacy FHM plots were favored first, then legacy FIA plots, then deleted plots could be reconstructed, and, finally, if a cell remained empty, a new plot was generated. Ties at any level of favor were broken by choosing the plot closest to hexagon center. Moreover, the Food Security Act (7 U.S.C. 2276) forbids release of plot locations or of any information that makes landowners' identity discernible. If plots did appear at grid point, the plot grid could

Figure 3.—*Hierarchy of hexagons.*



be reconstructed, and landowners' privacy, if not the integrity of the data, would be at risk.

Fifth, cells have regular addresses if deducing the addresses of a cell's neighbors is easy. This characteristic is important in developing grids of varying resolution.

The additional criterion in Kimerling *et al.* (1999) was that "the grid system have a simple relationship to the traditional latitude-longitude graticule" (Kimerling *et al.* 1999, 273). This particular criterion seems in conflict with an earlier criterion in the same paper, namely that "areal cells have the same shape, ideally a spherical polygon with edges that are great circles" (Kimerling *et al.* 1999, 272). FHM did its best to adhere to both criteria. It used a variation of a grid based on great circles, but it numbered its plots according to the latitude and longitude of the grid point. Plot numbers had seven digits, where the first two digits were degrees of latitude, the next three digits were degrees of longitude, the sixth digit was the number of eighths of latitude, and the seventh digit was the number of eighths of longitude. For instance, a grid point at 30 degrees 1 minute north, 89 degrees 59 minutes west would be assigned an identification number of 3008918. With this numbering system, no obvious way existed to deduce this cell's neighbors.

FHM has offered its own reasons for accepting hexagons as the ideal cell. Among the benefits of a hexagonal network that D.L. Cassell cited in the FHM 1992 Activities Plan (Alexander and Barnard 1992) were the following:

1. It is spatially compact.
2. It provides uniform spatial coverage.
3. It is very flexible for altering the grid density.
4. It is less likely than a square grid to coincide with anthropogenic features.
5. It generally leads to smaller variance estimates than a random selection.

Moreover, the hexagon patch is superior to the square in computing contagion index (Parresol and McCollum 1997) because no quarrel need exist about whether to use rook's rule or king's rule, and the patch is superior to the random shape because no quarrel need exist about how to weight edge lengths.

Composing Hexagons

In the interest of brevity, the authors will call a FIA Phase 2 (P2) hexagon a C_1 and an historic FHM Phase 3 (P3) hexagon a C_{27} ; in general, a T -fold composition of FIA hexagons will be called a C_T . The variables H and K provide a convenient coordinate system, as figure 4 shows. The bold numbers in the upper left of each cell represent H , and the row and the plain text numbers in the lower right of each cell represent K , the path. This addressing scheme facilitates the hierarchy criteria, because to generate C_{16} hexagons, one must choose the conjunctions

of every fourth row and every fourth path of hexagon centers, and then perform Thiessen polygon expansions. In general, for a T -fold composition, T different possible starting points $(a_{0,0}, b_{0,0})$ exist. Additional grid points are selected by the following equation:

$$\{(a_{i,j}, b_{i,j})\} = \{a_{0,0} + i \cdot h + j \cdot (h+k), b_{0,0} + i \cdot k - j \cdot h\} \quad (1)$$

where i and j are index variables belonging to the set of integers.

Thus, when trying to generate C_{16} hexagons, solve $T = 16 = h^2 + hk + k^2$. From above, $(h,k) = (4,0)$. If the starting point $(a_{0,0}, b_{0,0})$ is $(0,0)$, the generated points include $(4,0)$ if $(i,j) = (1,0)$, $(8,0)$ if $(i,j) = (2,0)$, and $(8,-4)$ if $(i,j) = (1,1)$. The index variables may become as negative or positive as necessary to cover the domain.

Equation (1) may also be used to assign hexagons to panels according to the interpenetrating design, as described in such papers as Reams and Van Deusen (1999) or Roesch and Reams (1999). For the seven-panel P2 scheme, this equation was followed. Because 5 and 10 are not triangulation numbers, the equations were not followed in the 5- and 10-panel P2 schemes. Rather, P2 hexagons were numbered in panels 1 to 10 in interpenetrating fashion across the country, and then 10 panels were collapsed into 5 panels.

Forest Health Monitoring

Before 1999, the FHM program measured one plot per 64,800 hectares. Cells were hexagons numbered in four panels. The plan was to measure one panel per year plus a one-third overlap from the rest of the cycle. A State might have 240 FHM plots, so that 60 plots existed in each panel, and 20 plots existed in each overlap. Thus, in 12 years, a plot would be measured three times in its own panel and once in its own overlap.

When FIA and FHM were combined, P2 plots were to be put on a hexagonal grid of approximately 2,400 hectares per cell, while P3 plots were to be put on a grid approximately 1/16th as dense. The overlap would no longer be measured, and a fifth panel would be added. The hypothetical State with 240 FHM plots would have 400 P3 plots. Ideally, one-fifth of the plots would be measured each year.

Figure 4.—Hexagons numbered in rows and paths.

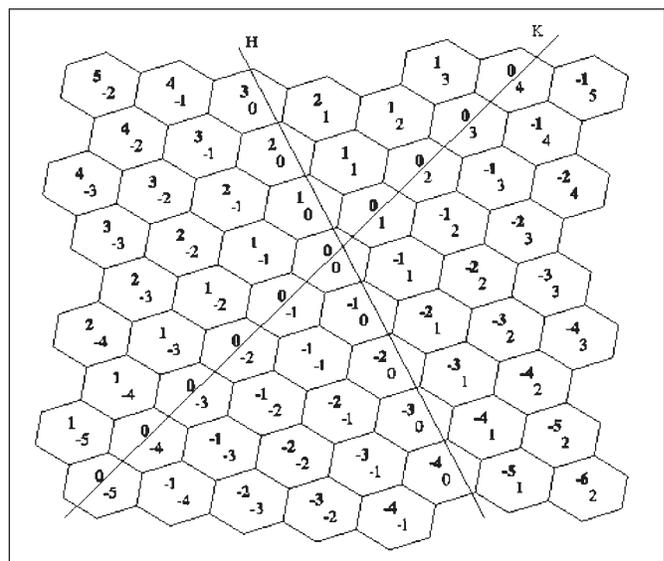
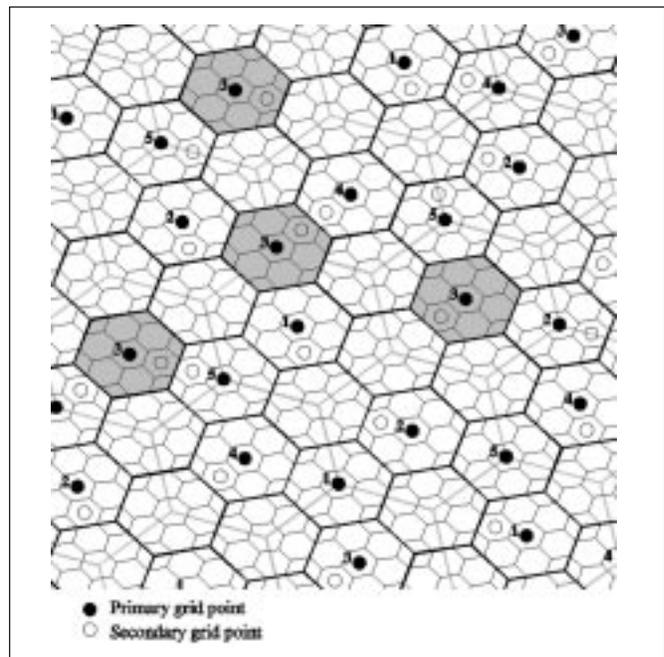


Figure 5.—Intensification method, panels 1–4, old FHM States.



Many States already had P3 plots on a grid of 64,800 hectares per cell. For the first four panels in those States, primary grid points (illustrated in this paper's figures by filled circles) were located at the centers of C_{27} hexagons, as figure 5 illustrates. Secondary grid points (illustrated in this paper's figures by open circles) were generated near the primary grid points. The secondary grid point would usually belong to the same P2 panel as

Figure 6.—Intensification method, panel 5, old FHM States.

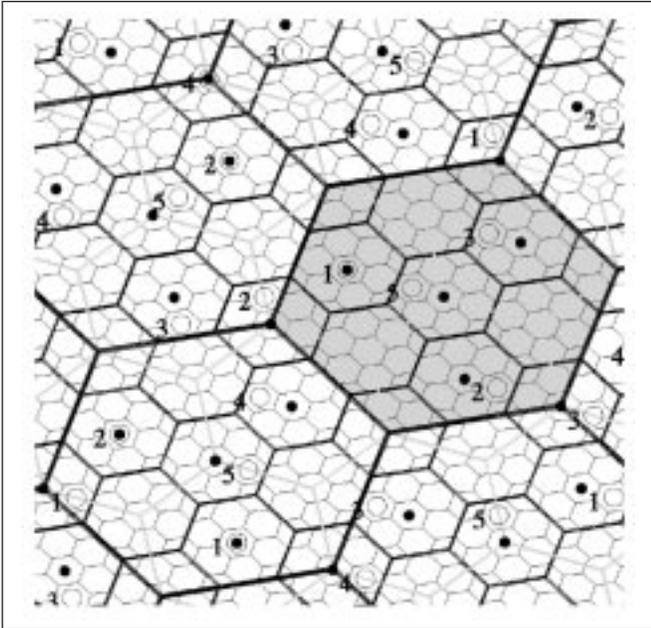
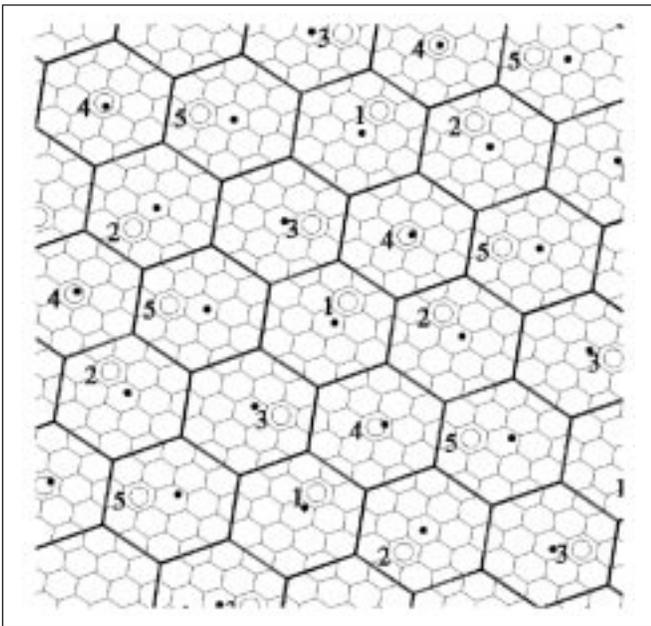


Figure 7.—Panel layout, new FHM States.



the intended panel of the P3 plot. Details on panel assignment may be found in Brand *et al.* (2000). The grid was intensified by locating three existing grid points of the same panel and then generating a fourth primary grid point of that panel. Again, a secondary grid point was generated. The result is a Y-shaped pattern, known to electrical engineers and linguists as a “wye,”

somewhat reminiscent of the mapped plot design (Alexander and Barnard 1992).

For the fifth panel in these States, the conjunctions at every ninth row and every ninth path of P2 hexagons were chosen as primary grid points, thereby unwittingly establishing C_{81} hexagons, as figure 6 shows. The reader may realize that the intensification method used in the old FHM States leads to an expected ratio of P2 grid points to P3 grid points of 16.2 to 1. Five out of every nine C_9 cells will be filled. In the C_{81} hexagon highlighted in figure 6, four C_9 cells that are filled form wye pattern. Around the perimeter of the C_{81} hexagon, three other C_9 cells that are filled exist, but each one is shared with two other hexagons. The four empty C_9 cells also form a wye.

For States new to FHM, for the most part, conjunctions of every fourth row and every fourth path of hexagons were chosen as primary grid points. Unwittingly, this selection process established C_{16} hexagons, which figure 7 illustrates. Panel assignment was based on the interpenetrating design (Roesch and Reams 1999, Reams and Van Deusen 1999). For States with little or no coastline, the expected ratio of P2 grid points to P3 grid points is 16.0 to 1.

Why Rip the Seams?

In the quilting hobby, “ripping the seams” means undoing stitches and sewing new patches rather than actually ripping. The reason the authors are ripping the seams is that a number of current difficulties exists. First, the actual ratio of P2 grid points to P3 grid points in many States is quite different from the expected ratio. Secondary grid points were overlaid on TIGER shapefiles (U.S. Census Bureau 2000) to see which States they were in and whether they landed on land or water. Table 1 shows the results.

The results show that States with proportionally large coastlines per unit area—Florida, Louisiana, and South Carolina, for instance—are undersampled. When the grid was first constructed, coastal water plots were eliminated with some method other than a TIGER land coverage. This method inadvertently eliminated some land plots; for security reasons, a map has been withheld. Some C_{16} cells are unsampled—from C_{16} center no P3 plot in any direction for at least 10,500 meters. Some of

Table 1.—Ratio of P2 to P3 hexagons in Southern States.

State	Old P3 States						New P3 States						
	Land only			All hexagons			Land only			All hexagons			
	P2	P3	Ratio	P2	P3	Ratio	P2	P3	Ratio	P2	P3	Ratio	
SC	3,252	193	16.85	3,453	204	16.93	LA	4,770	290	16.45	5,592	303	18.46
VA	4,295	261	16.46	4,613	271	17.02	FL	5,823	359	16.22	7,069	377	18.75
GA	6,291	383	16.43	6,413	388	16.53	TX	28,224	1,746	16.16	28,955	1,774	16.32
AL	5,464	337	16.21	5,655	348	16.25	AR	5,621	349	16.11	5,732	358	16.01
NC	5,266	325	16.20	5,800	338	17.16	KY	4,293	267	16.08	4,359	275	15.85
TN	4,462	276	16.17	4,538	282	16.09	OK	7,423	465	15.96	7,535	472	15.96

these unsampled C_{16} centers actually are land, and other C_{16} centers are water, the secondary grid point, if not the plot, easily could be land. One way to address this issue would be to extend the P3 grid to the exterior boundary of the United States, as indicated in TIGER files (U.S. Census Bureau 2000).

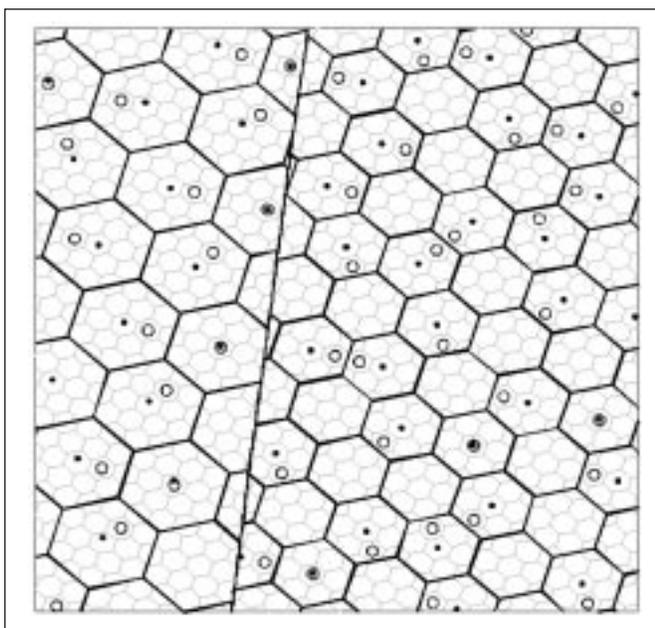
The second current issue has to do with borders—some C_{16} centers exist in new States that were not selected as primary grid points. Other C_9 centers exist in new States that were selected, but this resulted from the rule of the old States. Figure 8 illustrates examples of these anomalies. The authors' opinion is that either the grid or the documentation should be amended. Because the exact border rule is difficult to determine, the authors recommend that secondary grid points should be in the

same State as primary grid points, and primary grid points should be selected only under the rules of the State they are in rather than under the selection rules of neighboring States. This solution has another side effect. Currently, the primary grid point can be in one State, and the secondary grid point be in another, and both States are old or both States are new. The proposed solution could affect a number of plots along such State boundaries.

A third current issue is plot registration. Current Global Positioning System (GPS) readings do not necessarily agree with previously digitized plot locations, some to the point that many plots are outside their original assigned P2 hexagon. Several reasons exist for this problem. First, when plots were digitized, and no intention of one day laying them on a national grid existed. Quality assurance in digitizing maps may not have been what it should have been. Plots marked on county maps were only rough approximations of plot locations marked on aerial photographs. Second, when plot coordinates were collected with GPS units, the coordinates were transcribed to field sheets and then keypunched in the office, thereby providing at least two sources of transcription error. Third, in dense forest, GPS receivers can have difficulty maintaining a fix on satellites. Fourth, some miscommunication occurred regarding plot lists and plot selection rules.

A side effect of poor plot registration has been dropping plots because they are outside their assigned hexagons. On one hand, FHM does not want to needlessly drop past data, nor does FIA. On the other hand, FHM realized the importance of spatial dispersion among plots before FIA did. Keeping legacy plots and maintaining spatial dispersion, however, are conflicting goals. The authors believe that the best compromise would be to replace P3 plots if the field location is determined to be outside the assigned P3 hexagon. The P3 hexagon is the intended C_{16}

Figure 8.—Border issue, new FHM State vs. old FHM State.



hexagon in States new to P3 and is the intended C_9 hexagon in historical P3 States.

To ensure that P2 and P3 plots are in the correct hexagon, a new quality assurance effort will be instituted. First, plot coordinates should be digitized by “heads-up” screen digitizing. The authors have already implemented a program to do this in several Southern States. Second, county maps with latitude, longitude, and county boundaries clearly marked should be issued to field crews. Third, field crews need to make sure that they can download coordinates from GPS to their data recorders through supplied GPS cables.

Future Issues

The authors numbered all P2 hexagons into rows and paths, but might have renumbered the plots themselves. Field crews do not like it when plot 201 is in one side of the county, 202 is on the other side, 203 does not exist at all, and 204 is in the center. The lowest numbered row and path could be the first plot in the county, then plots could be incremented columnwise within row and then rowwise until all plots were numbered in logical fashion. It could be easy for the field crew to know the panel of a plot based on its plot number. The authors recognize the disadvantages of renumbering plots as well: database maintenance may become more difficult, and such a numbering scheme might be too easy for intruders (Office of Management and Budget 1994) to reconstruct.

Another application would be the construction of larger hexagons. Although hexagon maps are elegant, they may be legally risky. A C_1 map would reveal the P2 hexagon centers. A C_T map would not be risky as T grows large.

Conclusions

First, if the plot intensity for P2 or P3 changes yet again, global grid axioms should be followed. The axioms for intensifying the P3 grid have not been followed in the old States. With the current grid, either cells of unequal size exist or empty cells exist. According to the global grid axioms, there should be one plot in every cell, and cells should be equal area, regular hexagons.

Second, a favorite expression in the quilting hobby is “measure twice, cut once.” Failure to do so will turn a rationally planned patchwork quilt into a crazy quilt.

Literature Cited

Alexander, S.A.; Barnard, J.E. 1992. Forest health monitoring 1992 activities plan. EPA/602/R-93/002. Washington, DC: U.S. Environmental Protection Agency. 314 p.

Brackman, B. 1993. Encyclopedia of pieced quilt patterns. Paducah, KY: American Quilter's Society. 551 p.

Brand, G.; Nelson, M.; Wendt, D.; Nimerfro, K. 2000. The hexagon/panel system for selecting FIA plots under an annual inventory. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C., eds. Proceedings, 1st annual forest inventory and analysis symposium; 1999 November 2–3; San Antonio, TX. Gen. Tech. Rep. NC-213. St. Paul, MN: U.S. Department of Agriculture, Forest Service: 13–18.

Carr, D.; Kahn, R.; Sahr, K.; Olsen, T. 1998. ISEA discrete global grids. Statistical Computing and Statistical Graphics Newsletter. 8(2/3): 31–39.

Clarke, K. 2000. Criteria and measures for the comparison of global geocoding systems. Paper prepared for international conference on discrete global grids; 2000 March 26–28; Santa Barbara, CA. <http://www.geog.ucsb.edu/~kclarke/Papers/GlobalGrids.html>. (24 May 2004).

