USDA United States Department of Agriculture

Forest Service

Rocky Mountain Research Station

Proceedings
RMRS-P-56CD

July 2009

Forest Inventory and Analysis (FIA) Symposium 2008

October 21-23, 2008; Park City, UT



McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. **Forest Inventory and Analysis (FIA) Symposium 2008; October 21-23, 2008; Park City, UT**. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Abstract

These proceedings report invited presentations and contributions to the 2008 Biennial Forest Inventory and Analysis (FIA) Symposia, which was hosted by the Research and Development branch of the United States Forest Service. As the only comprehensive and continuous census of the forests in the USA, FIA provides strategic information needed to evaluate sustainability of current forest management practices across all ownerships. Symposium papers cover high priority and timely issue-based topics, including climate change, wildlife, fire, bio-energy, geo-spatial extensions, monitoring over time, integrating remote sensing and GIS applications, statistical and related quantitative solutions to emerging needs, and many others.

Keywords: climate change, fire, disturbance, environmental monitoring, national forest inventory (NFI), assessments, remote sensing

You may order additional copies of this publication by sending your mailing information in label form through one of the following media. Please specify the publication title and number.

Publishing Services

Telephone	(970) 498-1392
FAX	(970) 498-1122
E-mail	rschneider@fs.fed.us
Web site	http://www.fs.fed.us/rmrs
Mailing Address	Publications Distribution Rocky Mountain Research Station 240 West Prospect Road Fort Collins, CO 80526

Table of Contents

2008 Forest Inventory & Analysis Symposium

Fire, Disturbance, and Other Remote Sensing Applications to Support Large-Scale Forest Inventories I

Joint Simulation of Regional Areas Burned in Canadian Forest Fires: A Markov Chain Monte Carlo Approach
Steen Magnussen

Rapid Assessment of Wildfire Damage Using Forest Inventory Data: A Case in Georgia
Richard A. Harper, John W. Coulsten, Jeffery A. Turner

Cool Tools

Using FIA Data in the Forest Vegetation Simulator
John D. Shaw

EVALIDatorReports: Reporting Beyond the FIADB
Patrick D. Miles

Wildlife Habitat Applications I

Long Term Ecological Monitoring Program on the Kenai National Wildlife Refuge, Alaska: An FIA Adjunct Inventory
John M. Morton, Matthew Bowser, Edward Berg, Dawn Magness, Todd Eskelin

Monitoring and Modeling Terrestrial Arthropod Diversity on the Kenai National Wildlife Refuge
Matthew L. Bowser, John M. Morton

Fire, Disturbance, and Other Remote Sensing Applications to Support Large-Scale Forest Inventories II

Tree Crown Conditions in Virginia Before and After Hurricane Isabel (September 2003)
KaDonna Randolph, Anita Rose

Tracking Forest Land

Estimating Fine-Scale Land Use Change Dynamics Using an Expedient Photointerpretation-Based Method
Tonya Lister, Andrew Lister, Eunice Alexander

Development of Issue-Relevant State Level Analyses of Fragmentation and Urbanization
Rachel Riemann, Tonya Lister, Andy Lister, Dacia Meneguzzo, Sarah Parks

A Preliminary Investigation of Forest Carbon Changes Associated with Land-Use Change in Northern New England
Daolan Zheng, Linda S. Heath, Mark J. Ducey, James E. Smith

Wildlife Habitat Applications II

Quantification of Lewis's Woodpecker Habitat Using Forest Inventory and Analysis Data
Chris Witt

Remote Sensing to Support Local to Regional Monitoring and Management Needs

Integrating Landsat-Derived Disturbance Maps with FIA Inventory Data: Applications for State-Level Forest Resource Assessments

Sonja Oswalt, Chengquan Huang, Hua Shi, James Vogelmann, Zhiliang Zhu, Samuel N. Goward, John Coulston

Analyzing Landsat Time-Series Data across Adjacent Path/Rows and Across Multiple Cycles of FIA: Lessons Learned in Southern Missouri

Mark Nelson, Sean Healey, W. Keith Moser, Mark Hansen, Warren Cohen, Mark Hatfield, Nancy Thomas, Jeff Masek

Trends in Afforestation in Southern Missouri

W. Keith Moser, Mark D. Nelson, Mark H. Hansen, Sean Healey, Warren Cohen

National Inventory and Monitoring Applications

Overview of the National Inventory and Monitoring Applications Center (NIMAC)

Charles T. Scott

Design Tool for Inventory and Monitoring

Charles T. Scott, Renate Bush, Ken Brewer

Inventory of Trees in Nonforest Areas in the Great Plains States

Andrew Lister, Chip Scott, Steve Rasmussen

Effects of Plot Size on Forest-Type Algorithm Accuracy

James A. Westfall

Wisconsin State Forests Continuous Forest Inventory: A Look at the First Year

Randall S. Morin, Teague Prichard, Vern Everson, Jim Westfall, Charles Scott

FIA Map Products: Production and Assessment

Combining Forest Inventory, Satellite Remote Sensing, and Geospatial Data for Mapping Forest Attributes of the Conterminous United States

Mark Nelson, Greg Liknes, Charles H. Perry

Disturbance Impacts

An Assessment of the Relationship Between Emerald Ash Borer Presence and Landscape Pattern

Susan J. Crocker, Dacia M. Meneguzzo

Mountain Pine Beetle Infestations and Sudden Aspen Decline in Colorado: Can the Forest Inventory and Analysis Annual Inventory System Address the Issues?

Michael T. Thompson

Large Area Mapping Applications I

Harmonizing Estimates of Forest Land Area From National-Level Forest Inventory and Satellite Imagery

Bonnie Ruefenacht, Mark D. Nelson, Mark Finco, Ken Brewer

Enhancing FIA for Offset Monitoring

Investigation into Calculating Tree Biomass and Carbon in the FIADB Using a Biomass Expansion Factor Approach
Linda S. Heath, Mark Hansen, James E. Smith, Patrick D. Miles, Brad W. Smith

Estimating Diesel Fuel Consumption and Carbon Dioxide Emissions From Forest Road Construction
Dan Loeffler, Greg Jones, Nikolaus Vonessen, Sean Healey, Woodam Chung

Determining Landscape-Level Carbon Emissions from Historically Harvested Forest Products
Sean Healey, Todd A. Morgan, Jon Songster, Jason Brandt

Phase 3 Indicators: Past, Present, and Future

Lichens, Ozone, and Forest Health — Exploring Cross-Indicator Analyses With FIA Data
Susan Will-Wolf, Sarah Jovan

The Power of FIA Phase 3 Crown-Indicator Variables to Detect Change
William A. Bechtold, Kadonna C. Randolph, Stanley J. Zarnoch

Phase 3 Indicators: Data in Action

From Detection Monitoring to Evaluation Monitoring — A Case Study Involving Crown Dieback in Northern White-Cedar
KaDonna C. Randolph, William A. Bechtold, Randall S. Morin, Stanley J. Zarnoch

Can Live Tree Size-Density Relationships Provide a Mechanism for Predicting Down and Dead Tree Resources?
Christopher W. Woodall, James A. Westfall

Vegetation Inventory Data: How Much is Enough?
Bethany Schulz, Sonja Oswalt, W. Keith Moser

Combining Panel Data

Tree-level Imputation Techniques to Estimate Current Plot-Level Attributes in the Pacific Northwest Using Paneled Inventory Data
Bianca N. I. Eskelson, Temesgen Hailemariam, Tara M. Barrett

Opportunities to Improve Monitoring of Temporal Trends With FIA Panel Data
Raymond L. Czaplewski, Michael T. Thompson

Spatial-Temporal Models for Improved County-Level Annual Estimates
Francis A. Roesch

Predictive Geo-Spatial Modeling

Using Landsat Imagery and FIA Data to Examine Wood Supply Uncertainty
Curtis A. Collins, Ruth C. Seawell

Mapping Upland Hardwood Site Quality and Productivity with GIS and FIA in the Blue Ridge of North Carolina
Claudia A. Cotton, Stephen R. Prisley, Thomas R. Fox

Testing Ecoregions in Kentucky and Tennessee with Satellite Imagery and Forest Inventory Data
W. Henry McNab, F. Thomas Lloyd (deceased)

Biomass Estimation

Improving North American Forest Biomass Estimates from Literature Synthesis and Meta-Analysis of Existing Biomass Equations

David C. Chojnacky, Jennifer C. Jenkins, Amanda K. Holland

Landscape-Level Analyses I

Relationships between Forest Structure, Composition, Site, and Spruce Beetle Occurrence in the Intermountain West

R. Justin DeRose, James N. Long, John D. Shaw

Evaluating Cypress Sustainability—FIA Data "In The Hot Seat"

Mark J. Brown

Use of Damage Surveys and Field Inventories to Evaluate Oak and Sugar Maple Health in the Northern United States

Randall S. Morin, Christopher W. Woodall, Jim Steinman, Charles H. Perry

Biomass Applications

A Real-Time Web-Based Optimal Biomass Site Assessment Tool (BioSAT): Module 1. An Economic Assessment of Mill Residues for the Southern U.S.

Timothy M. Young, James H. Perdue, Andy Hartsell, Robert C. Abt, Donald G. Hodges, Timothy G. Rials

Mixed Multi-Scalar Methods to Assess Wood Biomass Availability on Family Forests in Virginia's Southside

M. D. Brinckman, J. F. Munsell

A Strategic Assessment of Biofuels Development in the Western States

Kenneth E. Skog, Robert Rummer, Bryan Jenkins, Nathan Parker, Peter Tittman, Quinn Hart, Richard Nelson, Ed Gray, Anneliese Schmidt, Marcia Patton-Mallory, Gayle Gordan

Utilizing Climate Information

Predicting Forest Attributes from Climate Data Using a Recursive Partitioning and Regression Tree Algorithm

Greg C. Liknes, Christopher W. Woodall, Charles H. Perry

Actual Evapotranspiration (AET) and Tree Species Richness in the Eastern U.S.A.

Weihong Fan, Richard H. Waring

Landscape-Level Analyses II

Summary and Findings of the 2006 BLM Forest Lands Report

Tim Bottomley, Jim Menlove

Tree Migration Detection Through Comparisons of Historic and Current Forest Inventories

Christopher W. Woodall, Christopher M. Oswalt, James A. Westfall, Charles H. Perry, Mark N. Nelson

From Genes to Ecosystems: Measuring Evolutionary Diversity and Community Structure with Forest Inventory and Analysis (FIA) Data

Kevin M. Potter

Fine Scale Mapping Applications

Combining FIA Plot Data with Topographic Variables: Are Precise Locations Needed?
Stephen P. Prisley, Huei-Jin Wang, Philip J. Radtke, John Coulston

Using Interpreted Large Scale Aerial Photo Data to Enhance Satellite-Based Mapping and Explore Forest Land Definitions
Tracey S. Frescino, Gretchen G. Moisen

System Improvements

The Role of Pre-Field Operations at Four Forest Inventory Units: We Can See the Trees, Not Just the Forest
Sara A. Goeking, Greg C. Liknes

A Stem-Map Model for Predicting Tree Canopy Cover of Forest Inventory and Analysis (FIA) Plots
Chris Toney, John D. Shaw, Mark D. Nelson

Correct County Areas with Sidebars for Virginia
Joseph M. McCollum, Dale Gormanson, John Coulston

Statistical Considerations

An Assessment of the Impact of FIA's Default Assumptions on the Estimates of Coarse Woody Debris Volume and Biomass
Vicente J. Monleon

Field Results for Line Intersect Distance Sampling of Coarse Woody Debris
David L. R. Affleck

A Comparison of Forest Height Prediction from FIA Field Measurement and LiDAR Data Via Spatial Models
Yuzhen Li

Calculation of Upper Confidence Bounds on Not-Sampled Vegetation Types Using a Systematic Grid Sample: An Application to Map Unit Definition for Existing Vegetation Maps
Paul L. Patterson, Mark Finco

Introduction

It's great to announce the proceedings of the FIA Symposium 2008! The Symposium has a long and rich history. Looking back to the 1980s, attendees were mostly mathematicians and statisticians boiling down a lot of ideas to compare designs for an annual forest inventory system. Those were exciting times as these meetings morphed into the FIA Symposium and the ideas harmonized. The amount of brain time that went into the annual system must be astronomical--far too many folks to mention here.

This latest Symposium was an international forum where we could discuss how FIA informs policy, make connections between our issue-focused analyses, exchange science and techniques development, showcase collaborative projects, foster partner-centric ventures, display the latest FIA science and tools, and communicate to the broader forest inventory community. In addition, a series of high priority, contemporary issue-based sessions were held concurrently. Some examples include climate change, wildlife, fire, bio-energy, geo-spatial extensions, monitoring over time, integrating remote sensing and GIS applications, statistical and related quantitative solutions to emerging needs, and many others.

The Organizing Board would like to thank all participants for taking the time to advance our FIA science! The Board also thanks those that traveled far, including the delegations from Canada and Mexico. Perhaps a theme for next time would be "monitoring across boundaries"?

Will McWilliams
Vicki Berrett
Ken Brewer
Renate Bush
Sally Campbell
Ray Czaplewski
Gretchen Moisen
Greg Reams
Paul Van Deusen

Fire, Disturbance, and Other Remote Sensing Applications to Support Large-Scale Forest Inventories I



Joint simulation of regional areas burned in Canadian forest fires: A Markov Chain Monte Carlo approach

Steen Magnussen[1]

***Abstract**: Areas burned annually in 29 Canadian forest fire regions show a patchy and irregular correlation structure that significantly influences the distribution of annual totals for Canada and for groups of regions. A binary Monte Carlo Markov Chain (MCMC) is constructed for the purpose of joint simulation of regional areas burned in forest fires. For each year the MCMC prediction is a binary vector with regions classified to a large fire year (LF) or a small fire year (SF). The regional area burned is then obtained from empirical quantile functions; separately for LF and SF years. The MCMC results were unbiased with respect to: the annual number of LF regions, national totals, and variances of area burned. Approximately 65% of the observed regional covariance was captured in the results.*

Keywords: binary correlation, multivariate simulation, marginal distribution, transition kernel.

Introduction

Forest fires affect forest resources and the global cycling of carbon and greenhouse gasses (Amiro et al. 2001, Bergeron et al. 2004, Gillett et al. 2004). They are a dominant driver in Canada's boreal forest carbon balance (Bond-Lamberty et al. 2007). Forecasting areas burned annually in forest fires (*BA*) at a regional and a combined regional scale is therefore important to predicting future greenhouse gas emissions (Kurz and Apps 2006, Kurz et al. 2008).

In Canada *BA* varies dramatically between years and regions. Large fires tend to occur during periods of stable high pressure (Skinner et al. 1999, Skinner et al. 2008). These atmospheric patterns are sub-continental in scale and may impose

[1] Canadian Forest Service, Natural Resources Canada, 506 West Burnside Road, Victoria, British Columbia V8Z 1M5, Canada. Phone: +1 250 363 0712, Fax: +1 250 363 0775, Email: steen.magnussen@nrcan.gc.ca

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

some regional synchronization in *BA*. Yet snow-cover or rain during the winter prevents the emergence of a strong temporal autocorrelation by saturating the forest fire fuels.

The simplest approach to forecasting *BA* is by recasting historic records. Recasting is attractive on grounds of expediency, simplicity, low costs, and transparency; however, this approach must take into account any interregional correlation structure. In Canada regional *BA* from 1955 to 1999 show an irregular pattern of weak and strong interregional correlations that exert a significant effect on the variance, and thus the shape, of distributions of sums of regional *BA*-values.

This study demonstrates a Markov Chain Monte Carlo (MCMC) procedure for joint forecasting of *BA* in 29 Canadian forest fire regions. The 29 regions account for about two-thirds of the areas burned in Canada. The rationale for the MCMC procedure rests with the fact that interregional correlations of *BA* are - to a large degree - shaped by a few years favorable to large fires.

Material and Methods

Data

Estimates of annual areas burned in forest fires (*BA*) from 1959 to 1999 in Canada's 29 forest fire regions were used as data for forecasting purposes (Magnussen 2008, Kurz and Apps 2006, Stocks et al. 2002).

Forecasting objectives

The objective is to forecast a sequence of *BA*-values for each of the 29 fire regions consistent with historic data from 1959-1999. Forecasted data should also conserve the regional correlations pattern so that the distribution of sums of regional *BA*-values matches the historic distribution of these sums.

Model premise

Regional correlations reflect the number of concurrent large values of *BA*. Consequently, a binary classification of region years to a large fire year ($LF = 1$) or a small fire year ($SF = 0$) is used to capture the regional correlations. To simplify the correlation structure it was decided to: *i*) form 29 balanced five-member ρ-cliques by maximizing the average within-clique correlation, and *ii*) assume that the *LF*-status of a region is only influenced by regions in the same ρ-clique. Accordingly, a two-stage process for the joint forecasting of *BA* is formulated: In stage one, the total number and regional allocation of *LF* years is

determined in a MCMC step (Robert and Casella 1999). In stage two, regional *BA*-values are drawn from empirical quantile functions, separately for *LF* and *SF* years.

Classification of region-years to LF or SF

The classification of region-years to either *LF* (1) or *SF* (0) was done by a *k*-means clustering ($k = 2$) routine. Following the classification, the probability (θ_i) that a *LF* year occurs in region *i* was estimated as:

$$\hat{\theta}_i = \frac{\sum_{j=1}^{41} LF_{ij}}{\sum_{i=1}^{29}\sum_{j=1}^{41} LF_{ij}} \qquad [1]$$

The MCMC (stage I)

Every forecast begins with a random draw of the number nLF^* of regions with a $LF = 1$ status. The draw is from a zero-truncated beta-binomial (Griffiths 1973) of nLF fitted to the classified data. An initial random allocation of the nLF^* is done with probability proportional to the regional probability of a *LF* year $\left\{\hat{\theta}_1,...,\hat{\theta}_{29}\right\}$ and modified in a sequence of switches $(s = 0,1,2,...)$ of *LF* status between two regions with opposite *LF*-status. The sequence maximizes the conditional likelihood of the allocation. Let $\mathbf{LF_0}$ denote the binary vector of the initial random allocation of regional *LF*-years. A switch involves two regions (*i* and *j*) for which $LF_i^* = 1$ and $LF_j^* = 0$, and $LF_i^* \to 0$ and $LF_j^* \to 1$. Let $\mathbf{LF}_s^*$ denote the vector of regional *LF* status after *s* switches, and let $\mathbf{LF}_{new}^*$ denote a proposed new configuration obtained from $\mathbf{LF}_s^*$ and applying a proposed switch. $\mathbf{LF}_{new}^*$ is accepted with a probability $\alpha_{s,s+1}$ $\left(\mathbf{LF}_s^* = \mathbf{LF}_{new}^*\right)$ or rejected with probability $1-\alpha_{s,s+1}$ in which case $\mathbf{LF}_{s+1}^* = \mathbf{LF}_s^*$. We have

$$\begin{aligned} &\alpha_{s,s+1} = \text{Min}\left[1, K\left(\mathbf{LF}_{new}^*, \mathbf{LF}_s^*\right)\right], U_{s+1}^* \sim \text{Uniform}[0,1] \\ &\mathbf{LF}_{s+1}^* = \begin{cases} \mathbf{LF}_{new}^* \text{ if } U_{s+1}^* \le \alpha_{s,s+1} \\ \mathbf{LF}_s^* \text{ if } U_{s+1}^* > \alpha_{s,s+1} \end{cases} \end{aligned} \qquad [2]$$

where U_{s+1}^* is a random draw from a uniform distribution on the unit interval [0,1] and $K(\cdot,\cdot)$ is defined in Equation 3

$$K\left(\mathbf{LF}^*_{new}, \mathbf{LF}^*_s\right) = \frac{\ell_p\left(\mathbf{LF}^*_{new}\right)}{\ell_p\left(\mathbf{LF}^*_s\right)} \times \frac{\ell\left(LF^*_{i,s}=1\,,\, LF^*_{j,s}=0\right)}{\ell\left(LF^*_{i,new}=0\,,\, LF^*_{j,new}=1\right)} \qquad [3]$$

with

$$\ell_p\left(\mathbf{LF}^*_{new}\right) \propto \ell\left(LF^*_{i,new} \mid LF^*_{i'\sim i,new}\right)\ell\left(\sum_{i'\sim i} LF^*_{i',new}\right)\ell\left(LF^*_{j,new} \mid LF^*_{j'\sim j,new}\right)\ell\left(\sum_{j'\sim j} LF^*_{j',new}\right)$$

$$\ell_p\left(\mathbf{LF}^*_s\right) \propto \ell\left(LF^*_{i,s} \mid LF^*_{i'\sim i,s}\right)\ell\left(\sum_{i'\sim i} LF^*_{i',s}\right)\ell\left(LF^*_{j,s} \mid LF^*_{j'\sim j,s}\right)\ell\left(\sum_{j'\sim j} LF^*_{j',s}\right) \qquad [4]$$

where ℓ denotes a likelihood and ℓ_p a pseudolikelihood and $i' \sim i$ denotes regions (i') in the same ρ-clique as region i. Likelihoods $\ell\left(LF^*_{i,new} \mid LF^*_{i'\sim i,new}\right)$ were estimated from maximum likelihood estimates (MLE) of clique-specific autologistic functions. Conversely, $\ell\left(\sum_{i'\sim i} LF^*_{i',new}\right)$ were estimated from MLE of a clique-specific probability mass function (binomial, zero-inflated-binomial, or beta-binomial). Finally $\ell\left(LF^*_{i,s}=\delta\,,\, LF^*_{j,s}=1-\delta\right)$, $\delta=\{1,0\}$ was estimated from the classified data as outlined in Congdon (2006, p 395).

After approximately 1000 accepted switches the Markov Chain reached a steady state so that the current value of the vector $\mathbf{LF}$ could be viewed as sampled from the joint distribution of regional *LF*-status (Robert and Casella 1999). As a safeguard, the vector $\mathbf{LF}^*_S$ after 3000 accepted switches was retained. A total of 41×100 random replicates of $\mathbf{LF}^*_S$ were generated, representing 100 replications of 41-year forecasts. Without a temporal autocorrelation, years and replicates are interchangable.

Forecasting BA (Stage II)

For a given *LF* forecast for region *i* the associated *BA* was determined as

$$BA^*_i = \begin{cases} \hat{F}_i^{-1}\left(u^* \mid LF^*_{i,S}=0\right),\ u^* \in \left[0, \hat{u}_{SFi}\right) \\ \hat{F}_i^{-1}\left(u^* \mid LF^*_{i,S}=1\right), u^* \in \left[\hat{u}_{SFi}, 1\right] \end{cases} \qquad [5]$$

where $\hat{F}_i^{-1}$ is the empirical quantile function of *BA* for region *i* (i.e. the inverse to the empirical distribution function), u^* is a random draw from a uniform distribution from a specified interval, and $\hat{u}_{SFi}$ is the MLE estimate of the regional cut-off quantile between *LF* = 1 and *SF* = 0 on the empirical quantile function. A

simulation study (Magnussen 2008) suggested the following endpoints for the quantile functions: $\hat{F}_i^{-1}(0) = 0.54 \times \mathrm{Min}(BA_i)$, and $\hat{F}_i^{-1}(1) = 1.35 \times \mathrm{M}\,ax(BA_i)$.

Results

Regional correlations of *BA* averaged 0.06 but were statistically significant (5% level) at a rate of 0.14. Only five regions (NB1, NF1, PQ1, PQ2, SK4) were not significantly correlated with at least one other region. A bootstrap simulation study confirmed that significant correlations were almost always due to a few concurrent years of large *BA*.

Examples of the classification of *BA* to *LF* (1) and *SF* (0) by the *k*-means procedure are in Figure 1. The average regional relative frequency of *LF* years was 0.19 but varied from a low of 0.05 in PQ4 to a high of 0.39 in SK2. The average cut-off point for classifying a *BA*%-value to *LF* was 0.7% (range: 0.04% in AB4 to 3.3% in AB3). Regional correlations of *BA* in shared *SF* years were, on average, about 84% below the correlations for the entire period of 41 years and the rate of significant correlations was consistent with the null hypothesis of a zero correlation.

In the MCMC forecasts the average rate of *LF* years was 2% below the rate in the classified data ($P = 0.12$, bootstrap *t*-test). The number of regional *LF* years forecasted for a 41-year period was positively correlated (0.77) with the observed number (Figure 2). There is a tendency to overestimate nLF_i in regions with lower frequencies of *LF* years and to underestimate in regions with higher rates. A bootstrap *t*-test with 36 degrees of freedom identified three regions (BC2, NS1, NWT1) with a significant difference $(0.018 \le P \le 0.026)$ between the classified and the forecasted number of *LF*-events during a period of 41 years.

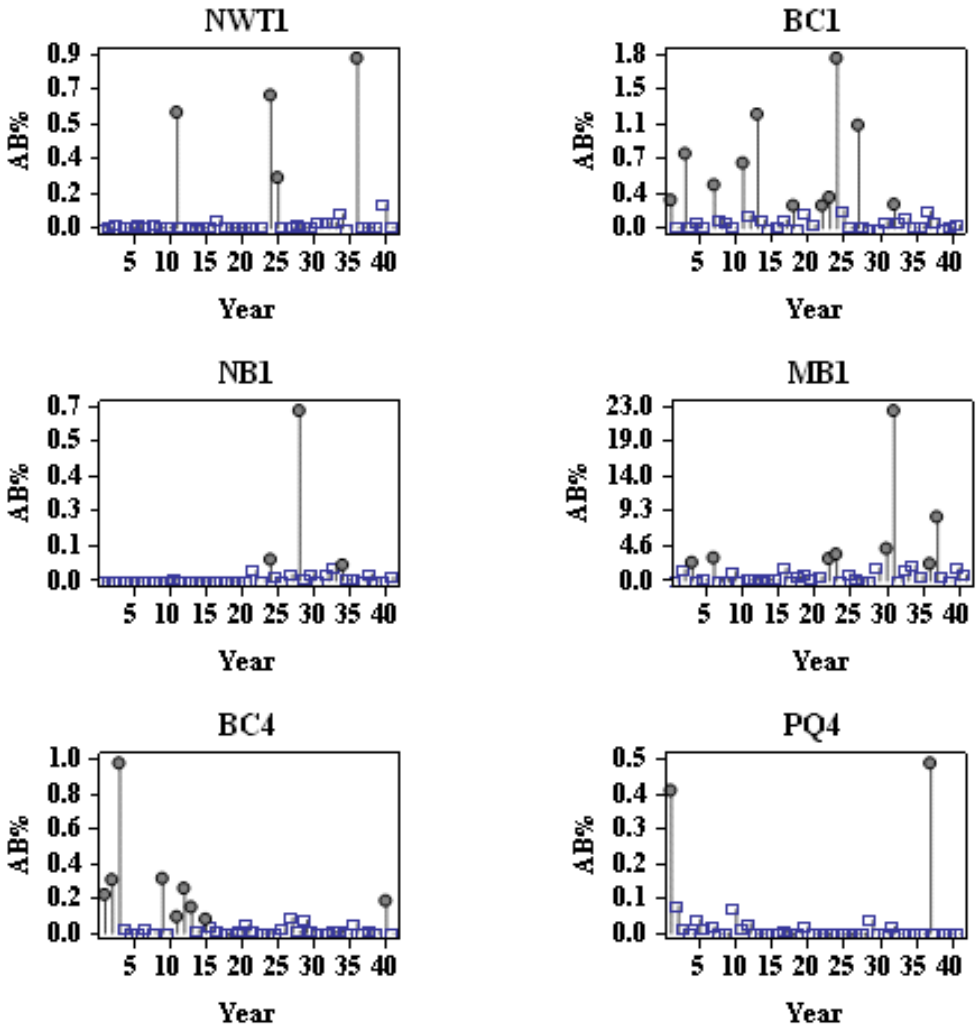


Figure 1: Percent of forested area burned annually (*BA%*) in six randomly chosen regions. Year 1 = 1959, year 41 = 1999. The classification of *BA%* to *LF* (large fire) or *SF* (small fire) is indicated by squares (*SF*) and circles (*LF*).

The bias pattern in Figure 2 carries over to regional correlations of *LF* years in the MCMC results (Figure 3) and created an inflation in cliques with a below average interregional correlation and vice versa. Across all regions, the average correlation of *LF* years was 0.04 in the forecasts and 0.05 in the data, and the relationship between the two sets of correlation coefficients was consistent with a slope of 1.0 and a zero intercept ($P = 0.16$).

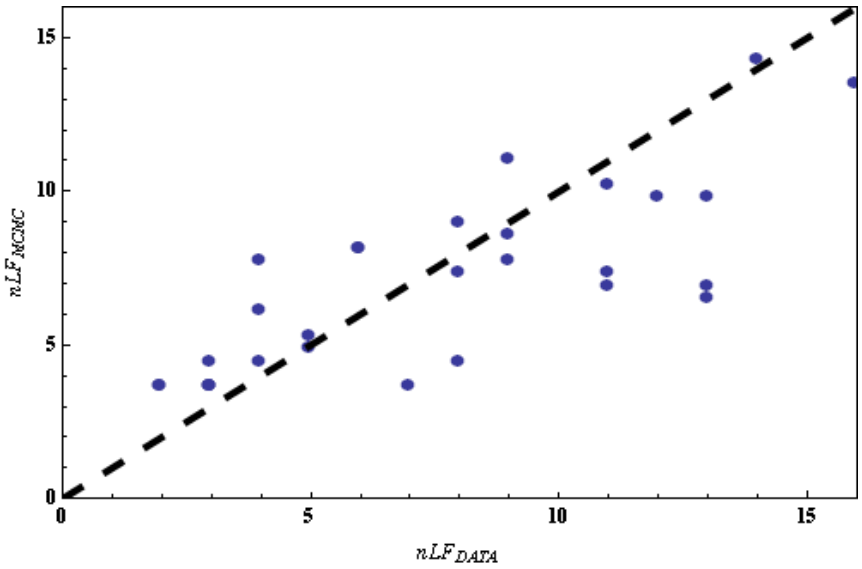


Figure 2: Forecasted total number of regional *LF* years during a period of 41 years (nLF_{MCMC}), plotted against the number in the classified data (nLF_{DATA}).

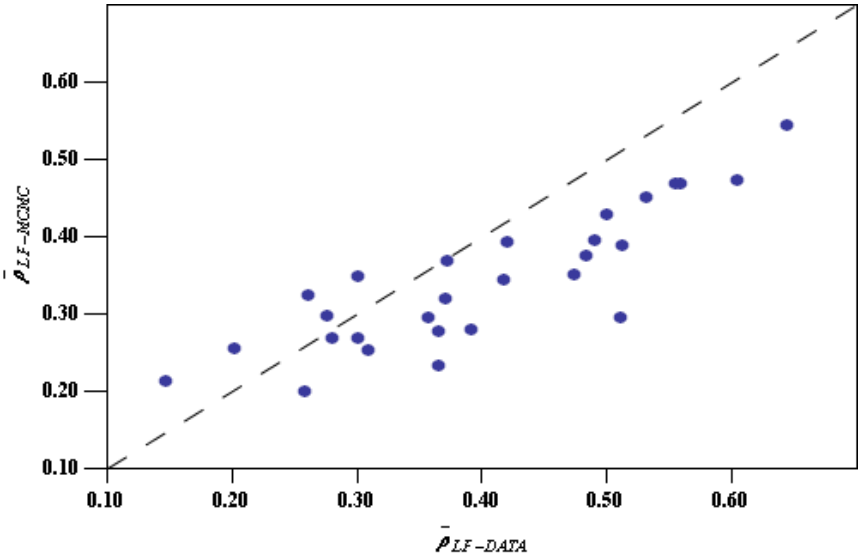


Figure 3: Forecasted average inter-regional correlation of *LF* years $\left(\overline{\rho}_{LF-MCMC}\right)$ plotted against the average correlation in the data $\left(\overline{\rho}_{LF-DATA}\right)$. The average is over regions in a ρ -clique.

The mean and variance of forecasted regional *BA* matched fairly closely their historic values. Scatter plots in Figures 4 and 5 convey a strong correlation (0.98) between forecasted and historic values. A linear relationship with a slope of 1.08 $\left(\pm 0.02\right)$ and an intercept not significantly different from zero ($P = 0.31$) captures the relationship. For all regions combined the average *BA* in the forecasts was 886 976 ha versus 817 308 ha in the data. The bias is attributed to the asymmetric capping of the empirical quantile functions. The standard deviation of the regional totals of *BA* was 849 596 ha but only 742 973 ha (-14%) in the MCMC forecasts.

The conversion to a binary variable (*LF*) and the ensuing attenuation of the regional correlations is the main factor behind the bias. Regarding regions as independent would generate a standard deviation of 486 990 ha (-43%). In other words, the MCMC procedure captured 65% of the regional covariance of *BA*.

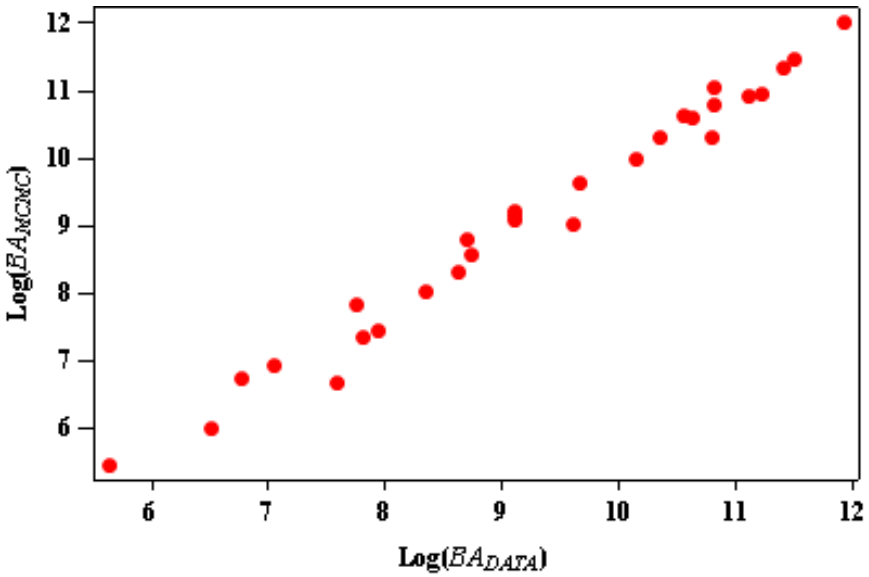


Figure 4: Forecasts of average annual regional *BA*-values $\left(BA_{MCMC}\right)$ plotted on a logarithmic scale against historic values $\left(BA_{DATA}\right)$.

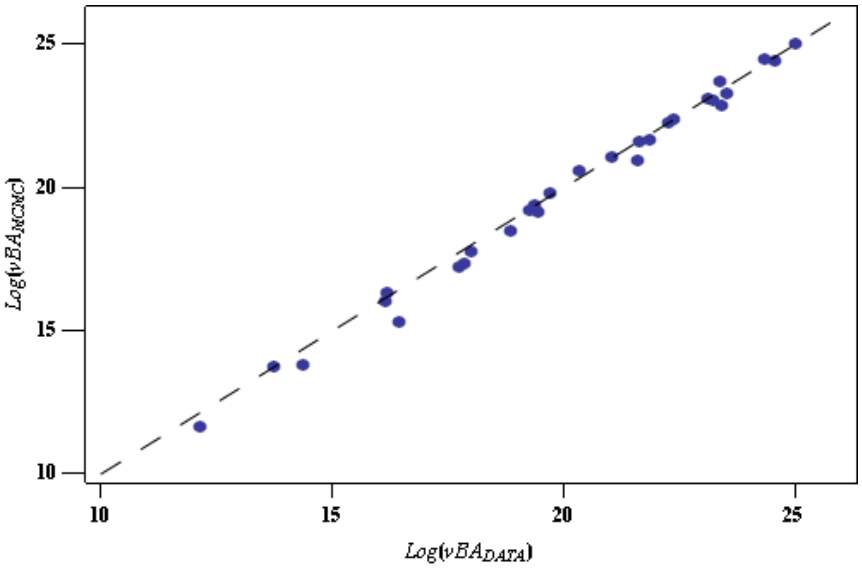


Figure 5: Forecasts of variance of annual regional *BA*-values $\left(vBA_{MCMC}\right)$ plotted on a logarithmic scale against historic values $\left(vBA_{DATA}\right)$.

Discussion and conclusions

A joint forecast of regional *BA* must take the apparent correlation structure into account or the variation of sums of regional values will be biased downward. Without a suitable multivariate distribution function, the task of a joint forecast becomes a complex challenge (Aalo and Piboongungon 2005, Carpenter and Diawara 2007).

A binary classification of *BA* as either large or small facilitates an interpretation of the regional correlation structure as it changes the focus from areas to years. Common environmental factors in region-years classified as *LF* may be identified. Modeling at the binary level also facilitates an integration of expected trends in *LF* years (Bergeron et al. 2004, Beverly and Martell 2003, Larsen 2007).

Modeling a regional distribution of correlated binary variables is commonly done by formulating the probability of an event in a region conditional on the number of events in some defined neighbourhood (ρ-cliques) composed of interdependent regions, usually a group of first-order spatial neighbours (Gilliland and Schabenberger 2001, Sherman et al. 2006).

The proposed MCMC procedure was simplified by conditioning on the marginal distribution of the total number of *LF* events in a year. Without this simplifying step, the transition kernel would have been considerably more complicated (Smith and Smith 2006). We surmise that our MCMC results reflect the constraints on the covariance structure inherent in all multivariate distribution functions (Johnson 1987). A restriction of first-order regional interactions to ρ-cliques limited our ability to capture the observed interregional correlation structure. The proposed MCMC approach is capable of reproducing the main features of observed marginal distributions and an irregular and patchy correlation structure.

References

Aalo, V.A.; Piboongungon, T. 2005. On the multivariate generalized gamma distribution with exponential correlation. In: Global telecommunications conference IEEE 3(28): 1229-1233.

Amiro, B.D.; Todd, J.B.; Wotton, B.M.; Logan, K.A.; Flannigan, M.D.; Stocks, B.J.; Mason, J.A.; Martell, D.L.; Hirsch, K.G. 2001. Direct carbon emissions from Canadian forest fires, 1959-1999. Canadian Journal of Forest Research 31: 512-525.

Bergeron, Y.; Flannigan, M.; Gauthier, S.; Leduc, A.; Lefort, P. 2004. Past, current and future fire frequency in the Canadian boreal forest: implications for sustainable forest management. Ambio 33:356-360.

Beverly, J.L.; Martell, D.L. 2003. Characterizing historical variability in boreal forest fire frequency: Implications for fire and forest management. In: Arthaudf, G.J.; Barrett,

T.M. (eds). Proceedings Systems Analysis in Forest Resources. Volume 7: Managing Forest Resources. p 3-14. Kluwer, Dordrecht NL. 326 p.
Bond-Lamberty, B.; Peckham, S.; Ahl, D.; Gower, S. 2007. Fire as the dominant driver of central Canadian boreal forest carbon balance. Nature 450: 89-92.
Carpenter, M.; Diawara, N. 2007. A multivariate gamma distribution and its characterization. Technical Report. Dep Math & Statistics. Auburn Univ.
Congdon, P. 2006. Bayesian statistical modelling. Wiley, Chichester, England. 571 p.
Gillett, N.P.; Weaver, A.J.; Zwiers, F.W.; Flannigan, M.D. 2004. Detecting the effect of climate change on Canadian forest fires. Geophysical Research Letters 31: L18211.
Gilliland, D.: Schabenberger, O. 2001. Limits on pair-wise associations for equi-correlated binary variables. Journal of Statistical Science 4: 279-285.
Griffiths, D.A. 1973. Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics 29: 637-648.
Johnson, M.E. 1987. Multivariate statistical simulation. Wiley, New York. 230 p.
Kurz, W.A.; Apps, M.J. 2006. Developing Canada's national forest carbon monitoring, accounting and reporting system to meet the reporting requirements of the Kyoto protocol. Mitigation and Adaptation Strategies in Global Change 11: 33-43.
Kurz, W.; Stinson, G.; Rampley, G.; Dymond, C.; Neilson, E. 2008. Risk of natural disturbances makes future contribution of Canada's forests to the global carbon cycle highly uncertain. Proceedings National Academy of Science USA 105: 1551-1555.
Larsen, C.P.S. 2007. Fire and climate dynamics in the boreal forest of northern Alberta, Canada, from AD 1850 to 1989. The Holocene 6: 449-456.
Magnussen, S. 2008. Joint Regional Simulation of Annual Area Burned in Canadian Forest Fires. The Open Forest Science Journal 17: 37-53.
Robert, C.P.; Casella, G. 1999. Monte Carlo statistical methods. Springer, New York. 507 p.
Sherman, M.; Apanasovich, T.V.; Carroll, R.J. 2006. On estimation in binary autologistic spatial models. Journal of Statistical Computation and Simulation 2: 167-179.
Skinner, W.; Skinner, M.; Flannigan, M.; Stocks, B.; Martell, D.; Wotton, B.; Todd, J., Mason; J.A., Logan, K.; Bosch E.M. 2008. A 500 hPA synoptic wildland climatology for large Canadian forest fires, 1959-1996. Theoretical and Applied Climatology 71: 157-168.
Skinner, W.; Stocks, B.; Martell, D.; Bonsal, B.; Shabbar, A. 1999. The association between circulation anomalies in the mid-troposphere and area burned by wildland fire in Canada. Theoretical and Applied Climatology 63: 89-105.
Smith, D., and Smith, M. 2006. Estimation of binary Markov random fields using Markov chain Monte Carlo. Journal of Computational and Graphical Statistics 15: 207-227.
Stocks, B.J.: Majson, J.A.; Todd, J.B.; Bosch, E.M.; Wotton, B.M.; Amiro, B.D.; Flannigan, M.D.; Hirsch, K.G.; Logan, K.A.; Martell, D.L.; Skinner, W.R. 2002. Large forest fires in Canada, 1959-1997. Journal of Geophysical Research 108: 1-12.

Rapid Assessment of Wildfire Damage Using Forest Inventory Data: A Case in Georgia

Richard A. Harper[1]
John W. Coulston[2]
Jeffery A. Turner[3]

Abstract: *The rapid assessment of damage caused by natural disasters is essential for planning the appropriate amount of disaster relief funds and public communication. Annual Forest Inventory and Analysis (FIA) data provided initial estimates of damage to timberland in a timely manner to State leaders during the 2007 Georgia Bay Complex Wildfire in southeast Georgia. FIA plots were selected from within the shape file (a polygon outline of the fire perimeter) of the burn area and processed with the National Information Management System. Forest area and total volume by stumpage products were compiled for the wildfire area. A mortality factor determined by the Georgia Forestry Commission was used to estimate the volume of damaged timber, and the value of damaged timber was assessed using Timber Mart-South stumpage prices.*

Keywords: FIA, forest disturbance, inventory, Mapmaker, stratification, wildfire.

Introduction

Large wildfires have been documented in southeast Georgia for more than 100 years. Climatic conditions create droughts that foster conditions for large wildfires in and around the Okefenokee National Wildlife Refuge and Wilderness Area. Wildfires, associated with droughts, were documented in 1844, 1860, 1910, 1932, 1954-1955, and 2007. The fire of 1932 began when a young boy started a fire to warm his hands. The intensity of the fire created gale force winds that worsened the damage. The drought occurring in the 1950s was severe, and the fires of 1954-1955 were peat fires that burned underground and therefore impossible to control. Analysis of peat samples back in the 1890s found it to be 85 percent combustible (Izlar 2007). In 2007, 9,500 wildfires burned about 504,000 acres in Georgia. Of particular note was the Georgia Bay Complex Wildfire which

[1] United States Forest Service, Southern Research Station, Forest Inventory and Analysis, 4700 Old Kingston Pike, Knoxville, TN 37919 USA; raharper@fs.fed.us

[2] United States Forest Service, Southern Research Station, Forest Inventory and Analysis, 4700 Old Kingston Pike, Knoxville, TN 37919 USA; jcoulston@fs.fed.us

[3] United States Forest Service, Southern Research Station, Forest Inventory and Analysis, 4700 Old Kingston Pike, Knoxville, TN 37919 USA; jturner02@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

burned 441,705 acres in southeast Georgia and destroyed nine homes (Georgia Forestry Commission 2007a). This fire started when a tree fell on a power line creating sparks that ignited dry, woody fuels.

The Forest Inventory and Analysis (FIA) program plays an important role in quantifying losses from broad-scale disturbances, and FIA personnel have several options available to perform rapid assessments of these disturbances. The options include Mapmaker analysis, analysis based on FIA database (the FIA public database), and analysis based on the National Information Management Systems (NIMS)—using estimation procedures documented in Bechtold and Patterson (2005). Each of these analytical methods requires different levels of effort and requires different turn-times. The key to rapid assessment is to provide required information for disaster relief in a timely fashion, which helps quantify losses and aid in recovery decisions.

The objective of this paper is to describe the assessment techniques used to quantify potential timber losses from the Georgia Bay Complex Wildfire.

Methods

On Memorial Day weekend 2007, the Georgia Forestry Commission (GFC) requested help in assessing damage from what was discovered to be the largest recorded wildfire in Georgia history (and the Southeastern United States). The GFC needed this information to respond to inquires from the media, State and Congressional representatives, and there were reports that the President of the United States would visit Georgia to view the damage. The first option was to use Mapmaker (Miles 2007), the FIA online database retrieval system tool, which would provide first-line estimate of the timber resource within the known burn area. GFC needed some assurance that their initial estimates were reasonable. A shape file was sent via email from the USDA Forest Service, Region 8, Southern Area Coordination Center, a key player in the National Incident Management System. The shape file had been constructed by firefighters through global positioning system (GPS) receivers. From the shape file, coordinates outlining the burn area were digitized using Arc Map and copied into the Mapmaker custom table retrieval polygon window. The polygon captured the plots from the Georgia 2004 annual inventory and provided the first estimate of total timber volume within the early extent of the burn area (fig 1).

Mapmaker queries separated volume into the major species groups of softwood and hardwood, sawtimber and poletimber, ownership class, age class, and stand origin. From these queries, tables were developed for forest stumpage products by softwood and hardwood. From young age classes, a table was developed to estimate the area of precommercial stands. Assumptions were made that precommercial planted pine stands ranged from 1 to 10 years old and precommercial natural stands ranged from 1 to 15 years old.

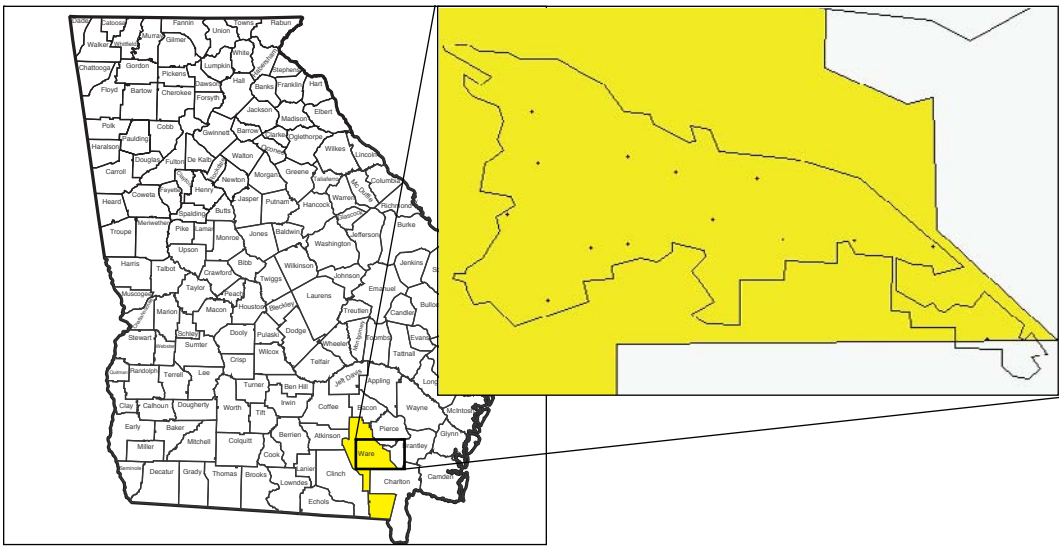

Ware
Charlton
Echols
Clinch
Pierce
Atkinson
Coffee
Bacon
Wayne
Brantley
Camden
Glynn
McIntosh
Berrien
Cook
Lowndes
Brooks
Thomas
Grady
Decatur
Colquitt
Mitchell
Miller
Early
Baker
Calhoun
Clay
Randolph
Terrell
Lee
Dougherty
Worth
Turner
Ben Hill
Irwin
Tift
Crisp
Dooly
Wilcox
Telfair
Pulaski
Dodge
Wheeler
Montgomery
Treutlen
Candler
Evans
Tattnall
Toombs
Bulloch
Jenkins
Emanuel
Johnson
Laurens
Twiggs
Wilkinson
Bibb
Crawford
Peach
Houston
Macon
Marion
Taylor
Talbot
Harris
Troupe
Meriwether
Upson
Pike
Lamar
Monroe
Jones
Baldwin
Washington
Jefferson
Burke
Richmond
Columbia
Lincoln
Wilkes
Greene
Morgan
Walton
Oconee
Clarke
Oglethorpe
Madison
Elbert
Hart
Franklin
Banks
Jackson
Barrow
Gwinnett
Forsyth
Cherokee
Dawson
Lumpkin
White
Stephens
Habersham
Rabun
Towns
Union
Fannin
Gilmer
Murray
Whitfield
Walker
Chattooga
Floyd
Polk
Haralson
Carroll
Heard
Coweta
Fayette
Clayton
Douglas
Fulton
De Kalb
Rockdale
Newton
Henry
Spalding
Butts
Jasper
Putnam
Hancock
Warren
McDuffie
Glascock
Taliaferro
Oconee
Greene
Oglethorpe
Elbert
Hart
Franklin
Stephens
Habersham
Rabun
Towns
Union
Fannin
Gilmer
Murray
Whitfield
Walker
Chattooga
Floyd
Polk
Haralson
Carroll
Heard
Coweta
Fayette
Clayton
Douglas
Fulton
De Kalb
Rockdale
Newton
Henry
Spalding
Butts
Jasper
Putnam
Hancock
Warren
McDuffie
Glascock
Taliaferro
Columbia
Lincoln
Wilkes
Greene
Morgan
Walton
Oconee
Clarke
Oglethorpe
Madison
Elbert
Hart
Franklin
Banks
Jackson
Barrow
Gwinnett
Forsyth
Cherokee
Dawson
Lumpkin
White



Figure 2: Aerial views showing the mosaic pattern of burn intensities across the landscape helped estimate areas of mortality. (Courtesy of the Georgia Forestry Commission).