Goodchild, M. 1994. Criteria for evaluation of global grid models for environmental monitoring and analysis. Handout from NCGIA Initiative 15; see Spatial Analysis on the Sphere: A Review, by Rob Raskin, NCGIA Technical Report 94-7.

Kimerling, A.J.; Sahr, K.; White, D.; Song, L. 1999. Comparing geometrical properties of global grids. Cartography and Geographic Information Systems. 26(4): 271–88.

-
- McCollum, J. 2001. Honeycombing the icosahedron and icosahedroning the sphere. In: Reams, G.A.; McRoberts, R.E.; Van Deusen, P.C., eds. Proceedings, 2nd annual forest inventory and analysis symposium; 2000 October 17–18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 25–31.
- Office of Management and Budget. 1994. Report on statistical disclosure limitation methodology. Statistical Policy Working Paper 22. 142 p. <http://www.fcsm.gov/working-papers/spwp22.html>. (24 May 2004).
- Parresol, B.; McCollum, J. 1997. Characterizing and comparing landscape diversity using GIS and a contagion index. *Journal of Sustainable Forestry*. 5(12): 249–261.
- Reams, G.; Van Deusen, P. 1999. The southern annual inventory system. *Journal of Agricultural, Biological, and Environmental Statistics*. 4(4): 346–360.
- Roesch, F.; Reams, G. 1999. Analytical alternatives for an annual inventory system. *Journal of Forestry*. 97(12): 33–37.
- Steinman, J. 2001. Integrating P3 data into P2 analyses: what is the added value? In: Reams, G.A.; McRoberts, R.E.; Van Deusen, P.C., eds. Proceedings, 2nd annual forest inventory and analysis symposium; 2000 October 17–18; Salt Lake City, UT. Gen. Tech. Rep. SRS-47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 137–140.
- U.S. Census Bureau. 2000. Census 2000 TIGER/line files [machine-readable data files]. Washington, DC: U.S. Census Bureau. <http://www.census.gov/geo/www/tiger/index.html>. (6 April 2005).
- White, D. 2000. Global grids from recursive diamond subdivisions of the surface of an octahedron or icosahedron. *Environmental Monitoring and Assessment*. 64(1): 93–103.

Models for Estimation and Simulation of Crown and Canopy Cover

John D. Shaw¹

Abstract.—Crown width measurements collected during Forest Inventory and Analysis and Forest Health Monitoring surveys are being used to develop individual tree crown width models and plot-level canopy cover models for species and forest types in the Intermountain West. Several model applications are considered in the development process, including remote sensing of plot variables and stand modeling with the Forest Vegetation Simulator. The modeling process is intended to be data driven, consistent with crown architecture and stand dynamics concepts, and compatible with multiple end-user applications.

Introduction

Canopy cover is an important forest stand variable that is used in a wide variety of applications, including assessment of wildlife habitat characteristics, stand competition status, and susceptibility to damaging agents. Crown width and canopy cover measurements have been collected periodically in the Intermountain West as part of Forest Inventory and Analysis (FIA) and Forest Health Monitoring (FHM) sample designs. These data have not yet been used to their full potential. New applications have been made possible with the implementation of the mapped plot design (Birdsey 1995, Hahn *et al.* 1995). Measured or modeled crowns can be stem-mapped to provide estimates of projected canopy cover that can be compared to other canopy cover estimates, such as those measured by field crews or derived from aerial imagery. Spatially explicit crown projections also permit development of specific overlap curves for stands of different composition and structure. Crown data from surveys conducted between 1980 and 1999 are being used to develop individual tree crown models and plot-level canopy cover models for most species and forest types that occur in the

Interior West FIA (IW-FIA) database. The suite of models generated by this research may lead to improvements in data collection methods, canopy cover estimates, and remote sensing applications.

This article describes the available data, analysis considerations, and process with which a comprehensive set of crown diameter and canopy cover models is being developed for trees and forest types in the Intermountain West. This modeling effort includes objectives, such as increasing field efficiency and improving accuracy of canopy cover estimates that are internal to the FIA program. Other objectives anticipate users' needs and build on recent research and applications that use crown and canopy cover data. For example, the canopy cover extension to the Forest Vegetation Simulator (FVS) produces overlap-adjusted canopy cover estimates, but cover estimates are based on the assumption that stems (and crowns) are in a random spatial arrangement for all stand compositions and structures (Crookston and Stage 1999). Data from mapped FIA plots can be compared to FVS estimates and may support the assumption of random arrangement or suggest alternative arrangements by forest type. Most crown modeling efforts (for example, Bechtold 2003, Bragg 2001) are focused on prediction of crown width based on stem diameter and other factors. Isolating and measuring individual trees using high-resolution imagery (for example, Gougeon 1995, Gougeon and Leckie 2003) is possible, however, increasing the ability to estimate stem diameters and stand basal areas using crown measurements. Therefore, models that are optimized for remote sensing applications—in other words, with stem diameter as the dependent rather than independent variable—will also be valuable.

Analysis Approach

The approach to crown modeling taken herein reflects the desire to anticipate end user needs and, at the same time, develop models that are based on an understanding of tree biology and

¹ Analyst, U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, Rocky Mountain Research Station, Ogden, UT 84401. Phone: 801-625-5673; fax: 801-625-5723; e-mail: jdshaw@fs.fed.us.

stand dynamics. Local parameterization of the models is desirable to the extent that the data permit. The analysis plan consists of a progression from simple individual tree crown models to more complex single tree models, and ultimately plot-level canopy cover models. Analysis will be conducted in two phases. Phase 1 will address individual tree crowns, and phase 2 will address plot-level cover models.

Mining Historic Data

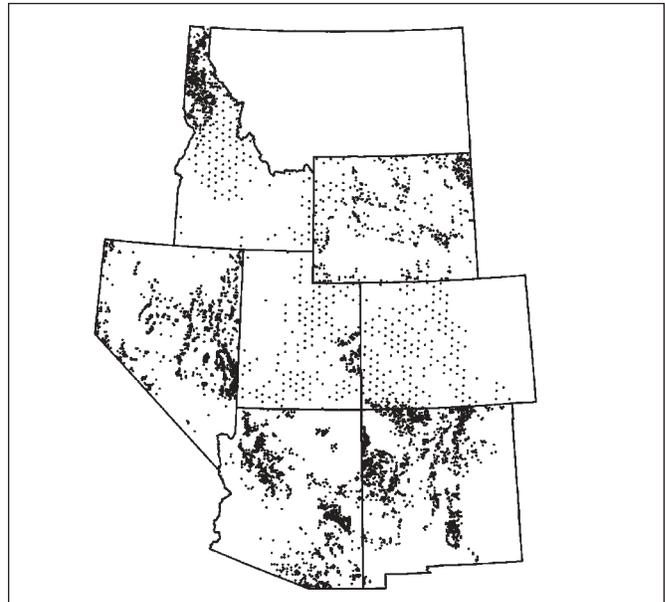
This study was conceived with the intent of using data collected from past surveys. The number of usable records was unknown, however. Twenty-two surveys conducted by FIA from 1980 to 1988 and by FHM from 1991 to 1999 were identified as potentially including crown width measurements based on field manual documentation. Crown width data from the oldest surveys were retrieved from digital tape archives. The surveys covered all, or major portions of, seven of the eight states in the IWFA analysis area: Arizona, Colorado, Idaho, Nevada, New Mexico, Utah, and Wyoming (fig. 1). No crown data from past surveys were available for Montana.

At a minimum, species, diameter, and two crown width measurements were required to constitute a usable record. Diameter was recorded to the nearest 0.1 inch in all surveys. For most species, diameter was measured at breast height (d.b.h.), but for woodland species (species of short stature and, commonly, a multistem habit), diameter was measured at the root collar (d.r.c.). For individuals with multiple stems, the method described by Chojnacky (1988) was used to calculate an equivalent diameter at root collar. Crown widths were measured to the nearest foot; further discussion of crown width data follows.

Other variables under consideration include number of stems (woodland species only), compacted or uncompact live crown ratio (measured or calculated), density measures (basal area or stand density index), elevation, and various geographic or political divisions (for example, National Forest unit or ecoregion).

The analysis data set includes 108,946 usable records for 59 species or species groups (table 1). The number of usable records varies widely by species, however, and complete sets of the potential additional variables could not be compiled for some species. This situation may limit the potential for geographic

Figure 1.—Locations of sample plots that include crown width measurements.



stratification for some species. For 17 species with a low number of observations, only preliminary models (or no models at all) may be developed. The distribution of observations appears to reflect the relative abundance of species across the area of interest, however, and less demand for models of rarer species is assumed. Some locally abundant species with limited geographic ranges, such as Arizona cypress (*Cupressus arizonica*), are not represented in the data. The analysis data set contains 43 species with $n \geq 30$, 38 species with $n \geq 50$, 33 species with $n \geq 100$, 22 species with $n \geq 500$, and 17 species with $n \geq 1,000$. Based on this distribution of observations and the behavior of the stem diameter-crown width relationship, developing reliable general crown width models for at least 32 species and local (in other words, geographically stratified) crown width models for at least 17 species should be possible.

Data Considerations

Because the data include observations for which two crown width measurements were obtained, some assumptions must be made about crown shape to determine the appropriate value for use in crown diameter modeling. FIA field manuals usually specify that the long crown axis should be measured first and

Table 1.—Species included in the crown width database and number of observations for each.

Species	n	Species	n	Species	n
Conifers		<i>Pinus flexilis</i>	991	<i>Cercocarpus montanus</i>	18
<i>Abies concolor</i>	815	<i>Pinus strobiformis</i>	49	<i>Cercocarpus breviflorus</i>	19
<i>Abies grandis</i>	1,818	<i>Pinus jeffreyi</i>	62	<i>Cornus nuttallii</i>	1
<i>Abies lasiocarpa</i>	2,563	<i>Pinus leiophylla</i>	16	<i>Olneya tesota</i>	95
<i>Abla var. arizonica</i>	385	<i>Pinus ponderosa</i>	7,088	<i>Platanus californica</i>	2
<i>Juniperus erythrocarpa</i>	21	<i>Pinus monophylla</i>	9,663	<i>Populus spp.</i>	72
<i>Juniperus californica</i>	20	<i>Pinus discolor</i>	270	<i>Populus angustifolia</i>	174
<i>Juniperus deppeana</i>	3,598	<i>Pinus monticola</i>	200	<i>Populus deltoides</i>	18
<i>Juniperus occidentalis</i>	263	<i>Pseudotsuga menziesii</i>	5,605	<i>Populus tremuloides</i>	3,822
<i>Juniperus osteosperma</i>	19,181	<i>Thuja plicata</i>	1,102	<i>Populus trichocarpa</i>	45
<i>Juniperus monosperma</i>	8,312	<i>Tsuga heterophylla</i>	437	<i>Prosopis spp.</i>	871
<i>Juniperus scopulorum</i>	2,748	<i>Tsuga mertensiana</i>	27	<i>Prunus spp.</i>	21
<i>Larix occidentalis</i>	569	Hardwoods		<i>Quercus spp.</i>	214
<i>Picea engelmannii</i>	2,962	<i>Acer macrophyllum</i>	20	<i>Quercus spp. (evergreen)</i>	2,634
<i>Picea pungens</i>	80	<i>Acer negundo</i>	25	<i>Quercus arizonica</i>	1,061
<i>Pinus albicalus</i>	153	<i>Acer glabrum</i>	128	<i>Quercus emoryi</i>	571
<i>Pinus aristata</i>	71	<i>Acer grandidentata</i>	49	<i>Quercus gambelii</i>	3,843
<i>Pinus edulis</i>	19,648	<i>Alnus rubrum</i>	12	<i>Quercus oblongifolia</i>	45
<i>Pinus contorta</i>	5,527	<i>Alnus rhombifolia</i>	5	<i>Robinia neomexicana</i>	33
<i>Pinus coulteri</i>	1	<i>Betula papyrifera</i>	118		
<i>Pinus engelmannii</i>	11	<i>Cercocarpus ledifolius</i>	784		

the short axis of the crown be measured perpendicular to the long axis and centered on the bole. In some field manuals, this method was specified in the text, but accompanying illustrations suggested that the second measurement should be taken where the crown was at its minimum width, regardless of the angle. The perpendicular measurement method was therefore assumed because measurement angles were not recorded.

When both crown width measurements are equal, the issue of diameter calculation is trivial. The issue becomes somewhat more important, however, as the difference between the two crown width measurements increases—in other words, crown shape is more eccentric. Typically, the value used in diameter crown width modeling efforts has been the arithmetic mean of

two (or occasionally more) crown width measurements (Bechtold 2003, Bragg 2001). This value has no mathematically intrinsic relationship, however, to crowns that are not round. If the true shape of a crown with differing width measurements is assumed to be an ellipse, crown area is calculated using equation 1.

$$K = \pi ab \tag{1}$$

where:

K is projected crown area, and

a and b are the major and minor radii of the ellipse.

Sensitivity to calculation of “average” diameter when the radii are unequal can be shown by a simple example (table 2). Three trees with crowns of varying eccentricity are used in the

Table 2.—Area and diameter calculations for three hypothetical crowns of varying eccentricity.

Wl	Ws	Wx	a	b	K	e	We	Percent area
8	8	8	4	4	50.3	0.00	8.0	1.000
10	6	8	5	3	47.1	0.40	7.7	0.938
12	4	8	6	2	37.7	0.47	6.9	0.750

Note: Wl and Ws are the long and short axes of the crown; Wx is the arithmetic mean of crown width; a and b are the long and short radii of the ellipse representing the crown; K is ellipse area in square feet; e is the eccentricity of the ellipse; We is the diameter of a circle with an area equal to the area of the ellipse; and percent area is the ratio of the area of the ellipse to the area of a circle with diameter Wx.

example. In the trivial case, both diameter measurements are equal (8 ft). In this case, eccentricity (e) equals 0, and area (K) equals 50.3 ft²—in other words, the area of a circle with a 4-ft radius. In the other cases, crown widths vary, but the arithmetic mean remains equal to 8 ft. As eccentricity increases, K decreases. The projected area of a crown with major and minor diameters of 12 and 4 ft is only 75 percent that of a round crown, although the arithmetic mean diameters are equal.

An argument can be made that this difference is biologically important, given that the projected crown area (and likewise, the surface area of the crown paraboloid) bears some relationship to the potential exposed photosynthetic area (Oliver and Larson 1990). Implications also exist with respect to packing of crowns in closed or nearly closed canopy stands. This suggests that considering the difference during analysis may be important.

Aside from the biological argument, another practical reason exists to consider projected crown area as the basis for calculating crown width. Determination of crown area is far more practical than determination of crown width in remote sensing applications. With the increasing availability of high-resolution imagery (in other words, ≤ 3 ft) and current image processing capabilities, isolating and measuring individual trees (for example, Gougeon 1995, Gougeon and Leckie 2003, Maltamo *et al.* 2004) is possible. The irregular nature of tree crowns makes it difficult (or computationally inefficient) to determine an “average” diameter for remotely sensed crowns. The area of individual crowns can be measured (or estimated) easily, however, whether by raster- or vector-based methods.

Based on the potential advantages of using crown area as the basis for measure instead of an arithmetic mean of diameters, a crown width value (W in table 2) was calculated from elliptical crown area using equation 2. In simple terms, crown width for an eccentric crown is defined as the diameter of a circular crown with a projected area equal to the projected area of the elliptical crown.

$$CW = 2(\sqrt{K/\pi}) \quad (2)$$

where:

CW is crown width, and

K is crown area according to equation 1.

By calculating crown width this way, one of the primary goals of this study—to produce compatible stem diameter-crown width and crown width-stem diameter models—is possible. Also,

crown width-stem diameter models are effectively “optimized” for the variable—crown area—that is more likely to be used in remote sensing applications.

Crown width data are notoriously variable, with hardwood species typically more variable than conifer species (Bechtold 2003, Bragg 2001). In the Intermountain West, woodland species (hardwoods and conifers), and especially those with multitemmed growth habits, tend to be the most variable. Analysis of the stem diameter-crown width relationship revealed three basic patterns into which each species could be grouped: (1) a well-behaved, apparently linear relationship, (2) a well-behaved, apparently nonlinear relationship, and (3) a highly variable pattern that masked the relative linearity of any underlying relationship, assuming one exists (fig. 2).

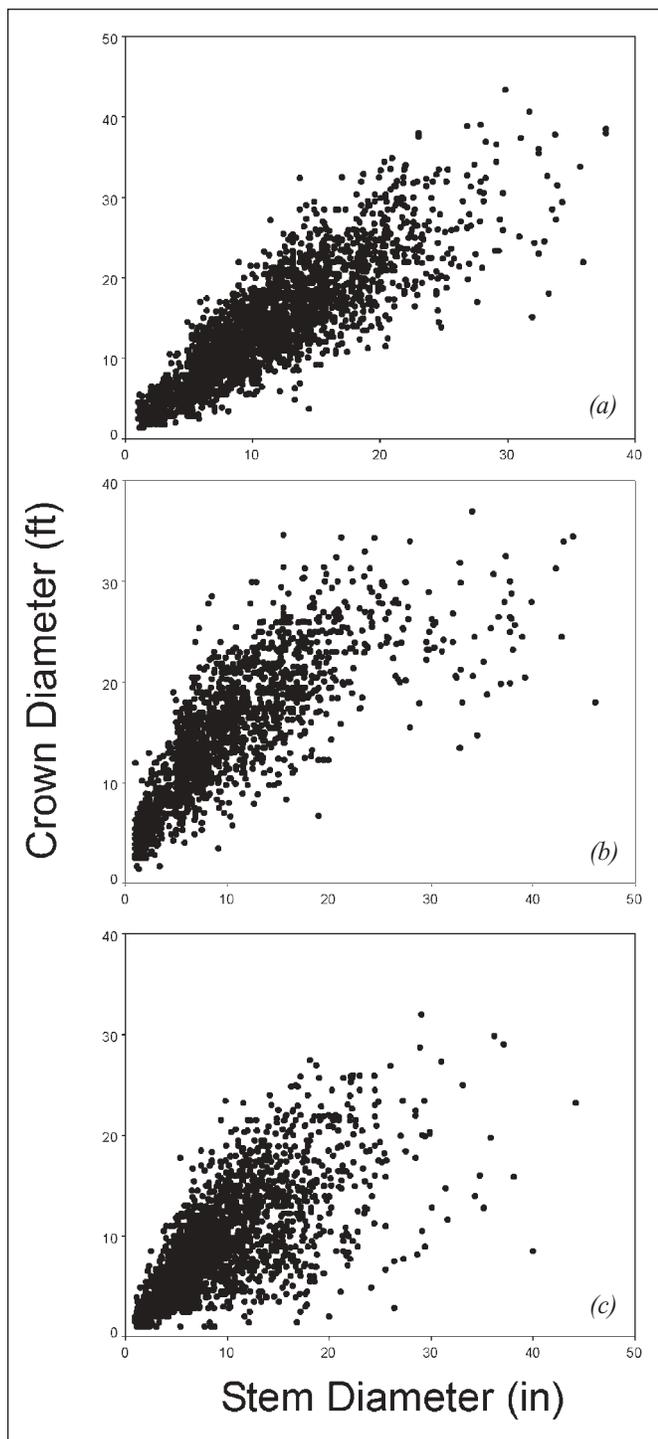
Scatter plots such as those in figure 2 show that the variance is not homogeneous across the ranges of diameter and crown width. For some species, the need for data transformation is obvious, but for others, such as ponderosa pine (fig. 2A), the need is not as clear. As a result, \log_{10} (diameter) and \log_{10} (crown width) were added to the data set. The decision of whether to use transformed or untransformed values in the models was deferred until the model fitting process, at which time the decision would be based on model behavior and residuals analysis.

Phase 1: Individual Tree Models

Phase 1 of the study involved the development of crown width models that relate crown width to stem diameter and other variables. Commonly, crown width studies produce “basic” models that predict crown width solely as a function of stem diameter and “complex” models that use one or more explanatory variables. The latter set of models assumes that the additional variables have been measured or can be calculated from other variables. Geographic location may be coded in the data set as a categorical variable (Crookston n.d.) or as a continuous variable such as latitude-longitude or Universal Transverse Mercator coordinates (Bechtold 2003). Significant variables may be selected using stepwise regression methods or by applying information efficiency criteria.

Simple equations used to predict crown width can have linear or nonlinear forms. Some investigators have made *a priori* decisions with respect to selection of linear or nonlinear models

Figure 2.—Three scatter plots that are representative of the patterns observed for species included in this study: (a) linear and well-behaved (*ponderosa pine*), (b) apparently nonlinear (*grand fir*), and (c) highly variable (*Rocky Mountain juniper*).



based on assumptions about the stem diameter-crown width relationship. Equations 3 and 4 are examples of simple crown width equations, and both models have been used with and without the intercept term (b_0).

$$CW = b_0 + b_1D \quad (3)$$

$$CW = b_0 + b_1D^{b_2} \quad (4)$$

where:

CW is mean crown width in feet,

D is d.b.h. for forest species and d.r.c. for woodland species, in inches, and

b_0 , b_1 , and b_2 are parameters to be estimated.

These equations are starting points for the model building process used in this study. As mentioned above, crown width data usually have heterogeneous variance, and transformations and weights have been used in previous studies. One such option is to use a log-transformed version of equation 4 that excludes the intercept term (equation 5). Although this equation provides a simultaneous transformation and linearization of the equation, the equation is technically (but perhaps not practically) inconsistent with crown width data because it produces a crown width of 0 at 0 d.b.h. Because crown width is measured at an unspecified height, trees shorter than breast height (4.5 ft) have a measurable crown width. The lack of an intercept term, however, should present no inconsistencies for species that have diameter measured at the root collar.

$$\log CW = b_1 + b_2(\log D) \quad (5)$$

By taking these considerations into account, the model building process used in this study will be data-driven and, at the same time, will attempt to develop the simplest appropriate model for a particular species. The steps may be summarized as follows:

1. Fit the linear model (that is, equation 3) to the data.
2. Evaluate residuals for homogeneity.
3. If transformation is warranted, refit model using transformed variables.
4. Evaluate residuals for linearity.
5. If nonlinear model is warranted, refit using nonlinear model (equation 4).
6. Evaluate performance of linear versus nonlinear model.

Using this process, developing the most parsimonious model for each species should be possible while ensuring that the model form best matches the data. The use of nonlinear models may be precluded by the lack of an adequate number of observations in some cases because the models may be “over-fitted” and unduly influenced by outlying observations. After the underlying pattern for a species is established, the influence of other factors on crown width can be weighed. This may require the assumption that the general stem diameter-crown width relationship for a species follows the same pattern in all parts of its geographic range.

Phase 2: Plot-Level Cover Models

The second phase of this study involves development of canopy cover models according to forest type or species compositional mixture. Overlap of crowns can be approximated on fixed-area, stem-mapped plot designs, such as the one currently used in FIA surveys (Birdsey 1995, Hahn *et al.* 1995). Measured or modeled crown diameters can be located according to stem location coordinates, allowing the calculation of a total canopy cover estimate that includes crown overlap. This should allow for improvement of the Canopy Cover Extension that is currently used with the FVS (Crookston and Stage 1999).

Crookston and Stage (1999) assumed a random stem distribution in their calculation of overlap-adjusted canopy cover (equation 6). This assumption may be inappropriate for some species, especially those that are shade intolerant or have been shown to exhibit crown shyness (for example, Long and Smith 1992). Equation 6 also has some practical limitations that are discussed below.

$$C_o = 100[1 - \exp(-xC')] \quad (6)$$

where:

C_o is adjusted canopy cover,

C' is unadjusted canopy cover, and

x is 0.01 for a random stem arrangement.

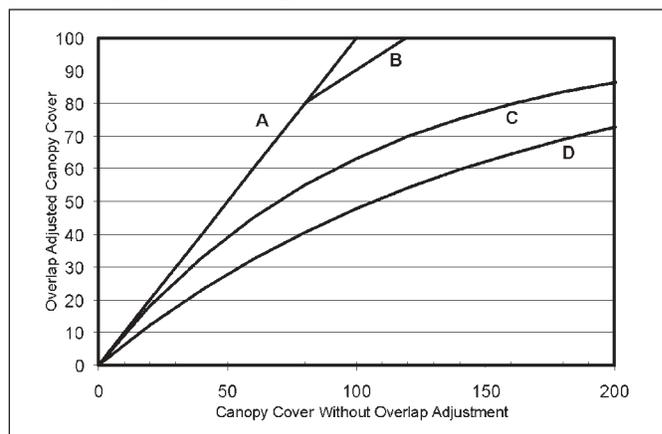
Crookston and Stage (1999) recognized the limitations of a fixed model and stated that the “ability to represent uniform distributions and some special attraction and repelling of canopies (to clump trees or clump openings, as the case may be) would

depend on empirical relations not currently available” (Crookston and Stage 1999, 2). They appeared somewhat pessimistic about prospects for improving the model, however, stating “experience shows that little accuracy would be gained by including more refinements” (Crookston and Stage 1999, 2). In any case, because such a simple model is unlikely to adequately represent all forest compositions and structures, exploring alternative models is sensible.

When considering the appropriate overlap model, consider also the function of the model in conceptual terms. The hypothesized space, in terms of the relationship between unadjusted and adjusted canopy cover, that the model should be capable of predicting can be determined using a few benchmarks and simple assumptions. When a sufficiently flexible model has been developed, the only remaining question is whether the patterns produced by different cover types can be distinguished (i.e., a statistically significant difference exists). The overlap relationship is explored in figure 3.

In figure 3, the x-axis represents unadjusted canopy cover, or, simply, the sum of the projected cover of all individual trees. The y-axis represents adjusted cover, or that which accounts for overlap of individual trees. Line A represents an obvious boundary, which represents the 1:1 relationship between unadjusted and adjusted canopy cover. In such a stand, trees might be evenly spaced, but more important, the crowns would be sufficiently plastic so that the projected canopy cover achieved 100 percent before any two individual crowns began to overlap.

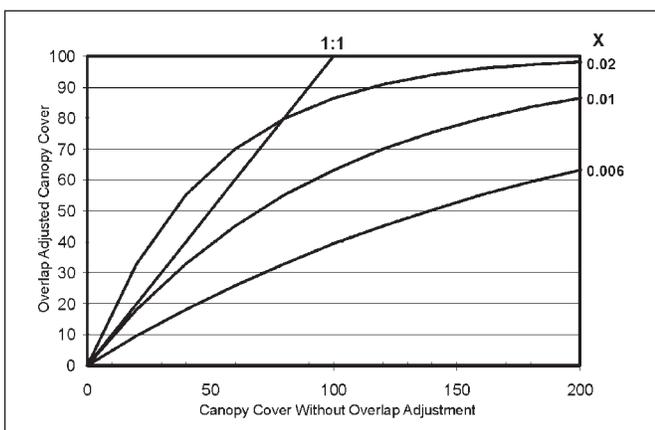
Figure 3.—Relationship between the sum of cover of individual trees (unadjusted canopy cover) and canopy cover that accounts for overlap (adjusted canopy cover).



Imagining plantation-grown trees behaving this way is easy, at least to a point. Mitchell and Popovich (1997) showed a 1:1 relationship and the point at which adjusted canopy cover breaks away from the 1:1 line for natural ponderosa pine stands in the Front Range of Colorado. Other species, such as lodgepole pine, have been shown to exhibit crown shyness and are unlikely to achieve 100-percent canopy cover in a mature, even-aged stand before understory reinitiation begins (Long and Smith 1992). Such stands may achieve their peak canopy cover at a relatively young age, because crowns are effectively trimmed back by abrasion caused by wind-driven sway as the stand grows taller (Long and Smith 1992).

Although achieving 100-percent cover without overlap may be theoretically possible, such an achievement is unlikely for most forest types; thus, an unknown boundary exists (fig. 3, line B) that probably varies by forest type. Line C in figure 3 represents the cover relationship for a random stem distribution, as modeled by Crookston and Stage (1999). This might be considered an average or typical model that lies somewhere between evenly spaced and clumpy crown arrangements. Therefore, the conceptual model can be completed by the addition of a lower boundary (line D) that represents some degree of clumpiness. The possible space occupied by the relationship between unadjusted and adjusted canopy cover is therefore bounded by lines A and D, with the likelihood that conditions do not exist in nature above line B.

Figure 4.—Curve behavior for three values of x in the equation $Co = 100[1 - \exp(-xC')]$, where Co is adjusted canopy cover, C' is unadjusted canopy cover, and x is 0.01 for a random stem arrangement.



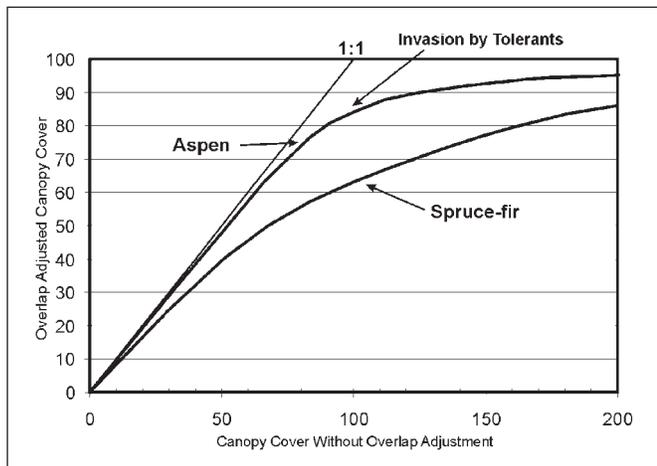
The primary limitations of equation 6 are that only a single parameter (x) exists in the model, and curves produced by the model are asymptotic to 100-percent adjusted canopy cover. By decreasing the value of x (fig. 4), curves representing increasingly clumpy crown arrangements can be produced but not the 1:1 cover relationship. Increasing x produces curves that cross above the 1:1 line and therefore represent an impossible condition—in other words, adjusted canopy cover that exceeds the sum of the individual trees.

Many options exist for producing canopy cover models with sufficient flexibility to reflect the conditions defined by the conceptual space in figure 3. Equation 6 may be modified, for example, by removing the constraint imposed by the asymptote of 100-percent cover. A flexible asymptote would permit the lower segments of some curves to closely follow the 1:1 line, although such curves could also cross into impossible space (adjusted cover > 100 percent). Mitchell and Popovich (1997) accomplished the transition between the 1:1 relationship and overlapping crowns using a segmented model. A full treatment of model options is not possible here, but the conceptual space in figure 3 can be likely modeled adequately.

The potential ability of a flexible canopy cover model can be explored by examining the stand dynamics that are expected to occur in contrasting forest types. The aspen and spruce fir types of the Intermountain West represent opposite ends of the shade tolerance range found among forest types of the region, with aspen being very intolerant, subalpine fir being very tolerant, and Engelmann spruce somewhat less tolerant than the fir (Long 1995). As with figure 3, certain benchmarks can be plotted in the space that represents the relationship between unadjusted and adjusted canopy cover in these two forest types (fig. 5).

Following fire, logging, or other disturbances, aspen commonly regenerate in large numbers by suckering. Regeneration on the order of 10,000 stems per acre or more is not uncommon (Long 1995). Considering that at maturity, perhaps in 50 years or less, the same stand will be at a density of a few hundred stems per acre, the sensitivity of aspen to competition is immediately apparent. Because of this sensitivity, expecting minimal crown overlap is logical, regardless of stand age. As noted earlier, an upper limit may exist to the amount of unadjusted cover that precludes adjusted cover from reaching 100 percent, at least as long as the stand remains pure and even-aged. Aspen stands are

Figure 5.—Hypothetical relationships between unadjusted and adjusted canopy cover for aspen (shade intolerant) and Engelmann spruce-subalpine fir (shade tolerant) stand types.



subject to invasion by a number of conifers that are more shade tolerant, however, including Engelmann spruce and subalpine fir (Mueggler 1987). Although the addition of more aspen canopy is unlikely, compositional change due to succession may increase unadjusted and adjusted canopy cover (fig. 5).

On the other hand, spruce fir stands tend to be clumpy. Engelmann spruce seedlings have difficulty surviving in open conditions and typically require shelter to regenerate successfully. This characteristic tends to influence the spatial arrangement of stems, making the distribution of crowns in spruce fir stands characteristically different from that in aspen stands. Modeling canopy cover for stands that behave similarly to those illustrated in figure 5 should be possible, based on stand composition and structure.

Conclusions

The modeling effort described in this article is multifaceted. Some of the anticipated outcomes are based on conceptual models, but the ability to achieve the desired results will depend, in part, on whether the data are sufficiently well behaved. Preliminary results suggest that the desired results can be achieved. Despite the large numbers of crown measurements available for analysis, comprehensive treatment of the species in the Intermountain West cannot be accomplished in this study.

Additional data are needed for species that are poorly represented in the collected database. Preliminary analysis also suggests that regional differences are important. Therefore, crown width data are needed for Montana because models developed for the other states may not be applicable there.

Acknowledgments

The author thanks Deborah Boyer, IWFLA, Ogden Forestry Sciences Lab, for her help with retrieving archived data files and Bill Cooke, Mississippi State University, and Nick Crookston, Rocky Mountain Research Station, Moscow Forestry Sciences Lab, for sharing their interest in and work with crown and canopy cover models.

Literature Cited

- Bechtold, W.A. 2003. Crown-diameter prediction models for 87 species of stand-grown trees in the Eastern United States. *Southern Journal of Applied Forestry*. 27(4): 269–278.
- Birdsey, R.A. 1995. A brief history of the “straddler plot” debates. *Forest Science Monograph*. 31: 7–11.
- Bragg, D.C. 2001. A local basal area adjustment of crown width prediction. *Northern Journal of Applied Forestry*. 18(1): 22–28.
- Chojnacky, D.C. 1988. Juniper, pinyon, oak, and mesquite volume equations for Arizona. Res. Pap. INT-391. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 11 p.
- Crookston, N.L. [N.d.]. Crown width equations for thirty-four northwest United States tree species. [Manuscript in preparation.]
- Crookston, N.L.; Stage, A.R. 1999. Percent canopy cover and stand structure statistics from the Forest Vegetation Simulator. Gen. Tech. Rep. RMRS-GTR-24. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 11 p.

-
- Gougeon, F.A. 1995. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing*. 21(3): 274–284.
- Gougeon, F.A.; Leckie, D.G. 2003. Forest information extraction from high spatial resolution images using an individual tree crown approach. Information Report BC-X-396. Victoria, BC: Canadian Forest Service, Pacific Forestry Centre. 26 p.
- 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.
- Long, J.N. 1995. The middle and southern Rocky Mountain region. In: Barrett, J.W., ed. *Regional silviculture of the United States*. New York: John Wiley & Sons: 335–386.
- Long, J.N.; Smith, F.W. 1992. Volume increment in *Pinus contorta* var. *latifolia*: the influence of stand development and crown dynamics. *Forest Ecology and Management*. 53: 53–64.
- Maltämo, M.; Eerikänen, K.; Pitkänen, J.; *et al.* 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sensing of Environment*. 90: 319–330.
- Mitchell, J.E.; Popovich, S.J. 1997. Effectiveness of basal area for estimating canopy cover of ponderosa pine. *Forest Ecology and Management*. 95(1): 45–51.
- Mueggler, W.F. 1987. Age distribution and reproduction of Intermountain aspen stands. *Western Journal of Applied Forestry*. 4(2): 41–45.
- Oliver, C.D.; Larson, B.C. 1990. *Forest stand dynamics*. New York: McGraw-Hill. 467 p.



Species Composition of Down Dead and Standing Live Trees: Implications for Forest Inventory Analysis

Christopher W. Woodall¹ and Linda Nagel²

Abstract.—The assessment of species composition in most forest inventory analysis relies solely on standing live tree information characterized by current forest type. With the implementation of the third phase of the U.S. Department of Agriculture Forest Service's Forest Inventory and Analysis program, the species composition of down dead trees, otherwise termed coarse woody debris (CWD), is now available to inventory analysts. To evaluate the possible contribution of CWD inventory data to forest ecosystem assessments, the species compositions of standing live and down dead trees for FIA plots across north-central States were compared within the context of forest inventory analysis. Results indicate that CWD species composition data may refine understanding of past tree mortality patterns in the context of stand development and species composition shifts. Further, CWD species composition data provide analysts with an additional categorical unit for inventory reports. Although use of CWD species composition data may be limited by measurement error and sparse sampling intensity, such data complement standing live tree data for a range of inventory analysis procedures.

Introduction

Forest types (FTs), otherwise known as forest cover types, are categories of forest defined by constituent vegetation (Eyre 1980, Helms 1998). The single attribute of forest vegetation often used as a delimiter of FT is the species composition of living forest biomass present in the stand/plot being typed (Eyre 1980, Helms 1998). Additionally, FTs may be defined by current

or potential vegetation (Daniel *et al.* 1979). The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture (USDA) Forest Service uses a definition of FT that deals mainly with the species composition of current tree biomass on a plot, "classification of forest land based on the species presently forming a plurality of the live-tree stocking" (Smith *et al.* 2001, 43). FT information has been used as a categorical variable for ecological analyses for decades and forms the basis of numerous forest reports produced by FIA and its cooperators (H. John Heinz III Center for Science 2002, Miles *et al.* 2003, Smith *et al.* 2001, USDA Forest Service 1965). Recent forest resource reports have placed additional emphasis on FT analyses (Heinz Center 2002, Smith *et al.* 2001) because changes in FTs across the United States may indicate effects of urbanization and climatic variations.

Because forest typing procedures usually include only living trees, the identifiable species composition of down dead trees is often omitted in forest inventories and subsequent analyses. Down and dead trees, otherwise known as coarse woody debris (CWD), serve as critical habitat for numerous flora and fauna. Flora use the microclimate of moisture, shade, and nutrients provided by CWD for regeneration establishment (Harmon *et al.* 1986). CWD provide a diversity (stages of decay, size classes, and species) of habitat for fauna ranging from large mammals to invertebrates (Bull *et al.* 1997, Harmon *et al.* 1986, Maser *et al.* 1979). Besides providing assessments of habitat, CWD may contain the history of the species composition of any particular stand, possibly refining understanding of mortality trends over time (i.e., succession). CWD studies to date often quantify only the volumes, sizes, and diameters of CWD with incidental information regarding CWD species composition (Goodburn and Lorimer 1998, Pedlar *et al.* 2002). Given the importance of CWD, a new categorical variable is proposed that may benefit CWD assessments and overall inventory analyses. "Coarse Woody Type" (CWT) may be defined as a

¹ Research Forester, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108. Phone: 651-649-5141; fax: 651-649-5140; e-mail: cwoodall@fs.fed.us.

² Assistant Professor, School of Forest Resources and Environmental Science, Michigan Technological University, Houghton, MI 49931. Phone: 906-487-2812; fax: 906-487-2915; e-mail: lmnagel@mtu.edu.

broad categorization of the species composition of the dead tree biomass in a forest stand. Because FIA inventory data may be used to determine both FT and CWT on selected inventory plots, the FT and CWTs may be used separately or in combination to refine understanding of forest attributes and stand dynamics.

The goal of this study was to determine if information on down dead tree species composition could be used to refine analytical procedures that have typically used only FT information. Specific objectives were to (1) assess difficulties in developing a CWD typing algorithm, (2) compare FT and CWT paired by individual plot and correlate them with the stand attributes of total stand basal area, stand age, and site index, and (3) link plot-level FT and CWTs to successional and stand development patterns regionally observed for common FTs.