It should be noted that this method assumes that timber was valued at "pre-fire" stumpage prices to estimate market value loss. However, post-fire stumpage value is discounted because of fire-caused timber damage, and due to increased volume introduced to the market (increased supply), which required salvage in a timely manner—regardless of fluctuating markets. It also is difficult for local industry to manufacture and market a dramatically increased supply.

The assessment using Mapmaker complemented the initial estimate of losses caused by the wildfire. The actual area of wildfire-affected stands grew beyond initial estimates, and because more time was available, a refined estimate was provided based on the expanded boundary and using the NIMS compilation system. Working with the raw data allowed flexibility needed to produce the best assessment with FIA plot data. The final shape file of the burn area was used to select actual plots from the Southern Research Station database (not fuzzed and swapped plots as is the normal procedure to comply with privacy laws) within the boundary of the burn area. The area estimation unit was established to the shape file allowing the expansion factors to directly represent the burn area. National Land Cover Data were used to stratify the FIA plots by forest/nonforest and the Okefenokee area to allow further refinement of the estimation unit and expansion factors for the area of interest. The data were compiled in NIMS to calculate forest volume of softwood and hardwood by ownerships, forest products categories, and precommercial forest area.

Results

Softwood and hardwood tables were developed by ownership classes and types of forest products. The detailed private ownership could be developed to compare with estimates reported by forest industry and individuals. The unadjusted value for all forest land within the burn area was estimated to be $284 million (table 1). Because the Okefenokee is by statute reserved forest land (not available for timber harvest), the timber volume and value were removed from the estimate leaving a

Table 1: Detailed breakout of total volume and value by ownership and forest products

Pulpwood	Volume	Unit value[a]	Total value
	cords	- - - - - - - - - - *dollars* - - - - - - -	
Softwood			
Okefenokee	1,302,821	$ 19.40	$ 25,274,731
Other federal	181,783	$ 19.40	$ 3,526,590
State	14,781	$ 19.40	$ 286,753
Industry	779,542	$ 19.40	$ 15,123,113
NIPF	166,750	$ 19.40	$ 3,234,948
Total	2,445,677	$ 19.40	$ 47,446,134
Hardwood			
Okefenokee	1,686,307	$ 21.07	$ 35,530,487
Other federal	370,864	$ 21.07	$ 7,814,105
State	7,179	$ 21.07	$ 151,271
Industry	88,451	$ 21.07	$ 1,863,659
NIPF	852	$ 21.07	$ 17,957
Total	2,153,654	$ 21.07	$ 45,377,479

Sawtimber	Volume	Unit value[a]	Total value
	million board feet[b]	- - - - - - - - - - *dollars* - - - - - - -	
Softwood			
Okefenokee	493,968	$ 262.00	$ 129,419,512
Other federal	93,015	$ 262.00	$ 24,369,861
State	9,887	$ 262.00	$ 2,590,488
Industry	83,864	$ 262.00	$ 21,972,402
NIPF	9,994	$ 262.00	$ 2,618,329
Total	690,727	$ 262.00	$ 180,970,592
Hardwood			
Okefenokee	37,902	$ 131.00	$ 4,965,221
Other federal	30,258	$ 131.00	$ 3,963,735
State	940	$ 131.00	$ 123,096
Industry	6,367	$ 131.00	$ 834,054
NIPF	—	$ 131.00	$ —
Total	75,466	$ 131.00	$ 9,886,107
Georgia Total			$ 283,680,312

[a] Timber Mart-South (2007).

[b] Thousand board feet, International log rule.

total value of $88 million. Softwood sawtimber made up the largest component ($52 million). Forest industry and other Federal lands had the largest total value, each totaling almost $40 million. GFC considered these total values and adjusted them based on estimated mortality factors to report the value of timber lost from fire damage. After the final analysis of the Georgia Bay Complex Wildfire, GFC reported the timber loss at $54 million.

Discussion

The initial assessment using Mapmaker had limits in the estimation. A large portion of the burn area was in the Okefenokee National Wildlife Refuge and Wilderness Area. By statute, the Okefenokee is reserved forest land and has no commercial timber value, i.e., it is not available for timber harvesting. Because of the hazards to field crews penetrating the Okefenokee area during recent data collection, there were a limited number of plots that accurately would facilitate a true assessment of timber volume within its boundary. The Okefenokee reserved forest land area was separated from other ownerships to allow flexibility in assessing damage for reporting.

The area estimation unit used to develop plot expansion factors by Mapmaker was the Southeast Survey Unit for Georgia. This somewhat skewed expansion factors for plots within the burn area. Because privacy laws require plots to be "fuzzed and swapped," there were possibilities that plots on the edge of the burn area do not represent the timber resource within the burn area—an introduced bias. Most timberland outside the Okefenokee is private ownership, and there could be no stratification of ownership types within the private group (privacy laws) using Mapmaker.

Conclusion

The initial response to the Georgia Bay Complex Wildfire using Mapmaker gave a quick comparison to on-the-ground estimates before reporting to policy makers and the media. As the wildfire continued to expand, FIA staff was able to refine estimates, and use plot data directly from the NIMS compilation system. Area stratification allowed segregation of tree volume estimates within the Okefenokee National Wildlife Refuge and Wilderness Area, where limited plot data were available, therefore, limiting the focus of the study to commercial timberland and a break-out of private ownerships. Even though some estimates had high sampling errors, the data offered a comparison to general volume per acre ground estimates.

Responses to the Georgia Bay Complex Wildfire provided insight for improvement of future rapid response to catastrophic events. Flexibility of annual FIA data compilation coupled with online tools, continue to improve and

provide better assessment of rapid response, which are suitable in a variety of applications. As awareness gains momentum among the growing and diverse FIA user groups and more public tools are developed, the FIA data offer opportunities for a variety of assessments that deal with the complexity of forest resources and their management.

Acknowledgments

The authors greatly appreciate suggestions made by Frank Green and James Johnson with the Georgia Forestry Commission. Their peer reviews were of substantial use. Firefighters, in addition to controlling the blaze, gathered the GPS coordinates surrounding the burn area; and the USDA Forest Service, Region 8, Southern Area Coordination Center compiled the shape file. The Georgia Forestry Commission and forest landowners worked together to assess mortality. Timber Mart-South, Frank W. Norris Foundation—at the University of Georgia—provided the 1st Quarter 2007 stumpage prices for Georgia. FIA field crews with GFC and the USDA Forest Service collected plot data. FIA staff responded during a holiday weekend to answer the call for assistance. This team of dedicated professionals are all part of this rapid response forest resource assessment.

References

Bechtold, William A.; Patterson, Paul L., Editors. 2005. The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS–80. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 85 p.

Georgia Forestry Commission. 2007a. Georgia wildfires of 2007 summary of facts and costs for recovery. [Online]. http://www.gfc.state.ga.us/GFCNews/GFCNewsArchive.cfm. [August 21, 2007].

Georgia Forestry Commission. 2007b. Wildfire damage assessment for Sweat Farm Road fire. [Online]. http://www.gfc.state.ga.us/GFCNews/GFCNewsArchive.cfm. [September 3, 2007].

Timber Mart-South, Frank W. Norris Foundation. 2007. Georgia Stumpage and Delivered Prices, 1st Quarter 2007. Athens, GA: University of Georgia, Warnell School of Forest Resources.

Izlar, B. 2007. Fire in the Okefenokee. Georgia Forestry Today. 3(1): 43.

Miles, Patrick D. 2007. Forest inventory mapmaker web-application version 2.1. [Online]. http://fia.fs.fed.us/tools-data/other/default.asp. [May 2007].

Cool Tools


Forest Inventory & Analysis Symposium
2008
Climate Change, Fire & Other Hot Topics

Using FIA Data in the Forest Vegetation Simulator

John D. Shaw[1]

ABSTRACT: *The Forest Vegetation Simulator (FVS) is a national system of forest growth models maintained by the USDA Forest Service. It is the official tool for stand growth projection on National Forest lands, but it is also used widely on other ownerships. Model extensions and post-processors permit FVS users to perform a broad range of functions, including silvicultural manipulations, wildlife habitat analyses, and fuel treatment evaluations. Because FIA data were made available in FVS-ready format through the FIA Mapmaker interface, an increasing number of users have been using FVS as their tool of choice for compilation of FIA data at the plot level. With the transition from Mapmaker to FIDO, users who have built analysis systems around this data availability have lost access to new data. Due to the need to update FIA-FVS data translation, there is an opportunity to re-design the system to eliminate prior limitations and take advantage of recent developments in FIA and FVS. Select capabilities of FVS, and potential modifications and enhancements to the FIA-FVS linkage are discussed.*

Keywords: Forest Vegetation Simulator, FVS, FIA Mapmaker, database, data access, ODBC

Introduction

Open access to Forest Inventory and Analysis (FIA) data has resulted in a greatly increased user base in recent years. As the number and diversity of users has increased, so has the demand for access to the data in different forms. Currently, FIA data are served through Forest Inventory Data Online (FIDO; http://fia.fs.fed.us/tools-data/), which allows users to generate reports using pre-defined and customized queries, and through the FIA datamart, which provides field-measured and computed data for inventory plots in download files (see http://fiatools.fs.fed.us/fiadb-downloads/datamart.html). Access to these data is critical to users who wish to conduct analyses that are beyond the scope of FIDO. Among these users are those who desire to project FIA plot conditions forward using forest growth simulators or other models.

The Forest Vegetation Simulator (FVS) is a national system of forest growth models maintained by the USDA Forest Service (Johnson 1997, Dixon 2002). It is the official tool for stand growth projection on National Forest lands, but can be used on land of any ownership. Model extensions and post-processors permit FVS users to perform a broad range of functions, including silvicultural

[1] Rocky Mountain Research Station, Forest Inventory and Analysis, 507 25th Street, Ogden, UT 84401. Ph. +001-801-598-5902 Email: jdshaw@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

manipulations, wildlife habitat analyses, and fuel treatment evaluations. A flexible programming environment also allows users to produce customized variables and output tables. This allows users to extract summary information from inventory data without having to manipulate the data directly.

FIA data were originally made available in FVS-ready format through the FIA Mapmaker interface until Mapmaker version 2.1. With the availability FIA data in FVS-ready format, an increasing number of users have been using FVS as their tool of choice for compilation of FIA data at the plot level. Many of these users were already familiar with FVS capabilities, and had an interest in using FIA data. The access to data in FVS-ready format facilitated their use of the data by eliminating the need to develop their own compilation methods. This accessibility led to a substantial amount of use; Miles (2008) reported that there were 2,386 downloads of FVS-ready data over a 4-year period.

In the Mapmaker release, FIA data are provided in the file formats introduced with the Suppose interface (Crookston 1997): a location file (.LOC) that contains information about the stands in an inventory location or project (i.e., those included in the Mapmaker download), a stand list file (.SLF) that contains stand-level data and refers to the files containing tree-level data, and one or more tree data files (.FVS), each of which contains the data from an individual stand (Dixon 2002). In the case of FIA data, each .FVS file contains data for all trees recorded on an individual FIA plot. In the data coding, FIA subplots may be treated as FVS plots, and the FIA plot is considered a "stand".

With the transition from Mapmaker to FIDO, users who built analysis systems around this data availability have lost access to newer FIA data. The most recent available data are from 2005, and there have been many user requests for the most current data. Because of this demand, there is a need to update the FIA-FVS data translation process. There is also an opportunity to re-design the translation process to eliminate prior limitations and take advantage of recent developments in FIA and FVS.

One of the most important FVS developments, and one that is key to the FIA-FVS data link, is the implementation of Version 2.0 of the FVS Database Extension (Forest Vegetation Simulator Staff 2003). The new database version eliminates the need for multiple input files, replacing the SLF and FVS files with three tables that can be accessed using Open Database Connectivity (ODBC). These tables may be contained in Microsoft Access, Microsoft Excel, or SQL-based relational databases such as Oracle (Forest Vegetation Simulator Staff 2003). One important advantage of this structure is that the database can include fields that are not used by FVS, but which may be important to the user for other purposes (e.g., last treatment dates or local cover type designations).

Taking advantage of additional data and new program features can be accomplished with relative ease, but there are several issues that should be

recognized by FIA developers and FVS users. Once these issues have been addressed, the necessary enhancements to the FIA-FVS linkage can be developed.

What FIA Developers Need to Know about FVS and FVS Users

FVS is the nationally supported forest growth modeling framework for the USDA Forest Service. It is maintained by the Forest Service Forest Management Service Center (http://www.fs.fed.us/fmsc/), and the program and source code is freely available. FVS is actually a collection of forest growth models, known as FVS variants, that are run under a common interface called Suppose (Johnston 1997, Crookston 1997). Many of these variants trace their lineage to the Stand Prognosis Model (Stage 1973, Wykoff and others 1982), but other models such as TWIGS and ORGANON are the growth engines of some variants. Variants typically cover specific geographic areas, with limited overlap (Figure 1). There are variants for all U.S. forests, with the exceptions of interior Alaska, Hawaii, and U.S. territories. However, not all species, or even all common species within a given geographic area may be included in the local variant. Model updates are frequent and ongoing; new variants for interior Alaska and Maine are in development, and Prognosis-based variants are being developed to replace some that are currently TWIGS-based.

FVS requires only species, diameter, and the number of trees per acre as minimum data. Unmeasured tree characteristics, such as height and crown ratio, are estimated with dubbing submodels. However, if variables such as height and crown ratio are measured, they are used in various submodels. These and other variables, such as periodic increment, may also be used to calibrate specific submodels if certain criteria, such as a minimum number of observations, are met. Internally, all variants include single-tree, distance-independent growth models. Stand density and tree rank affect growth and mortality. Although inter-tree distance is not used explicitly, diameter growth models and the mortality routines are sensitive to within-stand variability as represented by varying density among plots in the stand.

What sets FVS apart from other growth models is its capability to simulate silvicultural operations and their effect on future stand development. Through the use of keywords and custom scripts, users can implement a wide range of silvicultural operations spanning simulation periods of up to 300 years. One of the most common uses is to evaluate growth and yield implications of multiple management options through a series of “what if” simulations. In addition to modeling growth responses, model extensions for insect, disease, and fire permit users to model stand resistance and resilience in the face of anticipated disturbances.

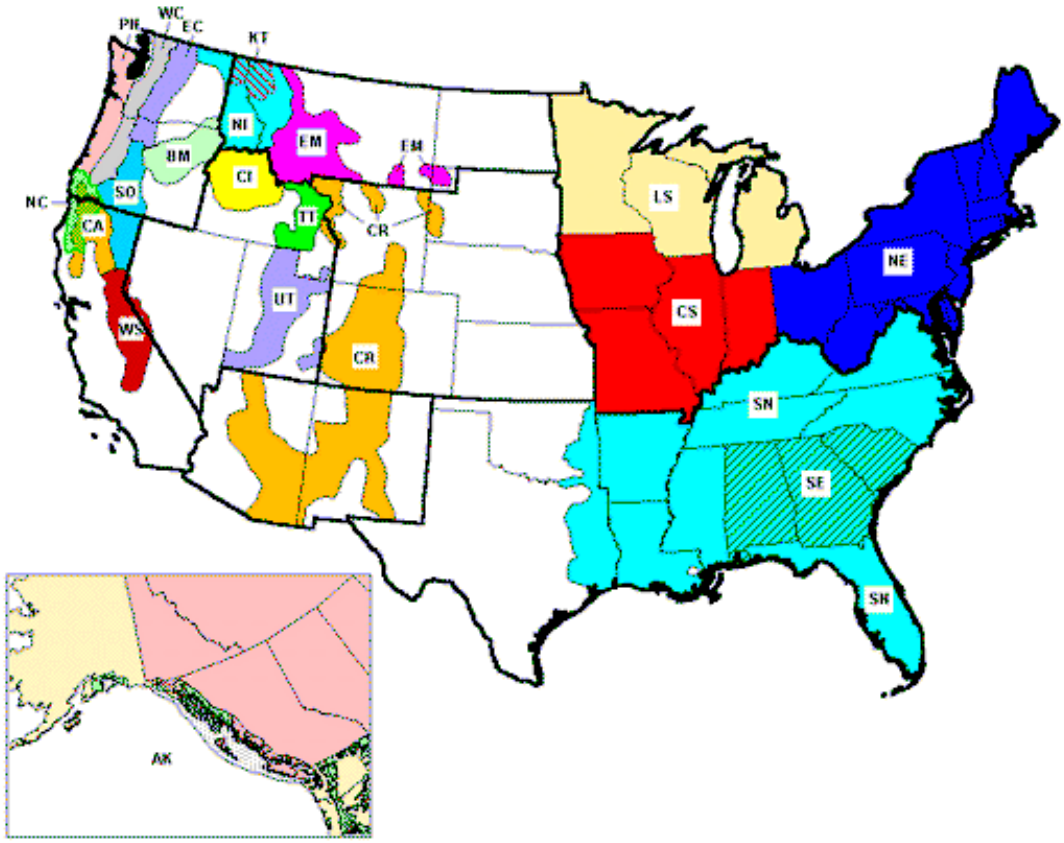


FIGURE 1: Map of FVS variant coverage in the coterminous 48 states and Alaska from the FMSC web site (http://www.fs.fed.us/fmsc). AK = Southeast Alaska/Coastal British Columbia, BM = Blue Mountain, CA = Inland California/Southern Cascades, CI = Central Idaho, CR = Central Rockies, CS = Central States, EC = East Cascades, EM = Eastern Montana, KT = Kootenai/Kaniksu/Tally Lake, LS = Lake States, NC = Klamath Mountains, NE = Northeast, NI = Northern Idaho/Inland Empire, PN = Pacific Northwest Coast, SE = Southeastern (superceded by SN), SN = Southern, SO = South Central Oregon/Northeast California, TT = Tetons, UT = Utah, WC = Westside Cascades, WS = Western Sierra Nevada.

The core FVS user base is primarily made up of silviculturists and vegetation management planners, but the user base has greatly expanded since the introduction of the Suppose interface. Suppose eliminated the need for the manual scripting and file management that was necessary in earlier versions of FVS, allowing users with little or no programming experience to run relatively complex simulations. The Fire and Fuels Extension (Reinhardt and Crookston 2003) has been used not only as a treatment evaluation tool, but also as an educational tool that managers have used to demonstrate the effects of proposed fuel treatments. Other important user groups include wildlife managers, economists, remote sensing specialists, and educators. Entry-level skills in FVS are now part of many university forestry curricula (Shaw and Long 2002).

Connections Between FIA and FVS

Currently there are no formal connections between the FIA and FVS programs, but there is ample opportunity to expand on informal connections that have been developed though various projects. As noted earlier, FVS is used as a compilation tool by some users of FIA data. Conversely, the data produced by the FIA program can be used to enhance the use and development of FVS. For example, Donnelly and others (2001) relied heavily on FIA data for development of the Southern variant of FVS. For variants that have already been developed, FIA data may be used for calibration or validation (figure 2). For example, since the initial data-gathering effort for the Southern variant, there have been as many as two full remeasurement cycles in some southern states. These data could be valuable for calibration and validation of the submodels developed for the first version of the variant. Validation applications may become extremely important for the next generation of FVS variants, which are to be climate sensitive. Certain submodels may require periodic adjustment as the effects of climate change become better known.

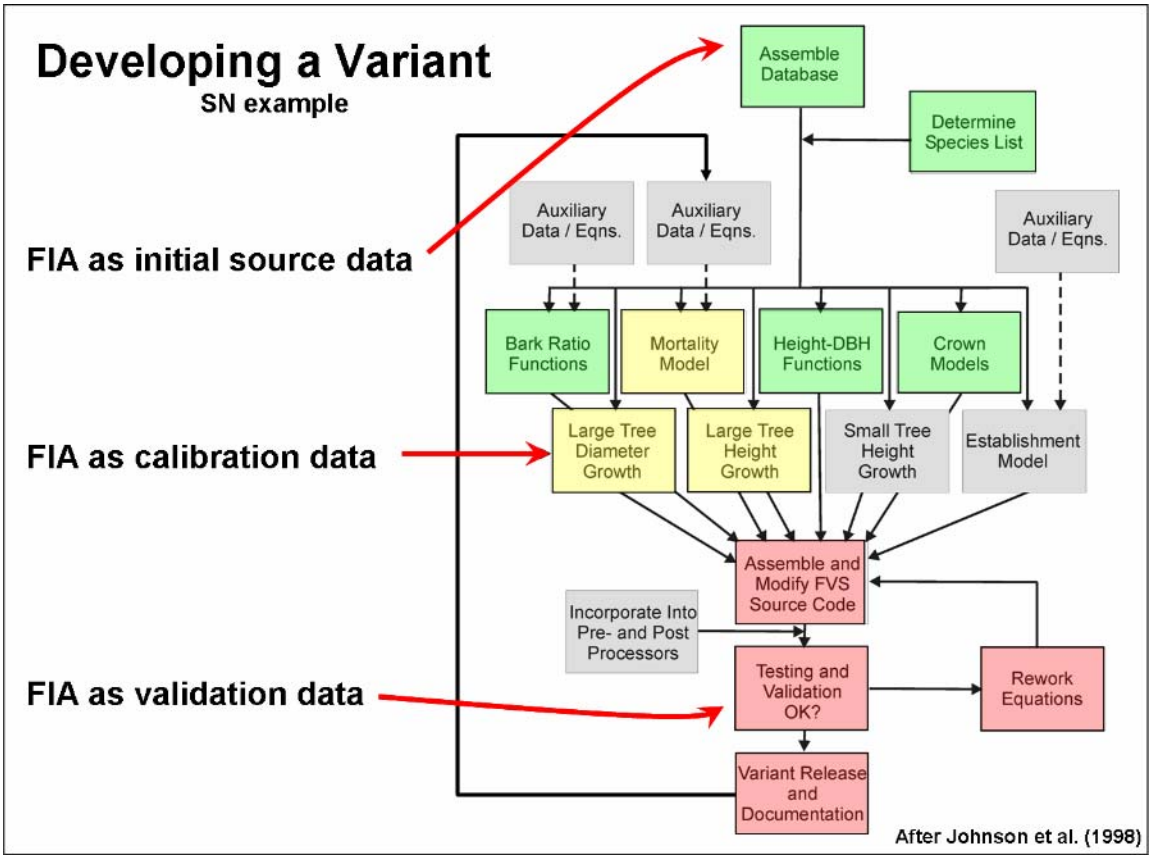


FIGURE 2: Generalized process for development of an FVS variant, and stages in the process where FIA data may be used.

FVS model validation is a current topic of interest among FVS users and researchers, so it would be beneficial to make the FIA database available to FVS users in the fullest extent possible. There are many variables in FIA data that are not usable by FVS directly, but which may be useful to users. Some of these variables may be used for data stratification, or as auxiliary variables during model testing. Because of the large number of FIA variables, it is impossible to comprehensively address all of the possibilities here. However, it is possible to discuss the variables that are used by all FVS base variants and discuss the feasibility of cross-walking them with FIA variables.

The remainder of this paper will describe the variables used in the base FVS variants, the corresponding FIA variables, and the issues and possible solutions that have been identified in cases where direct variable transfer is not possible. The list of variables includes all of those that are included in the StandInit and TreeInit tables in the database version of FVS input files (Forest Vegetation Simulator Staff 2008). Due to the large number of variables involved, they are listed in a table format with the FVS variable in the first column, the corresponding FIA variable(s) in the second column, and a description of the variable(s) and related notes in the last column. In the description column, the original FVS description appears in normal text. Additional text that describes the relationship between the FVS and FIA variables appears in bold.

TABLE 1: Variables in the predefined FVS_StandInit table structure, their corresponding FIA variables, and descriptions.

FVS variable	FIA variable	Description and notes
Stand_CN	PLOT.CN or COND.CN	Database control number. Required by Suppose 2.0 when populating stand lists. **FIA uses several control numbers as linkages between tables. Some may be usable in FVS.**
Stand_ID	Possible composite of INVYEAR, STATE, COUNTYCD, and LOC	Stand identification code. Required by Suppose 2.0 when populating stand lists. **The original translation of FIA data to FVS uses a composite stand ID that included the state, county, plot, and inventory year of the FIA plot. This variable was useful to users, because the general location of plots could be quickly identified during simulations.**
Variant	No FIA variable	The two character variant identification code. Required by Suppose 2.0 when populating stand lists. **In the original Mapmaker release, users were prompted to supply the name of the variant that would be used with the FIA data. It is possible that this step can be made transparent to the user by mapping variant coverage areas (see Discussion below).**
Inv_Year	MEASYEAR	The stand's inventory year corresponding to IY(1) in FVS. Required by Suppose 2.0 when populating stand lists. **FIA uses an inventory year variable (INVYR) in the PLOT table, but this represents the panel or subpanel to which the plot belongs. The actual year in which the data were collected is recorded in MEASYEAR. The latter is the most appropriate variable for use as FVS Inv_Year.**

TABLE 1: Variables in the predefined FVS_StandInit table structure, their corresponding FIA variables, and descriptions.

FVS variable	FIA Variable	Description and notes
Groups	Many possible	A list of Grouping codes, also separated by spaces, tabs, carriage returns, or newlines. Used by Suppose 2.0 when populating stand lists**. In the Mapmaker release, users were allowed to pick from a limited list of group codes that were populated from FIA variables. It would be possible to expand this list and increase users' flexibility (see Discussion below).**
AddFiles	n/a	A list of Addfile names (.kcp), separated by tabs, carriage returns, or newlines, which will be inserted into simulation file as one or more components. Used by Suppose 2.0 when stands are added to simulations. **Not applicable to FIA data at this time.**
FVSKeywords	n/a	A list of FVS keywords, separated by spaces, tabs, carriage returns, or newlines, which define the FVS run. Used by Suppose 2.0 when stands are added to simulations. **Some keywords may be used to override FVS default values with the values of certain variables that are stored in the FIA database.**
Latitude	LATITUDE	Latitude in degrees of the stands location. **The FIA program is prohibited by law from releasing precise plot coordinates to the public. In most cases, the coordinates in the public database are within one mile of the true location. These are referred to as "fuzzed" coordinates.**
Longitude	LONGITUDE	Longitude in degrees of the stands location. **See Latitude for treatment of FIA coordinates.**
Region	ADFORCD	USDA-FS (National Forest) Region code. **The first two places in the FIA variable ADFORCD specify the National Forest region.**
Forest	ADFORCD	USDA-FS National Forest code. **The last two places in the FIA variable ADFORCD specify the National Forest.**
District	No FIA variable	USDA-FS District code. **Not used by FIA, but mappable from FIA coordinates.**
Compartment	No FIA variable	USDA-FS Compartment code. **Not used by FIA, but mappable from FIA coordinates. Compartment size may limit accuracy when fuzzed coordinates are used.**
Location	See notes	Location Code representing the Region/Forest/District/Compartment codes and corresponds to KODFOR in FVS. When specified, Location takes precedence over Region, Forest, District, and Compartment. **See notes on component variables above.**
Ecoregion	ECOSUBCD	Bailey's Ecoregion code (not yet used by FVS). **FIA records the Bailey's Ecoregion subsection code (ECOSUBCD) according to the map update by Cleland and others (2005). These codes may be truncated to the Ecoregion level, depending on FVS needs at the time of implementation.**
PV_Code or Habitat Type	HABTYPCD1, HABTYPCD2	PV_Code identifies the potential vegetation. It is often the Habitat type or Plant association code. The two names shown are synonymous. **The FIA database includes thousands of Habitat Type and Plant Association codes, but many are represented by few or no plots in the database. These codes used by FIA are under review and may require extensive review before cross-walking.**

TABLE 1: Variables in the predefined FVS_StandInit table structure, their corresponding FIA variables, and descriptions.

FVS variable	FIA Variable	Description and notes
PV_Ref_Code	See notes	Potential vegetation reference code for the PV_Code. **See notes on PV_Code / Habitat Type above.**
Age	STDAGE, BHAGE, TOTAGE	Stand age in years. **FIA uses STDAGE to store stand age, but the methods used to populate this variable vary regionally and for periodic vs annual inventories. Age data for individual trees may be found in the BHAGE or TOTAGE variables in the TREE or SITETREE tables, depending on the species and FIA work unit.**
Aspect	ASPECT	Aspect in degrees. **FIA records ASPECT at the Condition and Subplot level.**
Slope	SLOPE	Slope in percent. **FIA records SLOPE at the Condition and Subplot level.**
Elevation	See ElevFt	Stand elevation represented in 100's of feet for all variants except AK were it is elevation in 10's of feet. (see ElevFt below)
ElevFt	ELEV	Elevation in feet. When specified, ElevFt takes precedence over Elevation. **The FIA variable ELEV is recorded in feet, and corresponds to the FVS ElevFt variable.**
Basal_Area_ Factor	DESIGNCD	Basal area factor corresponding to BAF in FVS. **FIA uses a design code (DESIGNCD) variable, located in the PLOT table, that defines fixed vs variable-radius design, diameter breakpoints, and the number of subplots in the design. For single-condition plots, FVS plot design variables can be populated directly. For multi-condition plots, translation is more complex (see Discussion).**
Inv_Plot_Size	DESIGNCD	The inverse of the fixed plot size in acres. **See notes for Basal_Area_Factor.**
Brk_DBH	DESIGNCD	Breakpoint DBH in inches. **See notes for Basal_Area_Factor.**
Num_Plots	DESIGNCD	Number of plots represented in FVS. **See notes for Basal_Area_Factor.**
NonStk_Plots	STATUSCD (SUBPLOT)	Number of non-stockable plots. **Some FIA subplots may not have tree data associated with them because they were not sampled or because they sampled non-forest conditions. Non-sampled subplots are identified with a status code (STATUSCD) in the SUBPLOT table. Depending on how users desire to treat FIA data in FVS, the FVS NonStk_Plots variable may be used.**
Sam_Wt	EXPNS or derived variables	Sampling Weight used to compute the average yield tables and other weighted averages**. For users who are interested in population estimates, FVS sampling weight could be used to store the appropriate expansion factor for an FIA plot, given the area of interest. These factors may have to be calculated on the fly and populated for each data query.**
Stk_Pcnt	STATUSCD (SUBPLOT)	Stockable percent. **See notes above for NonStk_Plots, and discussion on treatment of conditions.**
DG_Trans	No FIA variable	Diameter growth translation code. Code 0 for increment cores and code 1 for remeasurement data. **DG_Trans can be coded appropriately according to the FIA data source.**
DG_Measure	REMPER	Diameter growth measurement period. **The number of years between remeasurements of FIA plots is recorded in the REMPER variable in the PLOT table.**

TABLE 1: Variables in the predefined FVS_StandInit table structure, their corresponding FIA variables, and descriptions.

FVS variable	FIA Variable	Description and notes
HTG_Trans	No FIA variable	Height growth translation code. Code 0 for height growth and code 1 for remeasurement data. **HTG_Trans can be coded appropriately according to the FIA data source.**
HTG_Measure	REMPER	Height growth measurement period. **See description for DG_Measure.**
Mort_Measure	REMPER	Mortality measurement period. **See description for DG_Measure.**
Max_BA	No FIA variable	Maximum basal area. **Max_BA is used as part of the mortality routine in FVS, so there is no comparable variable in FIA data. FIA uses STOCKING variables in the TREE and SEEDLING tables that may be used to compute stocking on an area basis. Stand density index (SDI) is also used (see below).**
Max_SDI	SDIMAX	Maximum stand density index. **FIA uses maximum SDI values that are consistent with FIA computation methods and FIA forest types. These may or may not be the same maximum values that would be used for growth simulations or silvicultural objectives. The consistent definition and use of maximum SDI is being coordinated between FIA and the FVS staff.**
Site_Species	SISP	Site species code. **Site species codes are compatible with FIA species codes in SISP.**
Site_Index	SI	Site index. **FIA records site index in feet at a specified base age (SIBASE). These base ages are different than the base ages assumed by FVS in some variants. It may be necessary to cross-walk site index values because FVS base ages are fixed.**
Model_Type	n/a	Model type code. Only applies to CR and SE variants.
Physio_Region	n/a	Physiographic region code. Only applies to SE variant.
Forest_Type	FORTYPCD	Forest type code. **FVS forest type codes may or may not be compatible with FIA forest types. These will need to be cross-walked by variant.**
State	STATECD	FIA state code
County	COUNTYCD	FIA county code
Fuel_Model	*	Fire behavior fuel model
Fuel_0_25	*	Initial tons per acre of 0 to 0.25 inch fuel
Fuel_25_1	*	Initial tons per acre of 0.25 to 1 inch fuel
Fuel_0_1	*	Initial tons per acre of 0 to 1 inch fuel, if not using previous two fields
Fuel_1_3	*	Initial tons per acre of 1 to 3 inch fuel
Fuel_3_6	*	Initial tons per acre of 3 to 6 inch fuel
Fuel_6_12	*	Initial tons per acre of 6 to 12 inch fuel
Fuel_gt_12	*	Initial tons per acre of greater than 12 inch fuel.
Fuel_Litter	*	Initial tons per acre of litter
Fuel_Duff	*	Initial tons per acre of duff
Photo_Ref	*	Photo series reference number (1 – 32, see FFE documentation)
Photo_Code	*	Photo code (appropriate character strings depend on the photo series reference number, see FFE documentation)

* Fuel data are available only for Phase 3 FIA plots at this time. Phase 3 data were not included in the Mapmaker release, but it may be possible to include them in future translation programs.

Inclusion of these data will require a cross-walk process, based on the requirements of the FVS Fire and Fuels Extension.

Table 2. Variables in the predefined FVS_TreeInit table structure, their corresponding FIA variables, and descriptions.

FVS Variable	FIA Variable	Description and notes
Stand_CN	PLOT.CN or COND.CN	Same as Stand_CN in FVS_StandInit table. Not read by FVS, but may be used for querying purposes. **FIA uses several control numbers as linkages between tables. Some may be usable in FVS.**
Stand_ID	Possible composite of INVYEAR, STATE, COUNTYCD, and LOC	Same as Stand_ID in FVS_StandInit table. Not read by FVS, but may be used for querying purposes. **The original translation of FIA data to FVS uses a composite stand ID that included the state, county, plot, and inventory year of the FIA plot. This variable was useful to users, because the general location of plots could be quickly identified during simulations.**
Tree_ID	TREE	Unique tree identifier within FVS plot. **Because the FIA subplot is the equivalent of the FVS plot, FIA trees are uniquely identified by tree number within subplots. See Plot_ID note below.**
Plot_ID	SUBPLOT	Plot number in the FVS data. **FVS uniquely identifies plots within stands. Because the area represented by an FIA plot or condition is considered to be a "stand", the FIA subplot is the equivalent of the FVS plot.**
Tree_Count	TPA_UNADJ or 1	Number of trees represented by this data tree. **When plot data are reported on a per-acre basis (i.e., Basal_Area_Factor = -1 and Inv_Plot_Size = 1 in the FVS_StandInit table), Tree_Count is the per-acre expansion factor associated with the tree. When plot design data are supplied, Tree_Count is generally set to 1 (meaning that one tree of this species, diameter, height, etc was tallied on the plot).**
History	STATUSCD	In FVS, History Code 0-5 are live trees, 6 and 7 died during mortality observation, 8 and 9 died before mortality observation period. **The FIA variable STATUSCD distinguishes between live and dead trees, but other variables, such as MORTYR, may be used to assign the appropriate History code.**
Species	SPCD	Tree Species Code, can be the FVS alpha code, FIA numeric code or USDA plant symbol. **The FIA variable SPCD can be used without modification.**
DBH or Diameter	DIA	Diameter in inches. Diameter is an alias for DBH in the FVS tables. For woodland trees, diameter is measured at the root collar (DRC). **The FIA variable DIA can be used without modification.**
DG	No FIA variable	Diameter growth in inches (not tenths of inches). **The FVS DG variable may be calculated using the FIA variables DIA, TREE.PREV_DIA (previous diameter), and PLOT.REMPER (remeasurement period).**
Ht	HT	Height in feet. **In the case of trees with broken or missing tops, the FIA height variable includes the broken or missing portion. See HtTopK / ACTUALHT below.**

Table 2. Variables in the predefined FVS_TreeInit table structure, their corresponding FIA variables, and descriptions.

FVS Variable	FIA Variable	Description and notes
HtG	Float	Height Growth in feet. **FIA does not currently report height growth or provide both current and previous heights as national variables, partly because height measurements were not collected by some FIA work units in the past. However, height should become a national core variable in the future.**
HtTopK	ACTUALHT	Height to top kill is the height to the point of top kill of the tree in feet. **In FIA data, if ACTUALHT = HT, then the tree does not have a broken top. If ACTUALHT < HT, then the tree does have a broken or missing top.**
CrRatio	Float	If the number is 0-9 then it is considered a crown ratio code, according to the FVS documentation. **If the number is 10-99 the value is considered a percent live crown. In the past, the FIA variable for compacted crown ration (CR) has also contained a mixture of the coded and percentage crown ratios. At the time of the Mapmaker release, FVS appeared to not handle this situation correctly.**
Damage1	DAMTYP1	Damage Code, see the FVS documentation for details. **Although there are existing FIA damage variables, damage and severity codes are currently undergoing substantial revision within the FIA program. The new national coding scheme is anticipated to be implemented no sooner than 2011.**
Severity1	DAMSEV1	Severity Code corresponding to damage code 1. **See notes on Damage1**
Damage2	DAMTYP2	Second damage code. **See notes on Damage1**
Severity2	DAMSEV2	Second severity code. **See notes on Damage1**
Damage3	DAMTYP3	Third damage code. **See notes on Damage1**
Severity3	DAMSEV3	Third severity code. **See notes on Damage1**
TreeValue	Float	Tree Value Class Code 1 for desirable, 2 for acceptable, 8 for non-stockable and any other number represents a live cull. **FIA cull and growing stock variables are currently undergoing revision. It should be possible to develop a rule set for converting certain combinations of FIA cull variables into FVS codes.**
Prescription	No FIA variable.	Prescription code. **Prescription codes are used in FVS to describe trees that may be candidates for silvicultural manipulation. The FIA program does not assign similar identifiers to individual trees, so there is no corresponding variable.**
Age	BHAGE or TOTAGE	Age of the tree record. **In FIA data, individual tree ages can be found in the BHAGE or TOTAGE variables in the TREE and SITETREE tables.**
Slope	SLOPE (SUBPLOT)	Slope percentage on the plot where the tree was located. **The slope measured on the FIA subplot is comparable to plot-level slope in FVS.**

Table 2. Variables in the predefined FVS_TreeInit table structure, their corresponding FIA variables, and descriptions.

FVS Variable	FIA Variable	Description and notes
Aspect	ASPECT (SUBPLOT)	Aspect in degrees on the plot where the tree was located. **The aspect measured on the FIA subplot is comparable to plot-level aspect in FVS.**
PV_Code or Habitat Type	See PV_Code entry in Stand_Init table.	The potential vegetation code on the plot where the tree was located. **FIA habitat type and potential vegetation codes are recorded in the condition table, so they are more appropriately used in the Stand_Init table.**
PV_Ref_Code	See PV_Ref_Code in the Stand_Init table.	Potential vegetation reference code for the PV_Code
TopoCode	TOPO_POSITION_PNW	Topography Code 1=bottom, 2=lower, 3=mid slope, 4=upper slope, and 5=ridge top, on the plot where the tree was located. **Only the Pacific Northwest (PNW) FIA program uses a topographic position code that it comparable to the FVS TopoCode.**
SitePrep	TRTCD1, -2, and -3	Site Preparation code 1=none, 2=mechanical, 3=burn, and 4=road cuts/road fills/stockable road beds, on the plot where the tree was located

Tables 1 and 2 provide a general framework for translation of FIA variables into FVS-ready format. In some cases, there are several options for translation that may be user-defined or may be pre-set for ease of use, depending on users needs. However, some of these options are not yet common knowledge to users, so they will be introduced briefly here.

Choice of Plot and Tree Data Formats

FIA mapped (annual) design plots have four 1/24-acre subplots design for a total surface area of approximately 1/6 acre, resulting in an expansion factor of 6.02 trees per acres for tally trees $\geq$ 5.0 inches in diameter. Seedlings and saplings are measured on 1/300-acre microplots that are nested within each subplot, resulting in a per-acre expansion factor of 74.97. There are two ways to code this information in FVS:

A) Basal_Area_Factor = -24, Inv_Plot_Size = 300, Brk_DBH = 5.0, and Tree_Count = 1

B) Basal_Area_Factor = -1, Inv_Plot_Size = 1, Brk_DBH = null, and Tree_Count = 6.02 (trees $\geq$5.0 inches) or Tree_Count = 74.97 (trees <5.0 inches)

In example (A), the FVS-formatted data utilize the FIA plot design information and expansion factors are computed by FVS. The negative sign in front of the Basal_Area_Factor value indicates that the value is for a fixed-plot area, and not a basal area factor. In example (B), all trees are represented by their per-acre expansion factors and the FIA plot design specifics are ignored.

In terms of the computation of stand structure, volume, and other characteristics in FVS, both approaches produce identical results. However, in terms of growth, mortality, and other submodel functions, there may be differences between the two. The reason for this is that FVS uses both stand-level and plot-level variables in some submodels. In other words, a 10-inch tree that was recorded on a dense subplot may be grown slower than a 10-inch tree recorded on a sparse subplot. Although the magnitude of difference is unknown and will vary on every plot, preserving the information from intra-plot (subplot) variability may be desirable to some users. In addition, FVS preserves the subplot-level tally and passes that information to the Stand Visualization System (SVS; McGaughey 1997), allowing users to compare compositional and structural differences among subplots.

It is possible to make data available in either format (A) or format (B), depending on user needs. On single-condition, mapped-design plots, the preference for one format or the other makes little difference, except for the possible growth and mortality differences mentioned above. However, there are certain circumstances under which one format or the other may be preferable, or even necessary. The most important of these involves the treatment of multiple conditions that might occur on a single FIA plot.

Treatment of Multiple Conditions

Briefly described, conditions are delineated on FIA plots when part of the plot is occupied by forest and another part is occupied by nonforest land, water, or some situation exists that precludes sampling part of the plot (figure 3). Forested conditions are further divisible by reserved status, owner group, forest type, stand size class, regeneration status, or tree density. In some cases, these divisions equate to what would typically be called separate stands, but in other cases they are not.

Depending on user needs, the division of a plot by condition may or may not be important. If the objective of a simulation is to evaluate silvicultural options on a landscape, it may be important to exclude reserved lands from treatment options. The consideration of multiple-condition plots as a single condition that includes mixtures of forest and nonforest or different types of forest may misrepresent composition, structure, or growth potential at the plot level.

Other Considerations

It is not possible to anticipate all user needs, or even to consider all of the possibilities, given the complexity of the FIA database, the FVS base system, and

all of the FVS extensions. Undoubtedly, users who are intimately familiar with one of more of the programs will be able to identify issues not covered here.

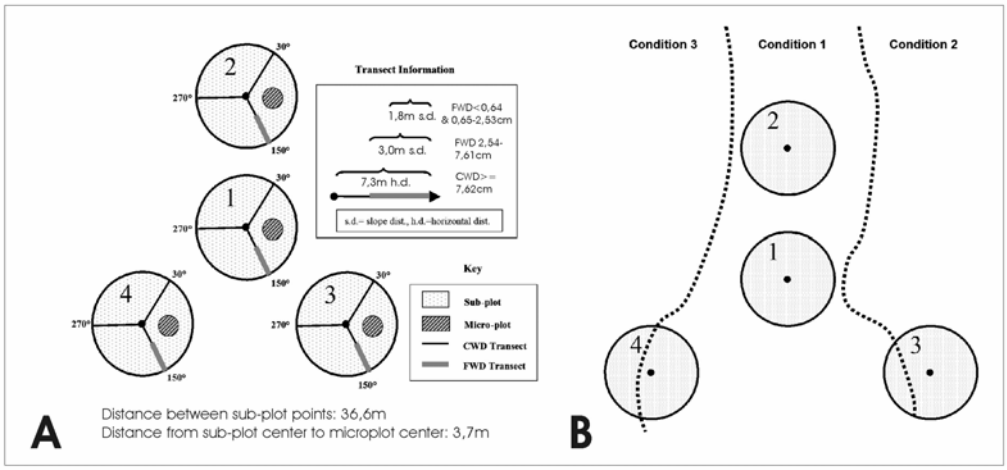


FIGURE 3: The FIA annual plot design (A) and example of mapping multiple conditions on the plot footprint (B).

Conclusion

In summary, the work that needs to be done on the next generation of FIA to FVS data translation can be described in three groups: opportunities, data needs, and open questions. Issues belonging each of these groups have been described, to some degree, in this paper, but the list is not exhaustive. The opportunities may be taken advantage of in one step, or may be implemented gradually. The data needs are relatively easily identified, and can be satisfied through further investigation. It is possible for FIA developers to begin to address these issues right away. The open questions will require solicitation of user input and priority ranking, because some answers may have important implications for data delivery and use. Once the open questions have been answered and addressed, the result should be a greatly enhanced outlet for FIA data.

Opportunities

- Export FIA data to FVS input files as a database
- Utilize the existing FVS map to assign default geographic variant

Data Needs

- Site index species and site index values must conform with FVS variant
- Stockable area should be computed consistently
- Damage and severity codes should be translated to FVS specifications where possible
- FIA habitat and community types need to be cross-walked to FVS types
- Diameter and height growth data should be provided when available

- Work toward allowing NFS volume equation computation for tree records
- Include tree defect data where available
- Include stem count for woodland species
- Include access to older periodic data
- Export seedling data when available
- Export dwarf mistletoe data when available
- Include fuel loading data from Phase 3 plots

Open Questions

- How to deal with multiple conditions?
- When to incorporate plot design codes?

Miles (2008) noted that "FVS users of FIA data make up a small but extremely important part of the overall FIA user community. It is important to maintain the delivery of FIA data to this user group." This paper is a first step toward restoring access to the most current FIA data for this user group. As this effort proceeds, input from all interested parties is welcome.

Acknowledgements

The author thanks D. Vandendriesche of the Forest Vegetation Simulator Staff for ideas and comments on this topic, and R.J. DeRose and S. Woudenberg for their reviews of this manuscript.

References Cited

Cleland, D.T.; Freeouf, J.A.; Keys, J.E., Jr.; Nowacki, G.J.; Carpenter, C.A.; McNab, W.H. 2004. Subregions of the conterminous United States. A.M. Sloan, tech. ed. Washington, DC: U.S. Department of Agriculture, Forest Service, presentation scale 1:3,500,000; colored. Also available on CD-ROM consisting of GIS coverage in ArcINFO format.

Crookston, N.L. 1997. Suppose: An Interface to the Forest Vegetation Simulator. In: Teck, Richard; Moeur, Melinda; Adams, Judy. 1997. Proceeding: Forest vegetation simulator conference. 3-7 February 1997, Fort Collins, CO. Gen. Tech. Rep. INT-GTR-373. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Dixon, G.E., comp. 2002. Essential FVS: A user's guide to the Forest Vegetation Simulator. Internal Rep. Fort Collins, CO: U. S. Department of Agriculture, Forest Service, Forest Management Service Center. 219p. (Last Revised: April 7, 2008)

Donnelly, D.; Lilly, B.; Smith, E. 2001. The southern variant of the Forest Vegetation Simulator. [available online: http://www.fs.fed.us/fmsc/ftp/fvs/docs/overviews/snvar.pdf]

Forest Inventory and Analysis Program. 2008. The forest inventory and analysis database: database description and users manual version 3.0 for Phase 2, revision 1. U.S. Department of Agriculture, Forest Service, Washington Office. [available

online: http://www.fia.fs.fed.us/library/database-documentation/FIADB_user_manual_v3-0.pdf]

Forest Vegetation Simulator Staff. 2003. Users Guide to the Database Extension of the Forest Vegetation Simulator Version 2.0. Fort Collins, CO: U. S. Department of Agriculture, Forest Service, Forest Management Service Center. 219p. (Last Revised: May 2008)

Johnson, R.R. 1997. A historical perspective of the Forest Vegetation Simulator. In: Teck, Richard, Moeur, Melinda; Adams, Judy, comps. Proceedings: Forest Vegetation Simulator Conference, February 3–7, 1997; Fort Collins, CO. Gen. Tech. Rep. INT–GTR–373. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

McGaughey, R.J. 1997. Visualizing forest stand dynamics using the stand visualization system. In: Proceedings of the 1997 ACSM/ASPRS Annual Convention and Exposition; April 7-10, 1997; Seattle, WA. Bethesda, MD: American Society of Photogrammetry and Remote Sensing. 4:248-257.

Miles, P.D. 2008. Forest inventory and analysis data for FVS modelers. Pp. 125-129 in: Havis, Robert N.; Crookston, Nicholas L., comps. 2008. Third Forest Vegetation Simulator Conference; 2007 February 13–15; Fort Collins, CO. Proceedings RMRS-P-54. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 125-129.

Reinhardt, E.; Crookston, N.L., 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.