Methods

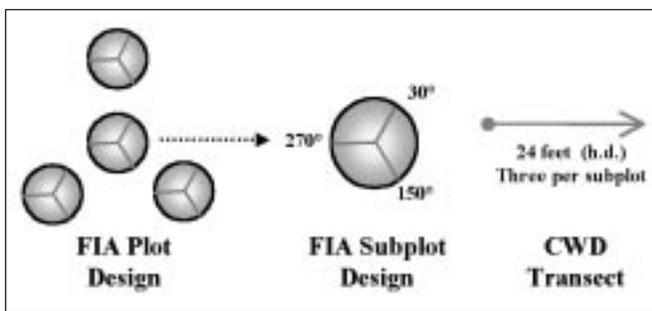
As defined by the FIA program, CWD are down logs with a transect diameter ≥ 3 in and a length ≥ 3 feet. CWD are sampled during a specific phase of FIA's multiscale inventory sampling design (USDA 2002). CWD are sampled on transects radiating from each FIA subplot center (fig. 1). Each transect is 24 feet long, three per subplot. Information collected for every CWD piece intersected on each of three 24-foot transects on each FIA subplot is transect diameter, length, small-end diameter, large-end diameter, decay class, species, evidence of fire, and presence of

cavities (fig. 1). Transect diameter is the diameter of a down woody piece at the point of intersection with a sampling transect. Decay class is a subjective determination of the amount of decay present in an individual log. Decay class 1 is the least decayed (freshly fallen log), while decay class 5 is an extremely decayed log (cubicle rot pile). The species of each fallen log is identified through determination of species-specific bark, branching, bud, and wood composition attributes (excluding decay class 5 CWD pieces). If a CWD piece is too decomposed to identify its species, a hierarchy of species identification is followed: species, species group, conifer or hardwood, or unknown.

CWD inventory data, along with corresponding tree and stand information, for this study were obtained from selected forested plots ($n = 345$) in the north-central States. Plots were sampled during the summers of 2001 and 2002. DeVries' line-intercept estimators were used to determine CWD volume per acre by species (DeVries 1986). A CWT was determined for each sample plot based on the species with the plurality of CWD volume per acre. Although a CWT algorithm may eventually be developed to readily determine CWTs, that objective was beyond the purview of this study. For this study, the CWT for each plot was determined by the species with the most cubic foot volume per acre using decay class, species, and log dimension information (volume per unit area estimators) of individual CWD pieces.

FTs were determined by field crews based on visual observations of the plot (USDA Forest Service 2002). Because numerous FTs may be present on any selected phase 2 plot, the FT for the condition class occupying the greatest proportion of the plot area was selected. If two or more FTs occupied the same area proportion, the FT of the proportion with the most basal area was selected. Both FTs and CWTs were broadly assigned to the following FT/CWT groups: pine, spruce/fir, oak/pine, oak/hickory, elm/ash/cottonwood, maple/ beech/birch, and aspen/birch (Smith *et al.* 2001). To accomplish the second objective of this study, all study plots were stratified into two classes for analysis: (1) plots that had different CWTs and FTs, and (2) plots that had no difference in CWTs and FTs (the species composition of down dead tree biomass is roughly equivalent to the species composition of the standing live tree biomass).

Figure 1.—Line-intercept coarse woody debris sampling design for the Down Woody Materials Indicator of the USDA Forest Service FIA program.



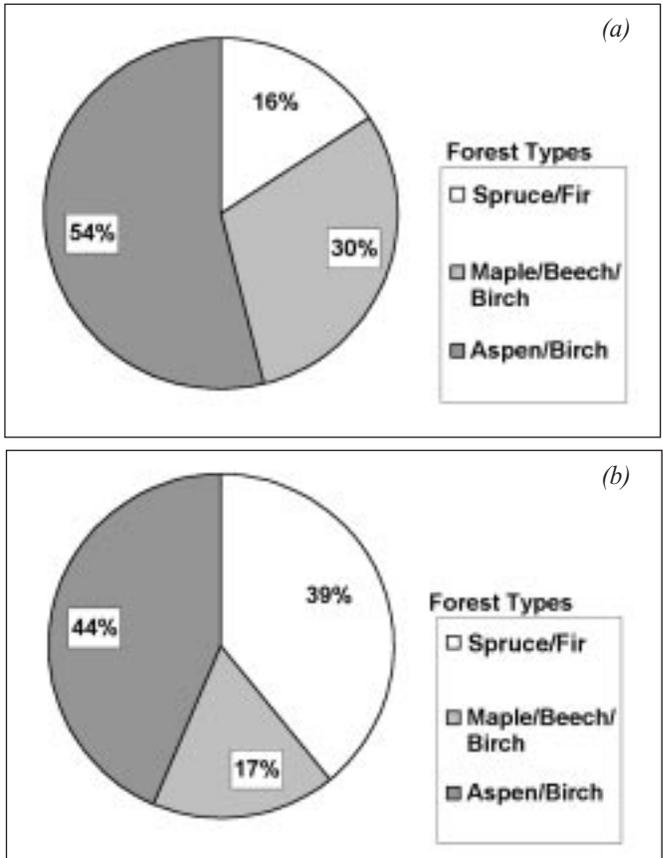
Results/Discussion

Preliminary determination of a CWT by using existing CWD data collected on FIA subplots provides an initial framework for developing a formal CWT algorithm. Many challenges exist to the development of a CWT algorithm using FIA data. First, the CWT for a forested plot may resemble no currently defined FT. For example, plots in the Southern United States may have a significant amount of large chestnut (*Castanea dentata*) down logs present on a plot dominated by standing northern red oak (*Quercus rubra*) trees. Therefore, in this case, the CWT would be chestnut, which no longer exists as a FT in the Southern United States. Second, a hierarchy of species identification may complicate typing algorithms. Field crews may readily identify the particular species of individual CWD pieces but may only identify other CWD pieces as unknown hardwood or conifer because of decay. Third, for decay class 5 logs, no species identification is possible because the logs are too decayed. For some plots, a majority of the CWD volume may be in decay class 5 and, thus, these null values would confound CWT efforts. Fourth, the effects and importance of CWD decay classes on the typing process need to be resolved. The species identification of a freshly fallen (decay class 1) CWD piece is more certain than the identification of a partially rotten (decay class 4) CWD piece. Fifth, latitude and climate may affect decay rates that may cause a spatial bias to CWT algorithms. Plots located in Minnesota or Wisconsin may have older logs of previous FTs that occupied the plot versus plots in Missouri where decay rates are faster with less chance of CWTs differing from that of current FTs. Thus, plots in more northerly latitudes or xeric sites may be more difficult to type. Finally, crew measurement error may affect CWD species identification in certain FTs. Some FTs, such as paper birch (*Betula papyrifera*), may have CWD that decays rather rapidly, while other FTs in adjacent areas may have CWD that is more resistant to decay. Therefore, field crews may have more uncertainty with species identification in paper birch forests than in other forests with more decay-resistant species such as black walnut (*Juglans nigra*).

For all 345 study plots, 52 percent displayed a difference between FT and CWT. The remainder of the plots (48 percent) showed no difference between CWT and FT. When the plots were examined in the context of three common FT groups of the Lake States (spruce/fir, maple/beech/birch, and aspen/birch),

distinct differences existed in FT and CWT comparisons between the conifer and hardwood FTs (figs. 2a–b). When considering the distribution of FTs between the two strata of difference/no difference, the proportion of plots in northern hardwood forests (maple/beech/birch and aspen/birch) that had a difference in FT and CWT (61 percent) was less than the proportion of plots with no difference in FT and CWT (84 percent) compared to spruce/fir FTs (figs. 2a–b). For spruce/fir forests, this trend was reversed: the difference in FT and CWT (39 percent) was greater than the proportion of plots with no difference in FT and CWT (16 percent) (figs. 2a–b). These results suggest that spruce/fir forests are more likely to have CWD of a different species from the FT than maple/aspen/birch forests, a result attributable to the regional maturation of aspen/birch FTs and understory development of more shade-tolerant climax spruce/fir forests (Kotar *et al.* 2002).

Figure 2.—Percentage of study plots by selected FT for North Central States (USDA Forest Service, FIA program) by study strata of (a) no difference between forest and coarse woody types and (b) differences between forest and coarse woody types.



For Lake State forests in particular, differences between FTs and CWTs may help elucidate successional and mortality trends occurring between maple/aspens/birch and spruce/fir forests. To determine if differences between FTs and CWTs were due to recent disturbances, the proportion of study plots in the two study strata were examined. As related to recent stand disturbances identified by field crews, 81 percent of plots that had a difference between FTs and CWTs showed no evidence of recent stand disturbances, and 83 percent of plots that had no FT and CWT differences had no recent stand disturbances. Due to the scarcity of recent stand disturbance events across a region, CWTs may not fluctuate in short time frames, especially within the sample plot sizes (24-ft radius) the FIA program uses. Rather, CWTs may potentially quantify species composition changes during extended years of stand development (FTs of the past, deceased forests in general). In addition, mature or old spruce/fir FTs are often maintained through small-scale, wind-, and disease-related gap dynamics (Frelich and Reich 1995), with remnant early successional species possibly comprising the CWD found on the forest floor. As stand development progresses in a northern hardwood FT, gap dynamics characterized by small windfall events perpetuate shade-tolerant species, such as sugar maple, that may have been present at stand initiation (Frelich 2002), resulting in the same CWT and FT over time. Both of these disturbance types are relatively small scale and may not be observed by field crews.

Mean proportions of plots having differences in FTs and CWTs among north-central States were examined (table 1). Northern latitude States (Minnesota, Wisconsin, and Michigan) showed more of a difference between FTs and CWTs than more southerly forests (i.e., Missouri). These results may be due to

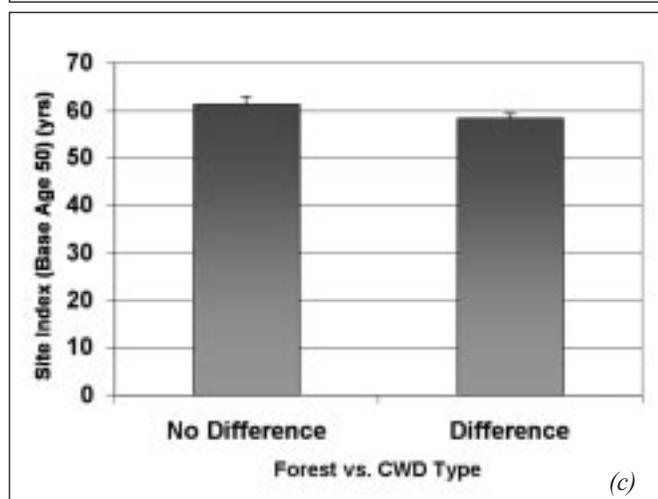
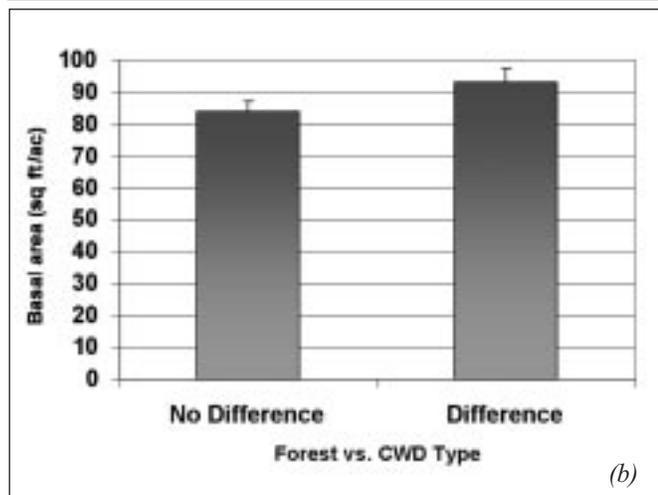
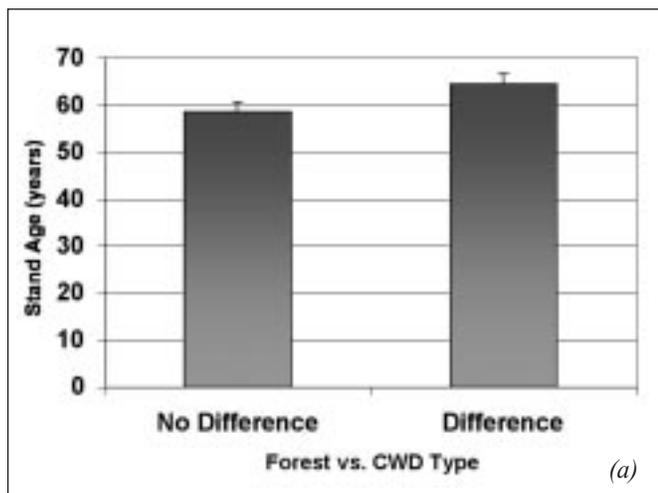
Table 1.—Percentage of CWD plots showing a difference in FT and CWT by North Central State (2001–02).

States	Plots with FT and CWT difference (%)
Indiana	7
Iowa	23
Kansas	33
Michigan	81
Minnesota	81
Missouri	34
Nebraska	13
Wisconsin	54

two reasons: (1) successional trends in spruce/fir forest climax types in Northern States, and (2) regional climatic gradients. For forests in high latitude/elevation and/or xeric regions of the United States, slow decay rates may preserve CWD pieces for decades thus exacerbating differences between down dead and standing tree species compositions. The results in table 1 also support the concept of successional shifts in Lake States forests causing differences in FTs and CWTs by FT. Shifts in CWTs and FTs, as suggested by results in this study, may not be related to recent stand disturbances but rather to long-term successional shifts. For FT groups in this study, the successional pathways of the hardwood forests of maples, aspen, and birches succumbing over time to developing spruce/fir forests may be evidenced by the prevalence of spruce/fir study plots having differences in their respective CWTs and FTs (figs. 2a–b).

The means and associated standard errors for stand-level variables of stand age (yrs), basal area (ft²/ac), and site index (base age of 50 yrs) were compared between the two study strata of FT and CWT differences/no differences. Plots that had a difference between standing live and down dead tree species composition were generally older stands, had greater basal area, and were on poorer quality sites than plots that had no difference in FTs and CWTs, although incorporation of summary statistics might alter those conclusions (figs. 3a–b). First, older stands (fig. 3a) are more likely to have disturbance and successional related mortality. These results may be justified by the fact that older stands have a longer time to accumulate CWD from a variety of species that may or may not be present in the current forest. Second, forests with greater levels of stand basal area (fig. 3b) may be more susceptible to density-related mortality. With greater levels of mortality over time, the greater the chance that the species composition of the CWD of a stand may not resemble the standing tree species. Third, forests on higher quality sites may have faster decay rates for CWD, less accumulation of CWD over time, and therefore less chance for a difference between FTs and CWTs. If a particularly high-quality site can grow trees faster (Assman 1970), the site may be able to grow more fungi and microbes to decompose CWD at faster rates. Overall, if stand and site attributes (density, site quality, or stand age) partially control the accumulation and decay of CWD, the hypothesis may be promulgated that examination of CWTs may indicate the past influence of stand/site attributes in forest stands.

Figure 3.—Means and associated standard errors for the stand-level attributes of (a) stand age, (b) basal area, and (c) site index for the study strata of no difference between forest and coarse woody types and differences between forest and coarse woody types.



Conclusions

Despite obvious difficulties and hurdles to developing CWD species composition typing algorithms, CWTs may afford inventory analysts with another categorical variable of analysis. Study results suggest that comparisons between FTs and CWTs may serve as an indicator of successional change at landscape scales. Additionally, FTs and CWTs may refine analysis of the complex relationships between stand/site factors and stand development. If the thesis statement—that CWD species composition indicates the historical mortality patterns of any particular stand—is correct, CWTs may afford opportunities to refine our understanding of CWD and its role as an indicator of forest health.

Literature Cited

- Assman, E. 1970. The principles of forest yield study. New York: Pergamon Press. 506 p.
- Bull, E.L.; Parks, C.G.; Torgersen, T.R. 1997. Trees and logs important to wildlife in the Interior Columbia River Basin. PNW-GTR-391. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 55 p.
- Daniel, T.W.; Helms, J.A.; Baker, F.S. 1979. Principles of silviculture. 2nd ed. New York: McGraw-Hill. 500 p.
- DeVries, P.G. 1986. Line-intersect sampling. In: Sampling theory for forest inventory: a teach-yourself course. Berlin: Springer-Verlag. Chapter 13.
- Eyre, F.H., ed. 1980. Forest cover types of the United States and Canada. Washington, DC: Society of American Foresters. 148 p.
- Frelich, L.E. 2002. Forest dynamics and disturbance regimes: studies from temperate evergreen-deciduous forests. Cambridge, UK: Cambridge University Press. 266 p.
- Frelich, L.E.; Reich, P.B. 1995. Spatial patterns and succession in a Minnesota southern-boreal forest. *Ecological Monographs*. 65: 325–346.

-
- Goodburn, J.M.; Lorimer, C.G. 1998. Cavity trees and coarse woody debris in old growth and managed northern hardwood forests in Wisconsin and Michigan. *Canadian Journal of Forest Research*. 28: 427–438.
- Harmon, M.E.; Franklin, J.F.; Swanson, F.J.; *et al.* 1986. Ecology of coarse woody debris in temperate ecosystems. *Advances in Ecological Research*. 15: 133–302.
- Helms, J.A., ed. 1998. *The dictionary of forestry*. Bethesda, MD: Society of American Foresters. 210 p.
- The H. John Heinz III Center for Science, Economics and the Environment. 2002. *The state of the nation's ecosystems: measuring the lands, waters, and living resources of the United States*. New York: Cambridge University Press. 288 p.
- Kotar, J.; Kovach, J.A.; Burger, T.L. 2002. *A guide to forest communities and habitat types of northern Wisconsin*. 2nd ed. Madison, WI: University of Wisconsin, Madison.
- Maser, C.; Anderson, R.G.; Cromack, K., Jr.; *et al.* 1979. Dead and down woody material. In: Thomas, J.W., tech. ed. *Wildlife habitats in managed forests: the Blue Mountains of Oregon and Washington*. Agric. Handb. 553. Washington, DC: U.S. Department of Agriculture: 78–95.
- Miles, P.D.; Brand, G.J.; Mielke, M.E. 2003. Minnesota's forest resources in 2002. *Resour. Bull. NC-225*. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 31 p.
- Pedlar, J.H.; Pearce, J.L.; Venier, L.A.; McKenney, D.W. 2002. Coarse woody debris in relation to disturbance and forest type in boreal Canada. *Forest Ecology and Management*. 158: 189–194.
- Smith, B.W.; Vissage, J.S.; Dar, D.R.; Sheffield, R.M. 2001. *Forest resources of the United States, 1997*. Gen. Tech. Rep. NC-219. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 190 p.
- U.S. Department of Agriculture (USDA) Forest Service. 1965. *Timber trends in the United States*. For. Res. Rep. No. 17. Washington, DC: U.S. Department of Agriculture, Forest Service. 235 p.
- USDA Forest Service, North Central Research Station. 2002. *Forest inventory and analysis national core field guide: Field data collection procedures for phase 3 plots*. Vol. II: 265–298.

Access and Use of FIA Data Through FIA Spatial Data Services

Elizabeth LaPoint¹

Abstract.— Forest Inventory and Analysis (FIA) Spatial Data Services (SDS) was established in May 2002 to facilitate outside access to FIA data and allow use of georeferenced plot data while protecting the confidentiality of plot locations. Modification of the Food Security Act of 1985 legislated the protection of information on plot location and ownership. Penalties were put in place for violations. Because of this change in the law, many customers have been served by FIA SDS, and demand for spatial analyses continues to grow. More than 130 requests for spatial data or other information have been received from academia, State and local governments, Federal agencies, and forest industry. This article describes how projects progress from inception to completion, including security concerns.

Introduction

Forest Inventory and Analysis (FIA) has been collecting data and reporting on the status of the Nation's forests for more than 70 years. Sample plots are located across the landscape and are revisited periodically. In recent years, plot locations have been recorded using Global Positioning System (GPS) technology, which has resulted in new uses for FIA data.

The Privacy Issue

With passage of the fiscal year (FY) 2000 Consolidated Appropriations Bill (Public Law 106-113), the Food Security Act of 1985 (7 U.S.C. 2276(d)) was modified, making it illegal to reveal information on FIA plot locations or ownership. Penalties for violating the law can include fines up to \$10,000 and/or a year in jail.

Even before the change in the law, FIA treated ownership information and plot location in a confidential manner for the following reasons: (1) to maintain goodwill because FIA relies on private landowners for access to their property, and they must understand that FIA is not concerned with any regulatory or taxation issues; (2) to eliminate unnecessary site visits that might damage or alter the plot; and (3) to ensure management decisions are not influenced by knowledge of a plot's location.

Background

FIA Spatial Data Services (SDS) was created in May 2002 to assist customers in accessing and using FIA data spatially within the bounds of existing legislation. FIA SDS also provides assistance with Geographic Information System (GIS) technology and in linking to non-FIA data, in addition to attempting to answer users' questions about the data. Although FIA SDS was established partly in response to the change in the privacy law, FIA's national management realized that users needed a single contact point for questions related to the FIA data confidentiality.

Almost one-third, 32 percent, of the requests for data have been from academia. Other consumers include other Federal agencies such as the Environmental Protection Agency and Bureau of Land Management as well as non-FIA U.S. Department of Agriculture Forest Service employees and State agencies. Academia and Federal agencies account for 63 percent of the data requests received in FY 2003.

Data Request Process

After customers initially contact FIA SDS, the data request is typically revised or refined for some period of time. Often, requesters are unaware of the kinds of data collected by FIA, how these data are collected, or the differences in data collection between States and FIA units and/or privacy issues related to the use of FIA data. Protecting the location of the plots and landowner privacy are the primary concerns of FIA SDS with

¹ Forester, Northeastern Research Station, U.S. Department of Agriculture, Forest Service, Newtown Square, PA 19073. Phone: 610-557-4049; fax: 610-557-4250; e-mail: elapoint@fs.fed.us.

every data request—for example, ensuring that the data provided cannot be back-engineered to reveal plot location or ownership information.

As part of refining a request, FIA SDS describes the relevant data to the requester with respect to the specific area of interest. Revising a data request, whether for security purposes or other reasons, can take weeks to months depending on the request's complexity, the requester's motivation level, the extent of communication between FIA SDS and the requester, and the work backlog at FIA SDS.

FIA data limitations are described while discussing the data request with the requester; for example, the variations in the quality of plot location data are pointed out. Originally, plots were georeferenced by digitizing their positions on aerial photographs or maps. These digitized positions are rarely as accurate as the GPS coordinates currently being collected for sample plots. Unfortunately, plot locations with no GPS coordinates still exist. GPS-derived plot locations also are subject to inaccuracies due to user error, signal degradation, or satellite configuration. Some of these problems should be resolved with the national standardization of GPS data collection methods and increased recognition of the importance of accurate locations.

Another limitation is that some areas of the Nation have few, if any, FIA plots. In the past, plots were not always established in wilderness areas or areas with no timberland, for example, western Texas and parts of California. Some areas, such as interior Alaska, had no plots established due to their remoteness.

Another issue that may affect a data request is that inventories in neighboring States may have occurred as many as 10 years apart, and, therefore, the data requested may not be current. However, FIA's recent transition to an annual inventory will result in more timely information.

After the data request has been refined and meets security requirements, the request is forwarded to the appropriate FIA program manager(s) for approval. Often, the security review by FIA SDS entails processing a portion of the data request and examining the results. Each FIA unit can refuse a data request or recommend additional refinement, use its own staff to process the request, or approve the request and redirect it to SDS for action.

When a request is approved, SDS completes its processing and forwards the results to the program manager(s) for final review. If the release of the data is approved, the results are

forwarded to the customer. If a request is denied, SDS will work with the customer to accept an alternative that best meets the customer's needs.

Revisions to the data request and communications among the requester, FIA SDS, and FIA management make up most of the time spent on a request. As data requests become more complex, the amount of time required to supply the data also will increase. Data requests that cover multiple States or FIA units can become more complicated due to differences in the data between States or units.

During FY 2003, 95 data requests were received, and an average of 5.5 requests were filled each month. Almost from its inception, FIA SDS has experienced a work backlog due to needed refinements in requests and/or communication time lags between SDS and the customer. In FY 2003, 67 percent of the requests received were filled within 4 weeks; the overall average fulfillment time was 7 weeks.

A customer can work with FIA SDS at its office at Newtown Square, PA, or submit a request at any regional FIA location.

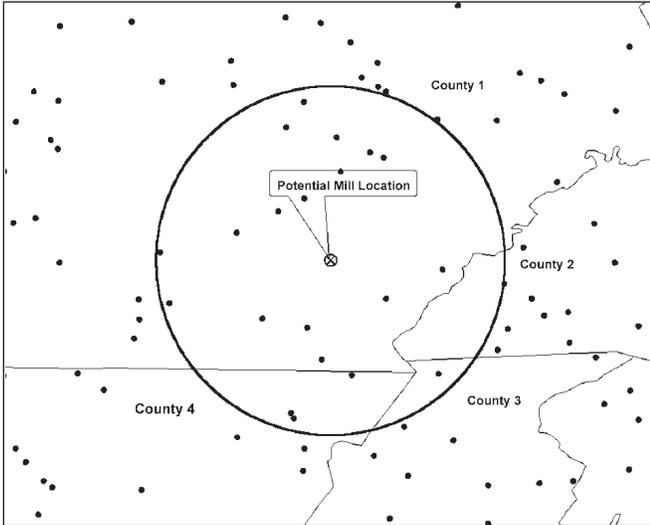
Examples of Requests

Simple Example

A common request is for a data summary for plots that fall within the requester's area of interest, for example, a summary of growth and removals within a given distance of a mill location. Because the data provided to the requester is summarized, no concerns exist related to disclosing information on plot location or ownership. The only security concerns are ensuring that the area of interest is of sufficient size to prevent plot location disclosure, and that the area of interest covers at least three private landowners. Current restrictions on FIA Mapmaker (http://ncrs2.fs.fed.us/4801/fiadb/fim_tab/wc_fim_tab.asp), an Internet-based data query and mapping application, prohibit circular retrieval with a radius less than 25 miles (1,256,637 acres). Generally, tables are created only if at least 12 forested plots fall within the area of interest; i.e., a circle with a minimum of 72,000 forested acres.

Creating tables of growth and removals at the county level can create security problems. Figure 1 depicts circular retrieval on a map containing county lines. If summarized data are provided

Figure 1.—Fifty-mile-radius circle retrieval around potential mill location. The single plot in lower right quadrant could reveal location or ownership information if county level summaries are provided. (Plots shown are fictitious.)



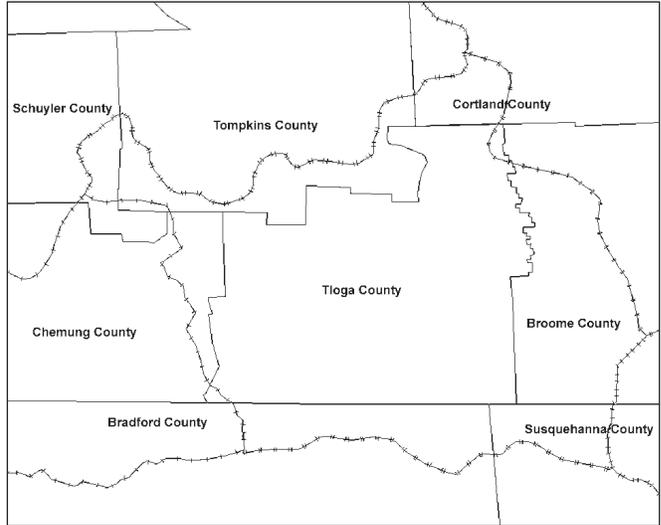
for the entire circle, the security issue is moot. If the customer wants county-level summaries, however, a security issue arises because overlaying the circular area onto the county layer creates a sliver polygon. The sliver may be small enough to locate the plot on the ground, or the owner of that area may be known. FIA SDS assumes that the customer has access to GIS technology and is able to overlay the circle onto the county layer.

Complex Example

A more complex request is one that entails overlaying plot locations onto a polygon layer to associate a polygon label or attribute with each plot—for example, the customer wants to assign a soil polygon to each FIA plot. The ideal solution is to provide summarized data for polygons and restrict the data to polygons with a minimum of three plots, which avoids creating sliver polygons. Figure 2 shows how overlaying a watershed layer onto a county layer can create polygons that would reveal plot location or ownership.

A possible solution to the security problem is to ensure that at least three privately owned plots exist in each polygon, although the three plots may be owned by the same person. Another solution is to make the polygon larger than common land holdings in the area of interest. This minimum size varies greatly between regions. For example, a polygon that encompasses

Figure 2.—Overlaying a watershed layer onto a county layer can create sliver polygons that can reveal plot location or ownership information; i.e., the southeast corner of Schuyler County and the northwest corner of Susquehanna County.



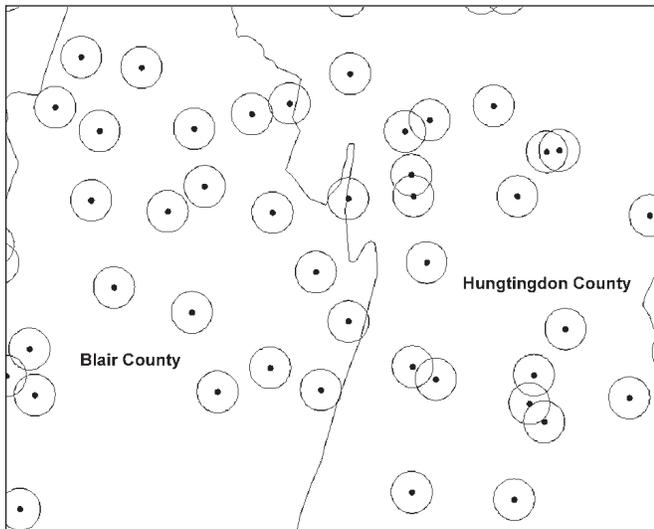
100 acres on the coast of Rhode Island likely will contain the required three private owners, but that same parcel size may be woefully inadequate in north-central Maine. Ideally, FIA SDS would provide data summarized for the polygons rather than provide the polygon ID for each FIA plot.

FIA SDS also handles requests involving remotely sensed imagery or other raster data. These are often data requesters sending in satellite imagery they have previously classified, SDS overlays the FIA plots and then returns information on the classes of the imagery. These requests are similar to the previous example with soil polygons. Providing summarized information for each class, or ensuring both a minimum of three privately owned plots in each class/county combination and that each class/county combination covers enough acreage to prevent disclosure of plot locations, avoids security issues.

Fuzzing and Swapping

To make using spatial data more accessible to customers, the FIA database has coordinates available for downloading. The coordinates associated with the plots are altered to protect ownership information and prevent locating plots on the ground.

Figure 3.—Plots with a 1-mile-radius circle depict the area from which a new “fuzzed location” would be randomly selected. (Plots shown are fictitious.)



The process of altering the plots entails “fuzzing” the plot locations and then swapping coordinates for a certain percentage of plots. Fuzzing involves creating a buffer area of 0.5 to 1.0 miles around each plot and randomly selecting a point within that circle as the “new” coordinate for that plot (fig. 3). This procedure is performed for all plots to prevent users from locating the true plot locations.

To protect landowner privacy, the location coordinates of up to 20 percent of privately owned forested plots are swapped with similar plots in the same county or supercounty (aggregation of two or more adjacent counties in the same State). For example, if plots A and B are selected as a swapping pair, plot A’s data will be assigned to plot B’s location, and plot B’s data will be assigned to plot A’s location. The plots are determined to be similar based on criteria established by each FIA unit. Plots usually will not be swapped outside their county.

Customers can download the fuzzed and swapped data and perform their own spatial analyses. In some cases, data requests are not significantly affected by using fuzzed and swapped coordinates, for example, when evaluating mill locations. In others, customers may want to use the fuzzed and swapped coordinates to examine and refine their data needs before submitting a data request.

Summary

FIA SDS has made great strides in improving access to FIA’s spatial data. FIA SDS was begun as a pilot project; since then, however, the need for this service has been demonstrated. By establishing FIA SDS, FIA has shown a commitment to its customers. As FIA’s commitment to spatial products continues to grow, so will the variety of spatial tools and spatial data available to FIA customers.

The Open-Source Movement: An Introduction for Forestry Professionals

Patrick Proctor, Paul C. Van Deusen¹, Linda S. Heath, and Jeffrey H. Gove²

Abstract.—In recent years, the open-source movement has yielded a generous and powerful suite of software and utilities that rivals those developed by many commercial software companies. Open-source programs are available for many scientific needs: operating systems, databases, statistical analysis, Geographic Information System applications, and object-oriented programming. Using “real world” examples, including applications employed by Federal agencies, we address the concerns associated with open-source software deployment: cost, security, software availability, and usability. The potential for application to U.S. Department of Agriculture Forest Service Forest Inventory and Analysis data is discussed.

The growing availability of open-source software is causing many businesses and organizations to consider its adoption. Open-source software has advanced to the point where it has become a viable alternative. “Open source” does not just mean free software that is distributed with its source code. For software to be considered open source, it must comply with 10 criteria of the Open Source Definition (Open Source Initiative 2004). The Open Source Initiative, a registered nonprofit organization, broadly oversees the certification of software distributed under a license that conforms to the Open Source Definition. This article will explore the nature of open source, compare it with similar proprietary corporate platforms, and address many of the concerns voiced by today’s information technology (IT) user. We believe the merits of open source allow for a formidable and attractive platform.

The Open-Source Philosophy

Raymond (2000) describes the major differences in the development paradigms between closed- and open-source software. He compares the former to the building of a cathedral, where the design, progress, and management of a software project are conducted under strict regiment in a group that is closed to non-members. Such models normally apply to corporate projects, although in the past they have also been applied to open-source software projects. By contrast, the development of the popular GNU Emacs editor (Free Software Foundation 2003a) exemplifies the open-source approach. This latter model is compared to a bazaar, which seems at first appearance to be chaotic and uncontrolled, but when the model is viewed with scrutiny, it more closely resembles the working of a diverse yet controlled system. Linus Torvalds was the first to popularize this open-source model with his “release early and often, delegate everything you can” (Raymond 2000, 2) philosophy. Torvalds is the creator of Linux, currently the most accessible and widely used open-source operating system. In this developmental model, users are often also contributors. One of the major keys to success of such ventures is that people contribute not because they were assigned to but out of love for the project.

Central to the open-source model and considered the core difference between the cathedral and bazaar models is Linus’s Law: “Given enough eyeballs, the bugs are shallow” (Raymond 2000, 9). In the cathedral model, bugs are insidious and often difficult to correct, if found at all, because of the limited number (and often high turnover) of programmers with access to the code. The bazaar model, however, draws on the talents of often thousands of “hackers”; with such a base to draw from, an insidious bug becomes something simply fixed not by the group as a whole but by the one or two people out of the many with the specific talent. “Release often,” then, becomes the vehicle for rapid development and evolution toward an unbreakable system.

¹ Supervisory Research Forester and Research Forester, respectively, U.S. Department of Agriculture, Forest Service, PO Box 640, Durham, NH 03824. Phone: 603-868-7612; e-mail: ltheath@fs.fed.us.

² Supervisory Research Forester and Research Forester, respectively, U.S. Department of Agriculture, Forest Service, PO Box 640, Durham, NH 03824. Phone: 603-868-7612; e-mail: ltheath@fs.fed.us.

The best example of the bazaar model, as Raymond (2000) points out, is Linux itself. The Linux platform is available in a number of “distributions” made by various software groups or companies that include Red Hat, Debian, and Yellow Dog. Although groups may package and sell the code, the source code is free and available to be compiled, and contributions are considered and encouraged from all. The Linux kernel—which is stable and often termed unbreakable—can scale from embedded devices (Embedded Linux Consortium 2003) to clusters running at supercomputer speed, including the world’s third fastest supercomputer as of June 2003 (TOP500 2003). Finally, as evidenced by the visionary GNU Project (Free Software Foundation 2003b) and the thousands of tools produced directly by members of the Free Software Foundation or under the GNU General Public License (Free Software Foundation 2003c), a large community of users have based their work on a Linux platform.

Security

As expected, security is a primary concern when switching to an open-source platform. The security measures available in open-source operating systems are comparable to those available in proprietary, closed-source operating systems such as Microsoft Windows (Microsoft 2004). Remote access to machines is controlled by a series of “ports,” each of which is assigned to a particular function (e.g., HTTP, FTP, Telnet). Access to these ports, in turn, is controlled by a firewall that blocks outside users and illegal ports. This user control (available in Linux distributions and in Microsoft Windows versions 2000 and XP only) is accomplished through a user name/password-based access system, which requires users to be verified by a system administrator before gaining access.

Because Microsoft Windows is the most used desktop operating system in the world, its exploitation by hackers is more likely for a number of reasons. First, more users in the form of individual desktop systems exist to “attack,” which makes an attractive target. Also, viruses and worms can spread more rapidly because of the large user base. Second, on Microsoft Windows systems, software, such as web browsers, are allowed to run scripts that, if the author is clever enough, can directly

access the operating system files—something that is not allowed on open-source Linux. Third, patches must go through a corporate testing and clearance process before being released to the public. This results in a long lag time until a resulting virus “cure” is built into the system itself. Typically, a patch, when finally released, is available exclusively through Microsoft servers. The code cannot be checked by outside sources because of its unavailability to the general public, and the reliability of the patch is based entirely on internal Microsoft control mechanisms. Although Microsoft has a full staff of software testers and security analysts, hackers consistently exploit Windows system vulnerabilities before these “holes” are discovered internally. Some recent examples include the Blaster and SoBig viruses (Cable News Network 2003). Unfortunately, Microsoft provides no means for users to assist in solving this problem other than to be aware of and follow Microsoft advisories. If Windows users want to address these security concerns, they often are required to look to third-party providers.

As mentioned above, the open-source community has a far less restrictive management system for vulnerabilities. Bugs are often discovered and patched by any of the numerous users involved in open-source development. Before code is put into practice, the code is checked and rechecked by a literally worldwide network of developers. Patches are quickly and freely distributed to anyone who wants them. Because the patches are open source, they can be hosted on any server, provided the server abides by the GNU General Public License (Free Software Foundation 2003c). The open-source community is always searching for new vulnerabilities, and community groups, such as the Linux Security Audit Project, exist for the sole purpose of finding and patching Linux vulnerabilities (Linux Security Audit Project 2003). In addition, efforts such as the National Security Agency’s Security-enhanced Linux project (National Security Agency 2003) provide even more protection if desired (Coker 2003).

Although debates occur about which distribution and patching system is more efficient or desirable, open-source solutions are in no way less secure than their proprietary counterparts. They clearly offer a well-documented and tested security alternative to proprietary operating systems.

Cost

Cost often is cited as a significant factor in the success of the open-source movement. Although prices of retail software continue to rise, open-source software remains entirely free or affordable to license and install. As table 1 shows, a number of retail closed-source packages have open-source counterparts, and the savings in using them can be immense (Newegg 2003). Although a cost advantage clearly exists to using open-source products, the argument can be made that the savings in retail cost is eclipsed by the time cost of retraining employees on new and/or unfamiliar applications.

Five to 10 years ago, when Linux was largely text-based, training users may have been costly. The Linux user interface has been redefined to be accessible to any user, however. A number of graphical interfaces are available to choose from, e.g., GNOME (GNOME Foundation 2003) and KDE (KDE e. V. 2003), all of which draw on industry-standard interfaces as their inspiration. Any user familiar with the Microsoft Windows operating system's graphical user interface could switch to the current Linux environment and find similar functionality. The same holds true for vital applications such as office productivity and photo-editing programs. Linux user interfaces will be familiar to Microsoft Windows users, and they also feature extensive online help. Also, databases based on the structured query language (SQL) must adhere to the SQL standard. Queries and databases written for retail programs, such as those from Oracle, can easily be migrated to the popular open-source database MySQL (MySQL AB 2003, Oracle 2004). Developers familiar with Oracle database products will find MySQL to be a similar, if not almost identical, environment. One major corporation that made the switch to open source was the Ernie Ball Guitar String Corporation. Ernie Ball's CEO, Sterling

Ball, disputed analysts' predictions of tremendous cost and user transition difficulties when migrating from Windows to Linux:

It's the funniest thing—we're using it for e-mail client/server, spreadsheets and word processing. It's like working in Windows. One of the analysts said it costs \$1,250 per person to change over to open source. It wasn't anywhere near that for us. I'm reluctant to give actual numbers. I can give any number I want to support my position, and so can the other guy. But I'll tell you, I'm not paying any per-seat license. I'm not buying any new computers. When we need something, we have white box systems we put together ourselves. It doesn't need to be much of a system for most of what we do. (Becker 2003)

Availability

One possible downside of open-source software is its lack of retail availability. Although more popular open-source packages are becoming available in stores and catalogs, most open-source software must be downloaded from the Internet. This often requires a high-speed connection or a long time waiting for downloads to complete. As high-speed Internet access continues to proliferate, this issue is becoming less of a problem. In fact, the online availability of open-source programs is actually becoming a benefit: no packaging materials are used, no shipping time is required to get the latest version of a program, and no money is wasted on programs that do not meet the user's needs. For users with high-speed connections, the available delivery mechanisms, such as apt-get (Chiba Industries 2003), RPM (RPM Community 2002), or yum (Duke University 2002), are superior to those of their retail counterparts.