Shaw, J.D.; Long, J.N. 2002. FVS Lessons on the web: A resource for users and instructors. Pp. 2-5 in Crookston, Nicholas L.; Havis, Robert N., comps. 2002. Second Forest Vegetation Simulator Conference; 2002 February 12–14; Fort Collins, CO. Proc. RMRS-P-25. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Stage, A.R. 1973. Prognosis model for stand development. USDA Forest Service, Intermountain Forest & Range Experiment Station. General Technical Report INT-137. Ogden, UT.

Wykoff, W. R.; Crookston, N.L.; Stage, A.R. 1982. User's guide to the Stand Prognosis Model. Gen. Tech. Rep. INT-133. Ogden, UT: U. S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 112 p.

EVALIDatorReports: Reporting Beyond the FIADB

Patrick D. Miles[1]

Abstract: *Tools for analyzing data collected by the U.S. Forest Service's Forest Inventory and Analysis (FIA) program are available in Microsoft Access© format. Databases have been created for every state, except Hawaii, and are available for downloading. EVALIDatorReports is a Visual Basic Application that is stored within each Microsoft Access© database containing FIA data for each state. EVALIDatorReports can generate reports for over 40 different types of estimates including forest area, number of trees, growing-stock volume, growth, removals, and mortality. The code for this Visual Basic Application can be viewed, modified, and extended. This approach provides advantages for users interested in creating customized data summaries utilizing external data and/or programs. The system does not require Internet access after the databases are downloaded to a personal computer. This paper contains one simple and three complex examples illustrating how items can be added to the list of classification variables or to the list of estimates within the EVALIDatorReports program.*

KEYWORDS: Forest inventory, database, reporting tool, forest statistics

Background

The needs of most customers of the U.S. Forest Service's Forest Inventory and Analysis (FIA) program can be met by Web applications such as the Forest Inventory Mapmaker program (Miles 2002) and more recently Forest Inventory Data Online (FIDO) (Wilson and Ibes 2008). The reporting capabilities of these tools are continually being improved but they will never be able to meet all the needs of power users. Power users – biometricians, statisticians, and analysts - will always require access to the underlying data.

FIA began providing FIA downloadable datasets in 1996. Initially, datasets for the eastern states were available in Eastwide database format (Hansen et al. 1992) while datasets for the western states were available in Westwide database format (Woudenberg and Farrenkopf 1995). In 2001 a standardized national format for all datasets became available and is known as the FIADB (Forest Inventory and Analysis Program 2008).

---

[1]Research Forester, U.S. Department of Agriculture, Forest Service, Northern Research Station, 1992 Folwell Ave., St. Paul, MN 55108; fax: 651-649-5140, phone: 651-649-5146, e-mail: pmiles@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Comma separated values (CSV) has been the standard format for dataset delivery. Users wrote scripts to read data from these CSV files into their programs or databases. Mistakes were frequent, such as loading a real number as an integer thereby inadvertently truncating or rounding the real number.

Beginning in February of 2008, FIA datasets became available in Microsoft Access© 2003 format in addition to CSV format. Microsoft Access© is available on most Forest Service computers and is used widely in the academic and research communities. Many databases and software programs can read and correctly interpret data from Microsoft Access© databases, thereby eliminating the formatting problem so often encountered with CSV files.

Microsoft Access© databases containing FIA data and reporting tools have been created for every state except Hawaii and are available for download from http://www.fia.fs.fed.us/tools-data/datamart.html. Each of these state databases also includes a form (i.e., a database tool) that allows users to import CSV data for additional states and create a multi-state database. Unfortunately, due to the two-gigabyte size limit for Microsoft Access© files, multi-state databases are typically limited to two or three states.

Data delivery in Microsoft Access© format creates an opportunity to ensure that results provided by the Forest Inventory Mapmaker and FIDO programs can be duplicated. To that end, a number of Structured Query Language (SQL) scripts were written and stored in the Microsoft Access© databases. These SQL scripts are useful to Microsoft Access© database users and help to verify that both the program used to create the CSV files and the import specifications used to load the data from the CSV files into the Microsoft Access© databases were working correctly.

While the delivery of FIA data in Microsoft Access© format was initially intended solely as a method of data transfer, it soon became apparent that there are additional opportunities to meet customer needs that could not be met by existing Web applications. The FIA program can provide personal open-ended databases with a reporting tool based on peer reviewed estimation protocols (Bechtold and Patterson 2005).

Advantages to having FIA data in personal database

There are several advantages to having FIA data in Microsoft Access©:

1) Archiving – data does not change
2) Availability – data is always available
3) Transparency – data and programs in one place
4) Scalability and flexibility – opportunity to add data and programs to the database

Archiving – A common complaint from users is that the numbers coming out of the Forest Inventory Mapmaker or FIDO have changed since the last time they ran a report. Mapmaker and FIDO use data stored in FIA's corporate Oracle database. The corporate database is not static. Corrections to field data or changes to compilation procedures

result in different answers from the reporting tools. A researcher with a personal copy of the database could choose when or if to update the information.

Availability – Network outages, denial of service attacks, funding cuts, and other unforeseen circumstances can limit access to FIA Web applications. Researchers can download and retain a static copy of the database to use in their research and publications.

Transparency – The code for the reporting tool (EVALIDatorReports) is stored inside the database and can be viewed and modified by the user. The code, data, and reporting tool are all located in a single Microsoft Access© file that can easily be copied and transferred.

Scalability – Users can add to the list of classification variables and the list of estimates. Three examples of extending the reporting capabilities of the database and reporting tools are provided here.

The EVALIDatorReports Reporting Tool

EVALIDatorReports has inputs and outputs that are essentially equivalent to those of the Forest Inventory Mapmaker and FIDO.

The user inputs the following information:

1) Geographic area of interest
2) Optional filters (for restricting the query to a specific ownership, species, etc.)
3) Attribute of interest (timberland area, number of trees, volume, etc.)
4) Classification variables to be used for page, column, and row headings

EVALIDatorReports outputs:

1) Table of population estimates
2) Corresponding table of sampling errors

EVALIDatorReports differs from the Forest Inventory Mapmaker and FIDO in that the user can add to the list of classification variables and to the list of attributes of interest.

Simple EVALIDatorReports Examples

To start EVALIDatorReports, open the Microsoft Access© database and click on the "Forms" object and then double-click on the "EVALIDatorReports" name (Fig. 1).

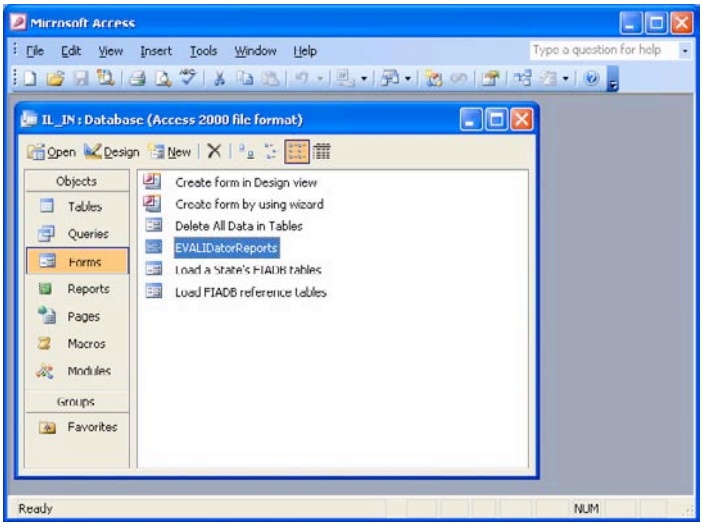


Figure 1: Starting EVALIDatorReports.

The user must first specify the variable to be estimated (Fig.2). In this example the estimate selected is “3 Area of timberland (acres)”. Once the user specifies this variable, a list of inventories available in the database for this estimate appears in the EvalID listbox. In this example, inventories of Illinois and Indiana for the year 2006 were selected from the available list. The retrieval can be further limited to national forest lands by placing the filtering clause “and cond.owncd=11” in the filtering textbox. The user must also specify the three classification variables for page, row, and column. Here a single page report is specified by picking “None” for the page classification variable. This single page report will have “Basal area live tree” classes for the row headings and “Stand-size” classes for the column headings.

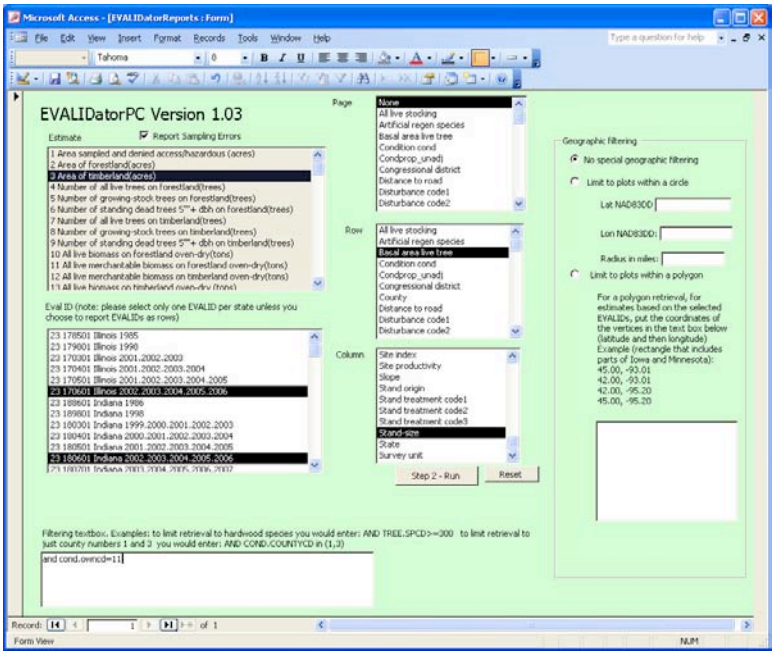


Figure 2: EVALIDatorReports input screen.

The resulting tabular reports (Tables 1 and 2) show that in 2006 there were 449,928 acres of national forest timberland in Indiana and Illinois plus or minus 3.85 percent (one standard deviation).

Table 1. National Forest timberland acres by basal area and stand-size, Illinois and Indiana, 2006.

	Stand-size			
Basal area	**Total**	**Large diameter**	**Medium diameter**	**Small diameter**
Total	449,928	368,894	64,532	16,502
0-40 sqft/ac	12,156	156	5,604	6,395
41-80 sqft/ac	74,137	44,740	19,290	10,107
81-120	211,918	181,456	30,462	-

sqft/ac				
120+ sqft/ac	151,718	142,542	9,176	-

Table 2: Sampling error percent for estimates in Table 1.

	Stand-size			
Basal area	Total	Large diameter	Medium diameter	Small diameter
Total	3.85	5.17	19.84	36.82
0-40 sqft/ac	42.78	101.38	67.74	56.24
41-80 sqft/ac	17.63	24.22	33.84	49.66
81-120 sqft/ac	9.62	10.52	32.95	-
120+ sqft/ac	10.78	11.2	49.43	-

Example 1: Adding a Runtime Classification Variable: Classification variables are stored in the table REF_PRC. To open the REF_PRC table in Microsoft Access© (Fig. 3) click on the "Tables" object and then double-click on the "REF_PRC" name.

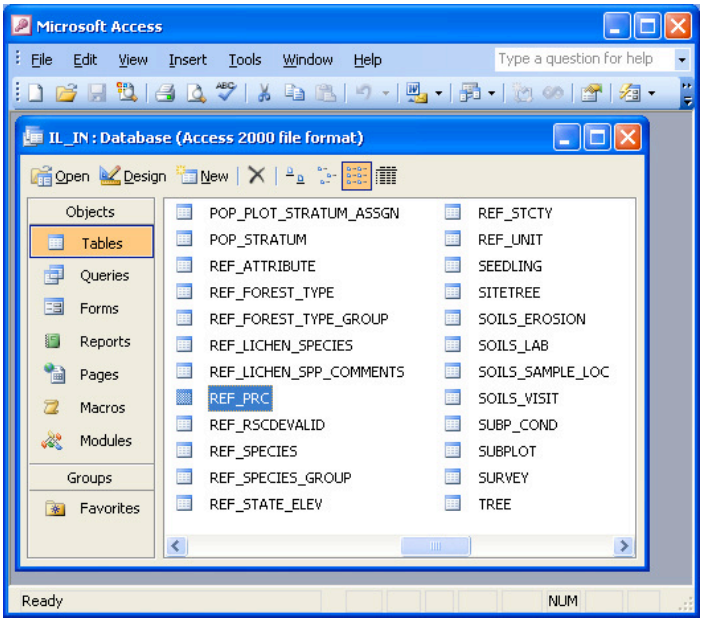


Figure 3: Opening the REF_PRC table in Microsoft Access©.

The values in the Page, Row, and Column lists in the EVALIDatorReports program are drawn from the column CLASSNM in the REF_PRC table (Table 3). In the simple example "Basal area live tree" and "Stand-size" were selected as the classification variables. In the Microsoft Access© table REF_PRC these are just 2 of the 54 classification variables available to the EVALIDatorReports program. A new item can be added to the classification variable list in EVALIDatorReports by adding a new record to the REF_PRC table.

Table 3: A subset of records in REF_PRC table.

CLASS NBR	COND TREE SEED	CLASSNM	FUNCTIONNM	PAGE CLASS	ROW CLASS	COL CLASS
3	COND	Stand-size	stdszcdLabel(cond.stdszcd)	Y	Y	Y
11	COND	Basal area live tree	baliveLabel(cond.balive)	Y	Y	Y

Basal area is commonly used for describing stands in the east. In the west, forest managers often prefer to use stand-density index (SDI). SDI (Long and Daniel 1990) is not currently available in EVALIDatorReports. But it can be added in two simple steps:

- Step 1 - Add the record in Table 4 to the REF_PRC table.
- Step 2 - Add function sdiLabel (appendix A) to the "Functions module" in the Microsoft Access© database (Fig. 4). To open the Functions module click on the "Modules" object and then double-click on the name "Functions module". Then copy the sdiLabel function code from Appendix A into the module. This sdiLabel function generates the classification variable at runtime. For each CONDITION record, all of the trees are selected and an SDI value is calculated for the condition. The resulting SDI is then assigned an SDI class.

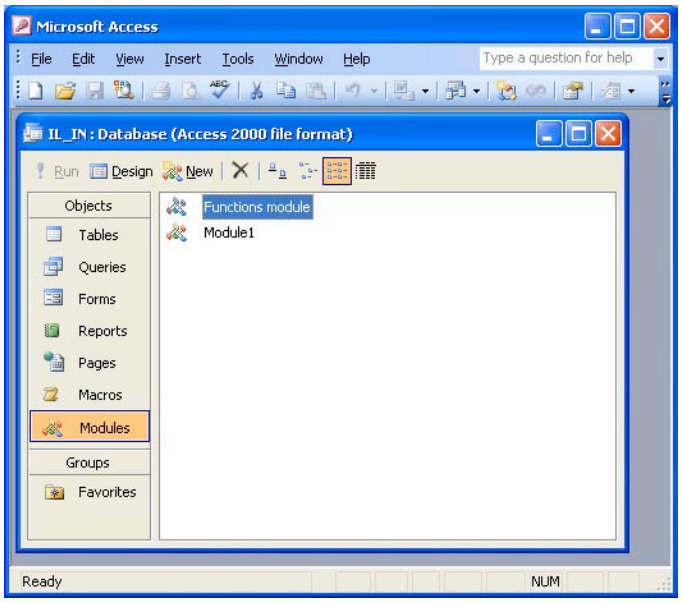


Figure 4: Opening the Functions module in Microsoft Access©.

Table 4: SDI classification variable record to be added to REF_PRC table

CLASS NBR	COND TREE SEED	CLASSNM	FUNCTIONNM	PAGE CLASS	ROW CLASS	COL CLASS
55	COND	SDI	sdiLabel(cond.plt_cn, cond.condid, cond.condprop_unadj)	Y	Y	Y

Now when the simple example is rerun using SDI as the row classification variable instead of "Basal area all live", the output tables provide a slightly different perspective to the land manager (Tables 5 and 6).

Table 5. National forest timberland acres by basal area and stand-size, Illinois and Indiana, 2006.

	Stand-size			
SDI	**Total**	**Large diameter**	**Medium diameter**	**Small diameter**
Total	449,928	368,894	64,532	16,502
less than 100 SDI	15,300	5,771	5,604	3,925
100 to 199 SDI	203,310	153,153	40,487	9,671
200 to 299 SDI	196,777	181,205	12,665	2,907
300 to 399 SDI	28,891	23,115	5,776	-
400 to 499 SDI	5,032	5,032	-	-
500 to 599 SDI	618	618	-	-

Table 6: Sampling error percent for estimates in Table 5.

	Stand-size

SDI	Total	Large diameter	Medium diameter	Small diameter
SDI	3.85	5.17	19.84	36.82
less than 100 SDI	38.78	65.78	67.74	65.83
100 to 199 SDI	9.65	11.6	26.72	51.55
200 to 299 SDI	9.18	9.86	41.26	88.47
300 to 399 SDI	29.2	32.97	66.97	-
400 to 499 SDI	55.86	55.86	-	-
500 to 599 SDI	101.38	101.38	-	-

Example 2: Adding a User-Provided Classification Variable to the FIADB:The state of Minnesota often reports information by Department of Natural Resources (DNR) administrative boundaries. Within the Microsoft Access© database for Minnesota (MN.mdb) a new column can be permanently added to the PLOT table. In this case the new column is called AREA_NAME. By overlaying the plot coordinates (LAT and LON) on an administrative boundary shapefile (Fig. 5) provided by the Minnesota DNR, each plot can be assigned to an administrative unit. The resulting administrative boundary codes are then assigned to the AREA_NAME column of the PLOT table.

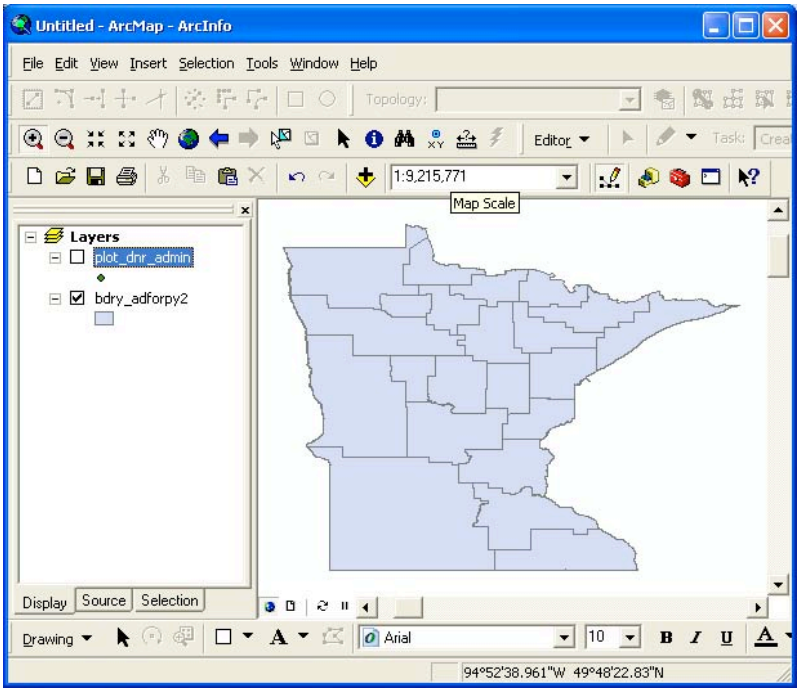


Figure 5: Minnesota DNR Administrative boundaries as viewed in ArcGIS.

Two additional steps are required before the EVALIDatorReports program can report information by administrative boundary. First a record must be added to the REF_PRC table. In this example the record in Table 7 is added to the REF_PRC table.

Table 7: AREA_NAME classification variable record to be added to REF_PRC table.

CLASS NBR	COND TREE SEED	CLASSNM	FUNCTIONNM	PAGE CLASS	ROW CLASS	COL CLASS
56	COND	AREA_NAME	area_nameLabel(plot.area_name)	Y	Y	Y

Then the function area_nameLabel (Appendix b) must be added to the "Functions module". Note that in this case the function simply assigns the area name class. A portion of the resulting output from this retrieval is displayed in Table 8.

Table 8: Area of timberland by DNR administrative area and stand-size, Minnesota, 2006.

	Stand-size	Large diameter	Medium diameter	Small diameter	Nonstocked
Total	15,112,725	4,115,614	5,564,086	5,193,133	239,892
Aitkin Area	757,616	214,109	303,031	229,298	11,178
Backus Area	865,762	284,596	333,551	237,369	10,246
Baudette Area	548,271	66,092	220,576	252,656	8,947
…	….	…	…	…	…

Example 3: Adding a Runtime Estimate: There is a difference between adding a classification variable and adding an estimate. In both cases a function must be added to the "Functions module" However, when adding a classification variable a record must be inserted into the REF_PRC table, while when adding an estimate a record must be inserted into the REF_ATTRIBUTE_ACCESS table.

The REF_ATTRIBUTE_ACCESS table has four columns. A single record of this table is presented in TABLE 9. The first column in the table is ATTRIBUTE_NBR. Currently there are 46 rows in the REF_ATTRIBUTE_ACCESS table so the rows are numbered 1 through 46. The second column is ATTRIBUTE_DESCR. The ATTRIBUTE_DESCR for the row depicted in Table 9 is "All live biomass on forestland oven-dry(tons)". The third column, VBA_SUMFROMWHERE, contains a segment of SQL code used by the EVALIDatorReports program to query the database. The fourth column, PEA_SURROGATE, is blank for all 46 rows that have been predefined. When adding a new row to the REF_ATTRIBUTE_ACCESS table the PEA_SURROGATE column should be filled with the ATTRIBUTE_NBR for the row that most closely reflects what is being estimated. The program will then use the surrogate's PEA records to identify the inventories that can be queried for the new estimate.

Table 9: A single record from the REF_ATTRIBUTE_ACCESS table.

ATTRIBUTE_NBR	ATTRIBUTE_DESCR	VBA_SUMFROMWHERE	PEA_SURROGATE
10	All live biomass on forestland oven-dry(tons)	SUM(tree.TPA_UNADJ*tree.drybiot* IIf(IsNull(tree.dia),PPP.adj_factor_subp, IIf(tree.dia<5,PPP.adj_factor_micr, IIf(IsNull(MACRO_BREAKPOINT_DIA), PPP.adj_factor_subp, IIf(dia<MACRO_BREAKPOINT_DIA, PPP.adj_factor_subp,adj_factor_macr)))))/2000 AS ESTIMATED_VALUE FROM TREE INNER JOIN (COND INNER JOIN ((POP_PLOT_STRATUM_ASSGN	

```
 INNER JOIN
 POP_STRATUM as PPP ON
POP_PLOT_STRATUM_ASSGN.STRATUM_CN = PPP.CN)
 INNER JOIN
 PLOT ON POP_PLOT_STRATUM_ASSGN.PLT_CN =
PLOT.CN) ON COND.PLT_CN = PLOT.CN)
ON TREE.PLT_CN = PLOT.CN
 WHERE ((TREE.STATUSCD)=1) AND
((COND.COND_STATUS_CD)=1) and
TREE.CONDID=COND.CONDID AND
 tree.TPA_UNADJ is not null
and tree.drybiot is not null and
```

The EVALIDatorReports program only allows the user to query those state inventories for a particular estimate where the FIA program says it is appropriate. It is not appropriate, for example, to estimate growth on forest land for inventories conducted prior to 1998 because trees were not always measured on unproductive and reserved forest land. For inventories conducted prior to 1998, it may only be appropriate to estimate growth on timberland. The POP_EVAL_ATTRIBUTE (PEA) table identifies which estimates can be computed for each state inventory. When new estimates are added either new rows must be added to the PEA table or an ATTRIBUTE_NBR for an estimate that has entries in the PEA table must be entered in the PEA_SURROGATE column of the REF_ATTRIBUTE_ACCESS table.

In this example, a user wants to compute all live aboveground tree biomass using Jenkins biomass equations rather than using the stored biomass values that were developed from FIA regional biomass equations. To use the stored biomass values, the researcher would select “10 All live biomass on forest land oven-dry(tons)” from the estimate list in EVALIDatorReports. The program would then use the tree biomass number stored in the DRYBIOT column of the TREE table.

To replace DRYBIOT with a value calculated using the Jenkins biomass equations the user would add a new row to the REF_ATTRIBUTE_ACCESS table (Table 10). The value of ATTRIBUTE_NBR would be 47. The value in ATTRIBUTE_DESCR would be “Jenkins All live biomass on forestland oven-dry(tons)”. The value in VBA_SUMFROMWHERE would be identical to the value from Table 9 with the exception that “tree.drybiot**”** would be replaced with the function call

"JenkinsBiomass(tree.spcd, tree.dia, tree.diahtcd, 0)". The value of PEA_SURROGATE would be 10.

Table 10: Jenkins biomass estimate in the REF_ATTRIBUTE_ACCESS table.

ATTRIBUTE_NBR	ATTRIBUTE_DESCR	VBA_SUMFROMWHERE	PEA_SURROGATE
47	Jenkins All live biomass on forestland oven-dry(tons)	SUM(tree.TPA_UNADJ*JenkinsBiomass(tree.spcd, tree.dia, tree.diahtcd, 0)* IIf(IsNull(tree.dia),PPP.adj_factor_subp, IIf(tree.dia<5,PPP.adj_factor_micr, IIf(IsNull(MACRO_BREAKPOINT_DIA), PPP.adj_factor_subp, IIf(dia<MACRO_BREAKPOINT_DIA, PPP.adj_factor_subp,adj_factor_macr)))))/2000 AS ESTIMATED_VALUE FROM TREE INNER JOIN (COND INNER JOIN ((POP_PLOT_STRATUM_ASSGN INNER JOIN POP_STRATUM as PPP ON POP_PLOT_STRATUM_ASSGN.STRATUM_CN = PPP.CN) INNER JOIN PLOT ON POP_PLOT_STRATUM_ASSGN.PLT_CN = PLOT.CN) ON COND.PLT_CN = PLOT.CN) ON TREE.PLT_CN = PLOT.CN WHERE ((TREE.STATUSCD)=1) AND ((COND.COND_STATUS_CD)=1) and	10

TREE.CONDID=COND.CONDID AND

tree.TPA_UNADJ is not null

and tree.drybiot is not null and

The researcher could then run EVALIDatorReports to generate biomass estimates based on either Jenkins (Table 11) or based on the regional biomass numbers (Table 12).

Table 11: Jenkins all-live biomass (oven-dry tons) on timberland Illinois and Indiana, 2006.

	Species group major		
State	**Total**	**Softwoods**	**Hardwoods**
Total	591,296,419	13,460,746	577,835,673
IL	292,274,945	5,790,358	286,484,587
IN	299,021,473	7,670,388	291,351,086

Table 12: Regional all-live biomass (oven-dry tons) on timberland Illinois and Indiana, 2006.

	Species group major		
State	**Total**	**Softwoods**	**Hardwoods**
Total	486,143,719	11,234,641	474,909,078
IL	234,606,405	4,812,103	229,794,301
IN	251,537,314	6,422,538	245,114,776

Future Developments

New permanent meaningful partnering opportunities may arise from providing FIA data and tools in a completely transparent and open-ended system. Microsoft Access© databases containing FIADB data and the EVALIDatorReports program were downloaded 543 times from the FIA website in the first 5 months. Interest has already been expressed by this small user community in developing a mechanism for sharing code for new classification and attribute variables. Usually future development is driven

by customer demand. In the case of the EVALIDatorReports program, future development may be driven by customer supply.

References

Bechtold, W.A.; Patterson, P.L. 2005. The enhanced forest inventory and analysis program: national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Forest Inventory and Analysis Program. 2008. The forest inventory and analysis database: database description and users manual version 3.0 for Phase 2, revision 1. U.S. Department of Agriculture, Forest Service, Washington Office. http://www.fia.fs.fed.us/library/database-documentation/FIADB_user_manual_v3-0.pdf.

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

Jenkins, J.C.; Chojnacky, D.C.; Heath, L.S.; Birdsey, R.A. 2003. National scale biomass estimators for United States tree species. Forest Science. 49(1): 12-35.

Long, J.N.; Daniel, T.W. 1990. Assessment of growing stock in uneven-aged stands.Western Journal of Applied Forestry. 5:93-96.

Miles, P.D. 2002. Forest inventory mapmaker web-application version 3.0. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. www.ncrs2.fs.fed.us/4801/fiadb/index.htm [Internet only]

Wilson, B.T;. Ibes, W.S. 2008. Forest inventory and data online (FIDO) web application, version 1.0.0r4. St. Paul, MN: U.S. Department of Agriculture, Forest Service, Northern Research Station. http://fiatools.fs.fed.us.

Woudenberg, S.W.; Farrenkopf, T.O. 1995. The Westwide forest inventory data base: user's manual. Gen. Tech. Rep. INT-GTR-317. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 67 p.

Appendix A – sdiLabel Function

```
Public Function sdiLabel(cn, condid, condprop_unadj)
Dim sSQL As String
Dim sdi As Double
   sSQL = "SELECT sum(nz(tpa_unadj,0)/ " & Str(condprop_unadj) &
"*(nz(dia,0)/10)^1.6)"
   sSQL = sSQL + " FROM tree where plt_cn=""" & cn & """ and condid=" &
Str(condid)

   Dim cnn As ADODB.Connection
   Set cnn = CurrentProject.Connection
   Dim rs As New ADODB.Recordset
   rs.Open sSQL, CurrentProject.Connection
      Do Until rs.EOF
         If IsNull(rs.Fields(0)) Then
            sdi = 0
         Else
            sdi = rs.Fields(0)
         End If
         rs.MoveNext
      Loop
   rs.Close
   cnn.Close
   Debug.Print sdi
   If (sdi < 100) Then
   sdiLabel = "0001 less than 100 SDI"
   ElseIf (sdi < 200) Then
   sdiLabel = "0002 100 to 199 SDI"
   ElseIf (sdi < 300) Then
   sdiLabel = "0003 200 to 299 SDI"
   ElseIf (sdi < 400) Then
   sdiLabel = "0004 300 to 399 SDI"
   ElseIf (sdi < 500) Then
   sdiLabel = "0005 400 to 499 SDI"
   ElseIf (sdi < 600) Then
   sdiLabel = "0006 500 to 599 SDI"
   ElseIf (sdi < 700) Then
   sdiLabel = "0007 600 to 699 SDI"
   ElseIf (sdi < 800) Then
   sdiLabel = "0008 700 to 799 SDI"
   ElseIf (sdi < 900) Then
   sdiLabel = "0009 800 to 899 SDI"
   ElseIf (sdi < 1000) Then
   sdiLabel = "0010 900 to 999 SDI"
   ElseIf (sdi < 1100) Then
   sdiclass = "0011 1000 to 1099 SDI"
```

```
    Else
    sdiLabel = "0012 1100+ SDI"
    End If
End Function
```

Appendix B – Area_nameLabel Function

```
Public Function area_nameLabel(area_name)
  Select Case area_name
Case "Aitkin Area"
area_nameLabel = "0001 Aitkin Area"
Case "Backus Area"
area_nameLabel = "0002 Backus Area"
Case "Baudette Area"
area_nameLabel = "0003 Baudette Area"
Case "Bemidji Area"
area_nameLabel = "0004 Bemidji Area"
Case "Blackduck Area"
area_nameLabel = "0005 Blackduck Area"
Case "Cambridge Area"
area_nameLabel = "0006 Cambridge Area"
Case "Cloquet Area"
area_nameLabel = "0007 Cloquet Area"
Case "Deer River Area"
area_nameLabel = "0008 Deer River Area"
Case "Detroit Lakes Area"
area_nameLabel = "0009 Detroit Lakes Area"
Case "Hibbing Area"
area_nameLabel = "0010 Hibbing Area"
Case "Lake City Area"
area_nameLabel = "0011 Lake City Area"
Case "Little Falls Area"
area_nameLabel = "0012 Little Falls Area"
Case "Littlefork Area"
area_nameLabel = "0013 Littlefork Area"
Case "New Ulm Area"
area_nameLabel = "0014 New Ulm Area"
Case "Orr Area"
area_nameLabel = "0015 Orr Area"
Case "Park Rapids Area"
area_nameLabel = "0016 Park Rapids Area"
Case "Red Lake Reservation"
area_nameLabel = "0017 Red Lake Reservation"
Case "Rochester Area"
area_nameLabel = "0018 Rochester Area"
Case "Sandstone Area"
area_nameLabel = "0019 Sandstone Area"
Case "Tower Area"
area_nameLabel = "0020 Tower Area"
Case "Two Harbors Area"
area_nameLabel = "0021 Two Harbors Area"
Case "Warroad Area"
```

```
area_nameLabel = "0022 Warroad Area"
Case Else
 area_nameLabel = "0023 Other "
End Select
End Function
```

Appendix C – JenkinsBiomass function

```
 Public Function JenkinsBiomass(p_spcd, p_dbh, p_diahtcd, p_wdldstem)
    Dim v_sppgrp     As String
    Dim v_b0, v_b1, v_bm, v_dbh, v_live_aboveground        As Double
    Dim pied, drcp, quga,drcq, stm            As Double

   v_dbh = 2.54 * p_dbh                  'Convert diameter from inches to cm
'-------------------------------------------------------------------------------
'Adjust diameter for woodland species that are measured at diameter root collar not dbh
    If (p_diahtcd = 2) Then
      pied = 0
      drcp = 0
      quga = 0
      drcq = 0
      If (p_wdldstem <= 1) Then
        stm = 1
      Else
        stm = 0
      End If
      If (p_spcd < 300) Then
        pied = 1
        drcp = v_dbh
      Else
        quga = 1
        drcq = v_dbh
      End If
      v_dbh = -6.818 + (1.0222 * v_dbh) + (1.8879 * stm) + (1.8971 * pied) - (0.0399 *
drcp) + (3.11 * quga) - (0.0689 * drcq)
      If (v_dbh < 2.54) Then
        v_dbh = 2.54
      End If
    End If
'-------------------------------------------------------------------------------
'Assign FIA species codes (spcd) to Jenkins species group codes (v_sppgrp)
Select Case p_spcd
    Case 350, 351, 352, 353, 354, 355, 740, 741, 742, 743, 744, 745, 746, 747, 748, _
                749, 752, 753, 754, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929
      v_sppgrp = "aa"
    Case 40, 41, 42, 43, 57, 60, 67, 68, 70, 71, 72, 73, 81, 211, 212, 220, 221, 222, 240,
241, 242
      v_sppgrp = "cl"
    Case 200, 201, 202
      v_sppgrp = "df"
    Case 310, 311, 312, 313, 315, 316, 317, 319, 320, 370, 371, 372, 373, 374, 375, 376,
377, 378, 379
      v_sppgrp = "mb"
```

```
   Case 323, 330, 331, 332, 333, 334, 336, 337, 341, 345, 356, 357, 358, 360, 361, 362,
367, 381, 391, _
            420, 421, 422, 423, 424, 430, 431, 450, 451, 452, 460, 461, 462, 463, 471, 481,
490, 491, 492, _
            500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 520,
521, 522, 540, _
            541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 555, 561, 571, 580,
581, 582, 583, _
            590, 591, 600, 601, 602, 603, 604, 605, 606, 611, 621, 631, 641, 650, 651, 652,
653, 654, 655, _
            657, 658, 660, 661, 662, 663, 664, 680, 681, 682, 683, 684, 690, 691, 692, 693,
694, 701, 711, _
            712, 720, 721, 722, 729, 730, 731, 732, 760, 761, 762, 763, 764, 765, 766, 769,
770, 771, 772, _
            773, 774, 852, 853, 854, 855, 856, 857, 858, 859, 860, 863, 864, 865, 866, 873,
874, 876, 877, _
            882, 883, 884, 885, 886, 887, 888, 890, 891, 895, 896, 897, 901, 906, 907, 908,
909, 910, 912, _
            913, 914, 915, 919, 931, 934, 935, 936, 937, 940, 950, 951, 952, 953, 970, 971,
972, 973, 974, _
            975, 976, 977, 981, 982, 986, 987, 988, 989, 991, 992, 993, 994, 995, 996, 997,
998, 999, 5091, 5092, 5093, 7211
     v_sppgrp = "mh"
   Case 314, 318, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413,
531, 800, 801, 802, _
        804, 805, 806, 807, 808, 809, 811, 812, 813, 815, 816, 817, 818, 819, 820, 821,
822, 823, 824, 825, _
        826, 827, 828, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842,
844, 845, 846, 847, 850
     v_sppgrp = "mo"
   Case 0, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115,
116, 117, 118, _
   119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135,
136, 137, 138, 139, 140, 142, 143, 144, 298, 299
     v_sppgrp = "pi"

   Case 90, 91, 92, 93, 94, 95, 96, 97, 98
     v_sppgrp = "sp"

   Case 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 230, 231, 232, 250, 251, 252, 260,
261, 262, 263, 264
     v_sppgrp = "tf"

   Case 50, 51, 52, 53, 54, 55, 56, 58, 59, 61, 62, 63, 64, 65, 66, 69, 300, 303, 304, 321,
322, 475, 755, _
            756, 757, 758, 768, 803, 810, 814, 829, 843, 902, 990
```

```
      v_sppgrp = "wo"
    Case Else
      v_sppgrp = "xx"
    End Select
'--------------------------------------------------------------------------------
'Assign coefficients based on Jenkins species group
    If (v_sppgrp = "aa") Then
      v_b0 = -2.2094
      v_b1 = 2.3867
    ElseIf (v_sppgrp = "mb") Then
      v_b0 = -1.9123
      v_b1 = 2.3651
    ElseIf (v_sppgrp = "mh") Then
      v_b0 = -2.48
      v_b1 = 2.4835
    ElseIf (v_sppgrp = "mo") Then
      v_b0 = -2.0127
      v_b1 = 2.4342
    ElseIf (v_sppgrp = "cl") Then
      v_b0 = -2.0336
      v_b1 = 2.2592
    ElseIf (v_sppgrp = "df") Then
      v_b0 = -2.2304
      v_b1 = 2.4435
    ElseIf (v_sppgrp = "tf") Then
      v_b0 = -2.5384
      v_b1 = 2.4814
    ElseIf (v_sppgrp = "pi") Then
      v_b0 = -2.5356
      v_b1 = 2.4349
    ElseIf (v_sppgrp = "sp") Then
      v_b0 = -2.0773
      v_b1 = 2.3323
    ElseIf (v_sppgrp = "wo") Then
      v_b0 = -0.7152
      v_b1 = 1.7029
    End If
    v_bm = Exp(v_b0 + v_b1 * Log(v_dbh))
'-----------------------------------------------------------------------------------------
'Calculate biomass
    v_live_aboveground = v_bm
    JenkinsBiomass = v_live_aboveground * 2.046 ' Convert kg to pounds

End Function
```

Wildlife Habitat Applications I



Long Term Ecological Monitoring Program on the Kenai National Wildlife Refuge, Alaska: An FIA Adjunct Inventory

John M. Morton[1], Matthew Bowser[1,2], Edward Berg[1], Dawn Magness[1,2] , and Todd Eskelin[1]

ABSTRACT: *Kenai National Wildlife Refuge (KENWR) has a legislative mandate "to conserve fish and wildlife populations and habitats in their natural diversity". To improve our understanding of spatial and temporal variation at the landscape level, we are developing the Long Term Ecological Monitoring Program (LTEMP) to assess change in biota on the sample frame used by the USDA Forest Inventory and Analysis program (FIA). Through a formal agreement with the FIA, we completed our baseline inventory of 259 permanent terrestrial plots systematically distributed at 5-km intervals across the 805,000-ha KENWR in 2004 and 2006. In addition to the forested vegetation sampled by the FIA, we sampled vascular and nonvascular plants on non-forested plots, and breeding landbirds, arthropods, and noise on all plots. All sampling methods are passive, nondestructive (to habitat), relatively inexpensive, and require ≤ 2 visits to a plot in a given sampling year. To date, we have recorded 647 species including one insect family and five insect species new to Alaska, two new sedges for KENWR, and a range expansion for Hammond's flycatcher. In collaboration with the FIA, we plan to resample 20% of plots every other year over a 10-year monitoring window. However, implementation of the monitoring phase was delayed as we complete species identification, develop novel ways of estimating species-specific detection probabilities, evaluate statistical power to detect change, and consider modifications to the proposed rotating panel design. Our approach provides a statistically-rigorous framework for landscape monitoring and modeling, yet maintains a great deal of design flexibility. Integration with the FIA ensures that LTEMP is cost effective, and the collocation of floral and faunal sampling permits additional species-habitat modeling and other explanatory spatial modeling. We believe LTEMP can serve as a template for agencies that are developing long-term monitoring programs of biodiversity at the landscape level.*

KEYWORDS: Kenai, National Wildlife Refuge System, Forest Inventory and Analysis, inventory, monitoring, occupancy modeling, Long Term Ecological Monitoring Program, diversity, arthropod, bird, plant

[1]U.S. Fish and Wildlife Service, Kenai National Wildlife Refuge, P.O. Box 2139, Soldotna, AK 99669.
[2]Department of Biology and Wildlife, University of Alaska, Fairbanks, AK 99775.

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

The National Wildlife Refuge System (NWRS), part of the U.S. Fish and Wildlife Service (USFWS), is composed of 38 million ha contained in more than 520 units in all 50 states and U.S. territories. The NWRS helps sustain more than 700 bird, 220 mammal, 250 reptile and amphibian, and 200 fish species including habitats for at least 250 species of threatened and endangered plants and animals (West et al. 2003). Although the NWRS is over a century old, its mission, *to administer a national network of lands and waters for the conservation, management and where appropriate, restoration of the fish, wildlife and plant resources and their habitats within the United States for the benefit of present and future generations of Americans*, was only recently established under the Refuge System Improvement Act of 1997 (Public Law 105-57). This federal legislation mandated that the "biological integrity, diversity, and environmental health of the system" be maintained; USFWS policy was subsequently codified as 601 FW 3 in 2001. Not only is this the first organic legislation to be enacted since the 1970s, Meretsky et al. (2006) point out that this is one of the most emphatic conservation directives ever written by Congress.

By policy (701 FW 2), most units in the NWRS have some sort of inventory and monitoring program for wildlife and habitats. Monitoring on refuge units has traditionally been focused on the species or taxon identified in their establishing legislation, usually the Migratory Bird Treaty Act or Endangered Species Act. However, given the spatial and taxonomic scope of the Refuge System Improvement Act, future monitoring of biological resources will be more complex and likely require new approaches. Indeed, Meretsky et al. (2006) cautioned that existing refuge monitoring programs should not automatically be "grandfathered". They recommended several ways that traditional monitoring programs could be modified to respond to this new legislation including the detection of biological changes even before the causes of those changes are recognized; use of ecological indicators to justify management actions; expand the scope of conservation monitoring, assessment, and management beyond refuge boundaries to encompass surrounding landscapes; consider the use of composite or multi-metric indexes and other methods to assess ecosystem condition; and, consider an NWRS-wide approach to developing ecosystem monitoring tools.

Additionally, it is recognized that new data will be needed to respond to the impacts of accelerated climate change. An *ad hoc* group of concerned USFWS biologists noted that "the fundamental impact of climate change…is the re-distribution of flora and fauna in response to changes in disturbance regimes and climatic envelopes. Because extirpation and extinction, colonization and invasion, and novel species assemblages are expected but unpredictable future outcomes, we need a more holistic approach to inventorying and monitoring biota. The USFWS cannot be expected to adapt strategically to climate change impacts without a better sense of *in situ* biodiversity. We speak almost cavalierly about the impacts of climate change and invasive species on biodiversity without any

knowledge of the full extent of species richness for the lands we manage. We strongly recommend developing a national program to inventory and monitor changes in the biological integrity, diversity, and environmental health of the NWRS consistent with the 1997 Refuge Improvement Act". More specifically, "objectives of this program would be scaled to detect changes in flora, fauna, and other metrics at the regional and/or national level. This program would consist of permanent sampling sites, ideally integrated and cost-shared with existing national monitoring programs such as the USDA Forest Inventory & Analysis (FIA) program" (Johnson et al. 2008).

Here we show how a formally-designated adjunct inventory to the FIA, the Long Term Ecological Monitoring Program (LTEMP) on the Kenai National Wildlife Refuge (KENWR), can be a viable approach for inventorying biodiversity at hierarchical scales. The programmatic goals of the LTEMP are to determine the occurrence and distribution of selected floral and faunal assemblages on KENWR and within selected habitats; develop explanatory statistical models to assess the effects of geological, biological, and anthropogenic factors on floral and faunal distributions; and assess trends in the occurrence and distribution of selected floral and faunal assemblages on KENWR. In this paper, we present preliminary results from our initial inventory of species richness on the KENWR.

Study area

The 805,000-ha KENWR is in south-central Alaska on the Kenai Peninsula, which is formed by the Cook Inlet and Prince William Sound (Fig.1). The KENWR shares boundaries with Chugach National Forest and Kenai Fjords National Park. The KENWR was established in 1941 as the Kenai National Moose Range but was renamed under the Alaska National Interest Lands Conservation Act (ANILCA) in 1980. This act also gave KENWR its mandates, of which the primary purpose is *to conserve fish and wildlife populations and habitats in their natural diversity.* Furthermore, ANILCA also defined fish and wildlife as *any member of the animal kingdom including without limitation any mammal, fish, bird, amphibian, reptile, mollusk, crustacean, arthropod or other invertebrate.*

Biodiversity is unusually high for this latitude (59°25'- 61°) because of the juxtaposition of two biomes on the Kenai Peninsula: the northern fringe of the Sitka spruce-dominated (*Picea sitchensis*) coastal rainforest on the eastern flank of the Kenai Mountains, and the western-most reach of boreal forest in North America on the western side of the Kenai Mountains. Forests on KENWR are dominated by white (*P. glauca*), Lutz's (*P. x lutzii*), and black spruce (*P. mariana*) with an admixture of aspen (*Populus tremuloides*) and birch (*Betula neoalaskana*). Extensive peatlands are interspersed among spruce in the Kenai Lowlands. Lichen-dominated tundra replaces mountain hemlock (*Tsuga mertensiana*) and sub-alpine shrub (primarily *Alnus* spp.) above treeline in the

Kenai Mountains and Caribou Hills. Topographic relief varies from sea level at Chickaloon Flats, a tidal estuary extending into the Cook Inlet, to 1800 m above sea level in the Kenai Mountains. The 1800-km^2 Harding Icefield straddles the Kenai Mountains along the boundary that separates Kenai Fjords National Park and KENWR.

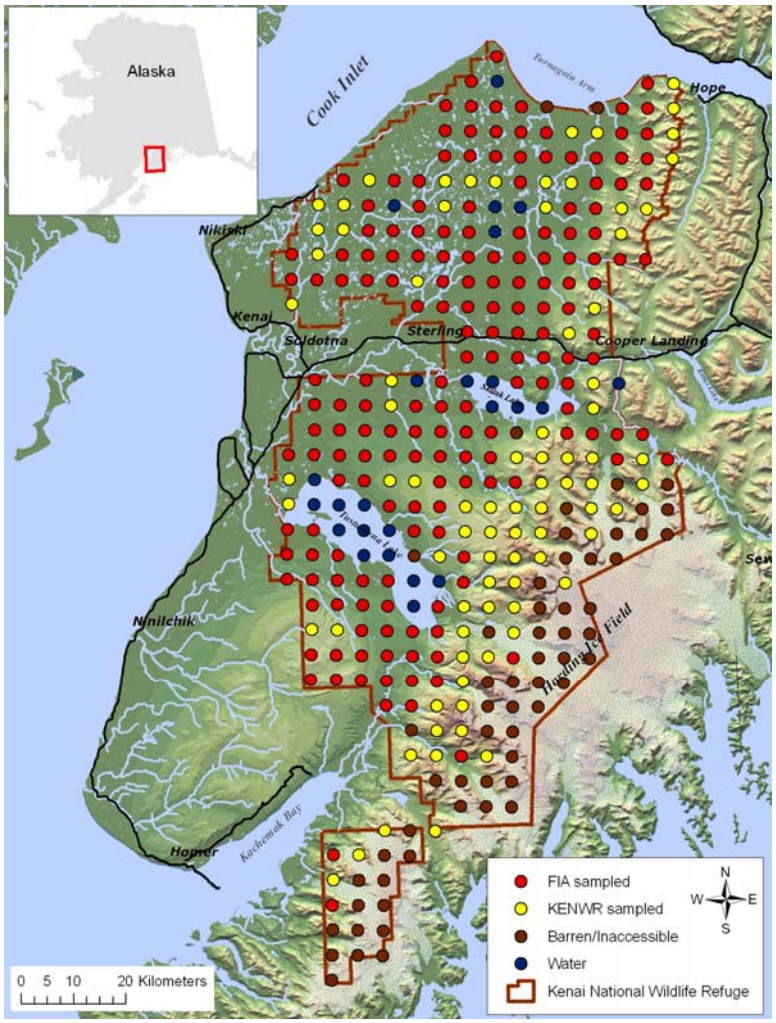


Figure 1: Distribution of 342 systematically distributed at 4.8 km intervals on the 805,000 ha Kenai National Wildlife Refuge.

The impacts of a warming and drying climate are well documented on the KENWR. The Kenai Peninsula was the epicenter of a spruce bark beetle (*Dendroctonus rufipennis*) outbreak that lasted over a decade and caused high mortality of Sitka, Lutz, and white spruce on 1.6 million ha in south-central Alaska. A run of warm summers since 1987 set the stage for this outbreak of unprecedented scale, suggesting that with a future warmer climate, fire and beetle kill may be more closely associated than in the past (Berg et al. 2006). Closed-basin lake levels have declined by as much as 1 m and many ponds shown on 1950 maps and aerial photos are now grassy pans with various degrees of black spruce

and hardwood invasion (Klein et al. 2005). The Harding Icefield has lost 21 m in elevation (Adageirsdottir et al. 1998) and 5% in surface area (Rice 1987) during 1950 - 1990s. Over this same period, treeline in the Kenai Mountains has risen, on average, 1 m per year (Dial et al. 2007).