Table 1.—*Popular closed- and open-source software packages and their retail prices (Newegg 2003).*

Closed source	Price (\$)	Open source equivalent	Price (\$)
Microsoft Windows 2000 Server	870	Linux	0
Adobe Photoshop	565	GIMP	0
Oracle (1 computer)	15,000	MySQL	0
Microsoft Office XP	297	Open Office	0
Total:	16,732	—	0

Code Accessibility

As mentioned in the first section, because of the open-source principles, code for open-source projects is freely available. Source code is the software component that is readable by humans before it is compiled into machine-readable code. Source code, considered intellectual property, is the component of software to which software copyrights apply. In retail products, source code is not openly available. For users adept at programming, being able to view the source code offers many distinct advantages: bugs can be fixed, features can be added, modules can be enhanced, and security features can be checked by outside sources.

“Real World” Applications

The real world applications of open-source software are numerous and diverse. Organizations and individuals are adopting the open-source platforms for a number of reasons: costs are reduced, capacity for customization is increased, licensing maintenance is eliminated, and security is easily maintained. A short list of organizations that use open source indicates the widespread acceptance of the technology. The following is a list of organizations that have given open source a central role:

- Amazon.com (Adelson 2002).
- Toyota USA (IDC 2001).
- Massachusetts Institute of Technology.
- Harvard University.
- U.S. Department of Energy (Weiss 2001).
- U.S. Navy (Orlowski 2003).

Each of these organizations cited reasons along the lines of those previously mentioned for switching to open-source software. Security, cost, software availability, and customization were all contributing factors. In some of these cases, immediate cost savings were as high as \$17 million (Adelson 2002). In the case of the U.S. Navy, the open-source code enabled the security customization required for specialized projects aboard nuclear submarines (Orlowski 2003). The Department of Energy has used open-source programs to create clustered supercomputers at an affordable price (Weiss 2001). On college campuses,

open-source software enables students to work with the source code and generally function on the leading edge of technology.

These real world success stories also are contributing to the viability of open source as a retail offering. Many hardware retailers, including Dell, IBM, and Target, are offering open-source-based hardware solutions to their customers. These solutions can range from “clean” systems with no retail software installed to default open-source installations to customized open-source platforms created for customers. These examples and the increasing demand for availability clearly indicate open-source software’s success.

Application of Open Source to Forest Analysis

Where possible, gradually replacing corporate software packages with their open-source counterparts would be a beneficial and exciting option. The result would be a decrease in cost, an increase in security and stability, and a more flexible computing environment. The easiest initial change would be to upgrade servers to open-source software. They could continue to interface with Microsoft Windows desktops for file sharing through Samba (Samba Team 2003) and act as servers for various FIA operations. This change would be largely transparent to the end user, especially because Forest Service servers are currently Unix-based. It would yield numerous benefits for the organization. Funding could be saved on software licenses for Oracle database software, Microsoft Windows operating systems, and other retail software. Additionally, use of the Linux kernel increases server stability and eliminates viruses, worms, and Trojan horses written to exploit Microsoft system and application vulnerabilities. Although upgrading systems to open source can be a significant and possibly daunting step, it can decrease IT overhead for an entire organization. Such an upgrade also establishes a niche at the forefront of a movement on the verge of changing the world of computing forever.

A clear example of open source being implemented successfully in a forest analysis project is Carbon On-Line Estimation (COLE) (Proctor *et al.*, in press). For this project, open-source development tools and practices are used exclusively. The result is a comprehensive data analysis solution produced at a fraction of the cost of using retail tools. Additionally, as COLE comes into

its own, it, too, will become a registered open-source project. This step will allow other developers to contribute to the development of COLE and enhance it to suit their own research. In short, the open-source development cycle will come full cycle.

Conclusion

The open-source movement is a useful and viable option in today's computing world. In nearly all areas, open source either meets or exceeds the features and quality of proprietary retail software. Most importantly, open source presents owners and managers with an alternative that alleviates many of the problems that currently plague the IT infrastructure of many organizations: security, licensing costs, viruses, and scalability. Perhaps the advantages of open source are best summarized by Sterling Ball on his company's transition to an all open-source office:

I'm not making calls to Red Hat (Linux) [for support]; I don't need to. I think that's propaganda.... What about the cost of dealing with a virus? We don't have 'em. How about when we do have a problem, you don't have to send some guy to a corner of the building to find out what's going on—he never leaves his desk, because everything is server-based. There's no doubt that what I'm doing is cheaper to operate. The analyst guys can say whatever they want. (Becker 2003)

The open-source revolution is clearly becoming a dominant force in computing, and the more its user base increases, the more it will gain power. Only time will tell if organizations will have the vision to take this powerful option to the next level.

Acknowledgments

This work was made possible by a grant from the USDA Forest Service Forest Health Monitoring Program.

Literature Cited

Adelson, A. 2002. Amazon.com: migration from UNIX to Red Hat Linux. Framingham, MA: IDC. http://www.redhat.com/whitepapers/services/Amazon_case_study.pdf. (26 September 2003).

Becker, D. 2003 (20 August). Rockin' on without Microsoft. CNET Networks. <http://news.com.com/2008-1082-5065859.html>. (26 September).

Cable News Network. 2003. Microsoft offers virus bounty. <http://www.cnn.com/2003/TECH/biztech/11/05/microsoft.bounty/index.html>. (5 November).

Chiba Industries. 2003. Unofficial APT repositories. <http://www.apt-get.org>. (26 September).

Coker, F. 2003. Kernel Korner—NSA security enhanced Linux. Linux Journal. July 7. <http://www.linuxjournal.com/article.php?sid=6837>. (26 September).

Duke University. 2002. yum, Yellow dog updater. <http://linux.duke.edu/projects/yum>. (26 September).

Embedded Linux Consortium. 2003. Embedded Linux Consortium home page. <http://www.embedded-linux.org>. (26 September).

Free Software Foundation, Inc. 2003a. GNU Emacs - GNU project. <http://www.gnu.org/software/emacs/emacs.html>. (26 September).

Free Software Foundation, Inc. 2003b. The GNU operating System. <http://www.gnu.org>. (26 September).

Free Software Foundation, Inc. 2003c. GNU general public license. <http://www.gnu.org/licenses/licenses.html>. (26 September).

GNOME Foundation. 2003. The GNOME project. <http://gnome.org>. (26 September).

-
- IDC. 2001. Toyota Motor Sales USA: Linux across the enterprise. http://www.redhat.com/whitepapers/services/IDC_Toyota_CS.pdf. (26 September 2003).
- KDE e. V. 2003. The KDE project. <http://www.kde.org>. (26 September).
- Linux Security Audit Project. 2003. Linux security audit project. <http://lsap.org>. (26 September).
- Microsoft, Inc. 2004. Microsoft Windows product home page. <http://www.microsoft.com/windows/default.msp>. (26 September).
- MySQL AB. 2003. The world's most popular open source database. <http://www.mysql.com>. (26 September).
- National Security Agency. 2003. Security-enhanced Linux. <http://www.nsa.gov/selinux/index.html>. [Date accessed unknown].
- Newegg. 2003. Price listings. <http://www.newegg.com>. (26 September).
- Open Source Initiative. 2004. Open Source Initiative home page. <http://www.opensource.org>. (13 January).
- Oracle Corporation. 2004. Oracle home page. <http://www.oracle.com>. (26 September).
- Orlowski, A. 2003 (6 August). Navy buys Linux on Apple kit. The Register. <http://www.theregister.co.uk/content/39/32211.html>. (26 September).
- Proctor, P.; Heath, L.S.; Van Deusen, P.C.; Gove, J.H. [In press.] COLE: A web-based tool for interfacing with forest inventory data. In: Proceedings of FIA science symposium; 2002 November 18–20; New Orleans, LA. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.
- Raymond, E.S. 2000 (August). The cathedral and the bazaar. <http://catb.org/~esr/writings/cathedral-bazaar/>. (26 September 2003).
- RPM Community. 2002. The RPM package manager. <http://www.rpm.org>. (26 September 2003).
- Samba Team. 2003. Welcome to the Samba web pages. <http://www.samba.org>. [Date accessed unknown.]
- TOP500. 2003. TOP500 supercomputer sites. <http://www.top500.org>. (20 October).
- Weiss, T.R.; Vijayan, J. 2001. Compaq, DOE, biotech firm to build \$150 million Linux supercomputer. Computerworld. <http://www.computerworld.com/softwaretopics/os/story/0,1080,56666,00.html>. (26 September 2003).

A Knowledge Base for FIA Data Uses

Victor A. Rudis¹

Abstract.—Knowledge management provides a way to capture the collective wisdom of an organization, facilitate organizational learning, and foster opportunities for improvement. This paper describes a knowledge base compiled from uses of field observations made by the U.S. Department of Agriculture Forest Service, Forest Inventory and Analysis program and a citation database of more than 1,400 bibliographic entries from the past quarter-century. This synthesis provides highlights of early novel uses from the 1930s through 1976, suggests evolving approaches toward comprehensive assessments, and refers to the usefulness of forward-looking efforts to document the types of users, available attributes requested, and information in demand.

Introduction

The challenges of assessing forest lands for their ability to provide products, services, and values to an increasingly diverse society have grown progressively complex. Integrated knowledge is essential when selecting relevant attributes for measurement and common procedures for data collection, management, and analysis. For any organization concerned with efficient collection and distribution of data about field observations, a strategic business plan that considers the multiple processes involved in addressing current and satisfying future customer needs will be necessary.

Knowledge management, a formal term with many definitions (Full Circle Systems 2003), provides a way for an organization to capture the collective wisdom about such processes, facilitate appropriate responsiveness to challenges, and foster innovation. Feedback from customers in the form of documented attributes of interest, the kinds of analysis requested, and the multiple and varied interpretations of data provides some of the knowledge

needed for long-term planning. The same is true for public agencies, whose supporters include not only the customary end users of data, but also legislators, nongovernmental organizations, businesses, and individuals. An agency's decisionmakers need information about promising new ventures when funding is increased or may need to take cost-cutting actions and periodically reassess priorities in years of lean funding.

This brief synthesis is intended to facilitate organizational learning of U.S. Department of Agriculture Forest Service Forest Inventory and Analysis (FIA) program staff, affiliates, and potential cooperators. Recent efforts now being used to capture what data are being used, what issues the data are addressing, and the FIA program's knowledge of data uses and users are problematic. The paper highlights a retrospective compilation of the last quarter-century's reports that used FIA-based field observations for novel uses (Rudis 2003a). Included are early efforts involving nontraditional uses, other disciplinary perspectives, and evolving approaches to conducting comprehensive forest resource assessments. Recent findings and new opportunities to assemble knowledge of unpublished data uses and users also are described.

Early Milestones

A search of the literature on FIA-associated data reveals an evolving program (Rudis 2003a, 2003b). When FIA surveys were initiated in the 1930s, the chief goal was to identify timber resources, such as lumber and naval stores. But almost from the beginning, a broader audience was attracted to the information provided, particularly land use, forest land area, and forest types. Not long after the first reports came out, this single-purpose forest survey became a source of spatial information for an array of users. Map displays always have been a popular cross-disciplinary feature of forest survey reports and continue to this day to serve a diverse audience.

In the 1950s, a second generation of reports included county-based representations of otherwise tabular FIA data,

¹ Research Forester/Landscape Ecologist, U.S. Department of Agriculture, Forest Service, Southern Research Station, Forest Inventory and Analysis Unit, 4700 Old Kingston Pike, Knoxville, TN 37919.

including tree distributions of individual species for Mississippi (Sternitzke and Duerr 1950), timber supplies for Florida counties (Larson 1952), and hickory timber volume in the South (Cruikshank and McCormack 1956). The audience for the data remained diverse, but reports of the time focused primarily on directly measured attributes of timber supply.

A decade later, the American Forestry Association sponsored a series of regional, community-based assessments of forest management across the country. This effort produced three book-length (300-plus page) reports that synthesized biological and physical attributes, ownership patterns, and geopolitical contexts for three regions, as represented by three States: California (Dana and Krueger 1958), Minnesota (Dana *et al.* 1960), and North Carolina (Pomeroy and Yoho 1964). Information assembled came from the forest survey as well as a much wider range of sources than was common in later decades. These reports served and may continue to serve as models of an accomplished synthesis from vastly different disciplines and sources. Although dated, they remain a treasure trove of information for people who want to compare historic land use, forest ownership patterns, and land management practices across regions.

Beginning in the 1960s, the forest survey began to expand into other disciplinary arenas. The then-pioneering concept of a “multipurpose” inventory focused on the feasibility of combining deer browse inventories with forest surveys in the Southeastern United States to address wildlife management concerns (Moore *et al.* 1960). In the 1970s, understory plants were used to identify potential wood productivity in the Pacific Northwest (MacLean and Bolsinger 1973) and forest range resources in the south-central United States (Pearson and Sternitzke 1974). Other reports were generated to document tree damage agents; e.g., laminated root rot (Gedney 1976). At the same time, growing recognition that more sociological information was needed to assess the availability of timber for harvest led to coordinated studies of nonindustrial owner intentions in the Northeastern United States (Kingsley and Finley 1975).

By the mid-1970s, laws were enacted requiring comprehensive forest resource assessments, which included reports of associated social issues and related resources such as range, recreation, water, and wildlife habitat. Efforts varied widely by region and are reported elsewhere (Rudis 1991, 2003a, 2003b).

Evolving Approaches Toward Assessments

Varied approaches toward forest resource assessments have been taken, and many were intended to be comprehensive. Rudis (2003a) provides details, but in brief, such approaches have ranged from those designed by (1) a singular discipline with a single data source for a single-discipline audience to address a single purpose, (2) a representative team of selected disciplines with a limited array of data sets to focus on a specific issue or topic, or (3) multiple disciplines and data sources to address a selected range of objectives. In the past quarter-century, FIA assessments have evolved from the first approach principally by making efforts to reach a broader audience.

Individual scientists and teams in selected disciplines also have made use of publicly available FIA data to address specific issues in subject matter journals; e.g., for modeling biogenic emissions (Wiedinmyer *et al.* 2001), urbanization of forest ecosystems (Kline *et al.* 2001), and conducting regional assessments of early successional habitat for wildlife (Trani *et al.* 2001). In recent years, awareness of the FIA program has reached the point where its data are cited commonly in national studies of forest resource issues; e.g., stewardship of private forest land (Best and Wayburn 2001). The prominence of studies that synthesize FIA data with other data sets cannot be overemphasized. Such studies often surface in widely read interdisciplinary journals; e.g., *Science* (Caspersen *et al.* 2000), newspapers, or other popular media.

Adapting and incorporating data and knowledge from other resource inventories and disciplines are hallmarks of a truly comprehensive forest resource assessment. The extensive time required to align data from disparate inventories and communicate relevant knowledge among scientists in other disciplines, however, is a common problem in preparing such assessments (Rudis 1993). Multidisciplinary forest assessments that focus on specific regions or issues are popular approaches toward streamlining the development of an integrated data set and an interdisciplinary synthesis.

An approach toward such an undertaking for an environmental analysis of land cover and land use in the early 1980s produced one of the first integrated data sets, now known as the GEOECOLOGY database (Olson *et al.* 1982). A landmark, multidisciplinary, team-oriented scientific effort conducted in

the 1990s assessed timber harvesting in the State of Minnesota (Jaakko Poeyry 1994), with support provided to collect additional data and analysis to fill in some of the then-recognized knowledge gaps. Narrowing the scope and streamlining data integration to complete the task in a time-efficient manner reduces the burden of commitment by individual team members and facilitates timely reporting of results. One recent approach in Arkansas employed a 6-month maximum analytical time frame and a series of reports on new FIA survey data by invited experts with different perspectives (Guldin 2001). Another approach relied on a 1- to 2-year analysis primarily of existing data, models, or published studies, and a multiple-team synthesis, e.g., the Southern Forest Resource Assessment (Wear and Greis 2002).

In all these approaches, common challenges are the limited time available to fill strategic cross-disciplinary information gaps and the paucity of protocols for modeling and analyzing data from several disciplinary perspectives. One interdisciplinary need, for example, is a way to link ecological land type classification systems with timber growth (Song 1994). Addressing such a seemingly intractable problem often is left to imaginative early-career researchers, most notably graduate students. I compiled and indexed abstracts of known graduate student reports (Rudis 2003b). I highlighted an array of new approaches to analyzing FIA data, which integrates the data with other relevant data sets or enables viewing the data through the lens of other disciplinary perspectives or concepts.

Knowledge of Data Users and Uses

Efforts to document uses of FIA data began in 1989 with an informal query of nontraditional uses; however, expanded efforts found many more novel uses. Over time, the list of citations has been updated as an online citation database (Rudis 2002-04). Citations include reports of studies that used FIA's regional, field-sample-based forest surveys, as well as graduate student reports, collected works, and selected documents concerning integrated assessments and multidisciplinary surveys. The list also includes representative timber resource assessments since

1975. The primary focus of this database is on nontraditional and original technical uses associated with FIA data from 1975 through 2001. Recent citations also include entries that reference other data collected on FIA plots from sampling protocols established by the Forest Health Monitoring Program (Mangold 1998).

To obtain knowledge of data uses that may not be associated with publications, current sources of information include tallies of data requests made through FIA customer service centers, which includes requester data. Requests for National FIA Spatial Data Services (<http://www.fs.fed.us/ne/fia/spatial/request.html>) indicate the types of customers that request spatial data retrievals. LaPoint (2005) noted that the largest group of requesters for this data in fiscal year (FY) 2003 was from academic institutions.

The complete list of the groups and percentages follow:

- Individuals from academic institutions—32 percent.
- Other Forest Service personnel—15 percent.
- Other Federal personnel—16 percent.
- Other State personnel—13 percent.
- FIA staff—7 percent.
- Forest industry—7 percent.
- Nongovernmental organizations—4 percent.
- National Forest System—1 percent.
- Others—5 percent.

For the same period, the Southern Research Station FIA recorded that 16 percent of data requests came from universities and a similar number came from environmental groups.² This type of data, when considered with additional information about the types of requesting organizations, attributes used, and periodic tracking by year, may provide valuable feedback for decisionmakers to discern topical issues, set data collection, analysis, and distribution priorities, and modify or retain attributes frequently requested by such users.

The Internet server that maintains the FIA MapMaker (http://www.ncrs2.fs.fed.us/4801/FIADB/fim_tab/wc_fim_tab.asp) is a Web-based application for generating tables and shaded maps, as well as a potential source of information about both FIA data uses and users. Security software automatically records the Internet Protocol address and domain name of the

² Carol Perry, Customer Service Representative, U.S. Department of Agriculture, Forest Service, Southern Research Station, Forest Inventory and Analysis Unit, 200 W.T. Weaver Blvd., Asheville, NC 28804. E-mail to author, April 16, 2004.

user, and can be programmed to tally the attributes requested. Domain names by themselves are not definitive, but they do provide clues to the broad categories of users. In FY 2002, the categories of domain names and percentages of individual accesses were the following (Miles 2002):

- Forest Service (fs.fed.us)—34 percent.
- Commercial firms—19 percent (forest industry—7 percent, other or unknown—12 percent).
- AOL.com—2 percent.
- Miscellaneous (net—12 percent).
- Academic institutions (edu)—12 percent.
- Government (gov) and nonprofit organizations (org) combined—1 percent.
- Other unknown—20.

Figure 1 illustrates the top 10 attributes requested from the FIA MapMaker Web site for FYs 2002 (Miles 2002) and 2003 (Miles 2003), other than county and State. A quick glance suggests that stand size and forest type are most requested, and that there has been some change in the frequency of attributes requested. Interpretation of figure 1 should proceed with caution, however, as attribute requests may be closely tied to the organization and availability of choices presented on the Web site.