Methods

Forest Inventory and Analysis Program (FIA)

The FIA consists of a nationally-consistent core program to "make and keep current a comprehensive inventory and analysis of the present and prospective conditions and requirements of the renewable resources of the forest and rangelands of the United States". The FIA can be enhanced at the regional, state or local level to address special interests (http://www.fia.fs.fed.us/). Vegetation data are stored in a well-defined database (Miles et al. 2001) and are available for public dissemination from the National FIA Database Retrieval System (http://www.srsfia.usfs.msstate.edu/scripts/ew.htm). However, coordinates for FIA plot locations are not available and are protected from Freedom of Information Act requests under the Agriculture Research, Extension, and Education Reform Act of 1998 (16 USC 1642(e)). The national FIA core consists of three phases:

Phase 1 uses remote-sensing data to classify land into forest and non-forest, and take spatial measurements such as fragmentation, urbanization, and distance variables. This phase has historically been done using aerial photography, but is changing to a system based on satellite imagery.

Phase 2 consists of a set of field sample points distributed across the landscape in a systematic grid of tessellated hexagons, each containing 2,360 ha. One FIA plot is located within each hexagon on which vegetation structure and composition are re-sampled every 10 - 15 years (Roesch and Reams 1999). Non-forest locations may also be visited to quantify rates of land use change.

Phase 3 consists of a small subset (~6%) of the Phase 2 plots which are visited during the growing season in order to collect an extended suite of ecological data including full vegetation inventory, tree and crown condition, soil data, lichen diversity, coarse woody debris, and ozone damage. Prior to 2000, these plots were part of the Forest Health Monitoring program.

As part of the implementation of the Alaska Coastal Inventory in south-central Alaska, the Pacific Northwest Research Station (PNWRS) established 176 FIA plots on KENWR

in 1999 (Fig.1). Consistent with Phase 2, these plots were distributed across the forested portion of KENWR at a resolution of 1 plot per 2,360 ha. By 2003, the PNWRS had completed the first inventory of these plots following field protocols of the Coastal Alaska Inventory (USFS 2002a).

However, the Coastal Alaska Inventory was implemented before the FIA national protocols were fully developed. Also, under a 2004 Memorandum of Understanding, the PNWRS agreed to reduce the frequency of re-sampling to ensure the integrity of 525,000 ha of Congressionally-designated Wilderness on KENWR. Consequently, there are several significant differences between what has been done on KENWR than in other parts of the U.S including systematically-distributed plots at regular intervals (versus randomly placed within tessellated hexagons); inventory of vascular and nonvascular flora in a 5.64-m radius column on each center subplot; and a monitoring design that specifies re-sampling 20% of plots every other year for 10 years (i.e., rotating panel design).

Sample Frame

To take advantage of inventorying and monitoring by the FIA, we adopted their sample frame by extending the existing FIA grid over the remaining non-forested portions of KENWR and overlaid additional sampling to inventory selected faunal assemblages on all points regardless of vegetation. This full sample frame, hereafter considered the LTEMP sample frame, is comprised of 342 plots systematically distributed at 4.8-km intervals across the 805,000-ha KENWR (Fig. 1). At this spatial resolution, the distribution of these plots is proportional to the availability of macro-habitat types (Table 1). This is an important attribute as data derived from the LTEMP sample frame are representative of the KENWR.

Table1: Comparison of the proportional distributions of LTEMP plots with macro-habitats on the Kenai National Wildlife Refuge.

Habitat	Plots (%)	Hectares (%)
Forest	161 (47)	382,790 (48)
Conifer	105 (31)	222,980 (28)
Deciduous	12 (4)	29,463 (4)
Mixed	44 (13)	130,347 (16)
Shrub/grass	26 (7)	57,392 (7)
Barren/sparsely vegetated	60 (18)	133,260 (17)
Wetlands	20 (6)	49,489 (6)
Snow/ice	51 (15)	117,348 (15)
Water	24 (7)	64,442 (8)
Σ	342 (101)	804,721 (101)

Vegetation Sampling of Forested Sites by FIA

The PNWRS sampled vegetation on 176 forested plots following field protocols used in the Coastal Alaska Inventory (USFS 2002a). Consistent with national Phase 2 protocols, each FIA site consists of four subplots to measure within-site variance (Fig. 2). Each subplot consists of three nested fixed-radius circular plots: 7.3-m tree plot, 5.64-m horizontal-vertical (HV) vegetation profile, and a 2-m seedling/sapling plot. Three 11.28-m transects intersect the center subplot to estimate downed wood (see USFS 2002a). National and regional FIA field protocols were generally similar except for relatively minor differences in measurement units and location of subplot sampling. The one obvious and significant discrepancy was the exclusive sampling of HV plots as part of the Coastal Alaska Inventory.

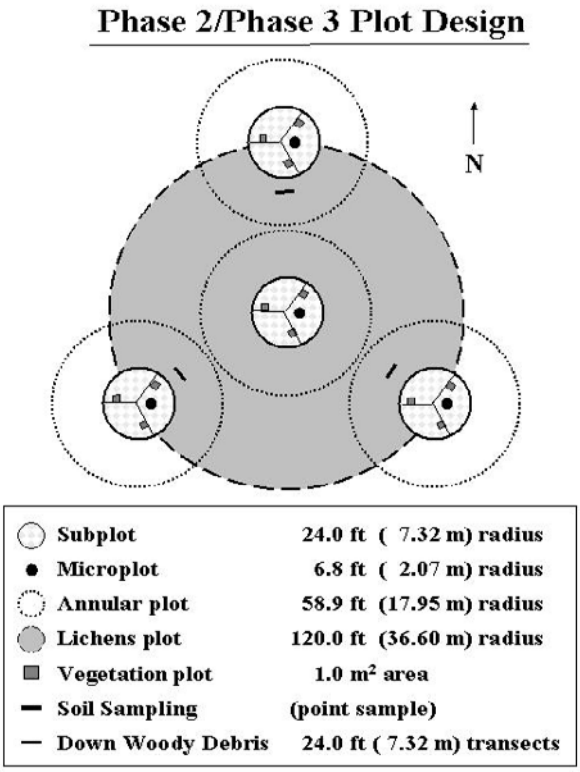


Figure 2: Sampling schematic at FIA grid point.

Horizontal-Vertical (HV) Plot: A 5.64-m radius (horizontal distance) circular plot was established at point center. The HV plot was used to determine the horizontal and vertical distribution, density, diversity, and composition of plants and non-living material. Vegetation was classified into strata or layers starting at ground level such that this plot represented a column with a base of 100 m^2. Vertical dimensions were estimated using the natural layer breaks observed on the plot. One HV profile plot was established for each polygon type that had a point meeting the following criteria: the polygon type at point center is vegetated (i.e., shrub, herbaceous, forest); the HV plot falls entirely within the polygon at point center (i.e., does not straddle polygon types); the polygon type at point center covers ≥ 90% of the area; and the point center does not fall in an inclusion.

Data collected on HV plots include shrubs, grasses, forbs, lichens, mosses, tree seedlings (trees < 2.5 cm DBH), and in some cases shrub-like trees not recorded on the tree plot. For example, Krummhol*z* and other stands where trees were growing in a twisted/stunted shrub-like form, were treated as shrub land. The HV record also included arboreal lichens, mosses, forbs and shrubs regardless of their substrate.

Adjunct Inventory Field Methods

Vegetation Sampling of Non-forested Sites by KENWR: We did not use the FIA cluster sampling design, where four subplots are surveyed at a site (Fig. 2). Instead, we sampled one plot, centered on the plot center and used sampling methods which better characterize non-forested vegetation (Fig. 3). Our objectives were to quantify the relative frequency of ground cover and species within different height strata at each site. Similar to the HV plot sampled by the FIA, we also recorded all vascular and nonvascular flora on a 100-m^2 circular plot with a 5.64 m radius from plot center. Voucher specimens or unknown plant material were collected at the site, but from locations > 10 m from plot center. We also took two stereo digital photographs on the north-south axis of each plot similar to PNWRS protocols for the HV plot.

Herbaceous and woody vegetative cover within the first 2 m above ground were sampled using a modified point-intercept technique (Dunn 1992). A version of this



Figure 3: Sampling alpine tundra using a modified line-intercept method in the Kenai Mountains. Ecologist Ed Berg kneels near the permanent monument that marks the center of a 5.64-m radius circular plot and four 10-m transects that radiate in the cardinal directions from plot center.

sampling approach has been adopted by Denali National Park. Four 10-m long transects were laid out in the cardinal directions from plot center using tightly-woven, braided nylon cord marked at 0.5-m intervals (n = 20 points per transect). A 2-m long, 13-mm diameter steel rod, marked at 1 m to separate the vertical column into two strata, was used as a sampling pin that is held vertically at each sampling point. Each plant taxon that touches the sampling pin ≥ 1 time was recorded within stratum at each point. Only one hit per point per stratum per taxon was recorded at each of the 80 sampling points. Consequently, a minimum of 40 tallies were recorded per stratum; the number of tallies recorded can exceed this value considerably depending on species richness.

Substrate categories were BARE GROUND, ROCK, LITTER, DEAD WOOD (> 25 mm diameter), WATER, SNOW/ICE, ASH/CHARCOAL, and LIVE VEGETATION. If a dead wood fragment/stick was < 25 mm in diameter or width, then it was recorded as LITTER. Rock particles < 13 mm in longest dimension were recorded as BARE GROUND; ≥ 13mm were recorded as ROCK. Any live vegetation recorded as a substrate was also recorded by species in the 1-m stratum. The total number of substrates recorded always equaled 20 per transect.

Arthropods: We sampled arthropod occurrence and relative abundance within the 5.64-m radius circular plot using a 30 cm (12 inch) diameter sweep net (BioQuip™ model 7112CP with 30 cm extension handle). We swept the entire plot for ≥ 5 min, ensuring that all substrates and microhabitats within reach (including overhead) were sampled. All specimens were placed in a single Nalgene® wide-mouth 250 ml bottle containing sufficient 95% denatured ethanol to cover the sample.

Birds: We sampled landbird abundance and occurrence using variable circular plot methods during the last three weeks in June. We adopted (with some modifications) the protocols used in the Alaska Landbird Monitoring Survey (Handel et al.1998), where horizontal distances to each bird were estimated at 1-min increments during a 10-min sampling interval using auditory or visual cues. Surveys were conducted 30 min after sunrise during the first 4 - 5 hrs of the morning, but only under specified conditions of good visibility, little or no precipitation, and light or no winds. We used a laser range finder to help with distance estimation and recorded unknown or questionable songs/calls with a Sony digital Walkman™ MZ-N10 minidisc recorder and a Saul Mineroff Electronics™, Inc., ATR55 mini-microphone boom.

Weather and noise: We measured wind speed (m/sec averaged over 30 sec), temperature (°C), and relative humidity with a Kestrel 3000 Pocket Weather™ meter at ground level. We used the Beauford wind scale to record wind speed at canopy level. We used the Larson Davis Model 720 sound meter to measure sound levels (Leq, Lmax and Lpeak) over a 5-min interval while birds were being counted. The sound meter was mounted on a tripod, oriented with the microphone pointed skywards, and placed ≥ 3 m from either observer to reduce recording of incidental noise.

Curation and Species Identification: In the lab, vascular plants were pressed, and lichens and mosses were stored in paper bags. Plant specimens were dried in a convection oven at 60° C until dry, typically 1 - 3 days. Arthropod specimens were processed and archived using appropriate curation methods. All arthropods present in the samples were sorted into orders and most were sorted to families. Species identifications for both plants and arthropods were made whenever possible but many specimens were shipped to >40 specialists for expert determinations. With the exception of small amounts of material lent to various systematists, all specimens remain in the collection of the KENWR (international collection coden: KNWR).

Results

Overview of Field Sampling Approach

Of 342 plots on the LTEMP sample frame, only 259 plots were deemed terrestrial and accessible; 59 plots were considered inaccessible and 24 plots were located on water. Of these 259 terrestrial plots, vegetation was sampled on 176 and 83 by FIA and KENWR staff, respectively. The forest vegetation was sampled as part of the Coastal Alaska Inventory during 1999 – 2002. We sampled the non-forested vegetation, primarily alpine tundra and lowland wetlands, in July - August of 2004 and 2006. A 2-person team typically spent ~ 2 hrs sampling vegetation on plot.

We inventoried breeding landbirds, arthropods, and noise on 255 terrestrial plots during the latter three weeks in June of 2004 (n = 152) and 2006 (n = 103); data from four plots were discarded for various reasons. Our sampling window was optimized to detect most breeding landbird species at this latitude, and arthropod and noise sampling were considered lesser priorities. We used a rapid approach, spending ~ 40 min on each plot. A 2-person team typically divided duties such that the primary bird observer conducted the 10-min bird survey while the other observer measured sound and weather parameters. While one person sampled for arthropods, the other would take plot photographs, record incidental wildlife sign, and take other notes. Because the plot and potential landing site were not at the same location, round-trip travel time between the helicopter and plot was typically 20 min such that a team could complete a plot every hour.

With a few exceptions, plots were accessed by helicopter. We used a Bell 206 Jet Ranger during wildlife surveys, “hop-scotching” two 2-person teams from point to point. We completed 4 – 6 plots per team per day, with hours of operation typically restricted to 0400 – 1000 due to constraints imposed by avian breeding activities. Because we were concerned about potential latitudinal (59°25’- 61°) and elevational (0 – 1800 m above sea

level) bias in breeding phenology over a 3-week sampling interval, we alternated starting points in the morning between the northern and southern part of KENWR and the western and eastern parts of KENWR although this protocol was contingent on weather constraints. We also shared an A-Star jet helicopter with PNWRS staff later in one summer to help complete sampling of non-forested plots.

FIA plots were previously monumented following protocols outlined in the Coastal Alaska Inventory (USFS 2002a). We monumented non-forested plots on the LTEMP sample frame (n = 83) with Bernsten® aluminum survey caps on 80-cm long aluminum break-away shafts that were pounded in so as to as flush as possible with the ground surface. Both the cap and shaft have magnets inserted to facilitate re-finding these plots with a metal detector during subsequent visits.

Data Derived from Adjunct Inventory

To determine species occurrence of vascular and nonvascular flora, and arthropods, we considered only those species detected in the 5.64-m^2 radius circular plot. To determine the occurrence of avifauna, we used only those bird species which were detected within 200 m of plot center.

To date, we have identified 647 species including 86 birds, 333 vascular plants, 78 nonvascular plants, 44 lichens and fungi, and 106 arthropods as part of LTEMP. As many as 54 species were recorded for a single LTEMP plot, but this value will certainly increase as outstanding taxonomic work is completed. The subset of forested HV plots that were sampled by the PNWRS yielded 217 vascular plant species and 33 nonvascular plant species.

We have spatial data (albeit limited for species with low detectability) for 90% of 96 landbirds and 67% of 499 vascular plants previously documented on the KENWR or thought to occur based on range maps. We identified two sedge species new to the KENWR (*Carex tenuiflora, C. dioica*) and a range expansion for Hammond's flycatcher (*Empidonax hammondii*). We found four exotic vascular plant species (*Lupinus polyphyllus, Poa trivialis, Potentilla gracilis, Taraxacum officinale*) on 3% of LTEMP plots.

From more than 15,000 arthropod specimens collected, we have identified 220 taxa representing 18 orders. Although the taxonomy of many arthropod and some nonvascular plant specimens are still outstanding, we have identified one insect family new to Alaska (Achilidae: *Epiptera* sp.) and five insect species new to Alaska (*Aphelopus albopictus*, *Aphodius aleutus*, *Delphacodes serrata*, *Gonatopus ashmeadi*, and *Paraliburnia kilmani*).

Lastly, we have established a benchmark for ambient sound during early morning hours in the absence of rain and high winds. Sound levels integrated over 5 minutes on 257 plots averaged 45.1 dBa (SE = 0.68) and ranged from 32 - 95 dBa.

Discussion

We have designed the LTEMP to be a comprehensive, albeit not exhaustive, spatial and temporal assessment of biodiversity (i.e., species richness) on the KENWR. In just two summers of inventorying, we have amassed a significant volume of spatial data for almost 650 species over 805,000 ha that continues to be processed and analyzed. Taxonomic work on nonvascular plants and insects is still being conducted. Nonetheless, the utility of the LTEMP as an adjunct inventory has already been demonstrated. We have expanded the known distributions of several species as part of this inventory. We have increased the known species richness on the KENWR to > 1,000 species. We have used these data to model bird distributions (Magness et al. 2008) and arthropod diversity (Bowser and Morton 2009).

We are already using the value of 3% to benchmark the occurrence of exotic flora on KENWR at the landscape scale. Although LTEMP plots represented only 10% of the plot data used to develop the first supervised vegetation classification of the Kenai Peninsula from LANDSAT imagery, they were disproportionately critical because of their complete spatial distribution (Lee O'Brien, pers. comm.). For landscape-level wildfire, which is still a major disturbance process on KENWR, we also expect to use vegetation and fuel load data from the FIA as pre-treatment data for wildfire and possibly as experimental controls for large prescribed burns.

The LTEMP will likely continue to expand as funding and field techniques permit. Unlike other similar programs such as the Long-term Ecological Monitoring Program at Denali National Park, the Alberta Biodiversity Monitoring Program, or the Forest Service's MultiSpecies Inventory & Monitoring Program (Manley et al. 2004), we employed a rapid sampling approach to assess species occurrence (primarily) as well as other taxon-specific population metrics. All sampling methods are passive, non-destructive (to habitat), relatively inexpensive, and require ≤ 2 visits to a plot in a given sampling year. Sampling efforts greater than this are likely to be too costly to sustain over the long term and may damage permanent plots by trampling or the introduction of exotic biota. Other critical attributes that make the LTEMP successful include

- permanent sampling sites to measure change;
- statistically robust sampling frame (systematic) to survive planned and unplanned habitat changes;
- data that are representative of the refuge unit;

- multi-taxa sampling and co-location of biotic and abiotic sampling to enhance both predictive and explanatory spatial modeling; and
- interagency cost-share.

It is critical to the success of this program, however, that we estimate species detection probabilities as the LTEMP is further developed to monitor spatial and temporal changes in species distributions (Bowser and Morton 2009). Species occupancy (which is essentially occurrence corrected for imperfect detection) will be our primary metric for assessing change in distributions of all taxa found on the LTEMP. Other metrics derived from repeated occupancy estimation include local rates of extinction and colonization, both of which would be desirable for evaluating shifts in species distributions. However, a major limitation in occupancy modeling as currently practiced is that detection is estimated by making multiple revisits within season to a plot (MacKenzie et al. 2006). This estimation procedure is expensive, particularly if helicopters are being used, and results in collateral damage to the plot. Alternatively, we are currently exploring spatial and temporal subsampling within a single plot visit to estimate detection probabilities (Bowser and Morton 2009).

We are also concerned that the PNWRS has abandoned the HV plot. Although the space occupancy concept behind the HV sampling methodology does not provide reproducible estimates of plant cover (van Hees and Mead 2000), the floral inventory generated on the HV plot appears to be fairly comprehensive. We developed the arthropod sampling to be spatially consistent with the floral community within the 100 m^2 plot. However, the PNWRS has chosen to use the Phase 3 sampling in lieu of the HV plot to be consistent with national protocols. We will have to consider changes to sampling design and protocols as we move towards the monitoring phase in LTEMP.

We had originally planned on extending this adjunct inventory into spatial and temporal monitoring by now. However, we have yet to complete the taxonomic identification of both arthropod and nonvascular plant specimens. Also, the synergistic value of co-locating additional inventory data on the LTEMP sample frame has become apparent as we explore spatial modeling approaches (e.g., Magness et al. 2008). Consequently, implementation of the monitoring phase has been postponed as we complete species identification, develop novel ways of estimating species-specific detection probabilities, evaluate statistical power to detect change, and consider modifications to the proposed rotating panel design (Bowser 2009, Bowser and Morton 2009). Assuming we can address these concerns in a timely way, we will re-sample 20% (n = 51) of the LTEMP plots every other year beginning in 2010 in collaboration with the FIA (Table 2).

In the future, the LTEMP sampling frame may also be used to sample the winter distribution of mammals with aerial digital photography, to apply more rigorous insect sampling methods for monitoring selected taxa, to collect soil samples to help develop better soil maps, and to determine baseline levels of ambient contaminants. We hope to

use the sound data as a basis for developing soundscape models. We have yet to analyze the structure and composition of the floral community, and to develop models to explain their relationship to faunal species distributions. In particular, we expect to examine potential silvicultural and fire management prescriptions to improve bird densities as estimated by Program DISTANCE (Laake et al.1993).

Table 2: Proposed rotating panel design for the Long Term Ecological Monitoring Program on the Kenai National Wildlife Refuge in which a panel of 51 plots is jointly re-sampled by both FIA and KENWR every decade.

PANEL	N		SAMPLING YEAR						
	FIA	KENWR	2004/2006		2010	2012	2014	2016	2018
			FLORA	FAUNA	ALL	ALL	ALL	ALL	ALL
1	35			X	X				
2	35			X		X			
3	35			X			X		
4	35			X				X	
5	35			X					X
Σ	175			X					
				X					
1		16	X	X	X				
2		16	X	X		X			
3		16	X	X			X		
4		16	X	X				X	
5		16	X	X					X
Σ		80	X	X					

Our approach provides a statistically-rigorous framework for landscape sampling, yet maintains a great deal of design flexibility. Integration with the FIA ensures that LTEMP is cost effective, and the co-location of floral and faunal sampling permits additional species-habitat modeling. Certainly the LTEMP approach and protocols are applicable to the other 15 refuges in Alaska which have similarly large landscapes; in fact, the 16 Alaskan refuges comprise 87% of the land base within the NWRS. On the other hand, almost 90% of the 540 units in the NWRS are < 50,000 acres (USFWS 2003). These refuge units are too small to benefit from the relatively coarse resolution of Phase 2 sampling by the FIA. However, this disadvantage can be overcome by increasing the sampling density within the FIA sample frame. Indeed, agency staff at Pictured Rocks National Lakeshore in Michigan have used exactly this approach to estimate fuel loads (Woodall and Leutscher 2003) and bird populations (Lieske 2006) for their unit while ensuring a landscape context. Consequently, we believe that LTEMP can serve as a template for agencies that are developing long-term ecological monitoring programs at the landscape level, particularly in the context of accelerated climate change, but still provide useful data at the local scale.

Acknowledgements

This work would not have been possible without the collaborative efforts of the PNWRS, FIA, particularly Sue Willits and Bill van Hees. LTEMP data management was provided by Mark Laker (KENWR). Several biologists and seasonal interns enthusiastically assisted in the field and lab including Toby Burke, Elizabeth Jozwiak, Charlotte Hockin, Doug Fisher, Susie Grimes, Suzy Medwell and Kimo Rogala. Additional funding was provided through Marathon Oil Corporation.

Literature Cited

Adageirsdottir, G.; Echelmeyer, K.A.; Harrison, W.D. 1998. Elevation and volume changes on the Harding Icefield, Alaska. J. Glaciology 44:570-582.

Berg, E.E.; Henry, J.D.; Fastie, C.L.; De Volder, A.D.; Matsuoka; S.M. 2006. Spruce beetle outbreaks on the Kenai Peninsula, Alaska, and Kluane National Park and Reserve, Yukon Territory: Relationship to summer temperatures and regional differences in disturbance regimes. Forest Ecology and Management 227:219-232.

Bowser, Matthew, L. 2009. Terrestrial arthropod biodiversity on the Kenai National Wildlife Refuge, Alaska. M.S. thesis. Dept. of Biology and Wildlife, University of Alaska, Fairbanks, AK. 141 pp.

Bowser, Matthew, L; Morton, John M. 2009. Monitoring and modeling terrestrial arthropod diversity on the Kenai National Wildlife Refuge. *In* McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2008.2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Dial, R. J.; Berg, E.E.; Timm, K.; McMahon, A.; Geck, J. 2007. Changes in the alpine forest-tundra ecotone commensurate with recent warming in southcentral Alaska: Evidence from orthophotos and field plots. J. Geophys. Res.112, G04015, doi:10.1029/2007JG000453.

Dunn, P. 1992. Long-term biological resource and threat monitoring of Hawaiian Natural Areas. Produced for the Hawaii Department of Land and Natural Resources, Division of Forestry and Wildlife, by The Nature Conservancy of Hawaii. Honolulu, HI.

Handel, Colleen M.; Matsuoka, Steve M.; Douglas, David C. 1998. The Alaska Landbird Resources Information System, Version 98.1. USGS Alaska Biological Science Center, Anchorage, Alaska. Available http://www.absc.usgs.gov/research/bpif/Monitor/alms2.html.

Johnson, Kurt A.; Morton, John M.; Anderson, Greer; Babij, Eleanora; Cintron, Gil; Fellows, Valerie; Freifeld, Holly; Hayum, Brian; Jones, LouEllyn; Nagendran, Meenakshi; Piehuta, Joe; Sterne, Charla; Thomas, Peter. 2008. Four key ideas to guide the Service's response to climate change: A white paper. Ad Hoc Climate Change Working Group, U.S. Fish and Wildlife Service, Arlington, VA. 28 pp. Available at https://intranet.fws.gov/climatechange/AdHoc_Climate_White_Paper_02-08-2008.pdf.

Klein, Eric; Berg, Edward E.; Dial, Roman. 2005. Wetland drying and succession across the Kenai Peninsula Lowlands, south-central Alaska. Canadian J. Forest Research 35:1931-1941.

Laake, J.L.; Buckland, S.T.; Anderson, D.R.; Burnham, K.P. 1993. DISTANCE Users Guide V2.0. Colorado Cooperative Fish & Wildlife Research Unit, Colorado State University, Fort Collins, CO.

Lieske, Patrick D. 2006. Impacts of site-specific habitat features on avian resource-use guilds in Pictured Rocks National Lakeshore, northern Michigan, USA. M.S. thesis. Northern Michigan University, Marquette, MI. 50 pp.

MacKenzie, Darryl I.; Royle, Andrew J. 2005. Designing occupancy studies: General advice and allocating survey effort. J. Applied Ecology 42:1105-1114.

MacKenzie, D. I.; Nichols, J.D.; Royle, J.A.; Pollock, K.H.; Bailey, L.L.; Hines, J.E. 2006. Occupancy estimation and modeling. Elsevier, NY. 324 pp.

Magness, Dawn R.; Huettmann, Falk; Morton, John M. 2008. Using Random Forests to provide predicted species distribution maps as a metric for ecological inventory & monitoring programs. Pages 209-229 *in* T.G. Smolinski, M.G. Milanova & A-E. Hassanien (eds.). Applications of Computational Intelligence in Biology: Current Trends and Open Problems. Studies in Computational Intelligence, Vol. 122, Springer-Verlag Berlin Heidelberg. 428 pp.

Manley, Patricia N.; Zielinski, William J.; Schlesinger, Matthew D.; Mori, Sylvia R. 2004. Evaluation of a multiple-species approach to monitoring species at the ecoregional scale. Ecological Applications 14:296-310.

Meretsky, Vicky J.; Fischman, Rob L.; Karr, James R.; Ashe, Dan M.; Scott, J. Michael; Noss, R.F.; Schroeder, Richard L. 2006. New directions in conservation for the National Wildlife Refuge System. BioScience 56:135-143.

Miles, P. D.; Brand, G.J.; Alerich, C.L.; Bednar, L.F.; Woudenberg, S.W.; Glover, J.F; Ezzell, E.N. 2001. The Forest Inventory and Analysis database: Database description and users manual version 1.0. Gen. Tech. Rep. NC-218. St. Paul, MN. USDA, Forest Service, North Central Research Station. 130pp.

Rice, B. 1987. Changes in the Harding Icefield, Kenai Peninsula, Alaska. M.S. thesis. School of Agriculture and Land Resources Management, University of Alaska, Fairbanks, AK. 116+ pp.

Roesch, F. A.; Reams, G.A. 1999. Analytical alternatives for an annual inventory system. J. Forestry 97:33-37.

U.S. Forest Service. 2002a. Field procedures for the Coastal Alaska Inventory. USDA Forest Service, Pacific Northwest Station, Forestry Sciences Lab & Region 10.

U.S. Forest Service. 2002b. Multiple species inventory and monitoring technical guide. Preliminary draft (4/15/03). 83 pp.

U.S. Fish and Wildlife Service. 2003. Annual report of lands under control of the U.S. Fish and Wildlife Service as of September 30, 2002. Division of Realty, USFWS, Arlington, VA.

van Hees, W. W. S.; Mead, B.R. 2000. Ocular estimates of understory vegetation structure in a closed *Picea glauca* / *Betula papyrifera* forest. J. Veg. Sci.11:195-200.

Viereck, L. A.; Dyrness, C.T.; Batten, A.R.; Wenzlick, K.J. 1992. The Alaska Vegetation Classification. General Technical Report PNW-GTR-286. USDA Forest Service, Pacific Northwest Research Station, Portland, OR. 278 pp.

West, Robin; Rundle, L.D.; Takekawa, John; Holle, D.G.; Kallin, S. 2003. The National Wildlife Refuge System: The next 100 years. Pages 80-92 *in* Transactions of the 68th North American Wildlife and Natural Resources Conference, Winston-Salem, NC.

Woodall, Christopher W.; Leutscher, Bruce. 2003. Extending and intensifying the FIA inventory of down forest fuels: Boundary Waters Canoe Area and Pictured Rocks National Lakeshore. Pages 159-169 *in* Proceedings of the Fifth Annual Forest Inventory and Analysis Symposium.

Monitoring and Modeling Terrestrial Arthropod Diversity on the Kenai National Wildlife Refuge

Matthew L. Bowser[1,2,3] and John M. Morton[1,2,4]

ABSTRACT: *The primary purpose of the Kenai National Wildlife Refuge (KENWR) is to "conserve fish and wildlife populations in their natural diversity," where "fish and wildlife" explicitly includes arthropods. To this end, we developed a Long Term Ecological Monitoring Program (LTEMP), a collaborative effort with the USDA Forest Inventory and Analysis (FIA) Program. In 2004 and 2006, we sweep-netted terrestrial arthropods on 255 100m^2 circular plots systematically distributed at 5-km intervals over the 805,000 ha KENWR. These samples yielded 15,136 specimens, which were sorted to families and to species when possible. The comprehensive spatial coverage of the LTEMP sampling design provided spatial data suitable for species distribution modeling, but we sought to improve upon this design by explicitly accounting for imperfect detection. We proposed a rotating panel design where each site would be visited once every ten years. Imperfect detection would be accounted for by spatial sub-sampling within plots. Using Monte-Carlo simulation, we assessed the proposed design of LTEMP for accurately monitoring changes in arthropod species distributions over time. Our simulations demonstrated that, for species that are likely to be collected in a single 50m^2 sweep net sample where they are present, the proposed LTEMP sampling design should provide accurate estimates of species distributions and local rates of colonization and extinction over the long-term. In order to document the landscape-scale patterns of arthropod diversity over the KENWR, we modeled arthropod family richness using random forest regression. Arthropod family richness data were obtained from LTEMP and topographic, temporal, and productivity variables were obtained from GIS datasets. The resulting map explained 22% of variation of diversity of sweep net samples. Highest diversity was predicted on the margins of coastal wetlands and in productive hardwood and mixed forests; lowest diversity was predicted at barren alpine sites.*

KEYWORDS: occupancy modeling, species distribution modeling, random forest regression, interpenetrating panel design, Monte-Carlo simulation.

[1]United States Fish & Wildlife Service; Kenai National Wildlife Refuge; PO Box 2139; Soldotna, Alaska 99669.
[2]University of Alaska Fairbanks; Biology & Wildlife Department; Fairbanks, Alaska, 99775.
[3]Matt_Bowser@fws.gov
[4]John_M_Morton@fws.gov

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

Motivation: A Broad Conservation Mandate

The primary purpose of the Kenai National Wildlife Refuge (KENWR) is "to conserve fish and wildlife populations and habitats in their natural diversity," where the term "fish and wildlife" was defined as "any member of the animal kingdom, including without limitation any mammal, fish, bird, amphibian, reptile, mollusk, crustacean, arthropod or other invertebrate," (Alaska National Interest Lands Conservation Act (ANILCA) of 1980).

Most of the "fish and wildlife" that the KENWR is charged to conserve are arthropods. They represent 80% of described diversity worldwide (Kremen et al. 1993). In Alaska, an incomplete, unpublished checklist of terrestrial arthropods included 6,500 species (unpublished checklist by Derek Sikes, University of Alaska Museum, Fairbanks, Alaska), roughly an order of magnitude greater than the number of vertebrate species of the state. Consistent with the world and the region, arthropods represent most of metazoan diversity on the KENWR.

Arthropods are generally more important than vertebrates in terms of ecosystem functioning in most terrestrial systems (Wilson 1987). In many ecosystems, herbivory by arthropods has a greater influence on primary production than vertebrates (Schowalter 2006). Most plants require pollination by insects (Tepedino 1979). Arthropods also serve as important predators (Snyder and Evans 2006), parasites, fungivores (Hopkin 1997), and decomposers (Wallwork 1983). Even a single species of arthropod can alter the composition and functioning of a landscape. For example, an outbreak of the spruce bark beetle, *Dendroctonus rufipennis* Kirby, has recently removed the overstory of vast forests of mature white spruce (*Picea glauca* (Moench) Voss) on the KENWR, resulting in proliferation of understory grasses (Boucher and Mead 2006). Many arthropod species, including the spruce bark beetle, are essential to the natural functioning of the KENWR.

Despite the mandate of ANILCA to conserve the natural diversity of arthropods, the fact that they represent most of biodiversity, and their immense ecological importance, little progress had been made toward finding out what arthropod species occur on KENWR. Insects, spiders, and other arthropods have been overlooked because of somewhat circular reasons: (1) none of the species on KENWR were known to be of conservation concern, (2) arthropod species are assumed to be free from risk of extinction unless contrary evidence exists and (3) they are thought to be prohibitively difficult to learn about or conserve.

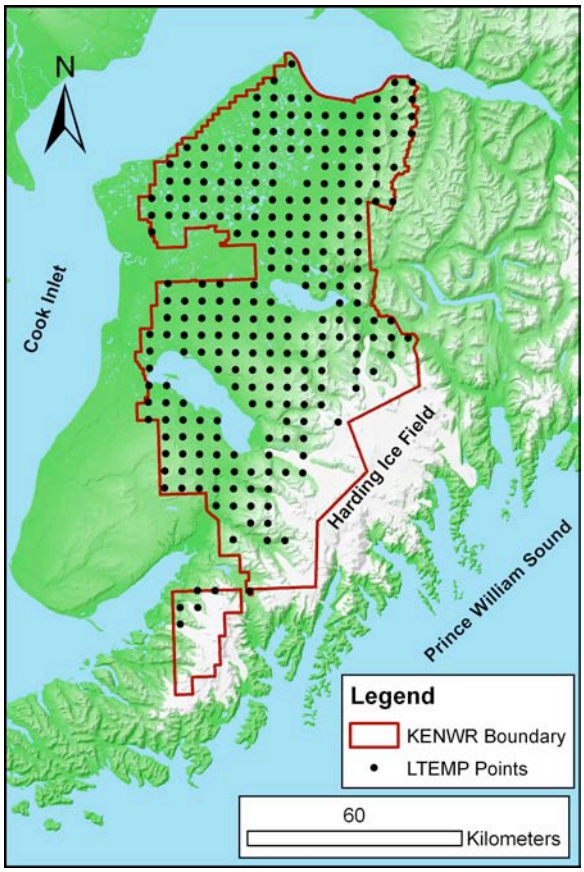


Figure 1: Map of the Kenai Peninsula showing the Kenai National Wildlife Refuge and LTEMP sampling locations.

The obvious first step is an inventory, a survey of the arthropod species that are present on the KENWR. However, a static inventory is inadequate because the composition of the fauna of the KENWR is changing and is expected to continue to change over the coming decades due to the responses of arthropods to a warming climate and due to additional introductions of exotic species. Mobile species are already moving up-slope and pole-ward as the climate warms (Gottfried et al., 1999; Parmesan 1996, 2006; Parmesan et al., 1999; Parmesan and Yohe 2003; Walther et al. 2002; Wilson et al. 2005), tracking their climatic preferences. As this process of re-distribution continues, some areas will become refugia (Gottfried et al. 1999) while others will serve as corridors (Hannah et al. 2002). While the most vagile species may respond rapidly, less motile species may not re-distribute themselves quickly enough to track climate. Habitat loss and fragmentation of suitable habitat will further exacerbate this problem, sometimes preventing species from moving to more suitable areas. Many species are expected to be lost as this process continues unless actions are taken to facilitate the re-distribution of species (Williams et al. 2005). Positioned as we are at the beginning of the re-distribution of species due to accelerating climate change, the

need for accurate documentation of current distributions of species and subsequent monitoring of species distributions is increasingly being recognized (Guisan and Thuiller 2005; Magness et al. 2008).

Long Term Ecological Monitoring Program (LTEMP) of the KENWR, a grid-based, multi-species, collaborative, inventory and monitoring framework undertaken through a memorandum of understanding between the USDA Forest Service Forest Inventory and Analysis (FIA) program (Figure 1), provided ample spatial data for documenting distributions of species (e.g. Magness et al. 2008). However, due to the sampling methods used, imperfect detection could not be accounted for.

When a species is not detected perfectly by field methods there is a possibility of reporting false absences (i.e., recording a species as absent where it was in fact present, but not observed). This leads to bias in estimates of the probability of occurrence of a species (MacKenzie et al. 2003, 2006). This issue of imperfect detection is especially relevant in a monitoring context. Unless it is accounted for, changes in the observed distribution of a species may be attributable to either (1) changes in the area occupied by a species or (2) changes in the likelihood that a species is detected due to changes in abundance, changes in seasonal phenology, etc.

Accounting for Imperfect Detection Using Occupancy Models

Substantial literature exists on the subject of accounting for imperfect detection (see MacKenzie et al. 2006 for a review). Occupancy models explicitly account for imperfect detection using repeated survey data to obtain unbiased estimates of presence/absence metrics (MacKenzie et al. 2003, 2006). Multi-season occupancy models can also be used to estimate local rates of colonization and extinction, potentially some of the most relevant monitoring metrics given expected distribution shifts and potential expansions of exotic species.

In order to accurately monitor species distributions of arthropods over time, we propose to modify the field methods of LTEMP so that imperfect detection can be explicitly accounted for.

Diversity as a Measure of Ecosystem Health

Species richness can be defined as the number of species distributions that overlap with the area of interest. Since ecosystem stability (McCann 2000), ecosystem function (Hooper et al. 2005) and resilience (Chapin et al. 2000) are generally correlated with diversity, species richness is itself considered to be a good indicator of the health of a system (Magurran 1988). Under this premise, taxon richness is often used as an indicator in environmental assessment studies

(e.g. Bechtel and Copeland 1970; Egloff and Brakel 1973; Wu 1982; Roth et al. 1994; Karr and Kimberling 2003).

Objectives

Our first objective was an initial inventory of the arthropod fauna of the KENWR. Our second objective was to develop methods for monitoring arthropod distributions over the long-term. Our third objective was to document landscape-scale patterns of arthropod biodiversity on the KENWR.

Methods

Sampling Methods

Study Area and Sampling Frame: Located in south-central Alaska, the 805,000 ha KENWR covers much of the western Kenai Peninsula (Figure 1) and consists mainly of boreal forest, lowland wetlands, and alpine habitats. The FIA program imposed a rectangular grid of sampling sites across the Kenai Peninsula with 4.8 km spacing between the sites. Of the sites within KENWR, 255 sites that did not fall on water or ice were included in the sampling frame of LTEMP. Sampling design and methods are described in more detail in Morton et al. (2009).

Plot Design and Field Methods: FIA crews established four circular, 5.64 m radius (100 m^2) sub-plots at each sampling site, with one central sub-plot centered on the site coordinates and three additional circular plots arranged in a triangle around the central sub-plot (Burkman 2005). LTEMP methods were focused on the central plots and excluded the three auxiliary plots.

Over the field seasons of 1999–2002, FIA field crews surveyed all of the 176 sites that had been determined to be forested. In 2004 and 2006, KENWR field crews sampled vegetation on the remaining 80 sites. Common to both FIA and KENWR field methods was collection of presence/absence data for all vascular plant species on the central 5.64m radius, circular plots.

Sampling of terrestrial arthropods was conducted by KENWR field crews over the field seasons of 2004 (152 sites) and 2006 (103 sites). Concurrent sampling of birds determined the seasonal (June 7–30) and daily (04:40 to 10:54 hours) sampling windows. A single sweep net sample was taken at each plot. For each sample, the collector swept a 30 cm diameter aerial insect net quickly back and forth over all vegetation and other substrates within reach over the entire circular plot. The contents of the net were then emptied into a vial of 80–90% ethanol.

Sorting and Identification: Arthropod specimens were processed and archived using appropriate curation methods. All arthropods present in the samples were sorted into orders and most were sorted to families. Species identifications were made whenever possible and many specimens were shipped to over 30 specialists for expert determinations. With the exception of small amounts of material lent to various systematists, all specimens remain in the arthropod collection of the KENWR (international collection coden: KNWR).

Occupancy Modeling of Species Distributions

Proposed Sampling Design: The proposed long-term sampling design for LTEMP is a rotating panel (interpenetrating panel) design where 51 sites (20%) would be sampled every other year so that each site would be visited once every ten years. Each panel would be representative of the KENWR as a whole.

Detection histories suitable for correcting for imperfect detection are most often obtained by visiting each site multiple times within a season. Because surveying all sites multiple times in one season would be cost-prohibitive, increase damage to the sites through trampling, and increase the likelihood of inadvertently introducing exotic species to the sites, we propose to modify arthropod collecting methods so that detection probabilities can be estimated from spatial subsamples taken on a single visit to a site. In the place of a single sweep net sample over the 100 m^2 circular plot, the plot would be split into two semicircles along a north-south axis and a separate sweep net sample would be taken from each 50m^2 semicircle. These spatial subsamples would provide a detection history suitable for occupancy modeling.

Monte-Carlo Simulations: Our general approach for evaluating LTEMP sampling regimes was to assess the performance of the proposed designs through Monte-Carlo simulation. These simulations were designed to answer the question of whether or not occupancy metrics could be estimated well; they were not designed to actually model species distributions. In the same way that a mean can be considered a special case of linear regression where there is only an intercept and an error term, the simple occupancy models we used estimated only occupancy metrics (occupancy, detection probability, and, in multi-season cases, rates of colonization and extinction) without considering additional variables (e.g. possible covariates) that would normally be included to produce species distribution models. These simulations were designed to answer the questions of interest without unnecessary complexity.

For each scenario considered, we (1) generated large numbers of simulated datasets that conformed to the proposed LTEMP monitoring design and had known parameter values, (2) fitted occupancy models using program PRESENCE (Hines 2007), and (3) compared the estimates obtained to the known parameter

values. We wrote scripts in the R programming language (R core development team 2008) scripts to generate datasets, write input files, fit occupancy models using PRESENCE, and extract results.

We generated datasets suitable for occupancy modeling in the following way. For a given set of scalar values of occupancy (Ψ, the proportion of the area occupied by a species) and detection probability (p), we first generated a list of occupancy states at n sites by specifying that the occupancy states at all sites were independently and identically distributed realizations of a Bernoulli process with a rate of Ψ using the rbern function in the Rlab add-on library (Boos et al. 2006). Similarly, we specified that detection events in each of the two sweep net samples from each site were independently and identically distributed realizations of a Bernoulli process with a rate of p. Multiplying the detection states by the occupancy states yielded detection histories with the specified parameter values of Ψ and p.

For single-season scenarios, we considered nine values of occupancy (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9) and the same nine values of detection probability. For each combination of parameters, we generated 100 datasets (a total of $9 \times 9 \times 100 = 8{,}100$ datasets). Using the same simulated datasets, we obtained estimates from program PRESENCE and naïve estimates of occupancy, where naïve estimates are simply the proportion of sites where a species is observed in any survey.

In order to assess how well rates of extinction and colonization could be monitored over the long-term by the proposed sampling design, we intentionally selected parameter values that were realistic. We set initial occupancy to 0.7, detection probability to 0.8, the local rate of extinction (ε) to 0.1, and the local rate of colonization (γ) to 0.05. Multi-season data were generated by simulating the Markovian processes of colonization and extinction over a specified number of seasons. For each site that was occupied at season t, there was a probability ε (the local rate of extinction) that it would become unoccupied by season $t + 1$; sites unoccupied in season t had a probability γ (the local rate of colonization) of becoming colonized by season $t + 1$. One hundred datasets were generated conforming to a population with the specified parameter values over 20 years. For each simulated population, estimates of the local rates of colonization were obtained using program PRESENCE every time the population would be sampled by the proposed LTEMP rotating panel design. Bowser (2009) provided a more detailed explanation of the methods used for generating datasets and fitting occupancy models using program PRESENCE, including the R scripts used.

Table 1: Families included in analyses.

Acanthosomatidae	Culicidae	Nabidae
Achilidae	Curculionidae	Otitidae
Acrididae	Delphacidae	Phlaeothripidae
Agromyzidae	Diapriidae	Phoridae
Anisopodidae	Dolichopodidae	Pipunculidae
Anobiidae	Drosophilidae	Platygasteridae
Anthocoridae	Dryinidae	Psilidae
Anthomyiidae	Dryomyzidae	Psyllidae
Anthomyzidae	Elateridae	Pteromalidae
Aphelinidae	Empididae	Pythidae
Aphididae	Encyrtidae	Rhagionidae
Apidae	Entomobryidae	Scarabaeidae
Argidae	Ephydridae	Scathophagidae
Asteiidae	Eulophidae	Scelionidae
Bethylidae	Eurytomidae	Sciomyzidae
Bibionidae	Formicidae	Scirtidae
Braconidae	Heleomyzidae	Sclerosomatidae
Cantharidae	Hemerobiidae	Sepsidae
Carabidae	Hypogastruridae	Simuliidae
Ceraphronidae	Ichneumonidae	Sminthuridae
Chamaemyiidae	Isotomidae	Sphaeroceridae
Chloroperlidae	Lathridiidae	Staphylinidae
Chloropidae	Lauxaniidae	Stratiomyidae
Chrysomelidae	Leiodidae	Syrphidae
Chrysopidae	Lithobiidae	Tabanidae
Cicadellidae	Lycidae	Tachinidae
Clusiidae	Lygaeidae	Tenthredinidae
Coccinellidae	Micropezidae	Tephritidae
Coenagrionidae	Miridae	Thripidae
Coniopterygidae	Muscidae	Torymidae

Throughout the simulation analyses, program PRESENCE often yielded fitted values of zero or one for at least one of the parameters estimated. When this occurred, it appeared as if one parameter was fixed at either zero or one while fitting was performed on the remaining variable(s), leading to erroneous results. In these situations, program PRESENCE issuing the warning, “numerical convergence was not reached,” in its output. The failure rate of a set of simulations was calculated by dividing the number of failed simulations by the total number of simulations and was expressed as a percentage. All other

summaries presented below exclude results from simulations where PRESENCE failed to converge, summarizing only simulations where PRESENCE converged successfully.

Random Forest Regression of Arthropod Family Richness

In order to create a continuous raster of arthropod family richness over the KENWR, predictions were made by random forest regression, a machine learning algorithm with high predictive accuracy (Breiman, 2001). For this exercise, taxonomic breadth was kept as broad as possible, although some groups were excluded due to practical constraints. Ninety families (Table 1) were included in analyses. Of the 255 sweep net samples available, two were rejected for the purposes of this regression analysis. In one case, a logistical error had caused spatial misalignment of the arthropod sample with other data and in another case, an avalanche dramatically altered a plot immediately before arthropods were sampled.

Table 2: Variables included in Random Forest regression.

Variable Name	Number of Variables	Description
Spatial		
latitude	1	
longitude	1	
Topographic		
elevation	1	Elevation from the Digital Elevation Model (DEM).
slope	1	Derived from DEM.
aspect	1	Derived from DEM.
curvature	1	Derived from DEM.
distance_ocean	1	Distance to the ocean.
Climate		
precipitation	13	Annual (precipitation_annual) and monthly (precipitation01-12) precipitation from the PRISM model.
temperature	13	Annual (temperature_annual) and monthly (temperature01-12) temperature from the PRISM model.
accumulation	1	Accumulated surface runoff based on PRISM annual precipitation data and DEM.
Vegetation		
NDVI	1	Normalized difference Vegetation Index calculated from 2002 LandSat 7 imagery.
land cover	1	Vegetation cover classes (Figure 2).
Temporal		
day	1	Julian day.
hour	1	Hours since midnight.
Historic		
years_post_fire	1	Years since last fire.
Total	39	

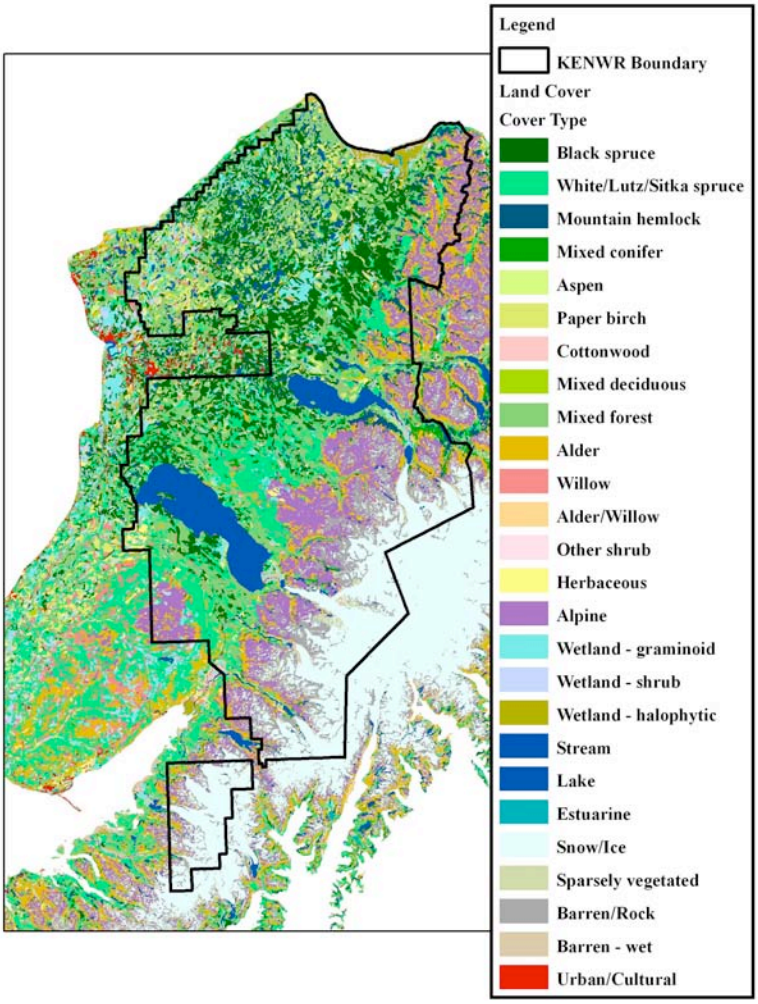


Figure 2: Kenai Peninsula land cover classification.