For future planning, customer service requests from FIA

MapMaker, other FIA-sponsored Web sites, and other customer service centers may supply insight into current and changing interests for already available online data, as well as information in demand but not currently available. Software may be applied to the discovery of knowledge from extensive records of customer requests. Cooley *et al.* (2000) provides an overview of the terminology and references for techniques to analyze Web user activities. Cooley *et al.* (1999) suggests initial data preparation of Web server data logs is a key to obtain more sophisticated information.

Acknowledgments

Carol Perry and Dennis Jacobs of the Southern Research Station, Elizabeth LaPoint of the Northeastern Research Station, and Pat Miles of the North Central Research Station reviewed an earlier draft and offered several valuable suggestions. Their help is greatly appreciated.

Literature Cited

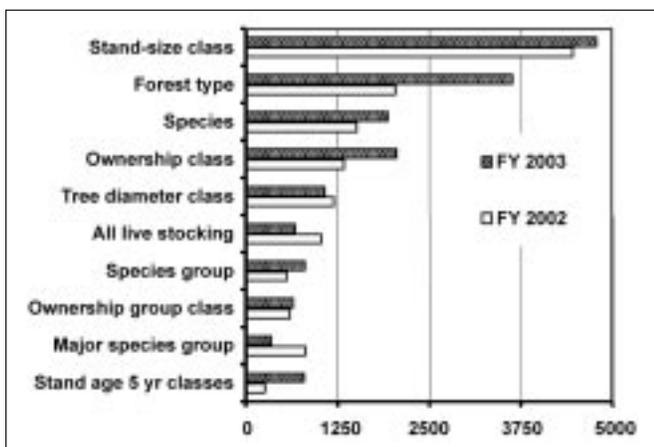
Best, C.; Wayburn, L.A. 2001. America's private forests: status and stewardship. Washington, DC: Island Press. 224 p.

Caspersen, J.P.; Pacala, S.W.; Jenkins, J.C.; *et al.* 2000. Contributions of land-use history to carbon accumulation in U.S. forests. *Science*. 290(5494): 1148–1151.

Cooley, R.; Mobasher, B.; Srivastava, J. 1999. Data preparation for mining World Wide Web browsing patterns. *Knowledge and Information Systems*. 1: 5–32. <http://maya.cs.depaul.edu/~classes/ect584/papers/cms-kais.pdf>. (6 April 2005).

Cooley, R.; Tan, P.; Srivastava, J. 2000. Discovery of interesting usage patterns from web data. In: Spiliopoulou, M.; Masand, B., eds. *Advances in web usage analysis and user profiling*. Lecture Notes in Computer Science. 1836: 163–182. <http://www.cs.umn.edu/research/websift/papers/lncs99.ps>. (6 April 2005).

Figure 1.—Top 10 attributes requested of the FIA database by number of requests, exclusive of county and State, fiscal years 2002 and 2003 from data supplied in Miles (2002, 2003).



-
- Cruikshank, J.W.; McCormack, J.F. 1956. The distribution and volume of hickory timber. Hickory Task Force Report No. 5. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station. 12 p.
- Dana, S.T.; Allison, J.H.; Cunningham, R.N. 1960. Minnesota lands: ownership, use, and management of forest and related lands. Washington, DC: The American Forestry Association. [Distributed to the trade by the Livingston Publishing Company, Narberth, PA.] 463 p.
- Dana, S.T.; Krueger, M. 1958. California lands: ownership, use, and management. Land ownership series. Washington, DC: The American Forestry Association. 308 p.
- Full Circle Systems, Inc. 2003. Knowledge management strategy. http://www.fullcirclesystems.com/knowledge_management_strategy.php. (6 April 2005).
- Gedney, D.R. 1976. The occurrence of laminated root rot on nonfederal timberland in northwest Oregon, 1976. Res. Note PNW-381. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station. 7 p.
- Guldin, J.M., tech. comp. 2001. Proceedings of the symposium on Arkansas forests: a conference on the results of the recent forest survey of Arkansas; 1997 May 30–31; North Little Rock, AR. Gen. Tech. Rep. SRS-41. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 125 p.
- Jaakko Poeyry Consulting, Inc. 1994. Final generic environmental impact statement study on timber harvesting and forest management in Minnesota. Tarrytown, NY: Jaakko Poeyry Consulting, Inc. <http://www.iic.state.mn.us/download/geis/main/geismain.htm>. (6 April 2005).
- Kingsley, N.P.; Finley, J.C. 1975. The forest-land owners of Delaware. Resour. Bull. NE-38. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 78 p.
- Kline, J.D.; Moses, A.; Alig, R.J. 2001. Integrating urbanization into landscape-level ecological assessments. *Ecosystems*. 4(1): 3–18.
- LaPoint, E. 2005. Access and use of FIA data through FIA spatial data services. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; McWilliams, W.H., eds. 2005. Proceedings of the fifth annual forest inventory and analysis symposium; 2003 November 18–20; New Orleans, LA. Gen. Tech. Rep. WO-69. Washington, DC: U.S. Department of Agriculture, Forest Service. 222 p.
- Larson, R.W. 1952. The timber supply situation in Florida. Forest Resour. Rep. 6. Washington, DC: U.S. Department of Agriculture. 60 p. and map.
- MacLean, C.D.; Bolsinger, C.L. 1973. Estimating productivity on sites with a low stocking capacity. Res. Paper RP-PNW-152. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station. 18 p.
- Mangold, R.D. 1998. Forest health monitoring field methods guide. Revision O. 429 p. On file with: U.S. Department of Agriculture, Forest Service, Southern Research Station, Forest Health Monitoring Program, P.O. Box 12254, Research Triangle Park, NC 27709.
- Miles, P. 2002. FY02 forest inventory mapmaker report. 18 p. On file with: U.S. Department of Agriculture, Forest Service, North Central Research Station, Forest Inventory and Analysis Unit, 1992 Folwell Avenue, St. Paul, MN 55108.
- Miles, P. 2003. FY03 forest inventory mapmaker report. 17 p. On file with: U.S. Department of Agriculture, Forest Service, North Central Research Station, Forest Inventory and Analysis Unit, 1992 Folwell Avenue, St. Paul, MN 55108.
- Moore, W.H.; Ripley, T.H.; Clutter, J.L. 1960. Trials to determine relative deer range carrying capacity values in connection with the Georgia forest survey. In: Proceedings of the 14th annual conference of the Southeastern Association of Game and Fish Commissioners. Columbia, SC: Southeastern Association of Game and Fish Commissioners. 14: 98–104.

-
- Olson, R.J.; Emerson, C.J.; Nungesser, M.K. 1982. Geoecology county-level environmental data for the United States, 1964–1979 [database]. http://www.daac.ornl.gov/VEGETATION/guides/geoecology_data.html. (6 April 2005).
- Pearson, H.A.; Sternitzke, H.S. 1974. Forest-range inventory: a multiple-use survey. *Journal of Range Management*. 27(5): 404–407.
- Pomeroy, K.B.; Yoho, J.G. 1964. North Carolina lands; ownership, use, and management of forest and related lands. Land ownership series. Washington, DC: American Forestry Association. [Distributed to the trade by the Livingston Publishing Company, Narberth, PA.] 372 p.
- Rudis, V.A. 1991. Wildlife habitat, range, recreation, hydrology, and related research using forest inventory and analysis surveys: a 12-year compendium. Gen. Tech. Rep. SO-84. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 61 p.
- Rudis, V.A. 1993. The multiple resource inventory decision-making process. In: Lund, H.G.; Landis, E.; Atterbury, T., eds. *Stand inventory technologies 92: proceedings of the stand inventory technologies: an international multiple resource conference; 1992 September 13–17; Portland, OR*. Bethesda, MD: American Society for Photogrammetry and Remote Sensing: 180–192.
- Rudis, V.A. 2002–04. The FIA citation database. http://srsfia1.fs.fed.us/fia_citation_database.php. (6 April 2005).
- Rudis, V.A. 2003a. Comprehensive regional resource assessments and multipurpose uses of forest inventory and analysis data, 1976 to 2001: a review. Gen. Tech. Rep. SRS-70. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 129 p.
- Rudis, V.A. 2003b. Fresh ideas, perspectives, and protocols associated with forest inventory and analysis surveys: graduate reports, 1974 to July 2001. Gen. Tech. Rep. SRS-61. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 49 p.
- Song, U. 1994. Use of a land classification system in growth and yield prediction on the Cumberland Plateau. Knoxville, TN: University of Tennessee. 53 p. M.S. thesis.
- Sternitzke, H.S.; Duerr, W.A. 1950. Tree distribution in Mississippi. Forest Survey Release 64. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 19 p.
- Trani, M.K.; Brooks, R.T.; Schmidt, T.L.; *et al.* 2001. Patterns and trends of early successional forests in the Eastern United States. *Wildlife Society Bulletin*. 29(2): 413–424.
- Wear, D.N.; Greis, J.G., eds. 2002. Southern forest resource assessment. Gen. Tech. Rep. SRS-53. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 635 p.
- Wiedinmyer, C.; Guenther, A.; Estes, M.; *et al.* 2001. A landuse database and biogenics emissions inventory for the State of Texas. *Atmospheric Environment*. 35: 6465–6477.

Comparison of Programs Used for FIA Inventory Information Dissemination and Spatial Representation

Roger C. Lowe¹ and Chris J. Cieszewski

Abstract.—Six online applications developed for the interactive display of Forest Inventory and Analysis (FIA) data in which FIA database information and query results can be viewed as or selected from interactive geographic maps are compared. The programs evaluated are the U.S. Department of Agriculture Forest Service's online systems; a SAS server-based mapping system developed by the Cooperative Extension Service at the University of Georgia (UGA); and three online applications developed at the D.B. Warnell School of Forest Resources, UGA: HTML and Java-based ESRI ArcIMS applications, HTML-based hypermaps, and a prototype Macromedia Flash-based application. Our study compared the following features in these Web applications: application scope, data resolution, number of available layers, number of variables, reporting capabilities, user interaction capabilities, and user interaction timing.

Much effort has been put into designing and implementing the nationwide U.S. Department of Agriculture (USDA) Forest Service Forest Inventory and Analysis (FIA) program. Numerous bits of information are collected in the field and later summarized and tabulated. The extensive information contained in the FIA database is a valuable source of forest land estimates to many public and private agencies. To facilitate sharing this data, several online programs have been developed that enable users to query different databases and generate various types of reports and maps.

Applications

Each application compared is briefly described below.

USDA Forest Service's Mapmaker

The USDA Forest Service's Mapmaker (fig. 1) is a server-side Web application that enables users to query, report, and map the most current FIA nationwide data available. Through a series of Web pages, the user enters the geographic area of interest, attribute of interest, query filters, and classification variables to query the FIA database and produce custom tables and maps.

USDA Forest Service's Ramiform

The USDA Forest Service Ramiform program (fig. 2) is a server-side application being developed by the USDA Forest Service at the North Central Research Station to serve nationwide FIA database information through the Web. This application enables users to query information derived from the 2002 Resources Planning Act (RPA) summary database, map the results to hexagon units, generate dot density displays, and tabulate area summaries by USDA Forest Service region for the query.

Georgia Statistics System

The Georgia Statistics System (GSS) (fig. 3) is the server-side Web application developed by the Center for Agribusiness and Economic Development, Cooperative Extension Service at the University of Georgia (UGA). GSS was developed to provide access to Georgia data from the Georgia County Guide (GCG), the multiple Farmgate Value Reports (FVR), and selected FIA database information. This application enables users to perform a cross-sectional, time series, analysis of county information.

Forest Maps Applications

The Fiber Supply Assessment group at the D.B. Warnell School of Forest Resources (WSFR), UGA, developed a set of applications consists of a large suite of Web pages designed for users to view, query, tabulate, and map the most current FIA database information for Georgia.

- The Unique Selected (Uni-select) maps for 13 Southern States (fig. 4) offers fast client-side access to multiback-

¹ GIS Analyst and Corresponding Author, Associate Professor, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602. E-mail: rcl7820@owl.forestry.uga.edu.

Figure 1.—USDA Forest Service Mapmaker, version 1.0.

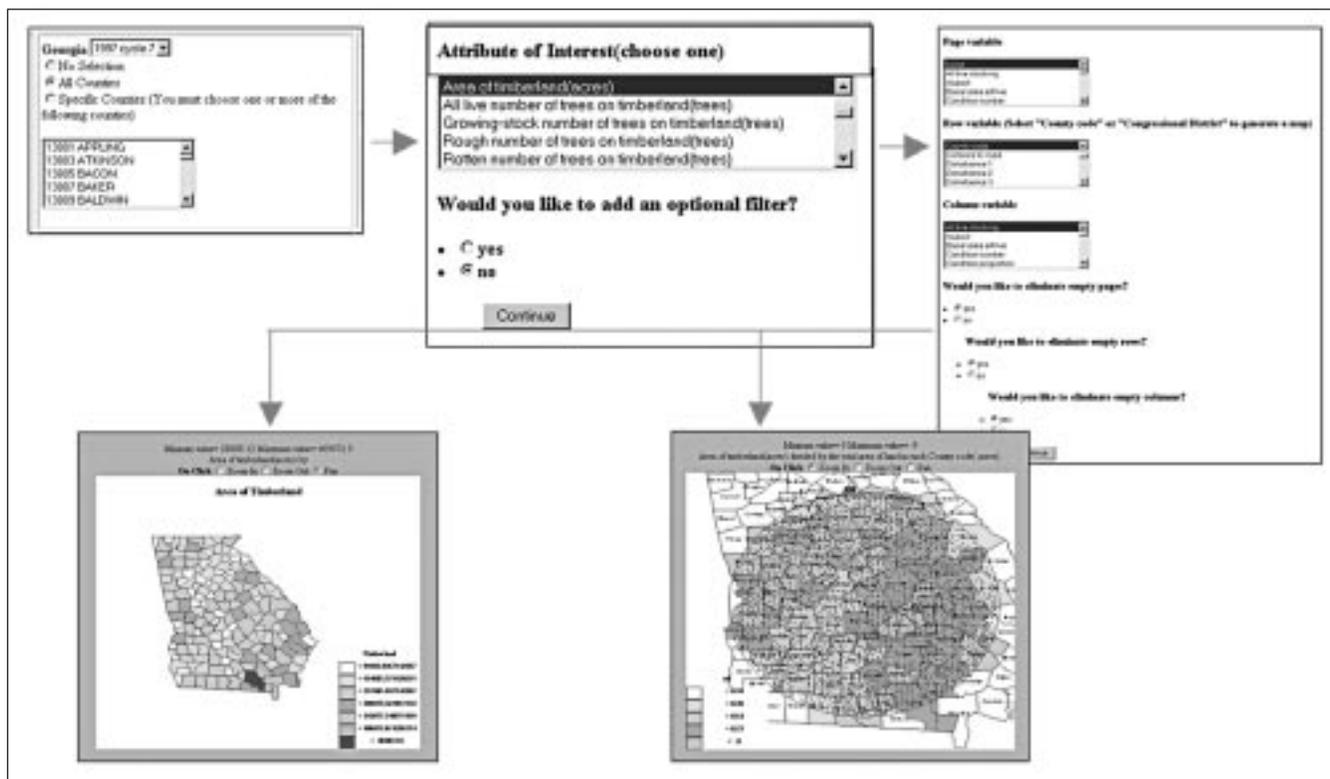


Figure 2.—USDA Forest Service Ramiform.

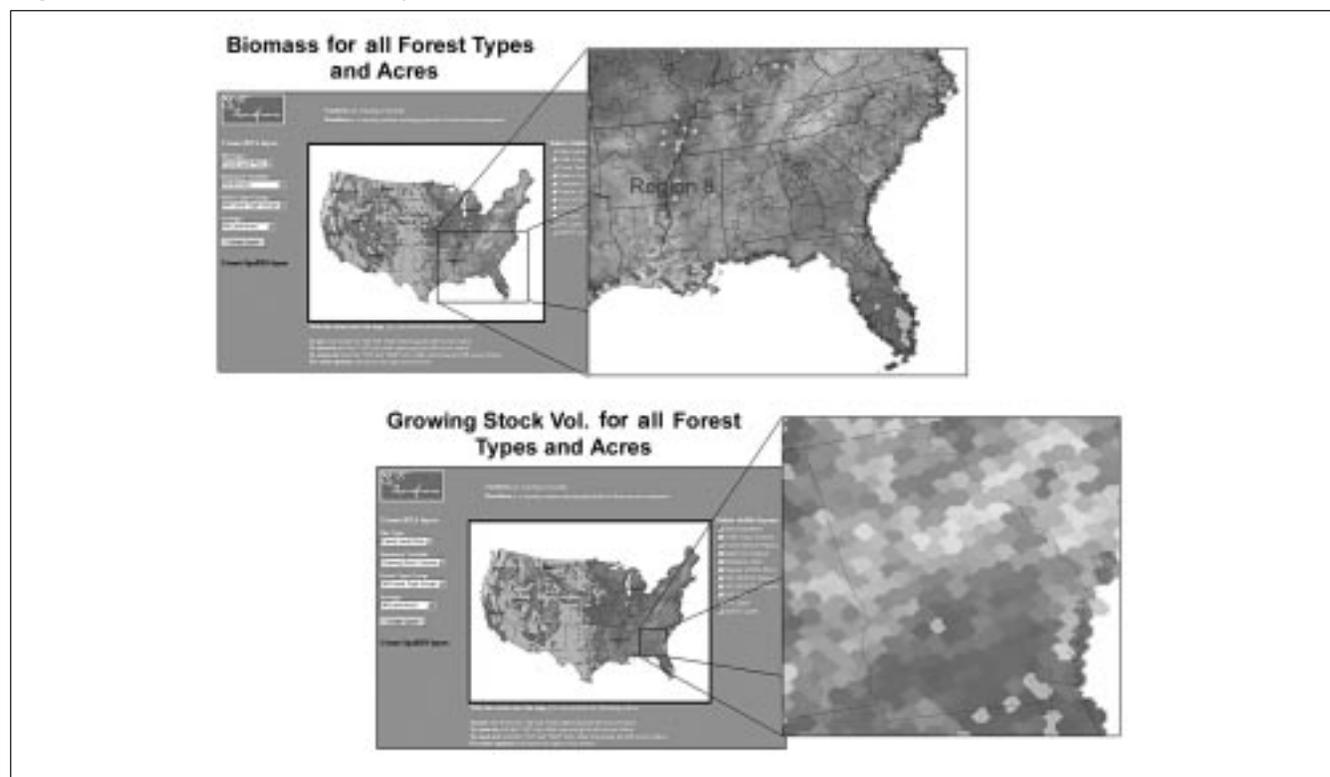


Figure 3.—Center for Agribusiness and Economic Development, Cooperative Extension Service, University of Georgia, Georgia Statistics System (GSS).