Extraction of Covariates: A 100 m × 100 m resolution prediction grid was imposed over the KENWR. Values of covariates that were available as raster datasets were extracted by resampling (bilinear interpolation) using ArcMap (Table 2). Covariates available as vector datasets were converted to raster datasets conforming to the prediction grid. For all covariates, values from pixels in which LTEMP sampling locations fell were extracted for use in fitting of random forest regression models.

Elevation data were extracted from a USGS digital elevation model (DEM) resampled from 30 m × 60 m to 30 m × 30 m resolution. Topographic variables were calculated from the DEM. The normalized difference vegetation index (NDVI) was calculated from a 30m × 30m resolution mosaic of LandSat 7 imagery taken in 2002 and made available through the Multi-Resolution Land Characteristics Consortium. Climate parameters were extracted from the 2 km resolution PRISM (Parameter-elevation Regressions on Independent Slopes

Model) raster datasets produced by the Spatial Climate Analysis Service at Oregon State University (SCAS/OSU). This dataset is currently the best climate coverage available for Alaska (Simpson et al. 2005). Vegetation types were extracted from a recent vegetation classification of the Kenai Peninsula (KENWR, unpublished data).

Random forest regressions were fitted using the randomForest package (Liaw and Wiener 2002) for R. For all random forest regressions, 5,000 trees were built. To select an optimal value of the 'mtry' parameter (the number of predictors randomly selected for consideration at each node), we ran random forest regressions for all values of mtry between 1 and 50 and selected the value of mtry that yielded the highest value of pseudo-R-squared and the lowest mean squared error. A random forest regression model was then fitted using this optimal value of the mtry parameter. We used this model to make predictions at all 100 m × 100 m pixels over the KENWR. For prediction, the time was set at median observed values from observed LTEMP data (June 18 at 8:00 am).

Results

Material Collected

The 255 sweep net samples yielded a total of 15,136 specimens, of which 9,961 were of the 90 families included in analyses. Over half of the arthropods collected (56%) were Diptera. Hemiptera (26%), Collembola (11%), and Hymenoptera (7%) also comprised substantial fractions of the specimens considered.

The abundance and frequency of specimens from each family varied greatly. Culicidae, with a total of 3,697 individuals collected and a frequency of 0.76 (i.e., collected at 76% of sites), was the most abundant and frequently collected family. Aphididae, Sminthuridae, Cicadellidae, Muscidae, Delphacidae, Empididae, Ichneumonidae, Simuliidae, Braconidae, Lauxaniidae, Phoridae, Biobionidae, Anthomyiidae, and Cantharidae were also relatively abundant and frequently collected. Ephydridae were abundant locally, one site on the margin of Chickaloon Flats yielding 133 of the 135 Ephydrid specimens collected, but they were generally infrequent, collected at only three sites. In contrast, Diapriidae were relatively common but were usually represented by few individuals at each site. Many taxa were rarely encountered and were represented by few specimens. Seventeen families (Isotomidae, Chloroperlidae, Achilidae, Nabidae, Phlaeothripidae, Anobiidae, Carabidae, Lathridiidae, Lycidae, Pythidae, Scarabaeidae, Scirtidae, Clusiidae, Dryomyzidae, Sepsidae, Aphelinidae, and Bethylidae) were represented by singletons.

Monte-Carlo Simulations

Short-Term Results: Using data from sampling of a single panel (a single sampling season), program PRESENCE had a high failure rate, especially when detection probability was low (Figure 3). The average failure rate was 24 failures out of every 100 simulations, with a high of 98 failures out of 100 simulations when both occupancy (Ψ) and detection probability (*p*) were 0.1. Program PRESENCE was most reliable when Ψ was between 0.3 and 0.7 and *p* was between 0.6 and 0.8.

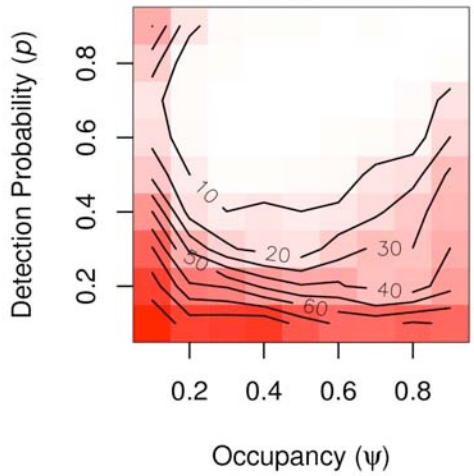
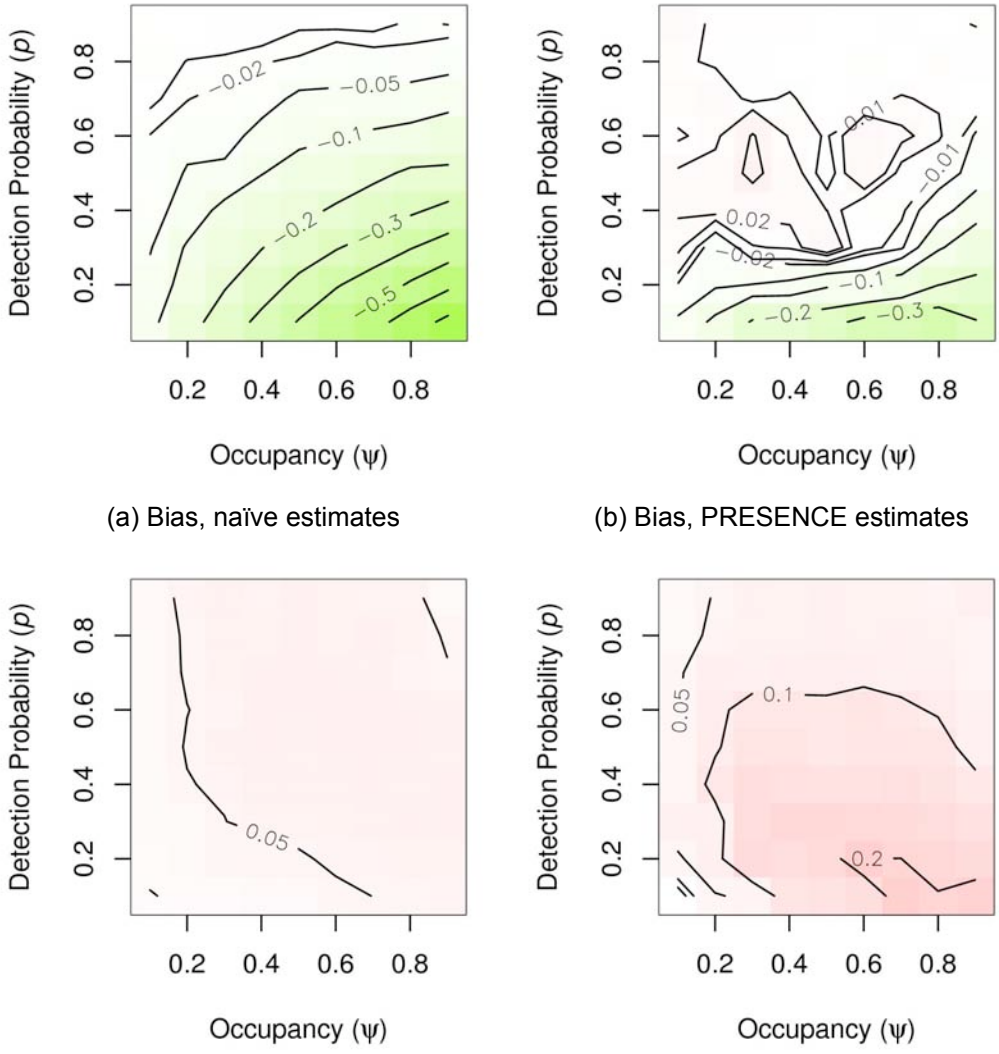


Figure 3: Failure rate of Program PRESENCE over a range of values of Ψ and *p* for the rotating panel design. Values are the percentage of simulations in which PRESENCE failed. Failures were defined as fitting of either Ψ or *p* with a value of 1 or 0. The colors of the pixels are graduated so that white represents a value of 0 and saturated red represents a value of 100%.

Naïve estimates of occupancy were nearly always biased low, with the magnitude of the bias increasing as *p* decreased and as Ψ increased (Figure 4(a)), with a maximum magnitude of average bias of -0.725. Only when detection probability was highest (0.9) were naïve estimates essentially unbiased. Program PRESENCE nearly always yielded less biased estimates of occupancy than the naïve estimates when it did not have convergence failures (Figure 4(b)). It produced generally unbiased estimates of occupancy as long as detection probability was greater than 0.3; when p was less than 0.3, estimates of Ψ tended to be biased low. The magnitude of bias was greatest when *p* was at its minimum value (0.1). Near this extreme of the parameter space, estimates of Ψ were highly variable.

Naïve estimates of occupancy were more precise, even if less accurate, than estimates obtained by program PRESENCE (Figure 4(c) and (d)). Standard deviation of naïve estimates was generally less than 0.05; standard deviation of estimates obtained from program PRESENCE were generally greater than 0.1 when detection probability was less than about 0.6.



(a) Bias, naïve estimates

(b) Bias, PRESENCE estimates

(c) Standard deviation, naïve estimates

(d) Standard deviation, PRESENCE estimates

Figure 4: Bias and standard deviation of naïve estimates and PRESENCE estimates of occupancy over a range of values of occupancy (Ψ) and detection probability (p) for a single-season estimate (i.e., a survey of one panel). The colors of the pixels are graduated so that white represents a value of 0, saturated red represents a value of 1, and saturated green represents a value of -1.

Long-Term Results: In these simulations, the rotating panel design did not yield reasonable estimates of local rates of colonization and extinction until some sites were visited twice, ten years into the sampling program. After this point, estimates of these change parameters were generally quite good. Estimates of occupancy were consistently unbiased (mean bias = 0.004), but not quite as precise as the desired maximum standard deviation of about 0.05 (SD = 0.062). Over time, the failure rate of program PRESENCE first rose dramatically from 0 at the second year to 67% at the fourth year, then remained high until the tenth year, when some sites were sampled a second time (Figure 5). After this point, the failure rate became negligible. Estimates of the change parameters ε and γ were initially biased high (Figures 6 and 7). Mean bias at the second year was 0.25 for ε and 0.61 for γ. This bias diminished slowly until the tenth year, when bias dropped quickly as some sites were sampled a second time. By the twelfth year and thereafter, estimates of these change parameters were essentially unbiased. After year twelve, precision was still not as good as is desirable, with standard deviations of about 0.1, but this large standard deviation was mostly due to a small number (2-9) of simulations each season where estimates of program PRESENCE were biased extremely high. When these outliers were removed, standard deviation of estimates of ε and γ were less than 0.02.

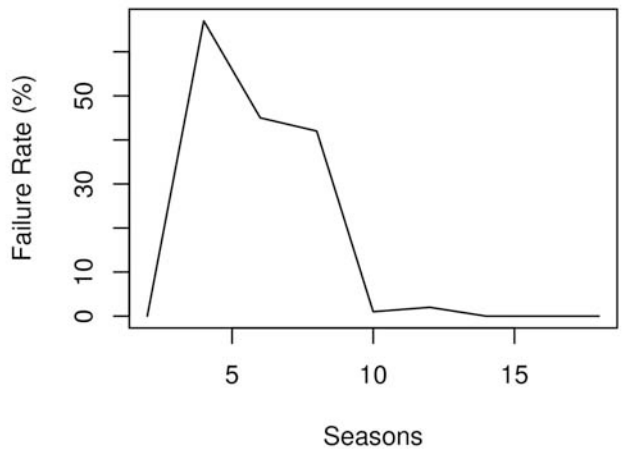


Figure 5: Failure rate of program PRESENCE over 20 years of simulations. Values are the percentage of simulations in which PRESENCE failed.

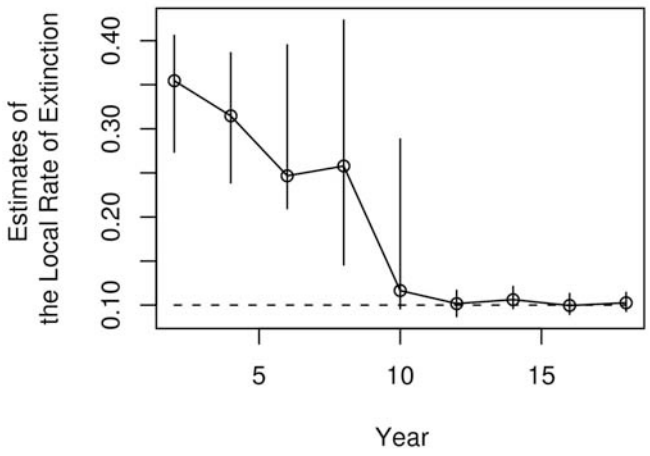


Figure 6: Estimates of the local rate of extinction over 20 years of simulations. Dashed line: true value (0.1). Circles and solid line: median of estimates with bars spanning from 25% to 75% quantiles.

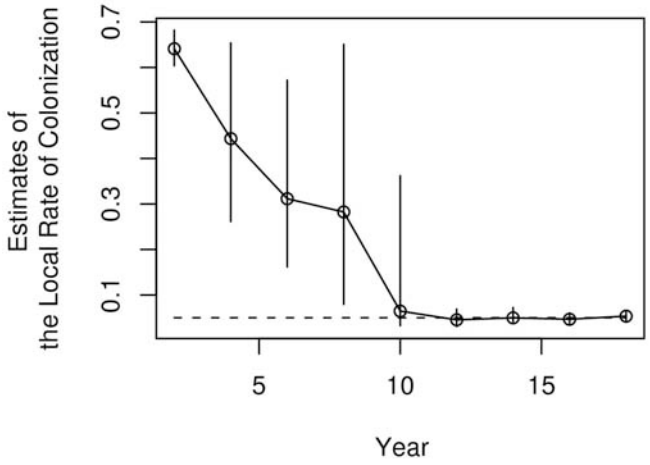


Figure 7: Estimates of the local rate of colonization over 20 years of simulations. Dashed line: true value (0.05). Circles and solid line: median of estimates with bars spanning from 25% to 75% quantiles.

Regression Results

Random forest regressions incorporating only information that was available as continuous raster data explained only 22% of the variation of family richness, but produced a reasonable map of predicted family richness (Figure 8). The land cover classification was the variable most often used by the decision trees for predicting arthropod diversity, followed by NDVI. The regression predicted lowest arthropod family richness at barren alpine sites. Highest richness was predicted in coastal wetlands and productive hardwood and mixed forests the KENWR (Table 3).

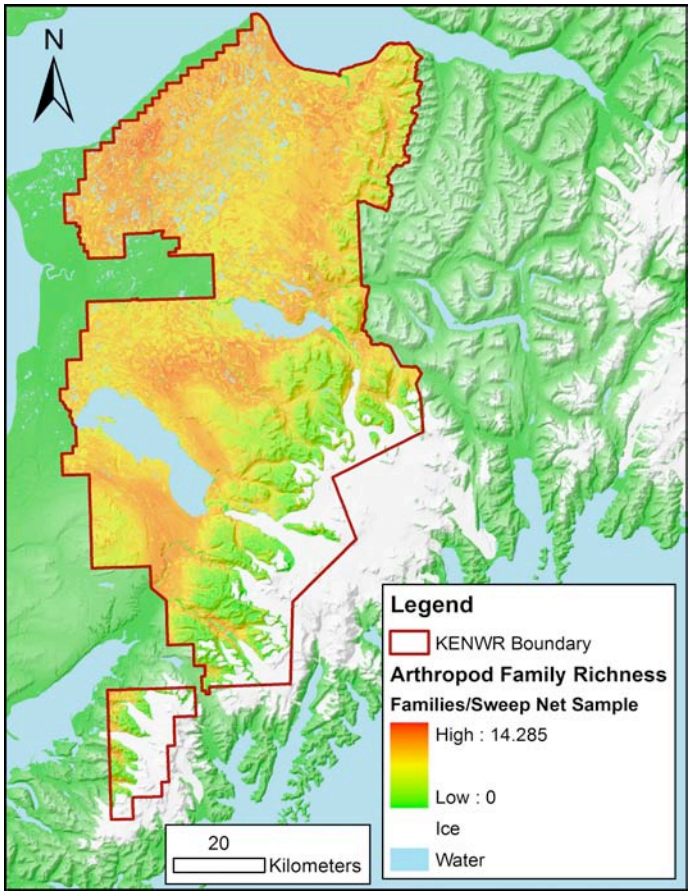


Figure 8: Expected family richness of terrestrial arthropods predicted by random forest regression.

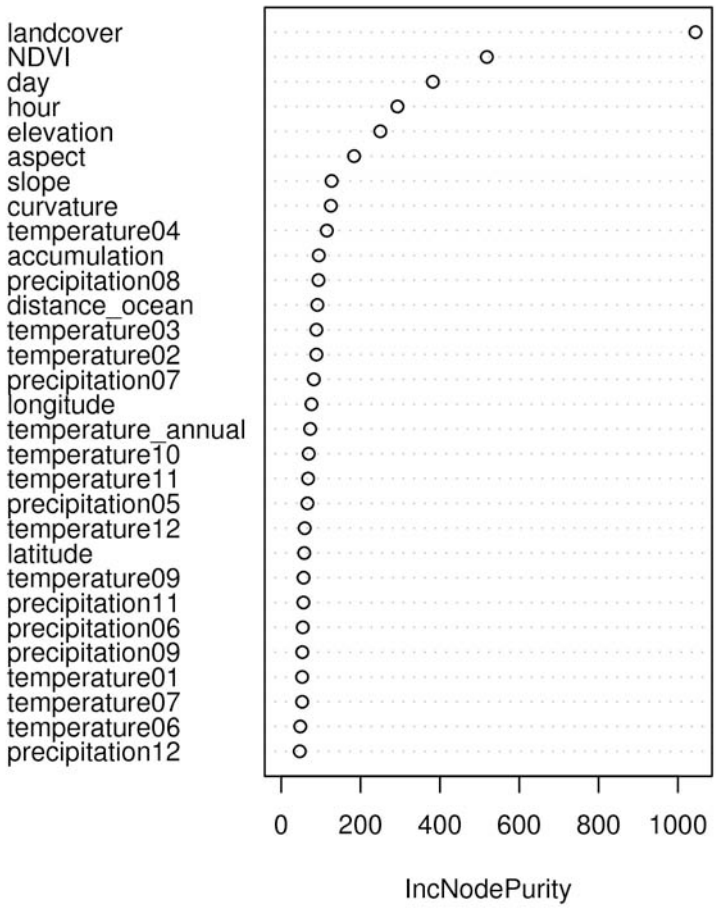


Figure 9: Importance of variables included in final random forest regression model. Variable names were defined in Table 2. IncNodePurity is the mean decrease in node purity, also known as the Gini splitting criterion (Breiman 2002).

Discussion

Utility of LTEMP for Monitoring Distributions

The proposed field methods of taking two spatial subsamples at each plot will allow for efficient monitoring of arthropod distributions on the KENWR while explicitly accounting for imperfect detection. Because precision was poor when detection probability was less than 0.6, the proposed methods would be most useful for species with detection probabilities of about 0.6 or greater. Although sampling of a single panel (51 sites) did not provide a large enough sample size

for modeling species distributions with good precision, the frequent sampling of this design would provide up-to-date information on a frequent basis while allowing a larger overall sampling frame than would be possible by sampling the same sites every season. Over the long term, the proposed methods would allow accurate estimation of species distributions and local rates of extinction and colonization.

Table 3: Predicted arthropod family richness by land cover type.

Land Cover Type	Mean	S.D.
Wetland - halophytic	9.03	0.90
Black cottonwood	8.88	1.17
Mixed forest	8.31	1.19
Paper birch	8.24	1.13
Wetland - shrub	8.15	1.44
Alder/Willow	8.05	1.54
Mixed deciduous	8.01	1.23
White/Lutz/Sitka spruce	7.99	1.33
Black spruce	7.79	1.09
Wetland - graminoid	7.76	1.24
Herbaceous	7.68	1.30
Urban/Cultural	7.54	1.34
Aspen	7.51	1.00
Willow	7.43	1.86
Mixed conifer	6.57	1.38
Alder	6.26	1.41
Other shrub	6.18	1.84
Mountain hemlock	6.06	1.16
Barren - wet	5.70	2.40
Alpine	4.76	1.26
Sparsely vegetated	3.69	1.48
Barren/Rock	3.33	0.85

In order to monitor distributions of arthropod less likely to be collected in individual sweep net samples (i.e., with detection probabilities < ~0.6), more than two sweep net samples should be taken at each site. Increasing the number of surveys at each site is generally the most efficient way to improve precision of occupancy estimates, especially for species with low detection probabilities (MacKenzie and Royle 2005). In the case of the proposed LTEMP design, where sample size is already large (~254 sites) and the number of samples taken at each site is at its minimum value (2), increasing the number of sweep net samples is certainly the best way to improve precision and enable monitoring of a larger

number of arthropod species. Since the sweep net sampling method yielded information on relative abundance of arthropod species, estimates of occupancy metrics may also be improved by incorporating this information as was done by Royle and Nichols (2003).

The reason that rates of local colonization and extinction could not be estimated well over the first eight years of simulations was that the rotating panel design provided little information pertaining to these parameters until some sites were sampled a second time, ten years into the proposed sampling program. Precision and accuracy of estimates of these change parameters would best be improved by addition of a number of sites that would be sampled in consecutive seasons to furnish information about changes in occupancy states over the short-term.

Patterns of Diversity

The good spatial coverage of the LTEMP sampling framework provided ample data for modeling arthropod family richness. This exercise yielded a map of coarse, landscape-scale patterns of arthropod diversity that will serve as a baseline for comparison with future arthropod diversity on the KENWR. It is also useful for putting small-area samples of arthropod diversity in the context of the KENWR as a whole.

Bowser (2009), using the same arthropod diversity data from LTEMP, found that temporal variable, productivity, and climate were the most important determinants of arthropod diversity on the KENWR. His analysis and the current analysis by random forest were consistent with McCoy (1990) in that, at about 60° latitude, maximum arthropod diversity occurred near sea level. The margins of coastal salt marshes and productive deciduous and mixed forest habitat types where random forest regression predicted highest arthropod diversity were relatively warm, productive sites; the alpine and barren habitat types where lowest arthropod diversity was predicted are relatively cold, unproductive sites. This apparent positive relationship between arthropod diversity and vegetative productivity is consistent with the species-energy theory of Wright (1983). He posited that, as the amount of energy moving through a system increases, the population sizes of living things supported by that system increase, reducing extinction rates.

Conclusions

The excellent spatial representation of the LTEMP sampling framework provided ample spatial data for modeling over the landscape of the KENWR. By correcting for imperfect detection in the future, we will be poised to monitor for changes in the distributions of species over time. Although monitoring was mostly

discussed in the context of the responses of species to climate, these same monitoring methods should serve well for monitoring the spread of exotic species and the subsequent responses of native biota in terms of distribution shifts.

Acknowledgments

This work would not have been possible without collaboration with the FIA program. Pat Doak and Diane Wagner (University of Alaska Fairbanks) greatly improved the quality of this work through their advice. Marathon Oil Corporation provided additional funding for this project. Charlotte Hockin (US Fish & Wildlife Service) performed a large part of the rough sorting of arthropod specimens. Todd Eskelin (US Fish & Wildlife Service) helped in various ways. Many experts helped with identifications.

Literature Cited

Bechtel, T. J.; Copeland, B. J. 1970. Fish species diversity indices as indicators of pollution in Galveston Bay, Texas. Contributions in Marine Science. 15:103-32.

Boos, Dennis D.; Brooks, Atina D.; Nychka, Douglas. 2006. Rlab: Functions and Datasets Required for ST370 class. R package version 1.3. [Online]. Available: http://www.r-project.org [August 25, 2008].

Boucher, Tina V.; Mead, Bert R. 2006. Vegetation change and forest regeneration on the Kenai Peninsula, Alaska following a spruce beetle outbreak, 1987-2000. Forest Ecology and Management. 227:233-246.

Bowser, Matthew L. 2009. Terrestrial arthropod biodiversity on the Kenai National Wildlife Refuge, Alaska. M.Sc. thesis. Fairbanks, AK: University of Alaska Fairbanks. 141p.

Breiman, Leo. 2001. Random Forests. Machine Learning. 45:5-32.

Burkman, Bill. 2005. Forest Inventory and Analysis: Sampling and Plot Design. FIA Fact Sheet Series. USDA Forest Service. [Online]. Available: http://fia.fs.fed.us/library/fact-sheets/data-collections/Sampling%20and%20Plot%20Design.pdf [March 26, 2008].

Chapin, III, F. Stuart; Zavaleta, Erika S.; Eviner, Valerie T.; Naylor, Rosamond L.; Vitousek, Peter M.; Reynolds, Heather L.; Hooper, David U.; Lavorel, Sandra; Sala, Osvaldo E.; Hobbie, Sarah E.; Mack, Michelle C.; Díaz, Sandra. 2000. Consequences of changing biodiversity. Nature. 405:234-242.

Egloff, D. A.; Brakel, W. H. 1973. Stream pollution and a simplified diversity index. Journal Water Pollution Control Federation. 45:2269-2275.

Gottfried, Michael; Pauli, Harald; Reiter, Karl; Grabherr, Georg. 1999. A fine-scaled predictive model for changes in species distribution patterns of high mountain plants induced by climate warming. Diversity and Distributions. 5:241-251.

Guisan, Antoine; Thuiller, Wilfried. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters. 8:993-1009.

Hannah, L.; Midgley, G.; Millar, D. 2002. Climate change-integrated conservation strategies. Global Ecology & Biogeography. 11:485-495.

Hines, James E. Program PRESENCE. [Online]. Available: http://www.mbr-pwrc.usgs.gov/software/presence.html [August 25, 2008].

Hooper, D. U.; Chapin, III, S. F.; Ewel, J. J.; Hector, A.; Inchausti, P.; Lavorel, S.; Lawton, J. H.; Lodge, D. M.; Loreau, M.; Naeem, S.; Schmid, B.; Stetälä, H.; Symstad, A. J.; Vandermeer, J.; Wardle, D. A. 2005. Effects of biodiversity on ecosystem funcitoning: a consensus of current knowledge. Ecological Monographs. 75:3-35.

Hopkin, Stephen P. 1997. Biology of the Springtails (Insecta: Collembola). New York: Oxford University Press. 330p.

Karr, James R.; Kimberling, Diana N. 2003. A terrestrial arthropod index of biological integrity for shrubb-steppe landscapes. Northwest Science. 77:202-213.

Kremen, C.; Colwell, R. K.; Erwin, T. L.; Murphy, D. D.; Noss, R. F. & Sanjayan, M. A. 1993. Terrestrial arthropod assemblages: their use in conservation planning. Conservation Biology. 7:796-808.

Liaw, Andy; Wiener, Matthew. 2002. Classification and regression by randomForest. R News. 2:18-22.

MacKenzie, Darryl I.; Nichols, James D.; Hines, James E.; Knutson, Melinda G.; Franklin, Alan B. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology. 84:2200-2207.

MacKenzie, Darryl. I.; Nichols, James D.; Royle, J. Andrew; Pollock, Kenneth H.; Bailey, Larissa L.; Hines, James E. 2006. Occupancy estimation and modeling. New York: Elsevier. 324p.

MacKenzie, Darryl. I.; Royle, J. Andrew 2005. Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology. 42:1105-1114.

Magness, Dawn R.; Huettmann, Falk; Morton, John M. 2008. Using random forests to provide predicted species distribution maps as a metric for ecological inventory & monitoring programs. pp. 209-229 in Studies in Computational Intelligence 122: Applications of Computational Intelligence. Springer Berlin. 424p.

Magurran, Anne E. 1988. Ecological diversity and its measurement. Princeton, NJ: Princeton University Press. 179p.

McCann, Kevin S. 2000. The diversity-stability debate. Nature. 405:228-233.

McCoy, Earl D. 1990. The distribution of insects along elevational gradients. Oikos. 58:313-332.

Morton, John M.; Bowser, Matthew; Berg, Edward.; Magness, Dawn; Eskelin, Todd. 2009. Long Term Ecological Monitoring Program on the Kenai National Wildlife Refuge, Alaska: An FIA Adjunct Inventory. in McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2008. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of agriculture, Forest Service, Rocky Mountain Research Station. [CD-ROM].

Parmesan, Camille. 1996. Climate and species' range. Nature. 382:765-766.

Parmesan, Camille. 2006. Ecological and evolutionary responses to recent climate change. Annual Review of Ecology and Systematics. 37:637-669.

Parmesan, Camille.; Ryrholm, Nils; Stefanescu, Constantí; Hill, Jane K.; Thomas, Chris D.; Descimon, Henri; Huntley, Brian; Kaila, Lauri; Kullberg, Jaakko; Tammaru, Toomas; Tennent, W. John; Thomas, Jeremy A.; Warren, Martin. 1999. Poleward shifts in geographic ranges of butterfly species associated with regional warming. Nature. 399:579-583.

Parmesan, Camille; Yohe, Gary. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature. 421:37-42.

R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. [Online]. Available: http://www.R-project.org [Febrary 29, 2008].

Royle, J. Andrew; Nichols, James D. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology. 84:777-790.

Roth, Dana S.; Perfecto, Ivette; Rathcke, Beverly. 1994. The effects of management systems on ground-foraging ant diversity in Costa Rica. Ecological Applications. 4:423-436.

Schowalter, Timothy D. 2006. Insect Ecology: An Ecosystem Approach. 2nd ed. London: Academic Press. 576p.

Simpson, James J.; Hufford, Gary L.; Daly, Christopher; Berg, Jared S.; Fleming, Michael D. 2005. Comparing maps of mean monthly surface temperature and precipitation for Alaska and adjacent areas of Canada produced by two different methods. Arctic. 58:137-161.

Snyder, William E.; Evans, Edward W. 2006. Ecological effects of invasive arthropod generalist predators Annual Review of Ecology, Evolution, and Systematics. 37:95-122.

Tepedino, J. V. 1979. The importance of bees and other insect pollinators in maintaining floral species composition. Great Basin Naturalist Memoirs.3:139-150.

Wallwork, John A. 1983. Oribatids in forest ecosystems. Annual Review of Entomology. 28:109-130.

Walther, Gian-Reto; Post, Eric; Convey, Peter; Menzel, Annette; Parmesan, Camille; Beebee, Trevor J. C.; Fromentin, Jean-Marc; Hoegh-Guldberg, Ove; Bairlein, Franz. 2002. Ecological responses to recent climate change. Nature. 416:389-395.

Williams, P.; Hannah, L.; Andelman, S.; Midgley, G.; Araújo, M.; Hughes, G.; Manne, L.; Martinez-Meyer, E.; Pearson, R. 2005. Planning for climate change: identifying minimum-dispersal corridors for the Cape Proteaceae. Conservation Biology. 19:1063-1074.

Wilson, Edward O. 1987. The little things that run the world (the importance and conservation of invertebrates). Conservation Biology. 1:344-346.

Wilson, Robert J.; Gutiérrez, David; Gutiérrez, Javier; Martínez, David; Agudo, Rosa; Monserrat, Víctor J. 2005. Changes to the elevational limits and extent of species ranges associated with climate change. Ecology Letters. 8:1138-1146.

Wright, David H. 1983. Species-energy: an extension of species-area theory. Oikos. 41:496-506.

Wu, Rudolf S. S. 1982. Periodic defaunation and recovery in a subtropical epibenthic community, in relation to organic pollution. Journal of experimental marine biology and ecology. 64:253-269.

Fire, Disturbance, and Other Remote Sensing Applications to Support Large-Scale Forest Inventories II



Tree Crown Condition in Virginia Before and After Hurricane Isabel (September 2003)

KaDonna Randolph[1]
Anita Rose[2]

***Abstract**: In September 2003, Hurricane Isabel made landfall in North Carolina as a Category 2 hurricane. As it moved inland, with sustained wind speeds of 37 to 69 miles per hour (59 to 111 km per hour) and gusts up to 91 miles per hour (146 km per hour), the hurricane caused widespread damage throughout Virginia and is a plausible explanation for adverse changes observed in tree crown condition since the hurricane. On average, trees measured before and after the hurricane showed a significant increase in foliage transparency. Increases in foliage transparency were greatest for loblolly pine, sweetgum, and the Coastal Plain region of the State. A significant correlation between tree size and increase in foliage transparency was not observed. This study highlighted the potential importance of foliage transparency as an indicator of tree damage from severe storms.*

Keywords: FIA, foliage transparency, forest health, hurricane damage.

Introduction

The U.S. Forest Service, Forest Inventory and Analysis (FIA) program has been conducting inventories of the Nation's forest land for 80 years. For most of this period, statewide inventories were completed approximately once every 6 to 8 years in the South and 11 to 18 years in the rest of the country (Gillespie 1999). Since the late 1990s, however, FIA has inventoried States on an annual basis, striving to complete an inventory cycle every 5 years in the East and every 10 years in the West.

The first set of such annual measurements began in Virginia in 1997 and concluded in 2001. This was the 7th statewide inventory of Virginia. The 8th statewide survey, also accomplished on an annual basis, was conducted during 2002 to 2007. As part of the 2001 (7th survey) report on the forests of Virginia, crown conditions were evaluated to identify potential forest health problems within the State (Rose 2007). Average crown density, foliage transparency, and

[1] United States Forest Service; Southern Research Station; Forest Inventory and Analysis; 4700 Old Kingston Pike; Knoxville, TN 37919 USA; krandolph@fs.fed.us

[2] United States Forest Service; Southern Research Station; Forest Inventory and Analysis; 4700 Old Kingston Pike; Knoxville, TN 37919 USA; anitarose@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

crown dieback were calculated for the softwood and hardwood species groups, as well as for individual tree species. When crown conditions from the 7th survey were compared to the forthcoming 2007 (8th survey) report (Rose in preparation), several changes were noticed. As expected, average values fluctuated for all three crown variables; however, the change for foliage transparency, especially among the softwoods, seemed atypical. Across the State, average hardwood transparency increased from 20 percent to 23 percent and average softwood transparency increased from 22 percent to 29 percent. Several individual species also showed large increases in average foliage transparency between the two surveys. Loblolly pine (*Pinus taeda* L.) average foliage transparency increased from 19 percent to 30 percent and the average for sweetgum (*Liquidambar styraciflua* L.) increased from 17 percent to 25 percent. The levels of foliage transparency measured during the 8th survey appeared unusual because the majority of both hardwood and softwood trees in the Southern United States typically have foliage transparencies of 25 percent or less (Randolph 2006).

Because hurricanes can cause substantial damage to tree crowns (Putz and Sharitz 1991), we considered the hurricane that struck Virginia during the 8th survey as a likely explanation for the changes noticed in crown condition. In September 2003 Hurricane Isabel made landfall on the Outer Banks of North Carolina as a Category 2 hurricane (figure 1). The storm caused widespread wind and flood damage across eight States from North Carolina to New York. Isabel

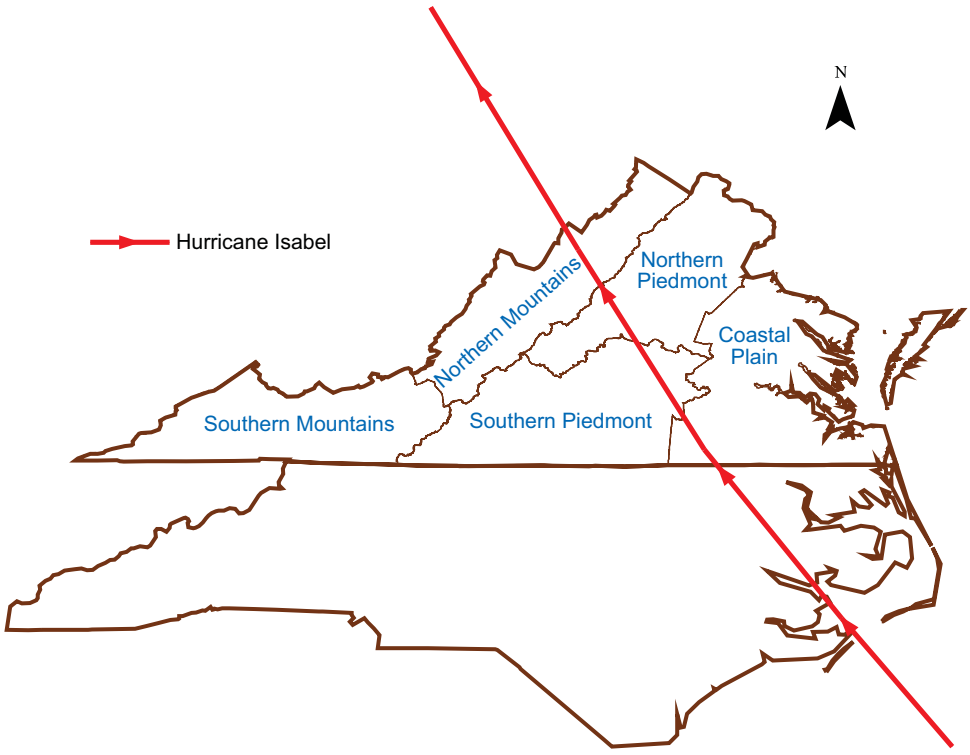


Figure 1: Path of Hurricane Isabel across North Carolina and Virginia, and the location of the five FIA units in Virginia.

passed through Virginia with sustained wind speeds ranging from 37 to 69 miles per hour (59 to 111 km per hour), with gusts up to 91 miles per hour (146 km per hour) (Beven and Cobb 2004). Rainfall from the hurricane averaged 4 to 7 inches (10 to 18 cm) over large portions of east-central Virginia, with the Shenandoah Valley in northern Virginia averaging 8 to 12 inches (20 to 30 cm) (Beven and Cobb 2004). A total of 77 counties and independent cities across Virginia were declared disaster zones, and the estimated economic loss was $925 million, greater than any of the other States through which Isabel passed (U.S. Department of Commerce 2004).

The purpose of this study was to examine trees that were measured both before and after the hurricane and explore the possibility that Hurricane Isabel was a contributing factor to the change in foliage transparency between the 7th and 8th surveys.

Methods

Data

Data for this study came from the FIA phase 3 plots measured in Virginia between 1997 and 2001 (7th survey) and between 2002 and 2007 (8th survey). FIA phase 3 plots are a cluster of four 1/24-acre (0.02 ha) circular subplots located across the landscape in a way such that each plot represents approximately 96,000 acres (38,850 ha) (McRoberts 2005). On each plot crown condition is assessed for every live tree ≥ 5.0 inches (12.7 cm) diameter at breast height (d.b.h.). Three crown condition variables are assessed: crown dieback, crown density, and foliage transparency. Crown dieback is the recent mortality of branches with fine twigs, whereas crown density and foliage transparency measure the amount of foliage and crown biomass on a tree (Schomaker and others 2007). Though foliage transparency and crown density are similar measures they cannot be interpreted as exact inverses. Crown density measures the amount of sunlight blocked by all biomass produced by the tree (both live and dead) in the crown, whereas foliage transparency measures the amount of sunlight penetrating only the live portion of the crown. Deductions are made from the maximum possible crown density for spaces between branches and other large openings in the crown. However, large gaps in the crown where foliage is not expected to occur are excluded from consideration when foliage transparency is rated.

Analysis

Individual trees from the 7th and 8th surveys were matched by plot and tree number and then divided into pre- and post-hurricane data sets. Trees measured before September 18, 2003 were assigned to the pre-hurricane data set and trees measured on or after September 18, 2003 were assigned to the post-hurricane data set. The mean, standard error, and 25th, 50th, 75th, and 90th percentiles were calculated for both the pre- and post-hurricane data sets. Change in crown condition was calculated for survivor trees as the difference between the post- and

pre-hurricane ratings. Trees that died between the first and second assessment were not included in this analysis. In this report we focus on the change in foliage transparency (δ_{ft}).

Paired t-tests were used to test the hypothesis that the mean of δ_{ft} equals zero at the State and FIA unit level for all trees combined, hardwood and softwood groups, and individual species groups. Scatter plots of δ_{ft} by d.b.h. and height, and average δ_{ft} ($\bar{\delta}_{ft}$) by crown position were produced to examine the relationship between δ_{ft} and tree dominance. In addition, $\bar{\delta}_{ft}$ based on only the remeasured trees was calculated for each plot with three or more live trees and mapped for visual inspection in relation to the path of Hurricane Isabel (U.S. Department of Commerce 2008).

Results

Pre- and post-hurricane assessments were matched for 1,492 live trees ≥ 5.0 inches (12.7 cm) d.b.h. on 74 plots. Hardwoods made up the majority of the trees assessed (67.9 percent). Chestnut oak (*Quercus prinus* L.), yellow-poplar (*Liriodendron tulipifera* L.), and red maple (*Acer rubrum* L.) were the most abundant hardwood species. Loblolly pine and Virginia pine (*Pinus virginiana* Mill.) were the most abundant softwood species. The time between the pre- and post-hurricane assessments was 4 years for one plot and 5 years for all other plots.

Across the State, there was a significant ($\alpha = 0.05$) increase in average foliage transparency pre- to post-hurricane for all trees combined, the hardwood and softwood groups, and several of the individual species groups (table 1). The increase in foliage transparency was greater among the softwoods ($\bar{\delta}_{ft} = 9.6$ percent) than among the hardwoods ($\bar{\delta}_{ft} = 5.7$ percent) and overall was greatest for sweetgum ($\bar{\delta}_{ft} = 14.4$ percent) and loblolly pine ($\bar{\delta}_{ft} = 12.0$ percent).

Changes in the percentiles of the foliage transparency frequency distributions followed the same general patterns as the changes in the average conditions. Statewide, median foliage transparency was 20 percent for all trees prior to the hurricane and 25 percent after the hurricane (table 2). Greater increases in the median were observed for loblolly pine and sweetgum (table 2). Since lower foliage transparency values typically indicate healthier trees, increases in the upper percentiles especially indicate that more trees had poorer conditions after the hurricane than before. Statewide, the 90th percentile increased from 25 percent to 45 percent for all trees combined and from 30 percent to 50 percent and 25 percent to 35 percent for the softwood and hardwood groups, respectively (table 2).

The main path of Hurricane Isabel passed through four of the five FIA units in Virginia (figure 1). Therefore, it was useful to examine the changes in foliage transparency at the unit level to help determine the extent to which Hurricane Isabel may have been the cause of these changes. For all trees combined, the greatest increase in average foliage transparency was observed in the Coastal

Table 1: Mean foliage transparency and other statistics[a] for all live trees ≥ 5.0 inches d.b.h. measured in Virginia before and after Hurricane Isabel, by species group

			Pre-hurricane		Post-hurricane		t-test
Species group	**Plots**	**Trees**	**Mean**	**SE**	**Mean**	**SE**	**p-value[b]**
	- - - number - - -		*percent*		*percent*		
Softwoods							
Virginia pine	16	135	25.7	1.55	34.9	3.95	0.0356
Loblolly pine	17	259	19.9	1.03	31.9	2.75	0.0005
Shortleaf pine	8	19	20.3	—	28.2	—	—
Eastern redcedar	7	14	18.6	—	27.9	—	—
Other softwoods	8	52	25.9	1.58	25.4	0.99	—
All softwoods	38	479	22.2	1.01	31.8	2.08	0.0002
Hardwoods							
Hickory	32	95	18.7	0.79	20.3	0.67	0.1792
Maple	37	156	19.4	0.70	27.5	1.83	0.0003
Tupelo	16	35	20.4	1.94	23.7	1.13	0.1540
Oak	53	386	19.9	0.44	23.3	0.89	0.0028
Sourwood	10	32	18.4	0.87	22.3	2.56	0.2344
Sweetgum	16	65	17.0	1.44	31.4	3.02	0.0001
Yellow-poplar	30	121	16.9	0.47	25.8	1.73	0.0001
Other hardwoods	37	123	19.4	0.76	26.2	1.82	0.0002
All hardwoods	69	1,013	19.1	0.36	24.8	0.98	0.0001
All trees	74	1,492	20.1	0.41	27.1	1.16	0.0001

SE = standard error; — = not presented due to insufficient sample.

[a] The mean and SE calculations consider the cluster of trees on plots. SE not presented for groups with < 20 trees.

[b] The probability of obtaining a large t-value under the null hypothesis that the difference between the two means equals 0. T-tests are not performed for species groups with < 10 plots.

Table 2: Foliage transparency frequency distribution percentiles for live trees ≥ 5.0 inches d.b.h. measured in Virginia before and after Hurricane Isabel, by species group

	Pre-hurricane percentiles				Post-hurricane percentiles			
Species group	**25**	**50**	**75**	**90**	**25**	**50**	**75**	**90**
				percent				
Softwoods								
Virginia pine	20	25	30	35	25	30	45	55
Loblolly pine	15	20	20	25	25	30	40	45
Shortleaf pine	20	20	20	25	20	25	30	40
Eastern redcedar	15	18	20	25	20	25	30	50
Other softwoods	20	25	30	35	20	25	30	30
All softwoods	20	20	25	30	25	25	40	50
Hardwoods								
Hickory	15	20	20	25	20	20	20	25
Maple	15	20	20	25	20	25	35	40
Tupelo	15	15	20	25	20	20	25	30
Oak	15	20	20	25	20	20	25	30
Sourwood	15	20	20	25	18	20	20	40
Sweetgum	15	15	20	20	20	30	35	45
Yellow-poplar	15	15	20	20	20	25	30	40
Other hardwoods	15	20	25	25	20	25	30	35
All hardwoods	15	20	20	25	20	20	25	35
All trees	15	20	20	25	20	25	30	45

Plain ($\bar{\delta}_{ft}$ = 15.8 percent) followed by the Southern Piedmont ($\bar{\delta}_{ft}$ = 7.1 percent), Southern Mountains ($\bar{\delta}_{ft}$ = 4.3 percent), Northern Piedmont ($\bar{\delta}_{ft}$ = 3.6 percent), and Northern Mountains ($\bar{\delta}_{ft}$ = 1.4 percent) (table 3). The increase in foliage transparency was significant at the 95 percent confidence level in all units except the Northern Mountains and was marginally significant in the Southern Piedmont (p-value = 0.06). Increases in the 90th percentiles were observed in the Coastal Plain, Southern Piedmont, and Northern Mountains (table 4).

Scatter plots of δ_{ft} by pre-hurricane d.b.h. and height showed no systematic pattern (figures 2 and 3). That is, increased losses of foliage (higher δ_{ft}) were not associated with larger and taller trees. Likewise, there was not a significant

Table 3: Mean foliage transparency and other statistics[a] for live trees ≥ 5.0 inches d.b.h. measured in Virginia before and after Hurricane Isabel, by FIA unit and species group

			Pre-hurricane		Post-hurricane		t-test
Unit and species group	**Plots**	**Trees**	**Mean**	**SE**	**Mean**	**SE**	**p-value[b]**
	- - - number - - -		*percent*		*percent*		
Coastal Plain							
Hardwoods	16	165	16.5	0.68	32.8	2.63	0.0001
Softwoods	13	199	19.0	0.88	34.4	3.27	0.0004
All trees	20	364	17.9	0.63	33.6	2.26	0.0001
Southern Piedmont							
Hardwoods	16	194	19.4	0.91	24.7	2.97	0.1053
Softwoods	9	145	22.8	1.57	32.4	5.07	—
All trees	16	339	20.9	0.80	28.0	3.67	0.0646
Northern Piedmont							
Hardwoods	11	204	19.1	0.76	23.1	1.10	0.0063
Softwoods	8	66	26.4	2.59	28.7	1.34	—
All trees	12	270	20.9	1.40	24.5	1.12	0.0122
Northern Mountains							
Hardwoods	16	280	20.4	0.62	22.1	0.85	0.1757
Softwoods	5	43	25.2	1.36	24.9	1.55	—
All trees	16	323	21.0	0.51	22.4	0.80	0.2200
Southern Mountains							
Hardwoods	10	170	19.1	0.53	23.9	1.13	0.0049
Softwoods	3	26	26.3	2.80	27.7	0.40	—
All trees	10	196	20.1	0.88	24.4	0.91	0.0096
Statewide							
Hardwoods	69	1,013	19.1	0.36	24.8	0.98	0.0001
Softwoods	38	479	22.2	1.01	31.8	2.08	0.0002
All trees	74	1,492	20.1	0.41	27.1	1.16	0.0001

SE = standard error; — = not presented due to insufficient sample.

[a] The mean and SE calculations consider the cluster of trees on plots.

[b] The probability of obtaining a larger t-value under the null hypothesis that the difference between the two means equals 0. T-tests are not performed for species groups with < 10 plots.

Table 4: Foliage transparency frequency distribution percentiles for live trees ≥ 5.0 inches d.b.h. measured in Virginia before and after Hurricane Isabel, by FIA unit

	Pre-hurricane percentiles				Post-hurricane percentiles			
Unit	**25**	**50**	**75**	**90**	**25**	**50**	**75**	**90**
				percent				
Coastal Plain	15	15	20	25	25	30	40	45
Southern Piedmont	15	20	25	30	20	20	35	55
Northern Piedmont	15	20	25	30	20	25	25	30
Northern Mountains	15	20	25	25	20	20	25	30
Southern Mountains	15	20	25	30	20	25	25	30

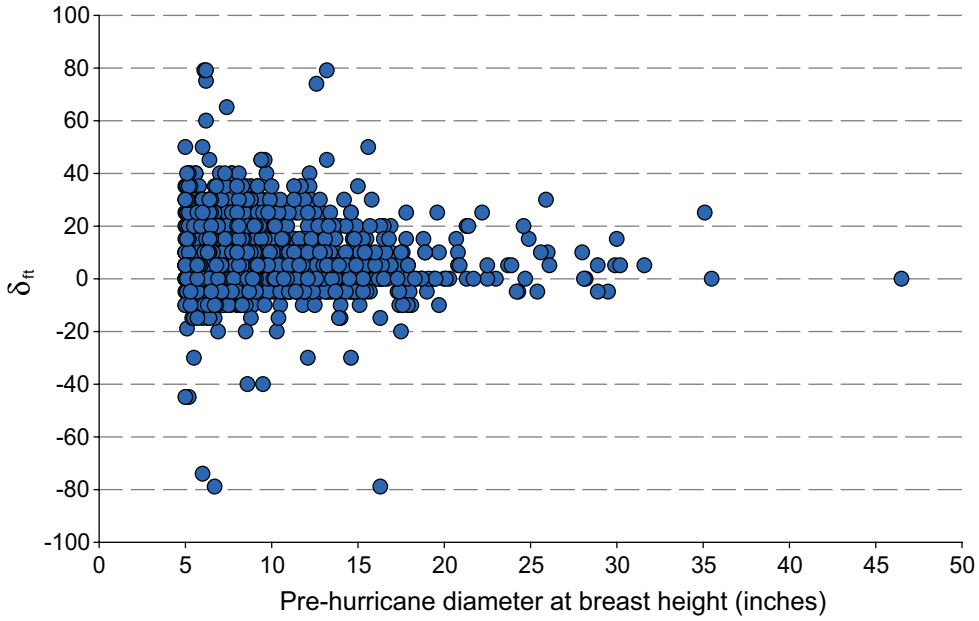


Figure 2: Change in foliage transparency for trees measured before and after Hurricane Isabel by pre-hurricane diameter at breast height.

difference in the average change in foliage transparency by crown position (figure 4). Across the State, $\overline{\delta}_f$ ranged from -6.3 percent to 31.3 percent on individual plots, with the highest $\overline{\delta}_f$ occurring on plots in the eastern half of the State (figure 5). In the Coastal Plain, $\overline{\delta}_f$ was 10 percent or more on 70 percent of the plots. In the Piedmont units, $\overline{\delta}_f$ was 10 percent or more on 18 percent of the plots. In the Mountain units only three plots (12 percent) had $\overline{\delta}_f$ of 10 percent or more.

Discussion

Even in the absence of a stressor, small changes in crown condition over time generally are expected due to the natural year-to-year variability in tree crowns. The significant increase in average foliage transparency for almost all species groups, along with the sizeable shifts in the 90th percentiles suggest that more

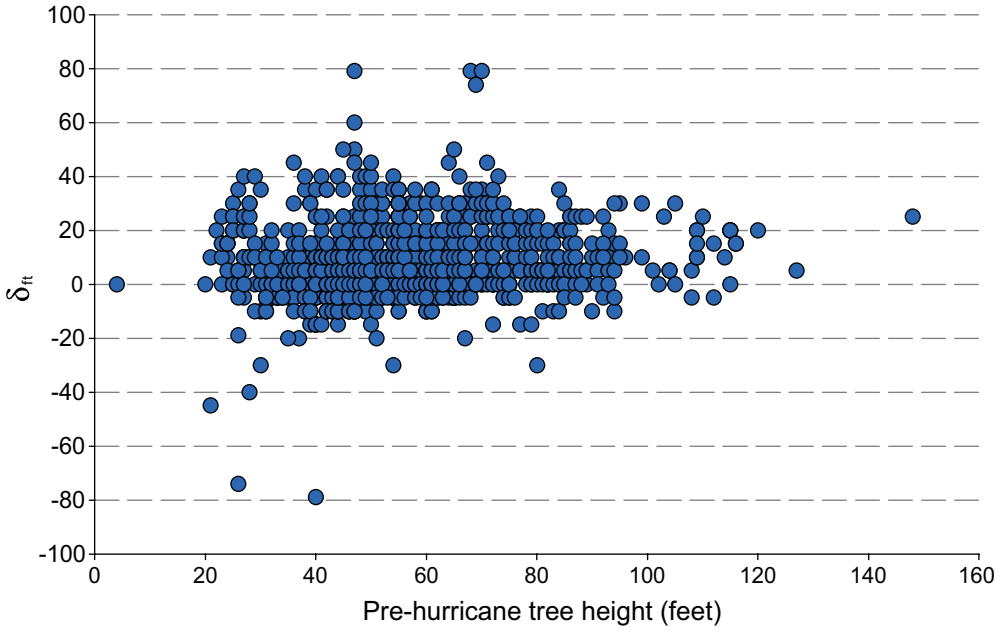


Figure 3: Change in foliage transparency for trees measured before and after Hurricane Isabel by pre-hurricane height.

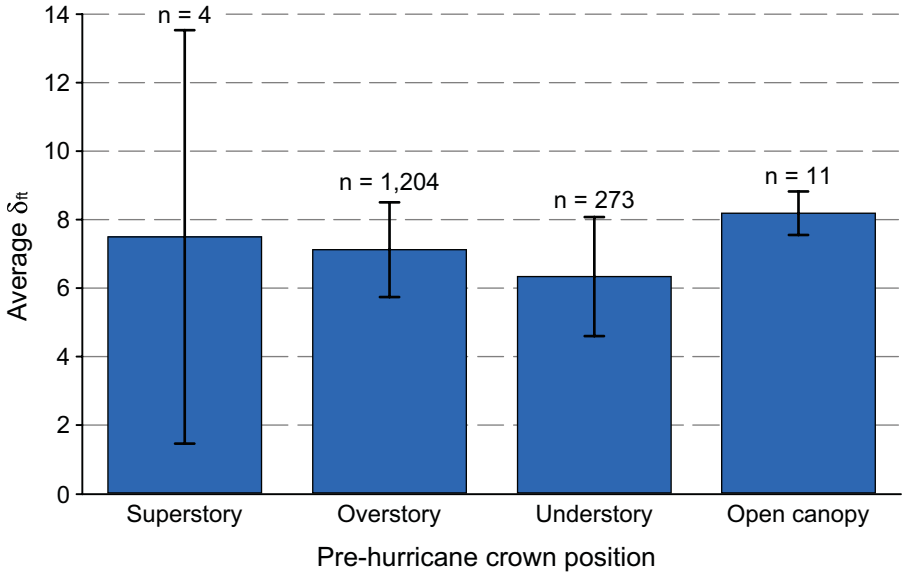


Figure 4: Average change in foliage transparency by pre-hurricane crown position, with standard error bars and sample size.

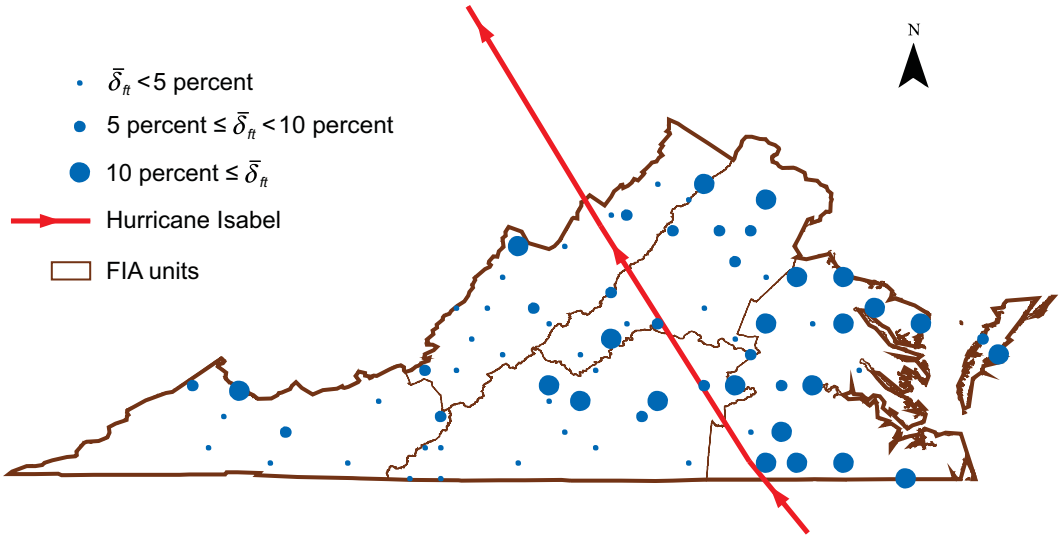


Figure 5: Change in percent foliage transparency for trees measured before and after Hurricane Isabel averaged at the plot level ($\bar{\delta}_{ft}$). Only plots with three or more live trees ≥ 5.0 inches (12.7 cm) d.b.h. are shown. Plot locations are approximate.

than just year-to-year variability was observed between the 7th and 8th surveys. Furthermore, the average change in crown density also indicated a significant loss in crown biomass at the State level as well as for most survey units and species groups (unpublished data). Thresholds for biologically significant changes have not been established for foliage transparency at this point in time. Therefore, changes in descriptive statistics highlight potential declining conditions.

The greatest changes in foliage transparency were observed in the eastern part of the State. Given the counter-clockwise rotation of hurricanes in the Northern Hemisphere and the north-westerly track of Hurricane Isabel, the greatest damage is expected on the northeast side of the hurricane's path. From the individual unit statistics and mapped plot averages, it is clear that this was the case in the Coastal Plain, though perhaps greater changes in foliage transparency were expected in the Northern Piedmont.

Boucher and others (2005) investigated the impact of Hurricane Isabel on a study site in Maryland, where individual trees on a 100 m by 100 m (328 ft by 328 ft) permanent plot were measured just prior to and again immediately after the hurricane passed through the area. After the hurricane they found a significant increase in the number of trees with severe damage. They also determined that trees with larger d.b.h. were more likely to suffer severe damage than smaller trees, and that for any given d.b.h., taller trees were more likely to be severely damaged than shorter trees. The present study did not find a significant correlation between d.b.h., height, crown position, and hurricane damage. However, Boucher and others (2005) considered damage as uprooting and leaning, and this study did not take those variables into account.

An increase in foliage transparency indicates a loss of foliage and hence, a loss of potential energy capture. The impact of foliage loss on tree growth in Virginia was not explored in this study; however, trees measured as late as 2006 still had elevated levels of transparency suggesting that trees experiencing foliage loss as a result of high winds and rain may take several years to recover. Despite the potential impact on growth, the loss of leaves may have prevented more serious damage by significantly reducing wind resistance (Putz and Sharitz 1991). Hedden and others (1995) reported that a 25 percent crown loss is more effective in preventing mortality than stem bending or branch streamlining at wind speeds below category 3 hurricanes, i.e., < 111 miles per hour (178 km per hour).

Conclusion

Quite often, studies of forest damage due to storms are limited in scale, with field measurements taken on subjectively and preferentially selected plots only after the storm has occurred. This preferential sampling typically emphasizes forest stands with unique characteristics, such as mature forests, or stands that have unusual features, such as rare species, making it problematic to determine the true scope of the damage across the landscape. In contrast, our approach utilized repeated measures on plots that were systematically distributed across Virginia. This allowed for the study of a wide range of stands across a variety of conditions before and after the hurricane.

The location of our study sites across the entire landscape and the observation that the greatest changes occurred in the Coastal Plain, lends support to the hypothesis that the changes in foliage transparency were due, at least in part, to the impacts of Hurricane Isabel. In addition to the crown condition indicators, a host of other individual-tree and plot-level variables are assessed on each plot. These include tree length, down woody material, and forest floor thickness. These data, along with auxiliary weather data (e.g. wind speeds and rainfall) are being explored to further understand and quantify the potential damage caused by Hurricane Isabel.

Acknowledgments

The authors wish to thank Bill Bechtold (Southern Research Station) and Dale Starkey (Region 8) for their helpful reviews of the manuscript. The authors also thank the Virginia Department of Forestry and the individuals who collected this data. This study was funded in part by the U.S. Department of Agriculture Forest Service, Region 8, Forest Health Monitoring Program (Analysis & Reporting Project #SO-AR-08-01).

References

Beven, J.; Cobb, H. 2004. Tropical cyclone report Hurricane Isabel. [Online]. http://www.nhc.noaa.gov/2003isabel.shtml. [November 16, 2007].

Boucher, D.H.; Rodick, C.L.; Bailey, J.N. [and others]. 2005. Hurricane Isabel and the forests of the Mid-Atlantic Piedmont and Blue Ridge: short-term impacts and long-term implications. pp. 201-208 in Sellner, K.G. (ed). 2005. Hurricane Isabel in Perspective. Chesapeake Research Consortium, CRC Publication 05-160, Edgewater, MD.

Gillespie, A.J.R. 1999. Rationale for a national annual inventory program. Journal of Forestry. 97(12): 16-20.

Hedden, R.L.; Fredericksen, T.S.; Williams, S.A. 1995. Modeling the effect of crown shedding and streamlining on the survival of loblolly pine exposed to acute wind. Canadian Journal of Forest Research. 25(5): 704-712.

McRoberts, R.E. 2005. The enhanced Forest Inventory and Analysis program. In Bechtold, W.A.; Patterson, P.L. eds. The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS–80. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station: 1-10.

Putz, F.E.; Sharitz, R.R. 1991. Hurricane damage to old-growth forest in Congaree Swamp National Monument, South Carolina, USA. Canadian Journal of Forest Research. 21: 1765-1770.

Randolph, K.C. 2006. Descriptive statistics of tree crown condition in the Southern United States and impacts on data analysis and interpretation. Gen. Tech. Rep. SRS–94. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 17 p.

Rose, A.K. 2007. Virginia's forests, 2001. Resour. Bull. SRS–120. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 140 p.

Rose, A.K. [In preparation]. Virginia's forests, 2007. Resour. Bull. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station.

Schomaker, M.E.; Zarnoch, S.J.; Bechtold, W.A. [and others]. 2007. Crown-condition classification: a guide to data collection and analysis. Gen. Tech. Rep. SRS–102. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 78 p.

U.S. Department of Commerce. 2004. *Service Assessment:* Hurricane Isabel, September 18-19, 2003. [Online]. http://www.weather.gov/os/assessments/pdfs/isabel.pdf. [October 25, 2007].

U.S. Department of Commerce, National Oceanic and Atmospheric Administration. 2008. Historical hurricane tracks. [Online]. http://maps.csc.noaa.gov/hurricanes. [July 22, 2008].

Tracking Forest Land



Estimating fine-scale land use change dynamics using an expedient photointerpretation-based method

Tonya Lister[1], Andrew Lister[1], and Eunice Alexander[1]

Abstract: *Population growth and urban expansion have resulted in the loss of forest land. With growing concerns about this loss and its implications for global processes and carbon budgets, there is a great need for detailed and reliable land use change data. Currently, the Northern Research Station uses an Annual Inventory design whereby all plots are revisited every 5 years and land use change matrices are estimated using a mapped plot design. These methods have great potential for providing the needed land use change data; however, for many states in the Northern region, these data will not be available until 2013 or later and the ability of these methods to capture finer scale changes, especially those due to urbanization, has not yet been tested. This paper presents an efficient photointerpretation-based change detection method that automates the work of gathering and loading images. A grid of photo plots is optimally created and overlain on the sample area, and land use change is recorded for two points in time by comparing digital imagery from 1998 and 2007. Results of a pilot test in Maryland show a net loss of forest land with losses due primarily to urban development and most gains in forest land coming from agricultural land uses. Forest losses are largely concentrated around Baltimore and Washington, DC. This pilot study indicates that about 75,000 photo plots would be needed to estimate land use change in Maryland at the county level. This would require approximately 125 hours, about 1.12 minutes per thousand acres, or roughly $1,500. The photointerpretation method presented here could be applied to other states and is well suited for land use change monitoring as the same points could be resampled when new imagery becomes available.*

Keywords: land use change, forest loss, photointerpretation

Introduction

Background

Several recent studies have predicted that urban expansion will continue to be a significant factor affecting forests in many areas of the country. For example, Nowak and Walton (2005) predicted that urban land in the United States would

[1]USDA Forest Service, Northeastern Research Station, 11 Campus Blvd., Suite 200, Newtown Square, PA 19073

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

nearly triple from 2000 to 2050 and that the percent of forest land in urban areas in Maryland would more than double to 37 percent by 2050. Similarly, Stewart et al. (2007) documented the current status of forests on the urban fringe, and highlighted potential impacts that continued urban expansion might have on them. Data from the National Resources Inventory (NRI) reported that more than 10 million acres of forest land were lost to developed land uses between 1982 and 1997 (U.S. Department of Agriculture 2000). Ecological impacts of urban expansion vary, but are generally related to loss of forest or other vegetative cover and increased edge habitat. Loss of forest cover has been shown to lead to loss of soil by both wind (Whiker et al. 2008) and water (Rice and Lewis 1991). It is generally accepted that the loss of topsoil not only lowers the productivity of agriculture and forest ecosystems, but also impacts aquatic ecosystems through sedimentation, nutrient enrichment, and other factors (Faulkner 2004).

The forests of the Chesapeake Bay watershed, many of which are in Maryland, provide benefits to wildlife and human populations. For example, they offer habitat for forest-dwelling species, protect drinking water, serve as buffers for estuarine species against sedimentation and nutrient enrichment, and provide economic and other benefits for humans (Sprague et al. 2006). Maryland state resource agencies are interested in assessing and monitoring land use change in these areas to understand the potential impacts of forest loss on the Chesapeake Bay and to generate needed information for urban planners, wildlife biologists, and other resource managers. Land use change data are also being used in modeling applications, forest resource projections, and carbon budgets (Woodbury et al. 2006).

Land cover products created with satellite imagery have been found to be inappropriate for land use assessments. For example, Irwin et al. (2007) found discrepancies between the satellite imagery-based National Land Cover Database (Homer et al. 2007) and a GIS dataset of land use derived from aerial imagery, in terms of both patterns and amount of developed land. These discrepancies are generally due to the fundamental difference between what a human interpreter can identify either on the ground or with photography (land use) and what an automatic, statistically based classification of satellite imagery can reveal (land cover) (Irwin et al. 2007).

The U.S. Forest Service's Forest Inventory and Analysis Program (FIA) conducts a continuous forest inventory using standardized methods that could be used to assess land cover dynamics. While recent advances in methodology have made it more feasible to monitor land use change using the existing FIA plot design, the data will not become available in Maryland and other states for many years. Furthermore, it is unclear if the sampling intensity of FIA plots is sufficient to meet precision requirements at the county level.

Objectives

To address these challenges, we created a flexible, inexpensive procedure with which to supplement FIA land use change estimates. The objectives of the study were to develop and implement a method with which to conduct a photo-based inventory of land use change in Maryland using FIA definitions. The goal was to not only obtain information useful to federal and state resource agencies, but also to convey information that can help practitioners implement this method in other states or in areas of interest (e.g., watersheds, wildlife areas, ecoregions).

Methods

Sample Design Development

To estimate land use change, we first had to develop a plot design. We decided the plot would consist of at least one subplot made up of a single point at which a photointepretation (PI)-based land use category would be assessed. From past experience, we determined that this type of plot is most amenable to rapid PI using FIA definitions. The NLCD change product (Homer et al. 2007) was used to determine the optimal subplot count and configuration and to assess various subplot arrangements. Although not PI-based, the NLCD data were used because they are the only spatially explicit and consistent land change data source that covers large analysis areas. The NLCD change product is a pixel-based GIS dataset in which each 30-m pixel is assigned a land cover change category based on comparisons of satellite imagery from circa 1990 and circa 2000. Focusing on forest loss, we first recoded the NLCD change product such that each pixel was labeled forest loss (1) or other (0). We then randomly generated 100 plots per county, with each plot consisting of an array of 25 subplots arranged in a square grid with 100-m spacing (Fig. 1). These plots sampled the NLCD change product-derived forest loss data using different numbers of subplots per plot. The sampling errors were determined for 10 randomly selected configurations of subplots for each subplot count category up to 10 subplots. Using these results, we calculated the total cost to achieve an acceptable level of precision (which we arbitrarily defined as generally having a sampling error no more than 20 percent of the county-level estimate) using the following equations:

$$N_{required} = ((t_{\alpha, n-1} * cv) / E)^2 \quad \text{equation 1}$$

and

$$Cost = a(N_{required}) + i*b(N_{required}) \quad \text{equation 2}$$

where $N_{required}$ = the sample size required to reach the desired precision, $t_{\alpha, n-1}$ = the critical value of the t distribution associated with a sample size of *n* at the 1-α confidence interval, cv=the coefficient of variation, E = the desired precision

expressed as the desired proportion of the mean that the confidence interval will represent -- in this case, 0.2, i = the number of subplots in the design, a = the cost in time required for the photointerpreter to switch between plots -- in this case, 1 second, and b = the time required to complete a single subplot -- in this case, 6 seconds.

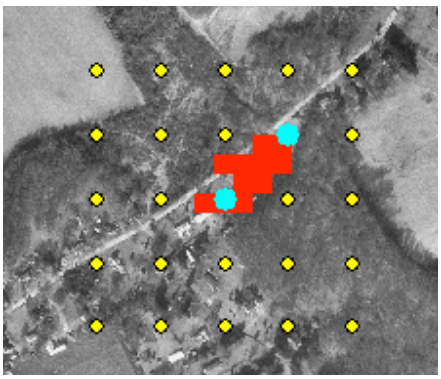


Figure 1. The 25-subplot design used in the first part of the study involving the NLCD change product. The subplots were arranged in a square grid with 100-m spacing. When subplots intersected the NLCD change category labeled forest loss (in red), they were counted, and the proportion of subplots counted in this manner was assigned to the plot for estimation purposes.

PI Methods

After determining which plot design minimized total cost and met precision requirements, we established a spatially balanced plot network consisting of 50,000 randomly selected plots across Maryland using a fractal-based tessellation approach described by Lister and Scott (2009). We divided the plots into 10 panels, each consisting of 5000 plots evenly distributed across the State. This paper discusses the results from one panel (5,000 plots) of data. These data were then used to reevaluate the number of plots needed for county-level estimates of land use change in Maryland, using the same methods as described above, only with the PI data instead of the NLCD change product data.

Land use category was assessed at two points in time (1998 and 2007) on each subplot by interpreting digital aerial imagery. The 1998 imagery consisted of panchromatic, leaf-on, 2-m-pixel resolution, digital orthophoto quadrangles (DOQs) from a state-level imagery dataset stored locally in an ArcGIS raster catalogue. The later date imagery consisted of color infrared, leaf-on, 1-m-pixel resolution, digital imagery from the National Agriculture Imagery Program collected for Maryland in 2007 and served over the Internet using a Web-mapping service (WMS). Land use categories used were based on an aggregation of more detailed FIA definitions (U.S. Forest Service 2005) including Forest, Agriculture,

Developed, and Other Nonforest. A single interpreter was trained and conducted all PI for this study.

To increase PI efficiency, an automation method was developed whereby an ArcGIS tool was used to subset imagery from the raster catalogue and the WMS to areas encompassing and slightly beyond the extent of the footprint of each plot. In other words, "snapshots" of imagery at a scale of 1:4000 were generated, with each image centered on the plot and containing sufficient detail for the interpreter to assess land use change. The two sets (1998 and 2007) of 5,000 images were stored locally, and displayed using a Microsoft Access form that we developed (Fig. 2). The form was designed to display the images and allow for data entry in such a way that the number of mouse clicks, wait time for image to loading, and data entry were minimized.

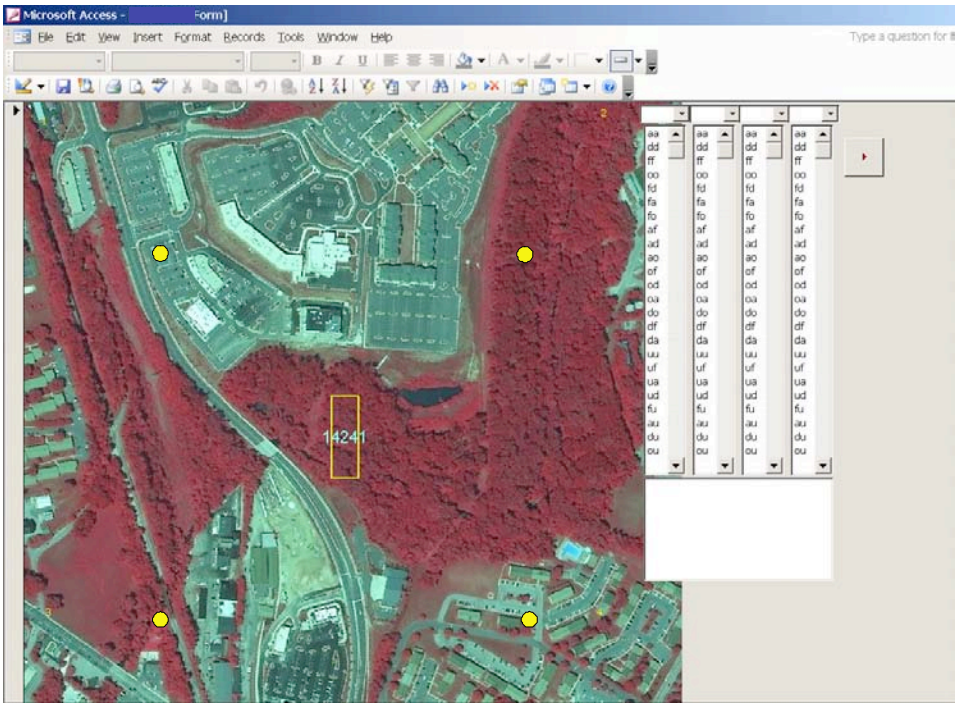


Figure 2. The MS Access form used to enter data and display imagery. Subplots are shown as dots at each corner of photo and rectangle in center is an acre area used as a reference guide. Data entry table for each subplot shown on right with all possible combinations of land use classes displayed.

Data from the 5000 plots were analyzed using the simple random sample estimator (Zar 1999), and estimates of the total areas of land use change categories were calculated, along with associated precision estimates. In addition, equation 1 was used to calculate the number of plots (and subplots) required to achieve acceptable precision, given a more realistic, optimized PI procedure and plot and sample design.

Results and Discussion

Pilot Plot Design Results

Figure 3 presents results of our evaluation of how various combinations of subplot counts and configurations affect sampling error, based on estimates of forest loss from the NLCD change product. As subplot count increased, large improvements in precision were observed until the subplot count reached 5 and then the rate of improvement was less pronounced. In other words, the change in the precision level after 5 subplots was not large enough to warrant the additional cost and time to add additional subplots into the final design. For plots with 3 and 4 subplots, we also graphed the average sampling error for those plots where the distance between subplots was maximized. In the best arrangements, subplots were located at the extremes of the subplot grid, where the intersubplot distances were maximized. One would expect this to be the case – subplots located farther apart are more likely to acquire different information about the landscape, making plot level summaries closer to the sample mean and thus lowering the variance of the overall estimate.

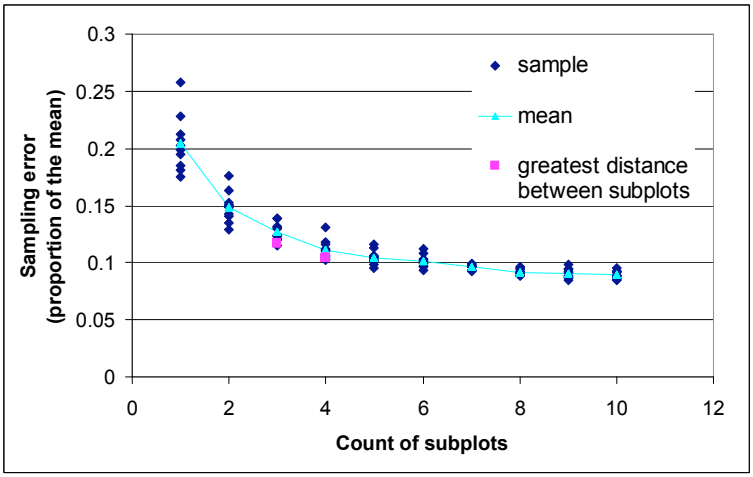


Figure 3. The relationship between subplot count (various configurations) and sampling error of estimates of area of forest loss from a set of plots that were intersected with an image depicting estimates of forest loss from the NLCD change product.

We conducted our cost analysis based on these results, using costs associated with between 1 and 5 subplots. With the cost function we chose, we determined that 3 subplots would be the optimal subplot count (Fig. 4). However, for our PI pilot study, we decided to use a 4-subplot design so as not to limit analysis opportunities. Subplots were arranged in the corners of a square pattern with vertical and horizontal distances of 500 m. This was the greatest practical distance given the constraints of the photo image resolution and size.

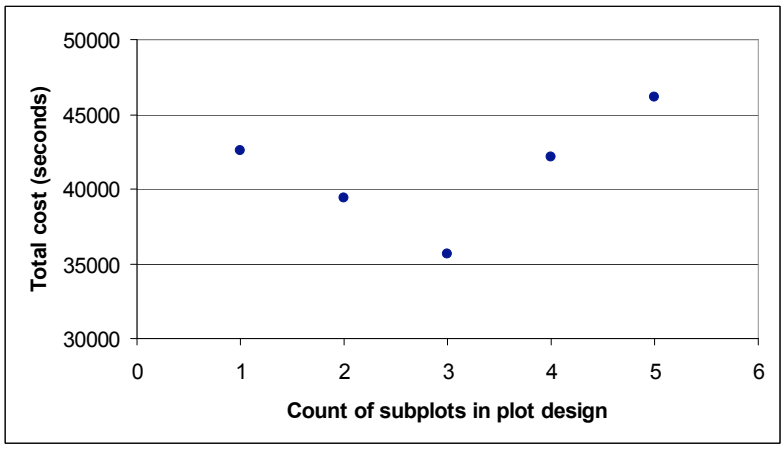


Figure 4. The relationship between cost and count of subplots. The optimal count was the point at which the cost was minimized.

Land Use Change Results

Land use change results from the PI pilot study show a net loss of 28,000 acres of forest land in Maryland from 1998 to 2007, which averages to be more than 3,000 acres per year (Fig. 5). The gross forest loss (66,000 acres) was primarily due to conversion to development, accounting for 91 percent of the total forest loss. Most forest gains were from agriculture (91 percent). The loss of forest land to development is an expected result, as Maryland experienced increases in population and housing densities during this period.

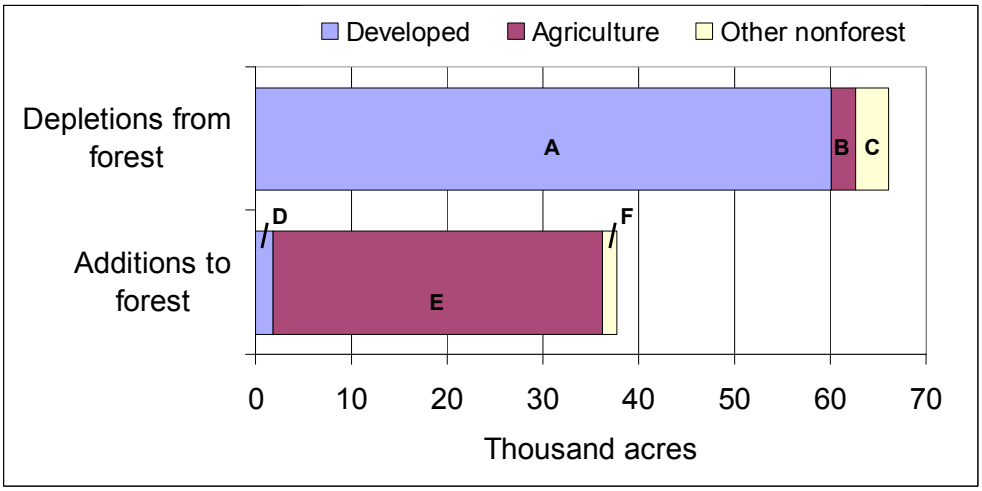


Figure 5. Estimates of areas of different land use change categories. Sampling errors are as follows: A: 7.9%; B: 35.3%; C: 32.1%; D: 33.3%; E: 9.9%; F: 44.7%.

Figure 6 shows the distribution of forest loss in Maryland between 1998 and 2007. There is a high proportion of forest loss plots in the growing suburbs of Baltimore and Washington D.C., an area of the state that has experienced the greatest pressure from urban expansion. For example, the highest proportion of forest loss plots is found in Prince George's county, which borders Washington D.C. From 2000 through 2007 more than 22,000 new housing units were approved for construction, making this one of the fastest growing counties in the state (Maryland Department of Planning 2007).

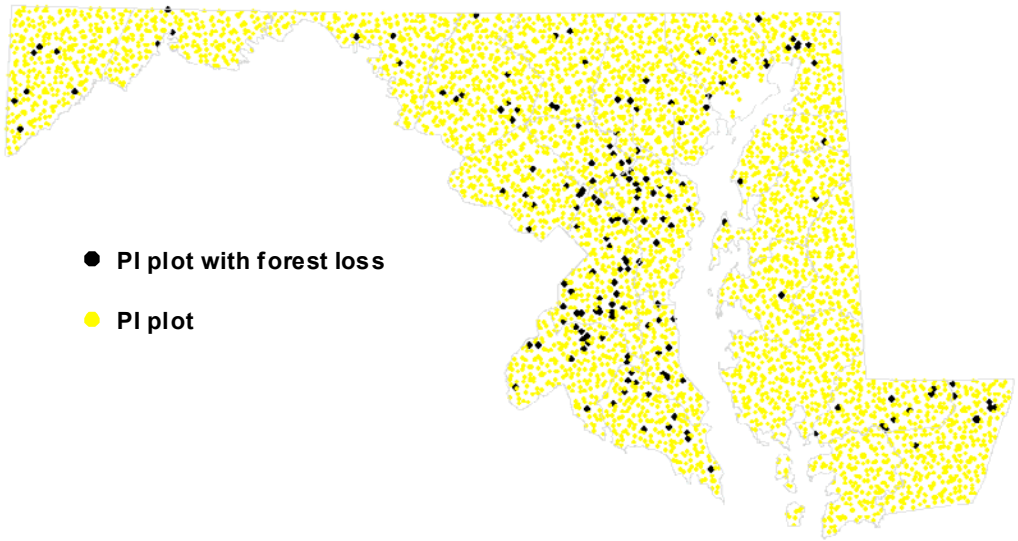


Figure 6. Distribution of land use plots highlighting plots showing forest loss, 1998-2007, Maryland.

Reevaluation of Study Design Using Pilot Results

During this study, we made improvements in image viewing and data entry methods that substantially lowered the cost (in terms of time) involved in switching between photos. With these new techniques, the cost of switching between plots in terms of file opening and image loading time was reduced. As our initial estimate of one second of time spent switching between photos neared zero, the time associated with doing a single plot of 4 subplots was not substantially different from doing 4 single-subplot plots. We therefore determined that the optimal sampling protocol for future work would be a single subplot plot.

Using results from the pilot study, we reevaluated the number of plots that would be necessary to estimate land use change at the county level in Maryland with acceptable precision (having a sampling error no more than 20 percent of the county-level estimate). Each county has a separate requirement for the number of plots needed, and there are some important things to consider when determining what plot density is best. If the end user's goals are monitoring and regional analyses, then a uniform density of plots across the state would be preferred. The plot density needed for each county ranged from one plot per 28 to 333 acres. The

counties where we would expect the greatest amount of forest land change as suggested by U.S. Census and Maryland planning data would be sufficiently sampled with a plot density of one plot per 91 acres, so we decided to use this as our plot density requirement across the state. Therefore, we estimated that 75,000 plots would be needed in Maryland for county-level estimates of land use change. Results from the study show that, on average, 10 points can be photointerpreted per minute. To complete the PI work for the whole state, this translates to approximately 125 hours—just over 3 weeks of full-time work, or approximately $1,500.

This method could be applied to other states, and cost could be easily estimated for a certain level of precision. Because images generated for the PI work are snapshots, there is no wait-time for new images to load and there is no need for network connections. More than 5,000 images can easily fit on a single DVD, making the procedure mobile and efficient. The method is also well suited for monitoring. When new imagery is flown, the same PI points can be measured allowing for time series analyses of land use change.

References

Faulkner, S. 2004. Urbanization impacts on the structure and function of forested wetlands. Urban Ecosystems. 7(2): 89-106.

Homer, C.; Dewitz, J.; Fry, J.; Coan, M.; Hossain, N.; Larson, C.; Herold, N.; McKerrow, A.; VanDriel, N.; Wickham, J. 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogrammetric Engineering & Remote Sensing. 73(4): 337-342.

Irwin, E.G.; Cho, H.J.; Bockstael, N.E. 2007. Measuring the amount and pattern of land development in nonurban areas. Review of Agricultural Economics. 2(3): 494-501.

Lister, A.J. and C.T. Scott. 2009. Use of space-filling curves to select sample locations in natural resource monitoring studies. Environmental Monitoring and Assessment 149(1-4): 71-80.

Maryland Department of Planning. 2007. New housing units authorized for construction (2000-2007). Available online at: <http://www.mdp.state.md.us/msdc/NEWHH/2007/annual/Table2A_2007.xls>. [November 25, 2008].

Nowak, D.J.; Walton, J. 2005. Projected urban growth (2000–2050) and its estimated impact on the US forest resource. Journal of Forestry. 103(8): 383-389.

Rice, R.M.; Lewis, J. 1991. Estimating erosion risks associated with logging and forest roads in northwestern California. Water Resources Bulletin. 27(5): 809-818.

Sprague, E.; Burke, D.; Claggett, S.; Todd, A., eds. 2006. The state of Chesapeake forests. Arlington, VA: The Conservation Fund. 115 p.

Stewart, S.I.; Radeloff, V.C.; Hammer, R.B.; Hawbaker, T.J. 2007. Defining the wildland urban interface. Journal of Forestry. 105:201-207.

U.S. Department of Agriculture. 2000. Summary Report: 1997 National Resources Inventory (revised December 2000). Natural Resources Conservation Service, Washington, DC, and Statistical Laboratory, Iowa State University, Ames, IA. 89 p.

U.S. Department of Agriculture, Forest Service. 2005. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 3.0. U.S. Department of Agriculture, Forest Service, Washington Office. Internal report. On file with: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, Rosslyn Plaza, 1620 North Kent Street, Arlington, VA 22209.

Whicker, J.J.; Pinder J.E.; Breshears, D.D. 2008. Thinning semiarid forests amplifies wind erosion comparably to wildfire. Journal of Arid Environments. 72(4): 494-508.

Woodbury, P.B.; Heath, L.S.; Smith, J.E. 2006. Land use change effects on forest carbon cycling throughout the southern United States. Journal of Environmental Quality. 35:1348-1363.

Zar, J.H. 1999. Biostatistical analysis. Upper Saddle River, NJ: Prentice Hall. 663 p.

Development of Issue-relevant State Level Analyses of Fragmentation and Urbanization

Rachel Riemann[1], Tonya Lister[2], Andy Lister[2], Dacia Meneguzzo[3], Sarah Parks[1]

Abstract: There has been considerable research concerning the extent and effect of urbanization and fragmentation and the importance of monitoring current and potential magnitudes of change is recognized. However, there are limited guidelines for interpreting fragmentation data or for their application for analysis and statewide planning efforts. In this study we take a first step toward developing a state-level analysis of urbanization and fragmentation that addresses three categories of information. Example maps, tables, and analyses are drawn from New York, Maryland, and Delaware. Landscape metrics calculated from various regional or national datasets were chosen for their relevance to issues of interest and other traits such as accuracy and consistency. Examples of results include maps accompanied by graphic and tabular analyses addressing several landscape factors that are increasingly impacting forest resources and the ecosystem services and products they provide. Where published guidelines are available, results include management-relevant maps in which the metrics have been translated into impacts on stream water quality, interior bird species composition, and other processes. From these elements a prototype structure can be developed for reporting on the status of fragmentation and urbanization in a state and across the region so that we can better understand our forest resource in the context of its surrounding landscape and the status of changes in its natural, social, or economic ecology.

Keywords: Urbanization, forest fragmentation, landscape metrics, state forest assessments, FIA reports.

Introduction

Forest land is a significant factor in the protection of surface and groundwater quality and is a major component of many increasingly threatened wildlife habitats. Forest land is also a resource heavily relied upon by people for recreation, timber, and nontimber products, and for more intangibles, such as aesthetics and intrinsic value.

As human population growth continues, many areas of the country are seeing developed land uses expand, often at the expense of forest land (Hammer et al.

[1]U.S. Forest Service, Northern Research Station, Troy, NY
[2]U.S. Forest Service, Northern Research Station, 11 Campus Blvd., Suite 200, Newtown Square, PA 19073
[3]U.S. Forest Service, Northern Research Station, St. Paul, MN

In: McWilliams, Will; Moisen, Gretchen: Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008; Park City, UT. Proc. RMRS-P-56CD.Fort Collins, CO; U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

2007, Nowak and Walton 2005, Robles et al. 2008, Stein et al. 2006). In addition to losing forest land to development, remaining areas of forest land are currently under pressure in many regions in terms of both fragmentation — being divided up into smaller and more disconnected pieces; and urbanization —an increasing proximity to urban development, population, and other anthropogenic pressures. Fragmentation causes changes in light, wind, and moisture microclimates, all of which provide an avenue for the introduction and spread of invasive plant and animal species. Fragmentation also introduces barriers to the movement of native species and degradation of native habitats (e.g., Belisle et al. 2001, Burke and Nol 2000, Cam et al. 2000, Herrmann et al. 2005, Rosenberg et al. 2003).

Urban development in or near forests can change local hydrology, increase recreation pressures, alter native species diversity, provide vehicles for the introduction of invasive species either by design or by accident, and often bring significant disturbance to the area (e.g., Airola and Buchholz 1984, Bastin and Thomas 1999, Heckscher et al. 2000, Iida and Nakashizuka 1995, McDonnell and Pickett 1990, Rudnicky and McDonnell 1989). Together, fragmentation and urbanization cause a disruption of the flow of material through the forest ecosystem, affecting both forest health and sustainability (e.g., Macie and Hermansen 2002). Researchers have documented varied impacts of forest fragmentation and urbanization on the probability of commercial forest management and timber harvesting (Wear et al. 1999, Munn et al. 2002, Kline et al. 2004), and on water quantity and water quality (e.g. Hunsaker et al. 1992, McMahon and Cuffney, 2000, Riva-Murray et al. in prep).

Forest fragmentation and urbanization are also inextricably linked to the effects of climate change. Since the dispersal and movement of forest plants and animals are disrupted by forest fragmentation, impacts of climate change on species and diversity losses can be magnified (McDonnell and Pickett 1990, Rodenhouse et al. 2008). Similarly, systems already under pressure from urbanization and fragmentation will be less resistant to the additional stresses imposed by climate change.

With the increasing fragmentation and urbanization of our landscape (Hobbs and Stoops 2002), interest has grown in the location, type, and magnitude of its potential impacts. The U.S. Forest Service's Forest Inventory and Analysis program (FIA) has begun to address landscape context and change in its reports containing state-level forest inventory results. Information on forest distribution and context is crucial for monitoring and assessment efforts like the U.S. Forest Service State and Private Forestry's statewide assessments, U.S. Geological Survey's national water quality assessments, the U.S. Environmental Protection Agency's wadeable streams assessments, and regional forest assessments like those done in Oregon and Washington.

Metrics and Data Sources

Thus, FIA is being asked to monitor the distribution, urbanization, and fragmentation characteristics of the forest over time, just as we monitor the

change in total forest land over time. To do this we must choose both metrics and the landscape data source(s) carefully.

Of all the fragmentation and urbanization metrics that could be calculated from the available data, it is important to identify a concise set of relevant, useful landscape metrics for fragmentation analyses. Avoiding redundant metrics (Riitters et al. 1995) and metric inconsistency over space and time (De Clercq et al. 2006) have been identified as useful ways to create a subset of potential landscape descriptors. When identifying useful metrics, insensitivity to both the spatial resolution and number of classes in the remotely sensed classification is desirable where monitoring over time is a goal (De Clercq et al. 2006). However, as De Clercq et al. (2006) and McAlpine and Eyre (2002) point out that in order to be truly useful for forest monitoring, individual metrics also need to be carefully chosen based on the particular questions being asked. In this study, we are interested in those landscape metrics that add value to the interpretation of forest inventory data because of their direct relationship to changes in the forest resource, our utilization of it, or its ability to provide ecosystem services and products. We have identified from the current literature accurate, consistent landscape descriptors and classification schemes (thresholds) that are most consistently related to forest ecological, social, or economic impacts of concern and can be more accurately and consistently calculated from available data sources.

The landscape data source used must also be carefully chosen because of its impact on both resulting values and interpretation of results (e.g., Riva-Murray et al. unpublished, and Riemann et al. in prep). Several studies have noted the impact of data source and have addressed this by applying different 'corrections' to the landscape dataset to more closely reflect conditions on the ground as they are typically seen by land managers or planners, or used by wildlife. For example, Heilman et al. (2002) included roads in their calculation of a forest intactness metric. Lister et al. (2005) removed patches smaller than a certain area and width to more closely match FIA definitions of forest land, and used local road density to relabel those forest or agriculture pixels in the 1992 National Land Cover Dataset (NLCD) dataset that were likely to be developed based on road density. In this study we have chosen to keep land cover and land use clearly separate, utilizing the 2001 NCLD for land cover, and U.S. Census-based datasets for factors relating to urban land uses, with a clear understanding of the development and limitations of each of these landscape datasets and any impacts on interpretation.

The goals of this paper are to:

- Provide a suite of reliable, interpretable, standardized fragmentation measures to authors and consumers of FIA state inventory reports
- Provide examples of how these measures can be added to state reports to enhance the interpretation of FIA forest inventory data and the understanding of the forest resource
- Describe briefly how these were derived from the best available landscape data sources, and provide links to complete metadata

The potential additions to state reports include maps, tables, and graphs that A) describe the current status of fragmentation and urbanization on forest land in general; B) begin to address its probable or potential impact; and C) combine landscape context information with data on FIA plots to characterize the forest affected.

Methods

Identifying Metrics to Use

We were specifically interested in interpretability of results with respect to impacts on the forest resource or on the forest's ability to provide ecosystem services and products. To be able to monitor over time and utilize any known relationships established in the literature, we have focused on choosing metrics based on their performance with respect to several criteria: A) consistency over time – e.g, robust to changes in dataset resolution; B) accuracy in comparison to what is observed on the ground or interpret as 'fragmentation'; and (C) representative of those characteristics of forest fragmentation and/or urbanization that have been shown in the literature to be relevant to the ecological, social, and economic impacts of concern.

Data Sources Used

We required the landscape data sources to be spatially continuous, available over broad areas, and of sufficient spatial resolution to meaningfully describe landscape processes of interest. The most widely available dataset meeting these criteria was the 2001 NLCD, a set of satellite image-based products produced by a consortium of federal agencies, led by the U.S. Geological Survey (Homer et al. 2007). These products are comprised of 30-m pixels, each labeled with a land cover category, percent impervious surface, and percent canopy cover estimates. In past studies, we used the U.S. Geological Survey's GAP datasets, which are similar to NLCD datasets, but are created with varying methods and slightly different goals. We chose NLCD over GAP data because they are produced with reasonably consistent methods and their accuracy and other properties over large areas are better understood.

With some caveats, particularly regarding impervious surface, NLCD captures land cover information reasonably well (Riemann et al. in prep). NLCD forest land was used for calculating forest pattern metrics even though it does not match the FIA definition of forest land (Ruefenacht et al. 2008). Thus the forest pattern metrics in this paper reflect the distribution of forest cover, not FIA forest land. Figure 1 illustrates the unit-level differences in forest area calculated by the two metrics in New York. It is evident that FIA forest percent is less than that reported by NLCD across most survey units, with a maximum difference between the two data sources of ~0.3 million acres (about 10 percent) in the Catskills-

Lower Hudson survey unit, an area with a substantial amount of residential development and forest/urban intermix (Radeloff et al. 2005). In the northeastern United States, NLCD's forest classification includes trees in residential areas, which would not fall under the FIA forest definition.

Derivation of urban and land-use data required integration of information from the U.S. Census Bureau. Road data were derived from the TIGER/Dynamap 2000 dataset (U.S. Census Bureau 2002). Metrics related to the levels and types of urbanization were drawn from the Wildland Urban Interface (WUI) database (Univ. of Wisconsin n.d., Radeloff et al. 2005), and from the U.S. Census Bureau block-level data (2002) on population, housing, and second home density compiled for the WUI project (Univ. of Wisconsin n.d.). Change in house density over time also was obtained from WUI database (Univ. of Wisconsin n.d., Hammer et al. 2004).

Hydrologic Unit Code (HUC) polygons were developed as part of the national hydrologic dataset and downloaded from the USDA Geospatial Data Gateway (USDA Nat. Resour. Conserv. Serv. n.d.). We used the HUC12 scale, which approximated that scale used in studies identifying the percent impervious thresholds chosen.