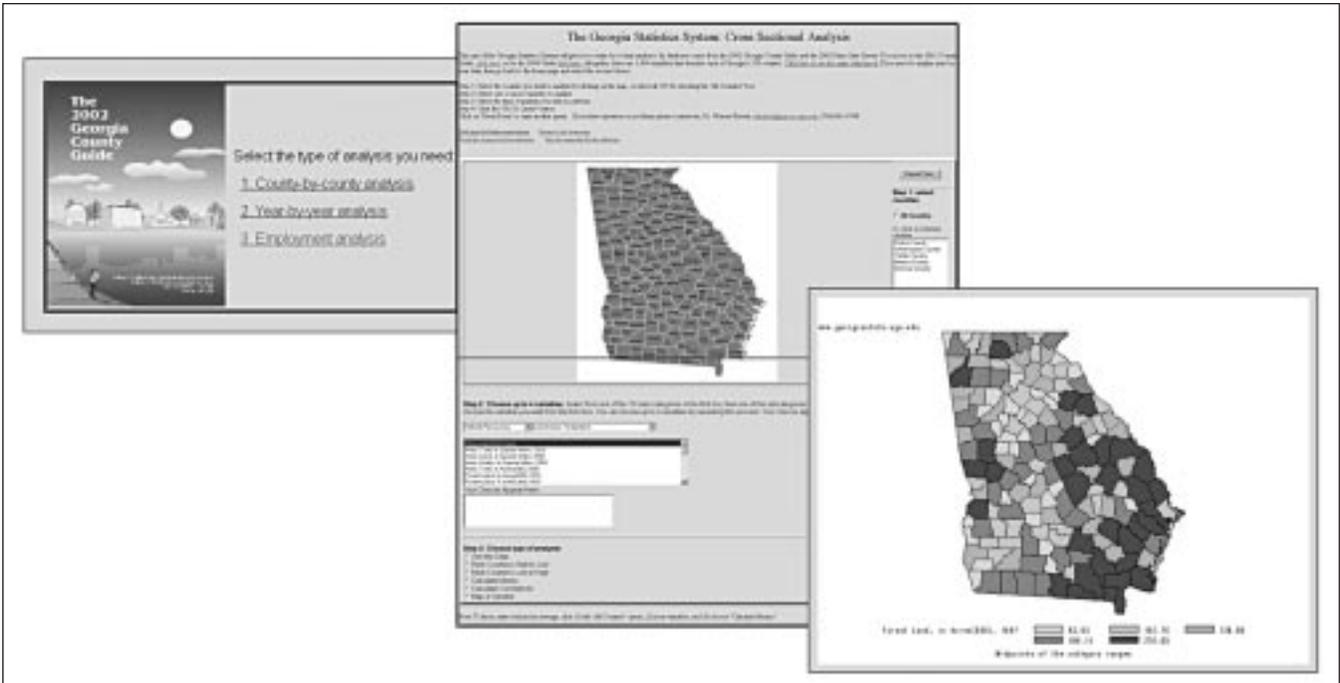
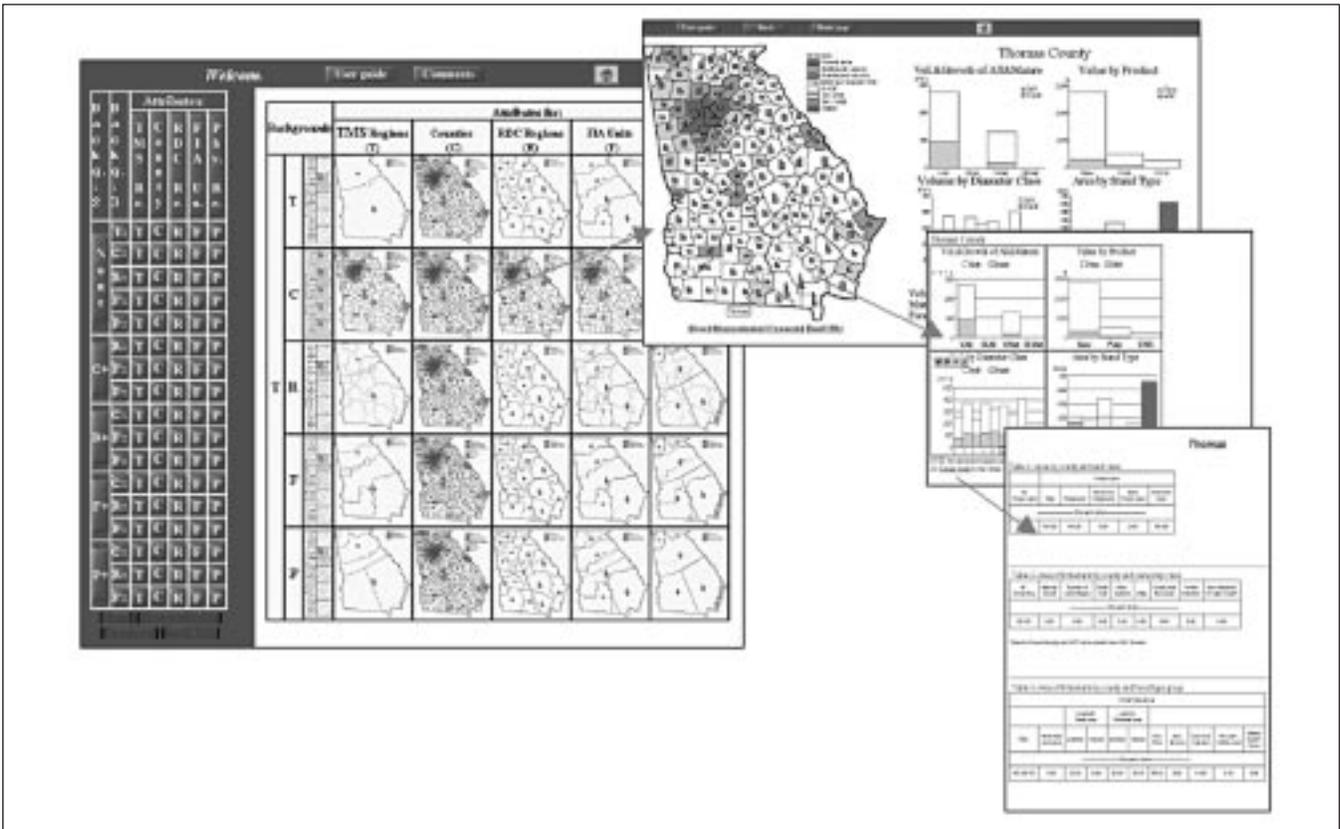


Figure 4.—D.B. Warnell School of Forest Resources (WSFR), University of Georgia. Uni-select maps.



ground statistics for cross-sectional selections of Georgia's Rural Development Centers (RDC) regions, FIA units, physiographic regions, Timbermart South regions, and counties.

- HTML- and Java-based Environmental Systems Research Institute, Inc.'s (ESRI's) ArcIMS® server-side platforms provide users with the ability to query and report FIA database information, and geographic information system capabilities such as buffering and subselections (fig. 5).
- The InFORM client-side application (fig. 6) makes available compiled FIA inventory information useful for high-level discussions and Georgia's inventory analyses. Users can display multiple FIA database variables as a choropleth map and tabulate FIA information for single, select, and subselections of counties.

Methods

The online applications were compared between November 10 and 15, 2003. Updates applied to the sites after this week may not be represented in the comparisons. All assessments were made from a home computer through a dial-up modem with a 49 kbps connection. Where applicable, at least 15 page loads were timed and averaged. After each page load, the Internet

browser was closed and the Internet cache cleared. Note that physical proximity to the application server computer may affect the page load timing. To minimize this effect and mitigate the impact of heavy online traffic, evaluations were conducted between 1:00 a.m. and 5:00 a.m. eastern standard time.

Results

The applications were compared on general features such as scope (geographic extent of analysis), data resolution (the smallest unit that can be mapped), and the number of layers and variables available for viewing and mapping were made (table 1). The two USDA Forest Service applications, Mapmaker and Ramiform, provide analysis capabilities at the national level; the others focus on Georgia. The Ramiform application's data resolution is the hexagon. The other applications use county-level resolution. ArcIMS can display 11 different layers, such as county and RDC boundaries, more than all the other applications. Ramiform makes 10 different layers available; the Uni-select maps have 5 layers that can be displayed on 5 backgrounds; and the remaining programs have single county boundaries. Mapmaker offers the largest number of variables, such as timberland and forestland area, that can be mapped or tabulated, followed by Ramiform, InFORM, Uni-select maps, GSS, and ArcIMS.

Figure 5.—D.B. Warnell School of Forest Resources (WSFR), University of Georgia, ArcIMS.

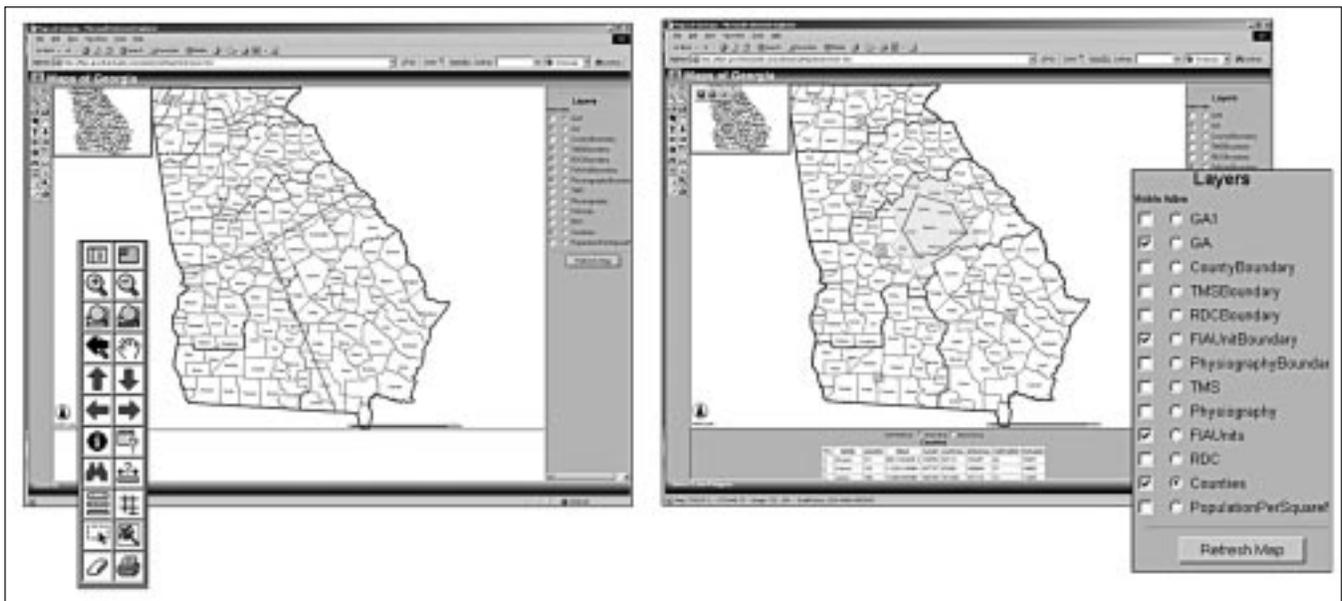
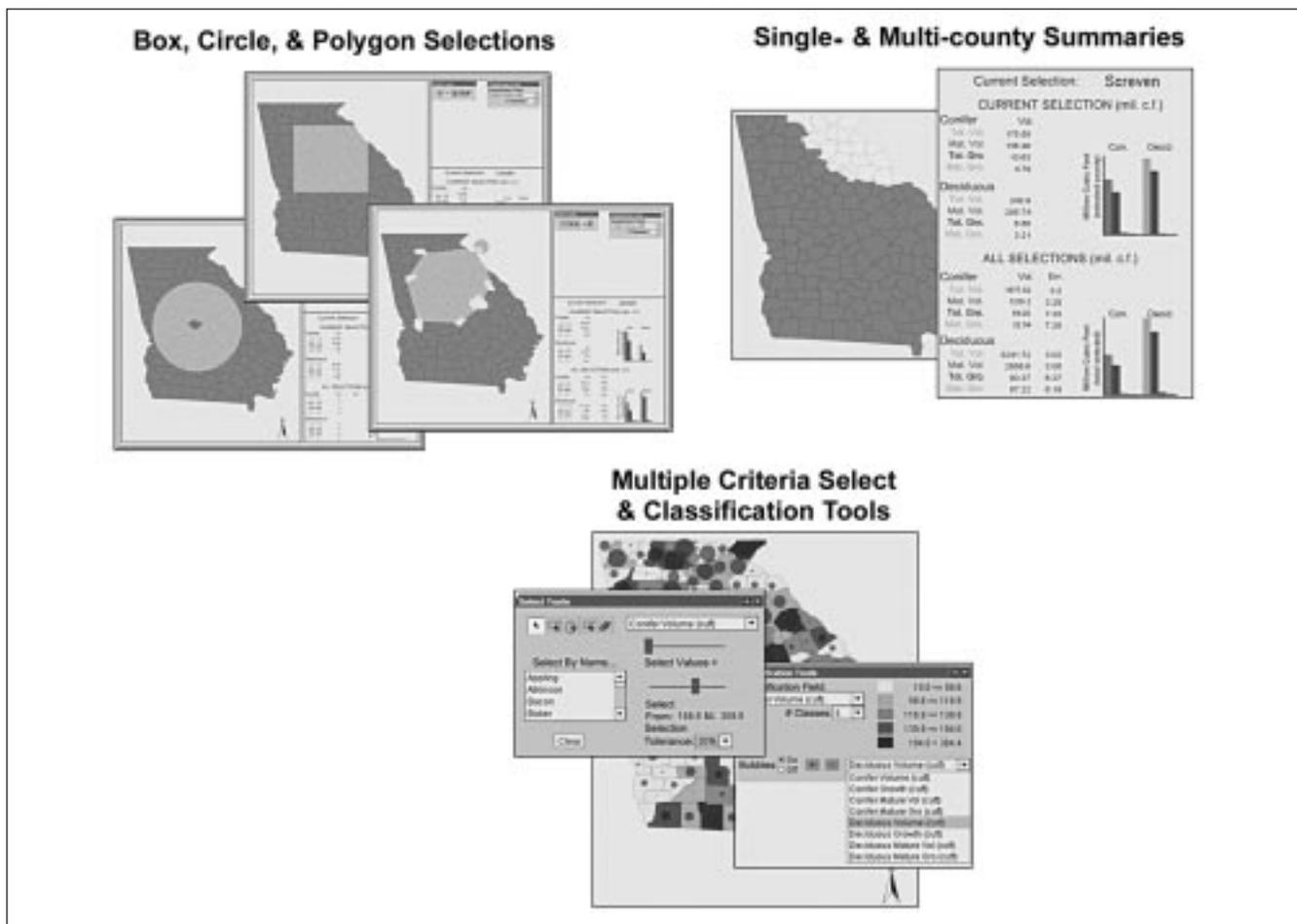


Figure 6.—D.B. Warnell School of Forest Resources (WSFR), University of Georgia, InFORM.



The applications were also compared on user interaction functionality, such as reporting, zoom, pan, variable display, and selection (table 1). All except ArcIMS have reporting capabilities. (Although ArcIMS can be programmed to offer reporting, its development team does not consider this functionality to be critical to the application’s development path.) Mapmaker, Ramiform, ArcIMS, and InFORM programs provide zoom and pan capabilities; Uni-select maps and GSS do not. ArcIMS site is the only application with buffer capabilities.

Ramiform, ArcIMS, and InFORM enable the user to display more than one variable at a time—for example, to show conifer and deciduous timberland area simultaneously. Users are able to simultaneously select more than one feature, such as a county or region, from a map, in Ramiform, ArcIMS, InFORM Mapmaker, GSS, and the Uni-select maps do not provide this capability. In addition to single click-and-select, InFORM and

ArcIMS enable the user to select multiple features with a user-defined box, circle, and polygon computing in real time summary statistics for the chosen selections. Ramiform had also a capability of selections, though without real time computing. Instead, a link function is used that was supposed to submit the selection to the Mapmaker (we were unsuccessful testing this external to the native application functionality). Furthermore, InFORM is the only application that allows for progressive increase in selections, in which selections of individual units can be added to circle selections by radius, which can be added to rectangle selections, which can be added to arbitrary selections, which can be altered by individual deselecting, and so forth.

Table 1 presents average times, in seconds, from multiple trials conducted with all the applications. The following events were timed: (1) initial load time, (2) layer load time, (3) selection time, and (4) report generation time. Ramiform had the longest

Table 1.—*Internet application comparisons.*

	D.B. Warnell School of Forest Resources (WSFR), University of Georgia (UGA)			USDA Forest Service		Center for Agribusiness and Economic Development, Cooperative Extension Service, UGA
	Uni-select	InFORM ¹	ArcIMS	Ramiform ¹	Mapmaker	Ga. stat.
Scope	Georgia	Georgia	Georgia	Nation	Nation	Georgia
Resolution	County	County	County	FIA Hex.	County	County
No. of layers	5	1	11	10	1	1
Simult. lyrs.	No	Yes	Yes	Yes	No	No
No. of variables	79 ²	90 ²	11	97	Numerous	43
Reports	Yes	Yes	No	Yes	Yes	Yes
User interaction						
Pan	No	Yes	Yes	Yes	Yes	No
Zoom	No	Yes	Yes	Yes	Yes	No
Buffer	No	No	Yes	No	No	No
Interactive selections						
Single	Yes	Yes	Yes	Yes	No	Yes
Box	No	Yes	Yes	Yes	No	No
Polygon	No	Yes	Yes	Yes	No	No
Circle	No	Yes	Yes	Yes	No	No
User-defined	No	Yes	Yes	Yes	No	No
Query	No	Yes	Yes	Yes	Yes	No
Interaction time (seconds)						
Initial load	2	2	10	43	6	2
Vector load	1	na	2	3	na	na
Raster load	na	na	4	17	na	na
Selection	na	1	4	4	na	na
Report	2	2	na	42	nt	7

¹ Applications under continued development.

² Includes report variables.

na: not applicable on the date of timing.

nt: event not timed; requires multiple queries.

initial load time for the application to fully load in the Web browser. The other applications loaded in less than 10 seconds. Vector data layers loaded quickly (less than 3 seconds) in those apps that provide that functionality. The raster layers in Ramiform took a bit longer (17 seconds). Selection time—how long it took to graphically display selected features—was relatively

short, fewer than 4 seconds for Ramiform, ArcIMS, and InFORM. The time required to generate a tabular report was less than 7 seconds for the Uni-select maps, InFORM, and CAED GSS; it took much longer—42 seconds—for Ramiform to create this type of report.

Summary

Our study evaluated six online systems developed to disseminate forest inventory information based on FIA data for the United States. The systems were assessed on scope, data resolution, simultaneous number of themes that can be displayed, reporting ability, interactivity, and speed in performing simple tasks. The Web applications currently being developed and maintained by the USDA Forest Service—Mapmaker and Ramiform—are national in scope and serve data at the county or RPA hexagon level. Applications based at The University of Georgia are statewide in scope and serve data at the county level. Mapmaker provides access to the largest number of variables for analysis than all the other applications, followed by Ramiform, InFORM, Uni-select maps, the Georgia Statistics System (GSS), and the D.B. Warnell School of Forest Resources (WSFR), University of Georgia's application, ArcIMS. Users can generate tabular output in each application except ArcIMS, although this functionality can be programmed. InFORM, Ramiform, Mapmaker, and ArcIMS provide zoom and pan capabilities; Uni-select maps and GSS do not. ArcIMS is the only one with buffer capabilities. Ramiform, Uni-select maps, and ArcIMS can display multiple layers simultaneously. InFORM, ArcIMS, GSS, and Ramiform provide interactive selection tools to select many features at the same time; the Uni-select maps and GSS provide individual county, FIA region, RDC region, and physiographic region selection functionality. Mapmaker is a query-driven application that does not allow user-driven selections from a map. Based on application event timing, the client-side applications load, select, and report data faster than the server-side applications.

Literature Cited

- Center for Agribusiness and Economic Development, Cooperative Extension Service at the University of Georgia. Georgia Statistics System. <http://www.georgiastats.uga.edu>. [Date accessed unknown.]
- Center for Agribusiness and Economic Development, Cooperative Extension Service at the University of Georgia. <http://www.agecon.uga.edu/~caed/staffreports.html>. [Date accessed unknown.]
- Fiber Supply Assessment Group, D.B. Warnell School of Forest Resources, University of Georgia. Fiber Supply Assessment about the forestry growth and yield. <http://growthandyield.com>. [Date accessed unknown.]
- Fiber Supply Assessment Group, D.B. Warnell School of Forest Resources, University of Georgia. Unique selected maps for 13 Southern States. http://growthandyield.com/main/maps_uni.htm. [Date accessed unknown.]
- Fiber Supply Assessment Group, D.B. Warnell School of Forest Resources, University of Georgia. The Internet Map Service for Forest of Georgia. http://growthandyield.com/main/maps_ims.htm. [Date accessed unknown.]
- Fiber Supply Assessment Group, D.B. Warnell School of Forest Resources, University of Georgia. minn_meeting1. http://growthandyield.com/inform/tty/minn_meeting1.html. [Date accessed unknown.]
- U.S. Department of Agriculture (USDA) Forest Service. Mapmaker, version 1.0. <http://ncrs2.fs.fed.us/4801/FIADB/index.htm>. [Date accessed unknown.]
- USDA Forest Service. Ramiform. <http://ncrs2.fs.fed.us/zope/ramiform>. [Date accessed unknown.]
- Center for Agribusiness and Economic Development, Cooperative Extension Service at the University of Georgia. Georgia County Guide. <http://www.agecon.uga.edu/%7Ecountyguide>. [Date accessed unknown.]