Forest Inventory and Analysis data were accessed from the internal FIA database (U.S. Forest Service n.d.).

Landcover Mosaic (LCM), Morphological Spatial Pattern Analysis (MSPA), and Forest Area Density (FDEN) datasets describe A) the mixture of agricultural/urban/natural land-cover type (Riitters et al. 2009); B) the structural element of which a forest patch is part (as described at European Commission, DG-Joint Research Centre, Institute for Environment and Sustainability, http://forest.jrc.it/biodiversity/Product); and C) the percent forest area, respectively, in the 15 ha (37 acre) local area surrounding each grid cell. Graphics showing this information for the continental United States are currently available at www.forestthreats.org/tools/landcover-maps. Geospatial datasets containing this information will be available shortly.

In this study we did not apply 'corrections' to the landscape data source. Instead we relied on careful qualification of 'forest cover' vs. 'forest land use' and the use of multiple metrics including specifically land use-based metrics. This approach provides more potential for application, as well as increasing ease of both calculation and interpretation.

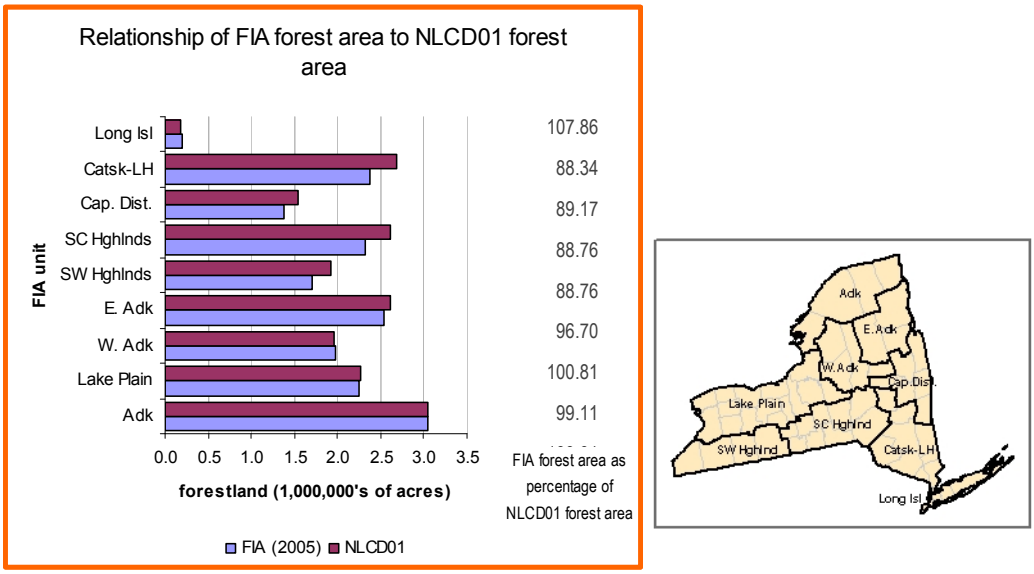


Figure 1: Relationship of FIA forest area to NLCD2001 forest area in New York, by FIA unit.

Processing

ArcMap software was used to prepare and analyze the input datasets. We automated most of the processing with ArcMap models to facilitate application of these processes to other datasets. Most of the spatial datasets shown in this paper have already been produced for the northeastern quadrant of the United States (the 20 states comprising the U.S. Forest Service's Northern Research Station).

Complete processing details for each spatial dataset are available at: http://www.fia.fs.fed.us/symposium/metadata/fragurban_metrics_metadata.doc. In general, data preparation included clipping input datasets to state or region boundaries and reclassifying the NLCD land cover data into appropriate land-cover groups (e.g., forest/nonforest, or forest/natural vegetation/water) for further geospatial analyses. A series of geoprocessing operations were then applied to the datasets to derive the landscape metrics, including vector-to-raster conversion and distance calculations (e.g. for distance to nearest road), calculation of patch areas, shrinking patch edges (for edge/interior calculations), area tabulations and/or continuous data summaries of the metrics by analysis unit (e.g., county, 10 km x 10 km grid, and watershed), and extraction of pixel values to FIA plots (e.g. patch size, distance to road, WUI class). For example, one ArcMap model used the Euclidean distance tool, the census roads, and the NLCD forest dataset for New York to assign each valid output pixel the distance to the nearest road. Subsequent steps in the model summarized the output of this step and created a table showing the frequency distribution of forest cover by distance to road category.

Thresholds to Facilitate Interpretation

Interpretation of the likely impacts of certain configurations of fragmentation and urbanization requires an understanding of their relationship to the ecological,

social, or economic response of interest. In addition, ecosystem response to fragmentation and urbanization does not always occur gradually across all levels of landscape change intensity. In some cases the observed response indicates a particular threshold of interest beyond which the rate or level of response is sharply different. In other cases, crossing artificial thresholds cause an impact, like in the case of passing legally allowable limits of forest cutting. Some studies have developed general management guidelines from these observed relationships between fragmentation measures and an ecosystem response. Where guidelines were available, we utilized this information to categorize our map and tabular outputs. In particular, in this paper the following guidelines were utilized in the interpretation or development of the maps and tables presented in the results:

- **Patch size and forest proportion:** Habitat requirements for wildlife vary by species. However, for reporting purposes it is often helpful to summarize forest-patch data using general guidelines. Many wildlife species prefer contiguous forest patches that are at least 100 acres. This patch area is often used as a minimum size still containing enough interior forest to be a source rather than a sink for populations of some wildlife species. Depending on your geographic region of interest or species of concern, this threshold could be customized. Some studies have found that in addition to patch size, the proportion of forest land in an area that extends beyond the patch can be used to develop habitat thresholds. Rosenberg et al. (1999) found that forest-patch size information can be used in relation to the amount of forest land in a surrounding 2500-acre area to develop habitat suitability models for certain species of interior forest dwelling birds. Their resulting matrix of ecosystem responses provides detailed information that can be applied, with some understanding of the quality of the landscape data source being used.

- **Forest edge**: While edge effects vary somewhat with distance, depending on the type of effect and species of vegetation or wildlife, (e.g., Chen et al. 2002, Rosenberg et al. 1999, Flaspohler et al. 2001), 100 to 300 ft (~30 to 90 m) is frequently used as a general range for the 'vanishing distance' or the distance into a patch where the edge effect disappears and interior forest conditions begin.

- **Impervious surface**: The amount of land area within a watershed that is impervious to water (pavement, buildings, parking lots, etc.) affects water quality. When water is able to pass through the ground, soil and vegetation act as a filter and improve water quality. As the proportion of impervious surface increases, however, many pollutants flow directly into the waterway. Impervious surface areas of 10 and 25 percent are generally recognized to be the thresholds above which small watersheds are impaired and impacted, respectively (Arnold 1996).

- **Human population density:** Population densities are generally recognized as having a negative effect on the viability and practice of

commercial forestry (Barlow et al. 1998, Kline et al. 2004, Munn et al. 2002, Wear et al. 1999). In this study we used thresholds identified by Wear et al. (1999), which showed that the probability of commercial forestry dropped from 75 to 25 percent as population density increased from 20 to 70 people per square mile. These thresholds were estimated based on data from the 2000 U.S. Census and other data from Virginia. As research results become available that are more timely and region-specific, these data will be used to create improved commercial forest probability maps.

- **House density:** Thresholds of house density used in the wildland-urban interface (WUI) and intermix definitions come from the Forest and Wildlife Ecology SILVIS laboratory at the University of Wisconsin-Madison. The WUI interface is defined as the area where human development meets rural or wildland areas. These thresholds were originally established to describe wildland firefighting guidelines (Radeloff et al. 2005).

The thresholds and guidelines presented are not an exhaustive list of what is available. Rather they represent examples in several important areas. More guidelines, including those of regional development and relevance, may be available now and more will be available in the future.

Results and Discussion

Forest fragmentation and urbanization data analyzed for inclusion in FIA state reports can be organized into three broad categories of information. The first are statistics, maps, and/or graphs that describe the landscape character and spatial pattern of forest land distribution in a state. Wherever possible, legend class breaks in these maps should be chosen with respect to known or suspected impact thresholds. The second category identifies the impact of forest fragmentation or urbanization on a particular ecological, social, or economic issue. These statistics, maps, and/or graphs frame the data with respect to the specific threshold(s) and scale(s) identified (and ideally established) in the literature and can thus be more directly interpreted for their probable impact on these issues. The third type of information that could be useful is the result of an overlay of the spatial context and urbanization information with data collected on FIA plots to assess the impact on different populations of the forest resource.

Table 1 provides a summary of the figures presented, in terms of the category of information, question addressed, input datasets used, and methods used in its creation. This is not an exhaustive list of the fragmentation or urbanization analyses desirable. Rather, the maps and graphs chosen represent some of the carefully chosen metrics that should be used as a starting point.

Table 1: Summary of figures presented in this paper.

Category of information (1, 2 or 3)	Question answered	Input Datasets Used	Summary of methods	Example figure in text
1	Where is forest land affected by underlying house densities greater than 6 per sq. km (15.54 per sq. mile)	1) U.S. Census 2000 house density data at the block level 2) NLCD2001 for the nonforest land cover mask	Block-level choropleth with 30m raster mask	2a
1	How much fores tland is affected by underlying house densities greater than 6 per sq. km	--same as above, plus: 3) FIA unit boundaries	Extraction of census block values to 30 m forest pixels Data summarization	2b
1	Where is the forest land affected by roads and to what extent	1) U.S. Census Bureau TIGER/Dynamap 2000 dataset [all roads] 2) NLCD2001 for nonforest land cover	Per-pixel distance calculation	3a
1	To what extent is forest land affected by proximity to roads	--same as above	Data summarization	3b
1	Where is a substantial proportion of the forest land occurring in patches less than 100 acres in size	1) NLCD2001 for forest pixels 2) 10 km x 10 km grid poly coverage	Patch area calculation Extraction of patch size values to 30 m pixels Data summarization to 100 sq. km grid cells	4a
1	How much forest land occurs in patches less than 100 acres	1) NLCD2001 for forest pixels	Patch area calculation Extraction of patch size values to 30 m pixels Data summarization	4b
2	What is the probable stream water quality as predicted by percent impervious surface alone?	1) NLCD2001 impervious surface layer 2) U.S. Census 2000 house density data at the block level 3) U.S. Census Bureau TIGER/Dynamap 2000 dataset 4) Hydrologic Unit level 12 polygons (HUC12) from the National Hydrologic dataset	Modeled percent impervious values for HUC12 basins Choropleth mapping of results	5
2	Where is forest land still suitable for an interior forest bird species such as scarlet tanager?	1) Land Management table from Rosenberg et al. 1999. 2) NLCD2001 for forest pixels	Patch area calculation Moving window analysis for percent forest in surrounding 2500 acre area Application of Rosenberg et al. (1999)'s table	6
2	What is the probability of commercial forestry occurring, and where?	1) NLCD2001 for forest pixels 2) U.S. Census 2000 population density data at the block level 3) Thresholds of	Extraction of population density values to 30 m pixel Application of Wear et al.'s (1999) thresholds as the legend	7

		commercial forestry probability by population density from Wear et al. 1999		
2	What is the probability of commercial forestry occurring within each county	--same as above, plus 4) County boundaries	Area tabulation to counties	8
3	How are species distributed with respect to the Wildland Urban Interface or Intermix Areas?	1) Wildland Urban Interface dataset 2) FIA plot-level data	Extraction of WUI code to FIA plots Data summarization	9

Characterizing forest distribution and context

In these examples, the metrics chosen relate to some aspect of urbanization or fragmentation that is suspected of, or has been documented to have an effect on the forest, its management, or on its ability to provide ecosystem services and products. Figures 2a through 4b provide examples of such maps and related graphs for New York. For example, figures 2a and 2b illustrate how much forest land is affected by underlying house densities greater than six houses per sq. km (15.54 per sq. mile), and where it occurs. Figures 3a and 3b show where and to what extent forest land is affected by roads. As Riitters and Wickham (2003) reported, this can be quite extensive. The distribution of forest land occurring in patches less than 100 acres is portrayed in figures 4a and 4b. One hundred acres is a threshold identified in the literature as an approximate minimum size for patches that contain enough interior forest area to be sources rather than sinks for wildlife populations. Other metrics and data sources providing valuable information with respect to understanding where fragmentation and urbanization impacts on forest land are occurring include: forest occurring within the WUI (Radeloff et al. 2005), changes in housing density over time (Hammer et al. 2004, Univ. of Wisconsin n.d.), forest land affected by edge conditions, forest connectivity for species requiring large ranges, and areas of forest where there is a substantial amount of second home development. The latter two maps can be depicted at the scale of 30 m pixels (e.g. depicting actual distance class to the nearest road), or at a summarized scale (e.g. 100 sq. km grid) depicting the proportion of the forest land in that pixel that is above or below a certain important threshold. In addition, the land-cover mosaic and spatial pattern metrics developed by Riitters et al. (2009), while not specific to a particular issue, do provide easily understandable, complementary, consistent and robust metrics of land-cover pattern that could also be analyzed with FIA plot data to describe some of the characteristics of those segments of forest land most under pressure from urbanization and fragmentation influences. All of the above information can also be tabulated, describing the proportion of forest land in each county that's affected by any one (or more) particular criteria (Table 2).

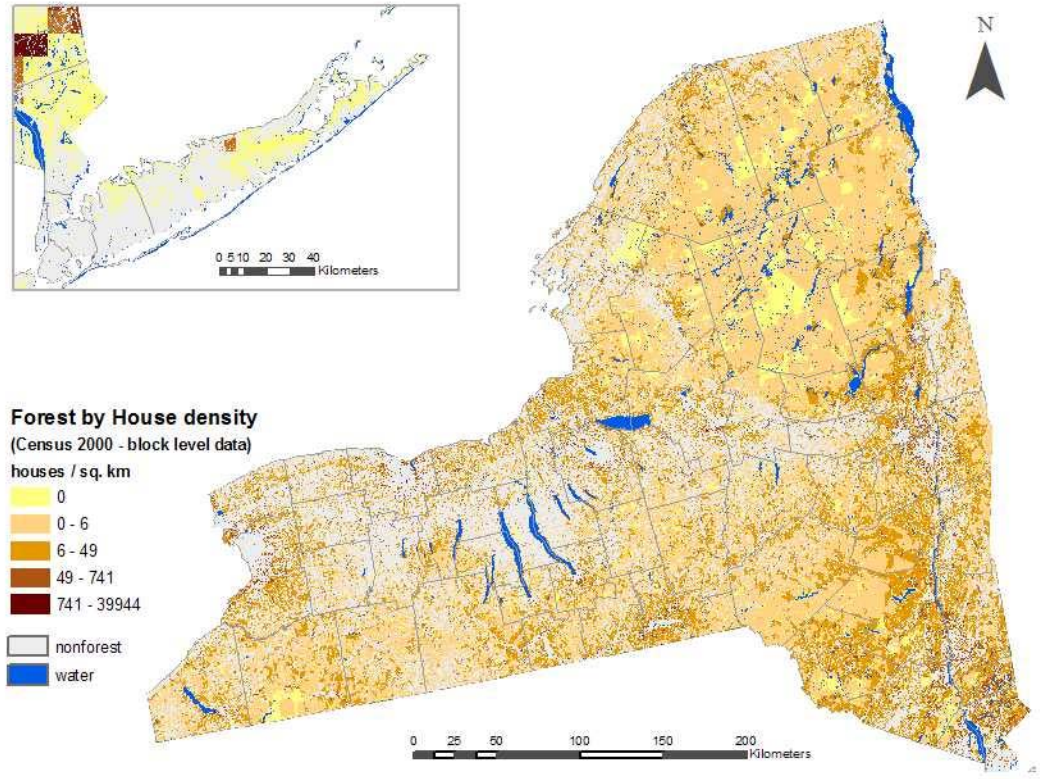


Figure 2a: Distribution of forest cover by house density, New York, 2001 (forest), 2000 (house density).

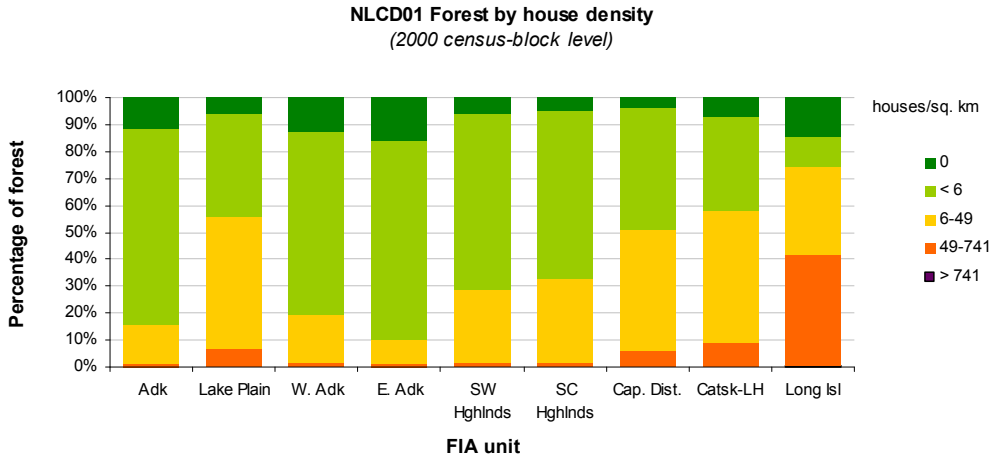


Figure 2b: Distribution of forest cover by house density, by FIA unit, New York, 2001 (forest), 2000 (house density).

Within FIA units in New York, between 10 and 73 percent of the forest occurs intermixed with house densities of ≥6 per sq. km. This represents the approximate density at which firefighting switches from 'wildland' to 'structure' firefighting techniques and costs (Radeloff et al. 2005). Forest intermixed with houses also represents areas of forest cover more likely to be in nonforest land

use, and/or more likely to be experiencing pressures from recreation, invasives, and other local human effects.

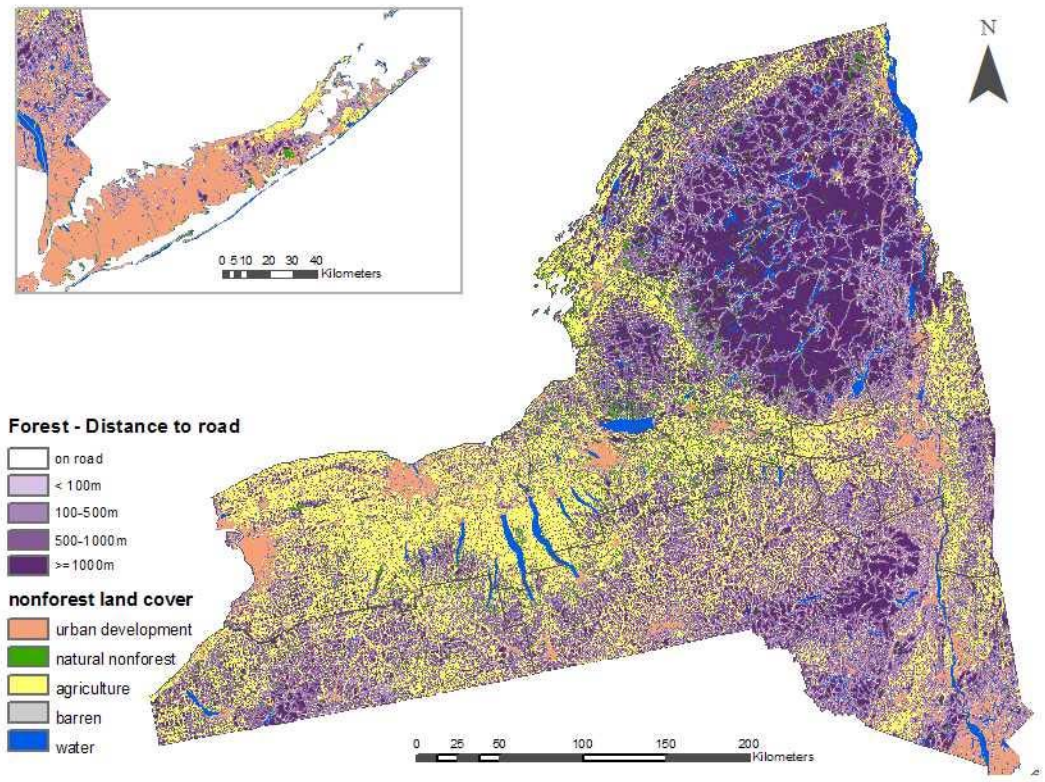


Figure 3a: Spatial distribution of forest cover by distance to road, New York, 2001 (forest), 2000 (roads).

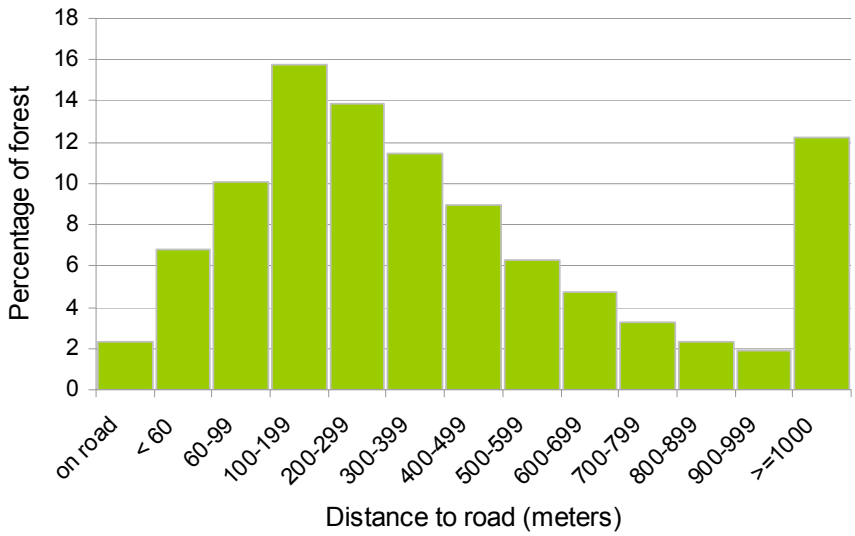


Figure 3b: Frequency distribution of forest cover by distance to road, New York, 2000.

In New York, nearly 60 percent of the forest is within 400 m of a road. According to Riitters and Wickham (2003), regions with more than 60 percent of their total land area within 382 m of a road may be at greatest risk of cumulative ecological impacts from roads.

Road effects distances range from 100 m for secondary roads (a rough estimate of a highly variable zone), 305 m for primary roads in forest (assuming 10,000 vehicles per day), and 810 m from roads in urban areas (50,000 vehicles per day) (Forman 2000). Using currently available road data, these thresholds could easily be applied state or regionwide to identify more specifically the location and magnitude of forest area affected by roads.

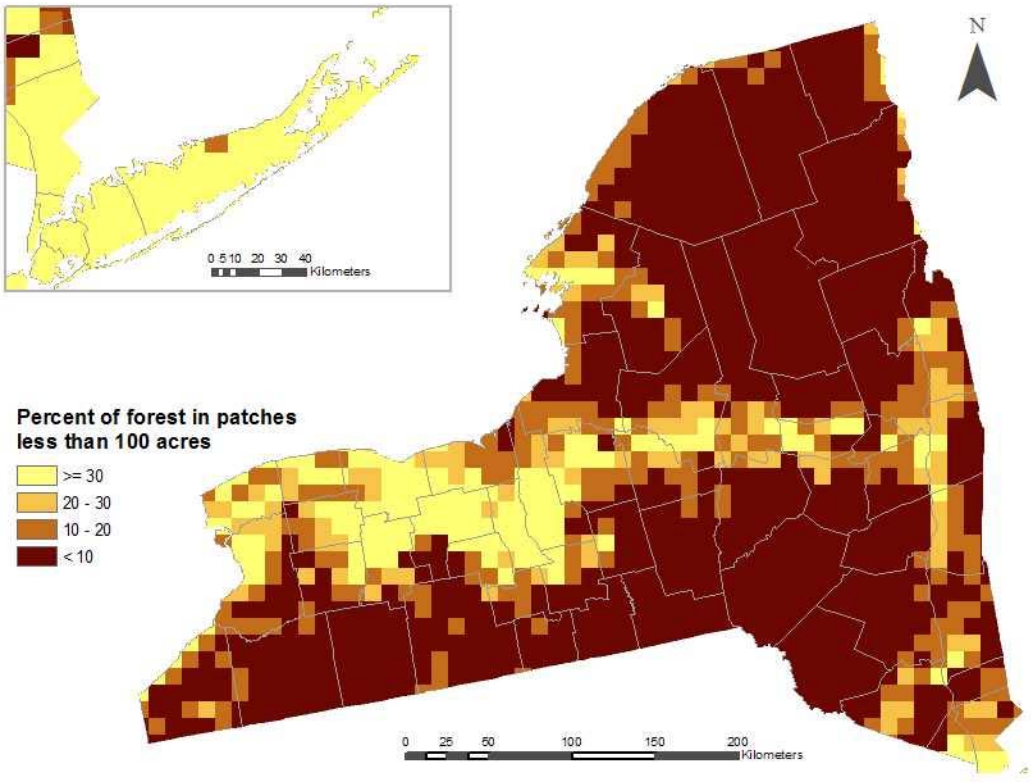


Figure 4a: Percent of forest cover in patches less than 100 acres, by 100 sq. km grid cell, New York, 2001.

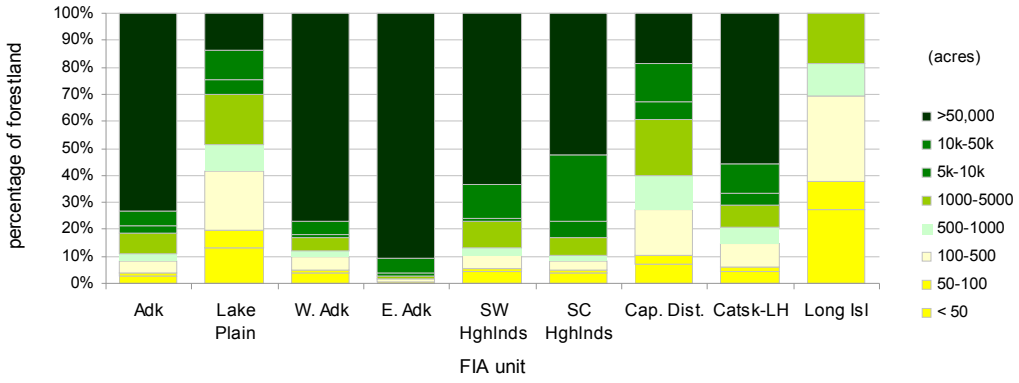


Figure 4b: Distribution of forest cover by patch size, by FIA unit, New York, 2001.

Areas with high proportions of forest area in small patches (patches <100 acres) occur along the river valleys in eastern and central New York, along the shores of Lake Erie and Lake Ontario, and over most of northwestern New York.

Table 2: The distribution of forest land with respect to several urbanization and fragmentation factors, expressed as a percent of the total land area in each unit.

FIA unit	% Forest land in unit	Forest land with house density > 6 per sq. km[a]	Forest land >90 m from an ag or developed edge[b]	Forest land located in patches >100 acres in size[c]	Forest land >300 meters from a road[d]	Forest land located in a block with population densities > 150/sqmi (390/sqkm)[e]
Adirondack	72	23	43	71	43	0
Capital District	59	54	28	59	24	1
Catskill-Lower Hudson	68	56	43	68	31	1
Eastern Adirondack	81	19	60	81	59	0
Lake Plain	41	60	14	41	18	1
South-Central Highlands	65	32	34	61	27	0
Southwest Highlands	64	35	30	64	29	0
Western Adirondack	68	25	36	68	42	0

[a] Approximating the forest land potentially affected by underlying development.
[b] Approximating the forest land undisturbed by edge conditions.
[c] Approximating the forest land with potentially enough core area for sustainable interior species populations.
[d] Approximating the forest land outside the effects of roads.
[e] Approximating the forest land not available for commercial forestry.

Table 2 shows that in the Adirondack unit, which is 72 percent forested, 23 percent of the land area (and 23/72 = 32 percent of the forest) is forest potentially affected by house densities greater than 6 per sq. km, 43 percent of the land area is in forest land that is far enough from an edge to be considered interior forest conditions. Most of that forest is in large patches (>100 acres), but only 60 percent of that forest (43/72) is greater than 300 m from a road (Table 2).

Identifying impact on a particular issue

A second type of analysis that would substantially enhance state reports is the use of published models or thresholds to depict the impact of forest fragmentation and/or urbanization on an ecological, social, or economic issue of particular concern in the state. This can only be done where such management-relevant information exists and data sources are available at the appropriate scale. Although few guidelines are definitive, useful information does exist in several areas so far, including three we have chosen to illustrate in this study: water quality (e.g. Arnold 1996), probability of occurrence of interior bird species (Rosenberg et al. 1999 and 2003), and commercial forestry (Wear et al. 1999). Application of these guidelines provides the user with not just a map of one aspect of fragmentation or urbanization but one that is already interpreted for probable or potential impact on a particular issue. Application of the thresholds established in these guidelines do not represent the final answer, rather they represent the best available knowledge of the impacts to date. To be most useful, such maps can and should be qualified for what they are presenting, as in the examples provided here.

Water Quality: As summarized in Arnold (1996), several thresholds have been identified for the amount of impervious surface that is correlated to a stream's water quality being *impacted* or *nonsupporting*. Applying these thresholds at the same scale as that identified in the literature (approximately HUC 12 basins), and with an understanding of the accuracy of the data source used, reveals a map of probable water quality (Fig. 5). In this map, percent impervious surface values are not calculated directly from the NLCD2001 data source because of known inaccuracies with percent impervious estimates at this scale when compared to photo-interpreted data (Riemann et al. in prep). Instead, basin-level percent impervious values are first modeled for HUC12 basins using the procedure identified in Riemann et al. (in prep).

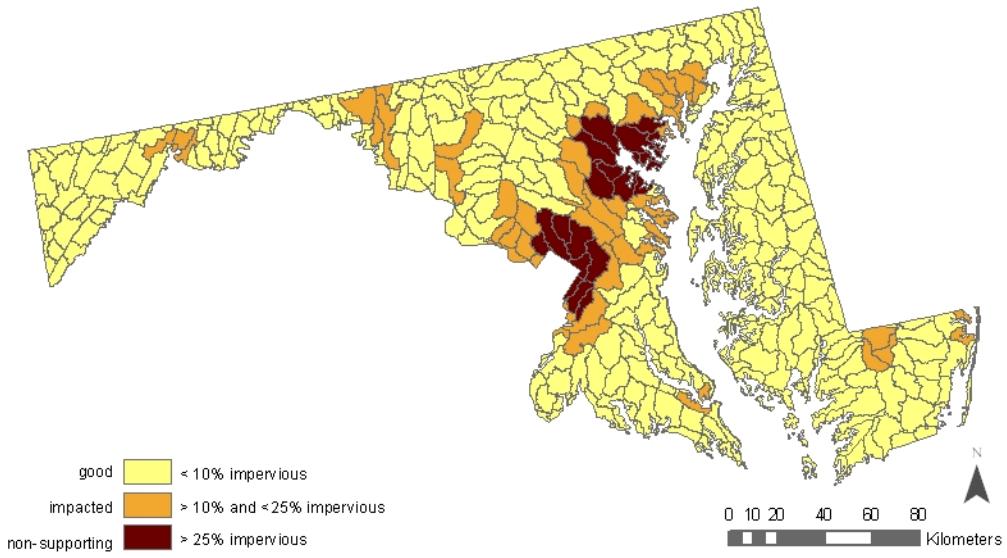


Figure 5: Probable stream water quality as predicted by percent impervious surface alone.

Impervious surface is highly correlated with stream water quality. This is due both to impacts of increased runoff from surfaces that add rather than filter pollutants, and due to its close relationship to increased levels of urbanization that are associated with multiple chemical, physical, and hydrologic changes (Arnold and Gibbons 1996). This example of using impervious surface information illustrates those watersheds in Maryland that are likely to be suffering from impaired stream water quality using thresholds fairly well established in the literature (e.g., Arnold 1996). The map is created by matching as closely as possible both the scale (watersheds no larger than HUC12) and data source (photo-intepreted or ground survey) that were used in the studies identifying the thresholds. Thus, Figure 5 is created using HUC12 watersheds and the NLCD2001 percent impervious layer modified by the observed relationship between the percent impervious in NLCD2001 and percent impervious values from photo-interpreted datasets (Riemann et al. in prep). Though it is an excellent and quantifiable land-use indicator, impervious surface is only one factor, and thus this map does not *predict* stream water quality in each watershed, but rather depicts the probable water quality absent of other mitigating or exacerbating factors. For more accurate information and suggestions for water quality improvement, watersheds in impacted and nonsupporting areas should be examined for mitigating factors that could be improved (e.g., additional forest land, additional tree cover in developed areas, additional forested stream buffer, restoring wetlands). Similarly, watersheds depicted as having very good or impacted water quality should be examined for any exacerbating factors (point sources, more grass than trees in developed areas, highly fragmented forests) that may reflect lower than depicted water quality, or identify landscape factors that could be addressed (Riva-Murray et al. in prep). Local management, regional assessment, and strategic planning efforts would all benefit from such information.

Probability of interior bird species: Rosenberg et al. (1999 and 2003) developed region-specific and species-specific guidelines that describe the probability of finding breeding individuals in a particular forest patch based upon patch size and the proportion of forest land in the surrounding 2500-acre landscape. Figure 6 depicts the results of applying the Atlantic Coast guidelines for scarlet tanager in Maryland and Delaware (Rosenberg et al. 1999). Patches with high habitat suitability have the same probability of supporting tanagers as a suitable unfragmented forest. Patches that are predicted to have a 25 percent lower probability of supporting tanagers are labeled as having moderate suitability, and patches which are 50 percent less likely to support tanagers relative to unfragmented forest are labeled as having low habitat suitability.

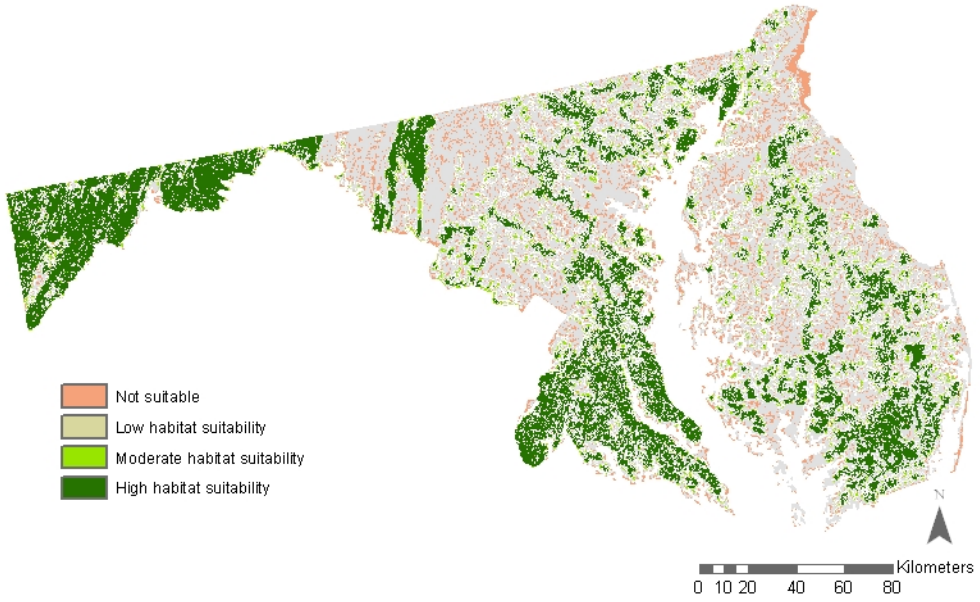


Figure 6: Forest land in Maryland/Delaware shaded by degree of habitat suitability for breeding scarlet tanagers (an interior forest species).

In addition to requirements of forest type, forest habitat suitability depends upon the configuration of forest land. For an interior bird species such as scarlet tanager, this suitability can be described as a function of patch size and the proportion of forest in the surrounding 2500-acre block. Forest type and other forest characteristics are not considered in Figure 6, but future versions of this analysis could easily include FIA modeled forest type and structure data. Future work is needed to study the accuracy of the percent forest and patch size data derived from NLCD2001.

Probability of Commercial Forestry: From a survey of experts in Virginia, Wear et al. (1999) developed a relationship between human population density and the probability that a patch of forest land is used for commercial forestry. More recent studies have reported this general relationship for other areas

(Barlow et al. 1998, Kline et al. 2004), however none have identified thresholds as clearly as Wear et al. (1999).

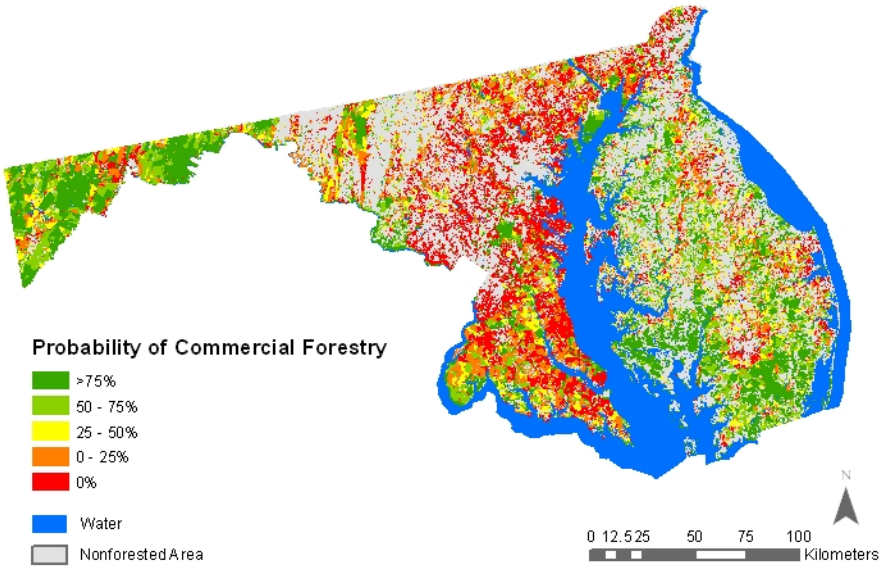


Figure 7: Forest cover by Probability of commercial forestry occurring (based on local population density effects and thresholds identified by Wear et al. (1999) in Virginia and applied in Maryland/Delaware).

Figures 7 and 8 show the relationship between local (block-level) population density and the probability that the forest land in that block will be used for commercial forestry, as developed by Wear et al. (1999).

Generally, harvesting and commercial forest management decline as forest landscapes become more populated and more urbanized. Other factors affecting timber management decisions include proximity to roads, distance to markets, ownership category, parcel size, and nontimber amenity value (see summary in Barlow et al. 1998). The base probability that the forest is under commercial use is 82 percent. Wear et al.'s (1999) study used data from Virginia circa 1991, so probability levels represent conditions in that state at that time. Actual probabilities may be different in Maryland/Delaware and will change as both forest treatments and people's perspectives evolve over time.

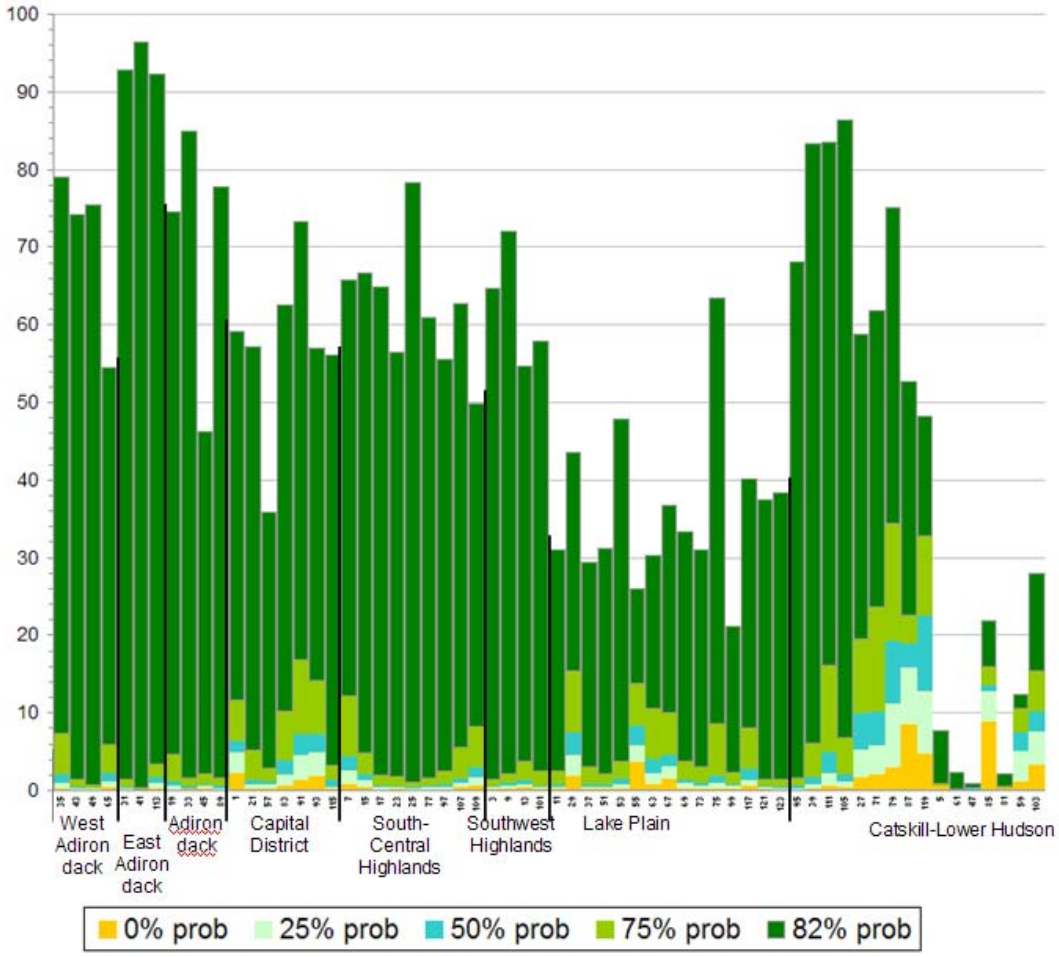


Figure 8: Percent forest in each county, shaded by the probability of forest being used for commercial forestry (based on local population density effects and thresholds identified by Wear et al. (1999) in Virginia and applied in New York, grouped by FIA unit, 2001 (forest), 2000 (population density).

In Figure 8, the dark green shade identifies that portion of the forest land in each county where the probability of commercial forestry is not impacted by human population densities and there is thus roughly an 82 percent probability of commercial forestry occurring. The height of the blue shade (including the orange and pale green) identifies that portion of forest land in each county where there is less than a 50 percent probability of commercial forestry occurring due to local human population densities.

Combining FIA Data with Geospatial Datasets

When data from FIA plots is overlaid with spatial context, forest pattern, and urbanization patterns, valuable information can be obtained. In this example, the forest occurring within the Wildland Urban Interface (WUI) designation represents one segment of forest land that is potentially impacted by urbanization. An analysis of FIA plot data (tree species, stand age, size class, invasive species,

lichens, etc.) against WUI class shows which species are being affected, whether different stand ages or size classes tend to be associated with forest in the WUI area, and whether invasive species or other indicators of ecosystem health are associated with the WUI area (Fig. 9).

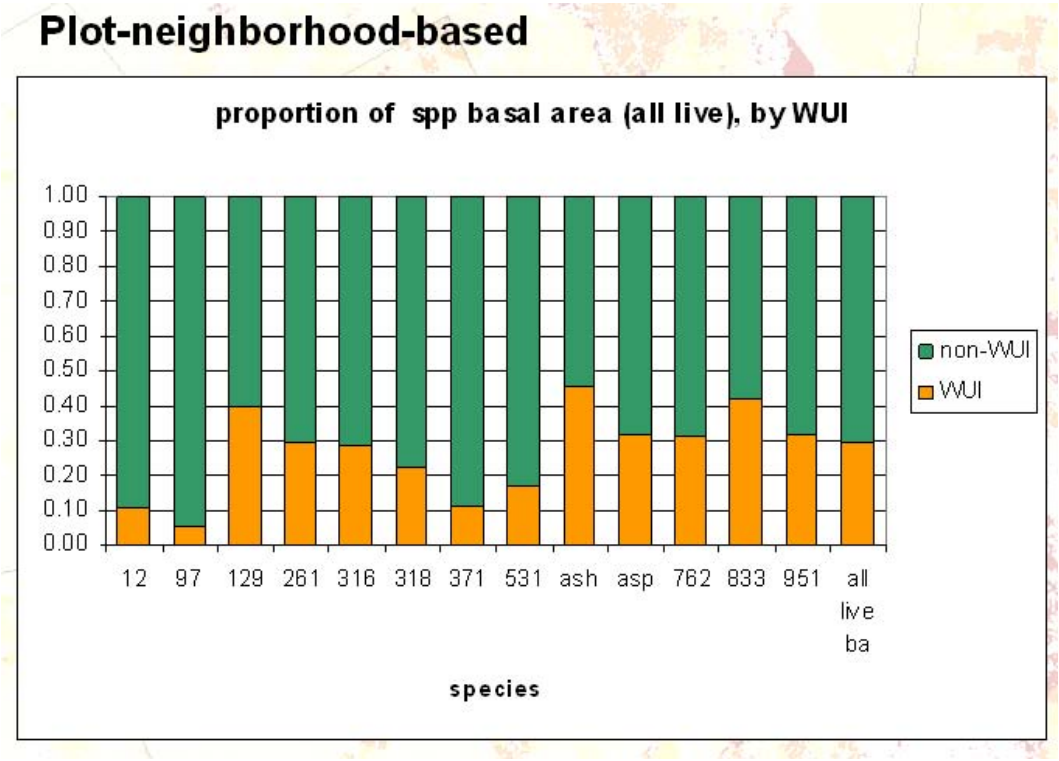


Figure 9: Proportion of species basal area (all live) occurring within the Wildland Urban Interface (WUI), New York, 2000-2005 (annual FIA plots), 2000 (WUI).

The effects of urbanization on forest land are highly dependent on the time that urban development has existed, particularly with large biomass systems such as forested ecosystems. Thus, when looking at a graph of stand size or invasive species vs. housing density or WUI, we might be looking at a resulting effect of that urbanization. More likely, we are describing areas where future changes are expected. Thus an analysis of which species or forest types are most influenced by WUI status, population density, edge conditions, and patch size, probably provides the best look at which are most likely to exhibit future change, experience health problems, suffer a decrease in habitat quality, or be less sustainable in terms of any of the above criteria.

Conclusions

Informed interpretation of forest inventory data requires information regarding the spatial pattern and spatial context in which the forest land occurs.

Characterizing forest land by its landscape context provides information that is currently absent from most FIA state reports. FIA should begin to consistently monitor and report those aspects of fragmentation or urbanization that affect the structure, function, health or sustainability of forests and/or its capacity and value for providing the ecosystem services and products on which we rely.

Information presented in this paper provides examples of what could be done. These have been chosen to provide a suite of relevant, consistent, standardized fragmentation and urbanization measures. Choice of which fragmentation or urbanization metrics to use in a particular state or region will depend on issue priorities, intended use, spatial scale of interest, and the accuracy requirements reflecting the intended use. Wherever both published guidelines and appropriate datasets are available, these can be used to generate relatively specific management-relevant maps and issue-focused analyses based on the best and most current available research. Qualifying statements accompanying the example maps clarify the assumptions and limits of what is expressed in the map while allowing a valuable look at our current best interpretation of impacts.

Using satellite imagery-based datasets in fragmentation analyses in conjunction with FIA data requires an understanding of the difference between the forest definition used by FIA and that expressed by the classified imagery. The relationship between FIA and NLCD2001 percent forest (Fig. 1) illustrates the potential magnitude of these differences. NLCD2001 data provides information on forest cover distribution largely independent of information on developed land uses and roads. This independence is sometimes noted in individual studies as land cover and land use are observed to have separate and independent effects on wildlife, water, and forest ecosystem processes. Thus having land cover and land use information separately available for landscape analyses will likely enable, rather than hinder, more specific application of results. The only caveat is that for the NLCD impervious cover variable, the land cover under trees and shadows is generally not included. As this may also be important, modeling percent total impervious cover to generate values closer to those derived from photo-interpretation may be necessary.

The metrics presented in this paper are simple to obtain. Their strength is in the use of multiple metrics simultaneously, in their relevance with respect to issues of concern, in the use of thresholds identified in the literature to aid interpretation, and in their analysis with FIA data.

References

Airola, T.M.; Buchholz, K. 1984. Species structure and soil characteristics of five urban sites along the New Jersey Palisades. Urban Ecology. 8: 149-164.

Arnold, C.L.J.; Gibbons, C. J. 1996. Impervious surface coverage: the emergence of a key environmental indicator. Journal of American Planning Association. 62: 243-258.

Barlow, S.A.; Munn, I.A.; Cleaves, D.A.; Evans, D.L. 1998. The effect of urban sprawl on timber. Journal of Forestry. 96(12): 10-14.

Bastin, L.; Thomas, C.D. 1999. The distribution of plant species in urban vegetation fragments. Landscape Ecology. 14(5): 493-507.

Bélisle, M.; Desrochers, A.; Fortin, M. 2001. Influence of forest cover on the movements of forest birds: a homing experiment. Ecology. 82(7): 1893–1904.

Burke, D.M.; Nol, E. 2000. Landscape and fragment size effects on reproductive success of forest-breeding birds in Ontario. Ecological Applications. 10(6): 1749–1761.

Cam, E.; Nichols, J.D.; Sauer, J. R.; Hines, J.E.; Flather, C.H. 2000. Relative species richness and community completeness: birds and urbanization in the mid-atlantic states. Ecological Applications. 10(4): 1196–1210.

Chen, J. Q.; Franklin, J. F.; Spies, T. A. 1992. Vegetation Responses to Edge Environments in Old-Growth Douglas-Fir Forests. Ecological Applications. 2(4): 387-396.

De Clercq, E.M.; Vandemoortele, F.; De Wulf, R.R. 2006. A method for the selection of relevant pattern indices for monitoring of spatial forest cover pattern at a regional scale. International Journal of Applied Earth Observation and Geoinformation. 8(2): 113-125.

Faulkner, S. 2004. Urbanization impacts on the structure and function of forested wetlands. Urban Ecosystems. 7(2): 89-106.

Flaspohler, D. J.; Temple, S. A.; Rosenfield, R. N. 2001. Species-specific edge effects on nest success and breeding bird density in a forested landscape. Ecological Applications. 11(1): 32-46.

Forman, R.T.T. 2000. Estimate of the area affected ecologically by the road system of the United States. Conservation Biology. 14: 31-35.

Hammer, R.B.; Stewart, S.I.; Winkler, R.L.; Radeloff, V.C.; Voss, P.R. 2004. Characterizing dynamic spatial and temporal residential density patterns from 1940-1990 across the North Central United States. Landscape and Urban Planning. 69: 183-199.

Hammer, R.B., V.C. Radeloff, J.S. Fried, Stewart, S.I. 2007. Wildland-Urban Interface housing growth during the 1990s in California, Oregon, and Washington. International Journal of Wildland Fire. 16: 255-265.

Heckscher, S.; Hornberger, K.; Mostrom, A.; Marsh, B. 2000. Biodiversity declining: structural and compositional changes in a Pennsylvania Piedmont Forest In: Proceedings of forest fragmentation 2000; 2000 September 17-20; Annapolis, MD. Alexandria, VA: Sampson Group, Inc. : 104-116.

Heilman, G.E. Jr.; 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.

Herrmann, H.L.; Babbitt, K.J.; Baber, M.J.; Congalton, R.G. 2005. Effects of landscape characteristics on amphibian distribution in a forest-dominated landscape. Biological Conservation. 123(2): 139-149.

Hobbs, F.; Stoops, N. 2002. Demographic trends in the 20th century. Census 2000 Special Reports. Washington, D.C.:U.S. Census Bureau.

Homer, C.; Dewitz, J.; Fry, J.; Coan, M.; Hossain, N.; Larson, C.; Herold, N.; McKerrow, A.; VanDriel, N.; Wickham, J. 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogrammetric Engineering & Remote Sensing. 73(4): 337-342.

Hunsaker, C.T.; Levine, D.A.; Timmins, S.P.; Jackson, B.L.; O'Neill, R.V. 1992. Landscape characterization for assessing regional water quality. In: McKenzie, D.; Hyatt, E.; McDonald, J., eds. Ecological indicators. New York, NY: Elsevier Applied Science Publishers: 997-1006.

Iida, S.; Nakashizuka, T. 1995. Forest fragmentation and its effect on species diversity in sub-urban coppice forests in Japan . Forest Ecology and Management. 73(1-3): 197-210.

Irwin, E.G.; Cho, H.J.; Bockstael, N.E. 2007. Measuring the amount and pattern of land development in nonurban areas. Review of Agricultural Economics. 2(3): 494-501.

Kline, J.D.; Azuma, D.L.; Alig, R.J. 2004. Population growth, urban expansion, and private forestry in Western Oregon. Forest Science. 50(1): 33-43.

Lister, A.; Riemann, R.; Lister, T.; McWilliams, W. 2005. Northeastern regional forest fragmentation assessment: Rationale, methods and comparisons with other studies, In: Proceedings of the fifth annual forest inventory and analysis symposium; 2003 November 17-21; New Orleans, LA. Gen. Tech. Rep.WO-69. Washington, DC: U.S. Department of Agriculture, Forest Service: 13-17.

Macie, E.A.; Hermansen, L.A., eds. 2002. Human Influences on forest ecosystems: the southern wildland-urban interface assessment. Gen. Tech. Rep. SRS-55. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 159 p.

McAlpine, C.A.; Eyre, T.J. 2002. Testing landscape metrics as indicators of habitat loss and fragmentation in continuous eucalypt forests (Queensland, Australia). Landscape Ecology. 17(8): 711-728.

McDonnell, M.J.; Pickett, S.T.A. 1990. Ecosystem structure and function along urban-rural gradients: an unexploited opportunity for ecology. Ecology. 71: 1232-1237.

McMahon, G.; Cuffney, T.F. 2000. Quantifying urban intensity in drainage basins for assessing stream ecological conditions. Journal of the American Water Resources Association. 36(6): 1247-1261.

Munn, I.A.; Barlow, S.A.; Evans, D.L.; Cleaves, D. 2002. Urbanization's impact on timber harvesting in the south central United States. Journal of Environmental Management. 64: 65-76.

Nowak, D.J. ;Walton, J. 2005. Projected urban growth (2000–2050) and its estimated impact on the US forest resource. Journal of Forestry. 103(8): 383-389.

Radeloff, V.C.; Hammer, R.B.; Stewart, S.I.; Fried, J.S.; Holcomb, S.S.; McKeefry, J.F. 2005. The wildland urban interface in the United States. Ecological Applications. 15: 799-805.

Riemann, R.; Riva-Murray, K. In prep. Assessing the utility of landscape characteristics calculated from NLCD datasets for watershed analysis: a case study of the impacts on interpretation of stream ecosystem responses.

Riitters, K.H.; O'Neill, R.V.; Hunsacker, C.T.; Wickham, J.D.; Yanke, D.H.; Timmins, S.P.; Jones, K.B.; Jackson, B.L. 1995. A factor analysis of landscape pattern and structure metrics. Landscape Ecology. 10(1): 23-39.

Riitters, K.H.; Wickham, J.D.; O'Neill, R.V.; Jones, K.B.; Smith, E.R.; Coulston, J.W.; Wade, T.G.; Smith, J.H. 2002. Fragmentation of continental United States forests. Ecosystems. 5: 815-822.

Riitters, K.H.; Wickham, J.D. 2003. How far to the nearest road? Frontiers in Ecology and the Environment. 1(3): 125-129.

Riva-Murray, K.; Riemann, R.; Murdoch, P.; Fischer, J.M.; Brightbill, R.A. (in prep). Landscape characteristics affecting streams in urbanizing regions of the Delaware River Basin (New Jersey, New York, and Pennsylvania, U.S.).

Robles, M.D.; Flather, C.H.; Stein, S.M.; Nelson, M.D.; Cutko, A. 2008. The geography of private forests that support at-risk species in the conterminous United States. Frontiers in Ecology and the Environment. 6(6): 301-307.

Rodenhouse, N.; Matthews, S.; Mcfarland, K.; Lambert, J.; Iverson, L.; Prasad, A.; Sillet, T.; Holmes, R. 2008. Potential effects of climate change on birds of the Northeast. Mitigation and Adaptation Strategies for Global Change. 13(5): 517-540.

Rosenberg, K.V.; Rohrbaugh, R.W. Jr; Barker, S.E.; Lowe, J.D.; Hames, R.S.; Dhondt, A.A. 1999. A land managers guide to improving habitat for scarlet tanagers and other forest-interior birds. Ithaca, NY: The Cornell Lab of Ornithology.

Rosenberg, K.V.; Hames, R.S.; Rohrbaugh, R.W. Jr.; Swarthout, S.; Barker, S.E.; Lowe, J.D.; Dhondt, A.A.. 2003. A land managers guide to improving habitat for forest thrushes. Ithaca, NY: The Cornell Lab of Ornithology.

Rudnicky, J.L.; McDonnell, M.J. 1989. Forty-eight years of canopy change in a hardwood-hemlock forest in New York City. Bulletin of the Torrey Botanical Club. 116: 52-64.

Ruefenacht, B.; Nelson, M.D.; Finco, M; Brewer, K. 2008. Harmonizing estimates of forest land area from national-level forest inventory and satellite imagery. In: McWilliams, W.; Moisen, G.; Czaplewski, R. comps. 2008 Forest Inventory and Analysis (FIA) Symposium; 2008 October 21; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. [Pages unknown].

Stein S.; McRoberts, R.E.; Nelson, M.D.; Theobald, D.M.; Eley, M.; Dechter, M. 2006. Forests on the edge: a GIS-based approach to projecting housing development on private forests. In: Aguirre-Bravo, C.; Pellicane, Patrick J.; Burns, Denver P.; and Draggan, Sidney, eds. Monitoring science and technology symposium: unifying knowledge for sustainability in the western hemisphere Proc. RMRS-P-42CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 736-743.

U.S. Census Bureau. 2002. U.S. Census 2000 TIGER/Line® files machine-readable data files. U.S. Census Bureau, Washington, D.C.

USDA Natural Resource Conservation Service. n.d. Natural resources geospatial data gateway. Washington, DC: U.S. Department of Agriculture, Natural Resource Conservation Service. http://datagateway.nrcs.usda.gov/GatewayHome.html. Accessed cc July 2008.

U.S. Forest Service. n.d. FIA data online. Washington, DC: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis Program. http://fiatools.fs.fed.us. Accessed cc July 2008.

University of Wisconson. n.d. SILVIS lab: wildland urban interface maps, statistics and data. Madison, WI: University of Wisconsin. http://silvis.forest.wisc.edu/Library/WUILibrary.asp. Accessed cc July 2008.

University of Wisconson. 2002. SILVIS lab: block-level population, housing and second home density data. Madison, WI: University of Wisconsin. ftp://ftp.silvis.forest.wisc.edu/SILVIS/data. Accessed cc July 2008.

Wade, T.G.; Riitters, K.H.; Wickham, J.W.; Jones, K.B. 2003. Distribution and causes of global forest fragmentation. Conservation Ecology. 7(2): 7.

Wear, D. N.; Liu, R.; Foreman, M.J.; Sheffield, R.M. (1999). The effects of population growth on timber management and inventories in Virginia. Forest Ecology and Management 118(1-3): 107-115.

A Preliminary Investigation of Forest Carbon Changes Associated with Land-Use Change in Northern New England

D. Zheng[1], L.S. Heath[2], M.J. Ducey[1], J.E. Smith[2]

Abstract: *Maine (ME), New Hampshire (NH), and Vermont (VT) are three of the four most heavily forested states in the United States. In these states, we examined how land-use change, at the Anderson Level I classification, affected regional forest carbon using the 30-m Multi-Resolution Land Characteristics Consortium 1992/2001 Retrofit Land Cover Change product coupled with county-level forest carbon stock densities and changes based on U.S. Forest Service, Forest Inventory and Analysis data during the 9-year period. Results indicate that about 1,100 km^2 of forests were newly developed from other land-cover types during 1992 and 2001 across the region, and about 3,100 km^2 of forests were converted to other cover types in the same period, resulting in an apparent net loss of 2,000 km^2 of forest. Thirty percent of land-cover changes occurred within 1.5 km of major roads. Forest land converted to nonforest land area change resulted in apparent carbon (C) loss of 26 million metric tons (10^{12} grams – teragrams (Tg)), nonforest land becoming forest land sequestered 1 TgC and forest land remaining forest land sequestered approximately 154 TgC. Consequently, the regional forests functioned as a carbon sink of 129 TgC over the entire 9-year period. All counties functioned as C sinks during the period, ranging from 0.07 Tg in Grand Isle, VT, to 12.5 Tg in Aroostook, ME. Spatially, 8 of the top 10 counties identified as C sinks were in ME and the other two in NH. In terms of forest carbon loss from deforestation alone, 8 of the top 10 counties were located in ME while the other two were in southeastern NH, where relatively high deforestation rates were detected.*

Keywords: NLCD land-cover maps, land-use change, afforestation, forest land remaining forest land, change detection.

Introduction

Global forests play a dominant role in the terrestrial carbon (C) cycle. They contain 86 percent of the earth's aboveground C and about 73 percent of the C in the world's soil (Post et al. 1982, Olson et al. 1983). Changes in land-use patterns affect C dynamics and balance (Dixon et al. 1994, Houghton 1995). Various models have predicted that the amount of C released from forest ecosystems annually to the atmosphere is positively related to the global deforestation rate (Alcamo et al. 1996, Yamagata and Alexandrov 1999). Turner et al. (1996) estimated that 45 percent of the potential forest cover of the conterminous United States had been converted to other land-cover types. Furthermore, spatial and

[1] Department of Natural Resources and the Environment, University of New Hampshire, Durham, NH 03824, USA, email: daolan.zheng@unh.edu, tel: 603-868-7696
[2] U.S. Forest Service, Northern Research Station, Durham, NH 03824, USA

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

temporal dynamics of forest ecosystems can vary substantially with human-introduced disturbances (such as road accessibility and urbanization). Although ownership can also affect spatial and temporal dynamics of forest ecosystems, it is not the focus of this study.

The Forest Inventory and Analysis (FIA) Program has been using systematic sampling schemes for surveying forest lands across the nation with periodic updates since the early 1930s before changing the survey method to an annualized approach in the early 2000s (Bechtold and Patterson 2005). The FIA field data provide accurate ground measurements for conducting statistical analyses, and for model development and validation; however, these point-based measurements are less suitable for conducting spatially explicit analyses across entire landscapes (Cramer et al. 2001, Zheng et al. 2003). Combining satellite observations with field-based natural resources inventory data will provide more consistent, reliable, and comparable analyses across both spatial and temporal dimensions for national-scale forest and carbon related studies (Nelson et al. 2002, Liknes et al. 2004, McRoberts et al. 2006, Zheng et al. 2008). A previous study demonstrated how land-cover data from different sources could be used for studying regional greenhouse gas dynamics (Brown et al. 2007). Recent collaborations among NASA, FIA, and other Forest Service and university partners indicate that potential benefits from linking the information have begun to be recognized by colleagues, scientific communities, and governmental agencies (Healey et al. 2007).

The National Land Cover Database (NLCD) provides national land-cover maps for 1992 and 2001 using 30-m Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) satellite data (Vogelmann et al. 2001, Homer et al. 2004). Although the classification methods and systems were not identical between the 2 years, one of the guiding principles in the NLCD 2001 map was to ensure that the second generation land-cover product maintained reasonable compatibility with NLCD 1992 map (Homer et al. 2004). Thus, the products are the best resources currently available for detecting land-cover changes between these years at regional and national scales. While direct pixel-to-pixel comparison between two datasets is not recommended (U.S. EPA 2008), the U.S. Geological Survey NLCD design team initiated research to devise an optimal way to compare the products. As a result, the team generated the NLCD 1992/2001 Retrofit Land Cover Change Product using a multistage processing method on the NLCD 1992 and 2001 datasets (MRLC 2008). We obtained the retrofit change product at the Anderson Level I (broader classification categories than those at the Level II used in the original NLCD 1992 and 2001 datasets; Anderson et al. 1976) for use in our regional study.

The states of Maine (ME), New Hampshire (NH), and Vermont (VT) are three of the four most heavily forested states in the country, about 73 percent forested in 1992 and 72 percent in 2001 based on the Retrofit product. While the three states accounted for 1.7 percent of the total land area in the conterminous United

States, they contained 4.6 percent of the total forest land. Therefore, the area has disproportionate importance in the nation's forest ecosystem carbon estimates. The overall goal of this study is to establish baseline information for area changes in land-cover types and forest carbon dynamics in the region. The specific objectives are to: 1) quantify land-use changes at the Anderson Level I in the three northern New England states between 1992 and 2001 focusing on the forest sector; and 2) illustrate how these changes affect countywide and regional forest area and carbon dynamics, as well as their spatial distributions (that is, related to distribution of roads).

Methods and Materials

Study area

The study area contains three northern states (ME, NH, and VT) of the New England area in the United States, with a total land area of 133,100 km^2. About 73 percent of the area is forested. The area is characterized by rolling hills, mountains, and a jagged coastline resulting from retreating ice sheets that shaped the landscape thousands of years ago. Elevation ranges from sea level to 1,917 m at Mount Washington in NH. Dominant forest types include: 1) spruce-fir in northern ME and at high elevation; 2) white/red/jack pine along the coast of southern ME and southeast of NH; and 3) maple/beech/birch in southwestern NH and western VT (Irland 1999). The area is classified as humid continental short and relatively cool summers and long, cold winters. Long-term annual mean temperature is about 4.4 ^{o}C. The average annual rainfall ranges from 500 to 1,000 mm.

Digital maps and data analyses

We downloaded the NLCD 1992/2001 Retrofit Land Cover Change Product provided by the Multi-Resolution Land Characteristics Consortium (MRLC) (MRLC 2008), and extracted the data for our study area. We compared our satellite-based forest area change estimates with those developed from FIA data during the corresponding period at the state level.

To simplify the calculation in carbon dynamics, we based carbon changes on area change categories of nonforest land becoming forest land, forest land remaining forest land, and forest land becoming nonforest land, by county. Because forest carbon is related to forest type, the most common forest type was identified for each county for carbon-related calculations. Since there was a 9-year interval for the area change estimates, we assumed the average age of new forests was 5 years, but a total of 9 years of growth occurs for the area of forest land remaining forest land. To calculate C loss for deforestation, we used the county-level change in forest area, multiplying the county mean forest C densities obtained from the most recent FIA data by a conversion factor of 0.8. This factor is based on the assumption that 80 percent of the nonsoil forest C (including live

tree, stand dead, understory, down dead wood, and forest floor) would be lost during conversion to nonforest (Smith and Heath 2008).

To examine the relationship between land-cover change and road distribution, we used the 2004 national major-road map from ESRI (2008). The map represents interstate, U.S. and state highways, major streets, and other major thoroughfares within the country. We clipped the roads for our study area and created a polygon cover identifying buffer zones within a distance of 1.5 km (one side) from all roads using the GIS function. The buffered roads were overlaid with the Retrofit change map to quantify the relationship.

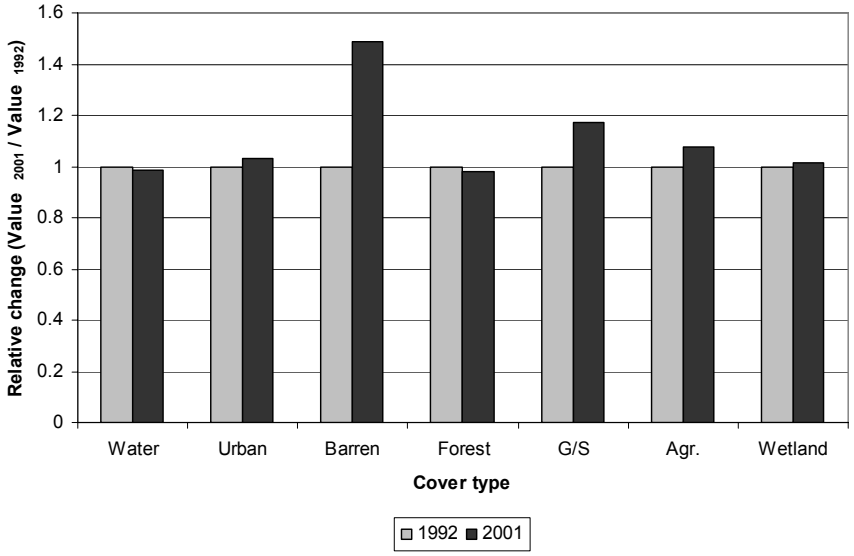


Figure 1. - Relative changes in area between 1992 and 2001 by the seven broad Anderson Level I cover categories from the Retrofit change product using 1992 values as reference (= 1). G/S = Grass/Shrub.

Results and Discussion

Area dynamics and spatial pattern

The most significant relative changes by land-cover types between 1992 and 2001 were found in barren (increased by 49 percent), followed by grass/shrub (17 percent), agriculture (7 percent), urban (3 percent), and forest (-2 percent) (Fig. 1). Across the region, area of water was estimated to decrease by 0.1 percent of total land area (Table 1). Percentage of forest land decreased from 73.3 percent in 1992 to 71.8 percent in 2001 by 1.5 percent of the total land area, at an annual rate of less than -0.2 percent. During the 9-year period, about 3.5 percent of total land area experienced cover-type change. About 91 percent of the land experiencing the change was related to forest. Although these changes occurred across the region and did not exhibit a specific pattern, we found 30 percent of these cover-type changes occurred within 1.5 km of major roads (Fig. 2). Such information is useful for future regional resource planning.

Table 1. - Statistics in 1992 and 2001 for the seven land-cover categories from the Retrofit change product of MRLC, in terms of percentages of total regional land area.

Cover Type	1992 (%)	2001 (%)
Water	4.5	4.4
Urban	4.4	4.5
Barren	0.3	0.4
Forest	73.3	71.8
Grass/Shrub	4.7	5.5
Agriculture	5.6	6.0
Wetland	7.3	7.4

During the 9-year period, about 3,100 km^2 of forest land changed to other types (Table 2). Most of the type changes, in order of amount of area changed, were from forest to grass/shrub (G/S), agricultural land, wetland, and urban (Table 3). About 1,100 km^2 of other types were converted to forests (Table 2), most of these conversions came from G/S, wetland, and agricultural land (Table 3). Other cover types with smaller percentage changes during the period were wetland (2 percent) and water (-1 percent). Overall, the region experienced an apparent net loss of 2,000 km^2 of forests during the 9-year period, at an annual rate of 220 km^2, which is less than 0.2 percent of total land area including water. Considering that all these cover types contain vegetation to some degree, potential uncertainties could be caused by mapping errors in satellite based products. For example, wetlands have proven difficult to map with satellite data because they may be rare in occurrence (4.6 percent in the region, (Stehman et al. 2003)) and their spectral and spatial characteristics are highly context-dependent (Wright and Gallant 2007). Among the 7 Anderson Level I categories in the 1992 NLCD map, wetlands have the lowest user accuracies using center pixel and mode agreement definitions (Stehman et al. 2003).

Spatial heterogeneity in net forest loss across the region between 1992 and 2001 was observed among the states. About 85.9 percent of regional net forest loss occurred in ME, followed by 8.4 percent in NH, and 5.7 percent in VT; this is due to relatively higher deforestation rates in ME, as well as its larger area (Fig. 3; Table 4). On a state basis, forest loss accounted for 2.1 percent, 0.7 percent, and 0.5 percent of the total areas in ME, NH, and VT, respectively, from 1992 to 2001. All counties in ME except Washington and Hancock in eastern ME exhibited some degree of forest loss, from 1 percent to 7 percent (Fig. 3). The other three counties in the region gaining forest area during the period were one in NH (Coos) and two in VT (Caledonia and Essex) (Fig. 3). All five counties gaining forest area exhibit small percentage gains (less than 1 percent) with the maximum of 0.7 percent in Washington County of ME. Two counties, Grand Isle and Lamoille, in VT showed no changes in forest area. Most of the remaining counties in VT and NH (excluding 3 counties with forest gains) had relatively low rates of forest loss (between 0 and 1 percent). The exceptions were two counties in southeastern NH: Hillsborough with 3 percent loss and Rockingham with 5 percent loss (Fig. 3).

Table 2. - Forest area change and carbon (C) dynamics during a 9-year period (1992-2001) in ME, NH, and VT by county (sorted by C sum). The numbers for total areas may slightly differ to those in Table 3 that were derived from regional summary due to rounding.

County	State	Area Status (km^2)			C Status (1000 ton)			C Sum
		Gained	Lost	Remain	Gained[a]	Lost[b]	Remain[c]	
Aroostook	ME	224	649	12119	291	5103	17342	12531
Piscataquis	ME	153	460	7889	196	3908	15620	11908
Penobscot	ME	122	282	6248	156	2148	12371	10379
Somerset	ME	162	696	7141	207	5477	14139	8870
Coos	NH	12	4	4136	15	34	8189	8170
Oxford	ME	25	40	4432	32	378	7339	6993
Franklin	ME	28	114	3611	36	995	7150	6190
Washington	ME	105	72	4469	137	519	6476	6093
Grafton	NH	10	16	3883	13	167	5766	5612
Hancock	ME	76	63	2842	99	520	4067	3646
Essex	VT	6	2	1517	8	17	3004	2995
Carroll	NH	7	9	2091	9	106	2860	2763
York	ME	0	38	1655	0	355	2741	2385
Orange	VT	3	5	1420	4	49	2352	2306
Rutland	VT	6	27	1739	8	298	2582	2292
Windsor	VT	4	46	2045	5	518	2798	2284
Merrimack	NH	14	32	1886	18	371	2580	2227
Caledonia	VT	4	3	1322	5	26	2189	2168
Washington	VT	2	5	1456	3	52	2162	2112
Orleans	VT	3	6	1278	4	57	2116	2063
Kennebec	ME	8	59	1497	10	545	2479	1945
Waldo	ME	9	51	1337	12	449	2214	1776
Cheshire	NH	8	25	1504	10	299	2057	1768
Cumberland	ME	4	37	1429	5	360	2122	1767
Windham	VT	3	27	1730	4	342	2086	1748
Bennington	VT	4	13	1447	5	168	1745	1582
Addison	VT	5	16	1001	6	156	1658	1508
Sullivan	NH	4	21	1141	5	226	1694	1474
Hillsborough	NH	15	57	1544	19	672	2112	1459
Lamoille	VT	1	1	966	1	11	1435	1425
Franklin	VT	3	4	943	4	42	1400	1363
Chittenden	VT	3	6	853	4	64	1267	1207
Belknap	NH	5	8	826	6	81	1227	1152
Androscoggin	ME	6	22	761	8	199	1260	1069
Lincoln	ME	4	38	812	5	367	1345	983
Rockingham	NH	3	59	1017	4	630	1510	885
Strafford	NH	0	15	617	0	158	916	758
Knox	ME	9	27	484	12	228	693	477
Sagadahoc	ME	1	11	383	1	107	569	463
Grand Isle	VT	0	0	44	0	0	65	65
Total		1061	3066	93515	1366	26201	153699	128864

[a] Carbon gain was estimated using carbon accumulation tables for afforestation (Smith et al. 2006), assuming the average age of the new forests in this 9-year period was 5 years.

[b] County-level carbon loss was estimated using average nonsoil forest carbon density by county from the latest FIA data, assuming that 20 percent of the nonsoil carbon remained after forest land became nonforest land.

[c] Carbon for forest land remaining forest land was estimated using carbon accumulation tables for reforestation (Smith et al. 2006) for the most common forest type in the county. The county-level carbon density was used to estimate the expected carbon growth.

Table 3. - Detected land-cover change (km^2) using the Retrofit change product in three northern New England states, U.S.A. G/S = grass/shrub.

	Water	Urban	Barren	Forest	G/S	Agric.	Wetland	Sum1992
Water	5817	2	16	8	5	4	72	5924
Urban	2	5773	1	16	4	48	14	5858
Barren	0	0	340	0	0	0	0	340
Forest	14	233	143	94470	1855	530	326	97571
G/S	1	18	4	695	5462	68	62	6310
Agric.	2	12	1	85	10	7284	15	7409
Wetland	3	5	0	266	41	18	9358	9691
Sum2001	5839	6043	505	95540	7377	7952	9847	133103

We compared our forest area changes detected from NLCD with those calculated from FIA during the corresponding years. Regional estimates from these sources were substantially different from each other. While the NLCD detected a forest net loss of 2,000 km^2 across the region during the period, FIA data showed a loss of 22 km^2 (Table 4). Compared at the state level, the satellite-based results overestimated the forest net loss by 24 percent in ME while underestimating the loss by 73 percent in NH, compared to FIA-based losses. For Vermont, even the sign of the change was opposite. Such a discrepancy may be caused by differences in forest definition and mapping errors. For example, forests in NLCD were defined as areas dominated by trees generally greater than or equal to 5 m tall and greater than or equal to 20 percent of total vegetation cover (Homer et al. 2004), whereas forest lands defined by the FIA are at least 10 percent stocked by trees of any size (Bechtold and Patterson 2005). Reconciliation of this issue requires better coordination and integrity between ground forest inventory and remotely sensed information in future, including 1) increasing plot density; 2) measuring all cover types; and 3) improving remote sensing techniques to reduce mapping errors.

Forest carbon dynamics and spatial pattern

During the 9-year period, regional afforestation sequestered a net 1.4 million metric tons (10^{12} grams – teragrams (Tg)) of carbon (C) and forest land remaining forest land sequestered approximately 153.7 TgC. Regional deforestation resulted in a loss of 26.2 TgC. As a result, the regional forests functioned as a carbon sink sequestering a total of 128.9 TgC (Table 2).

Spatially, all 40 counties in the region functioned as C sinks during the period ranging from 12.5 Tg (Aroostook in ME) to 0.07 Tg (Grand Isle in VT). Eight of

the top 10 counties were in ME and the remaining two were in NH (Table 2). The top 10 sink counties as a whole accounted for 62.4 percent of regional sequestered C during the period due to their larger size (on average 2.4 times larger than the regional mean) and higher forest cover percentages (on average 7 percent higher than the regional mean).

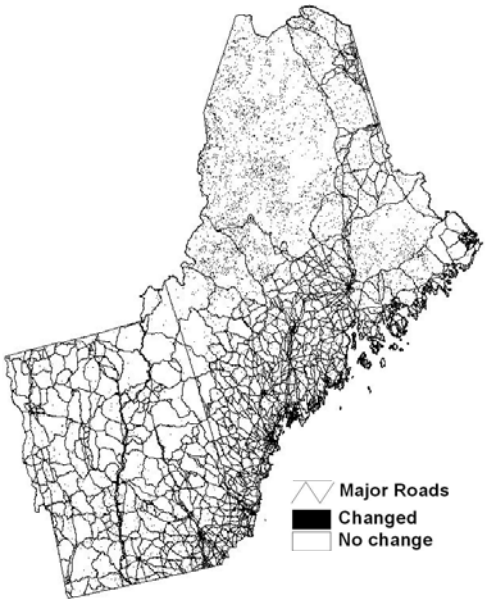


Figure 2. - Relationship of spatial distributions of the lands experiencing changes from one type to another (92-01) and the major roads distributions across the region.

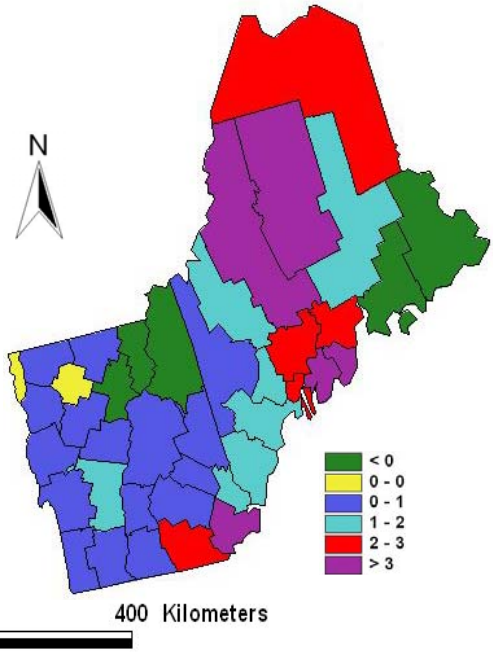


Figure 3. - Spatial pattern of regional forest area changes (92-01) in percentage (%, round to a whole number) at county level. Negative percentages (< 0) indicated gains in forest area.

In terms of C loss due to deforestation, 8 of the top 10 counties were located in ME. The remaining two counties were in southeastern NH where a greater rate of urbanization occurred. For example, areas used for urbanization and development in Rockingham County, NH, accounted for 18.9 percent of its total land area in 2001 compared to that of 4.5 percent for the region according to the Retrofit change product. From the afforestation perspective, 8 of the top 10 counties were located in ME and the remaining two in NH. Only three of the five counties with net forest gains during the 9-year period (Fig. 3) were in the top 10 carbon-sink-county list (Table 2) because 1) all the net gains in area were relatively small; and 2) carbon density for young trees (afforestation) was much smaller than the carbon density in mature forests lost to deforestation.

At the state level, ME had the leading numbers in all categories, both in area and C status: forest gained, lost, and forest land remaining forest land (Table 5) because of its much larger size of forest land than those in the other two states. Maine accounted for 60.1 percent of total regional sequestered C during the 9-year period, followed by 20.4 percent for NH and 19.5 percent for VT.

Table 4. - State-level comparison of forest area changes (km^2) between FIA and NLCD Retrofit change product based calculations in Maine (ME), New Hampshire (NH), and Vermont (VT): 1992-2001.

State	FIA	NLCD
ME	-1388	-1723
NH	-631	-168
VT	1997	-114
Total	-22	-2005[a]

[a] County level based summary.

Table 5. **-** State-level statistics of forest area change and carbon (C) dynamics during a 9-year period (1992-2001) in Maine (ME), New Hampshire (NH), and Vermont (VT). Sums may not match exactly due to rounding.

State	Area Status (km^2)			C Status (1000 ton)			C Sum
	Gained	Lost	Remain	Gained[a]	Lost[b]	Remain[c]	
ME	936	2659	57109	1207	21658	97927	77475
NH	78	246	18645	99	2744	28911	26268
VT	47	161	17761	61	1800	26859	25118

See Table 2 for [a], [b], and [c].

Conclusions

This study illustrates an approach to associate carbon changes with specific categories of cover-type changes. Our results could be used as a reference for monitoring future emissions of CO_2 and forest removals in the region. Spatial patterns identified from this study can provide useful information for improving our existing forest management strategies related to where and how much carbon could be enhanced or reduced. Future management strategies might also need to consider the effects of forest accessibility on greenhouse gas emissions. Our method is simple and straightforward because we used a consistent national land cover change product. Current limitations of this study are that it included neither estimates of carbon in harvested wood products, nor carbon changes from the soil pool. Although further coordination and integrity between field observations and remotely sensed information are needed to reduce potential uncertainties, this spatially explicit approach provides a way to associate changes in carbon with land-cover categories.

Acknowledgments

This study was funded in part by the USDA Forest Service through grant 05-DG-11242343-074. We are grateful to Rebecca Whitney and Elizabeth LaPoint for their GIS assistance.

Literature Cited

Alcamo, J.; Kreileman, G.J.J.; Bollen, J.C.; Van den Born, G.J.; Gerlagh, R.; Krol, M.S.; Toet, A.M.C.; de Vries; H.J.M. 1996. Baseline scenarios of global environmental change. Global Environmental Change. 6: 261-303.

Anderson, J.R.; Hardy, E.E.; Roach, J.T.; Witmer, W.E. 1976. A land use and land cover classification system for use with remote sensing data. USGS Professional Paper 964. Reston, VA: U.S. Geological Survey.

Bechtold, W.A.; Patterson, P.L., eds. 2005. The enhanced Forest Inventory and Analysis Program--national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Brown, S.; Harris, N.; Grimland, S.; Winsten, J.; Sampson, N.; Sohngen, B. 2007. Report submitted to the Wisconsin department of natural resources. Arlington, VA: Winrock International. 30 p.

Cramer, W.; Boudeau, A.; Woodward, F.I.; Prentice, I.C.; Betts, R.A.; Brovkin, V.; Cox, P.M.; Fisher, V.; Foley, J.A.; Friend, A.D.; Kucharik, C.; Lomas, M.R.; Ramankutty, N.; Sitch, S.; Smith, B.; White, A.; Molling, C.Y. 2001. Global response of terrestrial ecosystem structure and function to CO2 and climate change: results from six dynamic global vegetation models. Global Change Biology. 7: 357-373.

Dixon, R.K.; Brown, S.; Houghton, R.A.; Solomon, A.M.; Trexler, M.C.; Wisniewski, J. 1994. Carbon pools and flux of global forest ecosystems. Science. 63: 185-190.

ESRI. 2008. Data: The Raw Materials for Your GIS. http://maps.unomaha.edu/ Workshops/ Career/ESRI/index4.html. [10 October].

Healey, S.; Moisen, G.; Masek, J.; Cohen, W.; Goward, S.; Powell, S.; Nelson, M.; Jacobs, D.; Lister, A.; Kennedy, R.; Shaw, J. 2007. Measurement of disturbance and regrowth with Landsat and Forest Inventory and Analysis data: anticipated benefits from Forest Inventory and Analysis' collaboration with the National Aeronautics and Space Administration and university partners. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P. C.; McWilliams, W.H., eds. Proceedings of the seventh annual Forest Inventory and Analysis Symposium; 2005 October 3-4; Portland, ME. Gen. Tech. Report WO-77. Washington, DC: U.S. Department of Agriculture, Forest Service: 171-178.

Homer, C.; Huang, C.; Yang, L.; Wylie, B.; Coan, M. 2004. Development of a 2001 National Land-cover Database for the United States. Photogrammetric Engineering and Remote Sensing. 70: 829-840.

Houghton, R.A. 1995. Land-use change and the carbon cycle. Global Change Biology. 1: 275-287.

Irland, L.C. 1999. The Northeast's changing forest. Cambridge, MA: Harvard University Press.

Liknes, G.C.; Nelson, M.D.; McRoberts, R.E. 2004. Evaluating classified MODIS satellite imagery as a stratification tool. In: Mowrer, H.T.; McRoberts, R.E.; Van Deusen, P.C., eds. Joint proceedings of the 15th annual conference of the International Environmetrics Society and the 6th international symposium on spatial accuracy assessment in natural resources and environmental sciences; 2004 June 28 to July 1; Portland, ME. Online at http://www.spatial-accuracy.org/PDF/Liknes.pdf. [07 December].

McRoberts, R.E.; Holden, G.R.; Nelson, M.D.; Liknes, G.C.; Gormanson, D.D. 2006. Using satellite imagery as ancillary data for increasing the precision of estimates for the Forest Inventory and Analysis program of the United States of America. Canadian Journal of Forest Research. 36: 2968-2980.

MRLC. 2008. NLCD 1992/2001 Retrofit Land Cover Change Project. http://www.mrlc.gov/multizone.php. [09 October].

Nelson, M.D.; McRoberts, R.E.; Liknes, G.C.; Holden, G.R. 2005. Comparing forest/nonforest classifications of Landsat TM imagery for stratifying FIA estimates of forest area. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; McWilliams,W.H., Cieszewski, C., eds. The fourth annual Forest Inventory and Analysis Symposium. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 121-128.

Olson, J.S.; Watts, J.A.; Allison, L.J. 1983. Carbon in live vegetation of major world ecosystems. ORNL-5862. Oak Ridge, TN: Oak Ridge National Laboratory.

Post, W.M.; Emanuel, W.R.; Zinke, P.J.; Strangenberger, A.G. 1982. Soil carbon pools and world life zones. Nature. 298: 156-159.

Smith, J.E.; Heath, L.S.; Skog, K.E.; Birdsey, R.A. 2006. Methods for calculating forest ecosystem and harvested carbon with standard estimates for forest types of the United States. Gen. Tech. Rep. NE-343. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 216 p.

Smith, J.E.; Heath, L.S. 2008. Carbon stocks and stock changes in U.S. forests. In: U.S. agriculture and forestry greenhouse gas inventory: 1990-2005. Tech. Bull. 1921. Washington, DC: U.S. Department of Agriculture, Office of the Chief Economist: 65-80, C1-C7.

Stehman, S.V.; Wickham, J.D.; Smith, J.H.; Yang, L. 2003. Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: Statistical methodology and regional results. Remote Sensing of Environment. 86: 500-516.

Turner, M.G.; Wear, D.N.; Flamm, R.O. 1996. Land ownership and land cover change in the Southern Appalachian highlands and Olympic peninsula. Ecological Applications. 6: 1150-1172.

U.S. Environmental Protection Agency. 2008. NLCD Changes (NLCD 1992 versus NLCD 2001). http://www.epa.gov/mrlc/change.html [14 October].

Vogelmann, J.E.; Howard, S.M.; Yang, L.; Larson, C.R.; Wylie, B.K.; Van Driel, N. 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 and Remote Sensing. 67: 650-652.

Wright, C.; Gallant, A. 2007. Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data. Remote Sensing of Environment. 107: 582-605.

Yamagata, Y.; Alexandrov, G.A. 1999. Political implications of defining carbon sinks under the Kyoto Protocol. World Resource Review. 11: 346-359.

Zheng, D.; Heath, L.S.; Ducey, M.J.; Smith, J.E. In press. Quantifying scaling effects on satellite derived forest area estimates for the conterminous U.S. International Journal of Remote Sensing.

Zheng, D.; Prince, S.D.; Wright, R. 2003. Terrestrial net primary production estimates for 0.5^0 grid cells from field observations--a contribution to global biogeochemical modeling. Global Change Biology. 9: 46-64.

Wildlife Habitat Applications II



Quantification of Lewis's Woodpecker habitat using Forest Inventory and Analysis data

Chris Witt[1]

***ABSTRACT**: The Utah Department of Natural Resources' Division of Wildlife Resources (UDWR) placed Lewis's Woodpecker (Melanerpes lewis) on their Sensitive Species Tier II list due to declining populations and suspected local extirpations throughout the state. It is thought that the decline in burned coniferous forest has reduced the amount of suitable habitat for these birds, which are known to be closely tied to disturbed landscapes. The UDWR has identified several key components of Lewis Woodpecker habitat using an Ecological Integrity Table (EIT). An EIT identifies "key ecological attributes" such as breeding habitat, food, and their acceptable range of variation as measured by selected indicators. The main indicators for Lewis's Woodpecker breeding habitat were identified as plant association (i.e. forest type), tree canopy cover, and number of suitably-sized snags per hectare of habitat. Each of these indicators is measured in some fashion by FIA crews on every visited phase 2 forested plot. The indicators and their FIA surrogates were used in this analysis to quantify breeding habitat that occurs within and outside of Lewis's Woodpecker's preferred habitat of burned coniferous forests in Utah. We produced area estimates for each indicator thought to be important to Lewis's Woodpecker as well as estimates of forest land that provide all of the structural components in concert. We demonstrated the utility of FIA data for large scale habitat evaluation, identifying limiting attributes, and as a monitoring tool for habitats of sensitive species that use forested landscapes at some point of their life history.*

KEYWORDS: Lewis's Woodpecker, *Melanerpes lewis*, breeding habitat, FIA, monitoring, ecological integrity table

Introduction

Lewis's Woodpecker (*Melanerpes lewis*) occurs throughout western North America where open "park-like" forest stands or recent burns predominate and provide structure that facilitates the bird's mating and foraging strategies (Bock 1970, Sousa 1983, Tobalske 1997). A preference for burned areas has resulted in the species being categorized as a "burn specialist" (Bock 1970, Raphael and White 1984, Block and Brennan 1987, Saab and Vierling 2001). Large numbers of dead and decaying trees, presence of woody ground cover, and open canopy and sub-canopy found in burned pine forests are ideal for this species (Linder and Anderson 1998, Russell et al 2007, Saab et al 2007). Fire suppression policies over the past century have reduced the extent of burned forest landscapes in the

[1] United States Forest Service; Rocky Mountain Research Station; Interior West Forest Inventory and Analysis, 507 25th Street, Ogden, Utah 84401

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Intermountain West. This has resulted in a decline in preferred habitats of Lewis's Woodpecker. However, some unburned ponderosa pine, cottonwood, and aspen stands have been found to provide suitable structure and are used successfully by the bird (Medin and Clary 1991, Tobalske 1997, Vierling 1997, Linder and Anderson 1998, Saab and Vierling 2001). While the use of burned areas by this species has been well documented in some forest types, the qualitative nature of undisturbed forest structure used by Lewis's Woodpecker is still largely unknown, particularly in aspen stands. Although use of ponderosa pine is not uncommon, use of aspen habitats has not been well documented in Utah. From management and census perspectives it is important to know the potential of alternate habitats to provide suitable breeding sites for these birds, given the relative scarcity of recently burned forests in Utah (Figure 1).

The Utah Department of Natural Resources' Division of Wildlife Resources (UDWR) placed Lewis's Woodpecker on the Utah Sensitive Species Tier II List due to declining populations throughout the state as well as suspected local extirpations (Behle et al 1985, Sorenson 1986, Parrish et al, 1999, UDWR 2007). Herein we explore the utility of the Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture Forest Service as an accounting and monitoring tool for forest species habitat in Utah. We use potential breeding and foraging habitat in undisturbed (not burned or cut) forests for Lewis's Woodpecker as a test case.

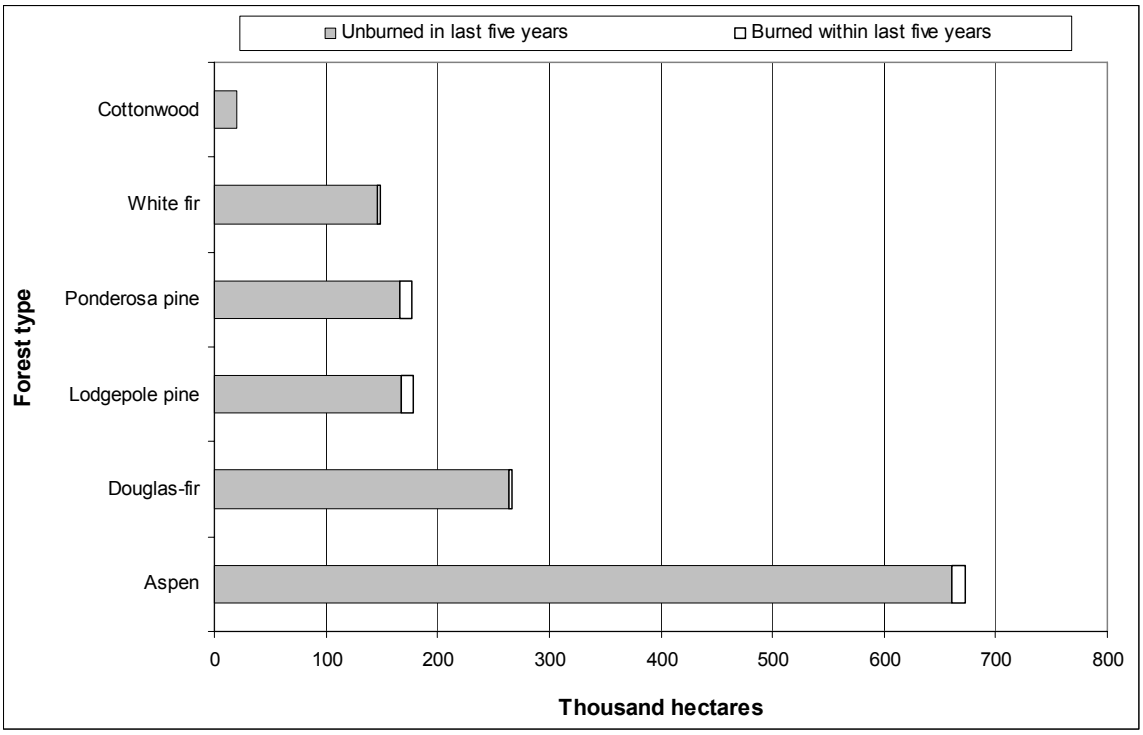


Figure 1: Estimated burned and unburned area of various forest types in Utah. Plots classified as burned have done so within five years of the field crew visit.

Forest Inventory and Analysis Data

FIA is responsible for assessing the status and trends of all forested lands in the U.S. (Gillespie 1999). FIA conducts inventory on all forested lands of the U.S. using a nationally standardized plot design at an intensity of approximately 1 field plot per 2,388 hectares (USDA 2007). The Interior West region of FIA (IW-FIA) is responsible for data collection and analysis in Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming.

In 2000, IW-FIA implemented a continuous annual inventory system (Gillespie 1997). Under annual inventory, approximately 10 percent of plots from the full sample set in a state are measured each year. Plots belonging to an annual panel are distributed evenly across each state so as to be free of geographic bias. States have been gradually phased into the annual system (Utah, 2000; Arizona, 2001; Colorado, 2002; Idaho, 2003; Montana, 2003; New Mexico, 2008), increasing geographic coverage of the Interior West over the past 8 years. Nevada partially employed the annual plot system in 2004 and 2005.

IW-FIA field crews measure and document well over one-hundred plot, condition and tree variables on forested plots. Many of these variables are "national core" variables and are collected in a consistent manner across all regions of the United States. IW-FIA also has several variables that are collected using protocols specific to IW-FIA in order to serve local needs and interests. Some of these variables are used in our analysis as well. For this analysis we used data gathered using annual inventory protocols in Utah during 2000-2006.

Methods

The UDWR has identified several key habitat components of Lewis's Woodpecker in Utah using an Ecological Integrity Table (EIT) (Oliver 2009). An EIT identifies "key ecological attributes", such as breeding habitat and food, and their acceptable range of variation as measured by selected indicators (Parrish et al 2003, UDWR 2005).These tables have been compiled for each species listed in the Utah Comprehensive Wildlife Conservation Strategy (Sutter et al. 2005). The primary indicators for Lewis's Woodpecker breeding habitat were identified as plant association (Bock 1970, Tobalske 1997, Saab and Vierling 2001), tree canopy cover, and number of suitably-sized snags per hectare of habitat (Sousa 1983). The primary indicator for Lewis's Woodpecker foraging habitat was percent cover of woody understory (Sousa 1983, Tobalske 1997). Each of these indicators is measured in some fashion by FIA crews on every visited forested plot. The indicators and their FIA surrogates used in this analysis are listed along with a brief description of methodologies used to acquire the data in table 1. In addition to the IW-FIA variables used to quantify habitat for Lewis's Woodpecker, data collected on land ownership are analyzed for the purposes of discussing where potential Lewis's Woodpecker habitat occurs in Utah and what

management potential exists. A detailed description of IW-FIA field protocols and definitions can be found in USDA (2007).

TABLE 1: Name and definition of habitat indicators from UDWR Ecological Integrity Table for Lewis's Woodpecker, along with their associated FIA variables and a basic description of how they are collected on the plot. Detailed protocols for collecting IW-FIA field data can be found in USDA 2007.

Key Ecological Attribute	Description	FIA surrogate	How measured
Plant Association	Major vegetation class. The two best rankings include ponderosa pine, aspen, cottonwood, disturbed pine, and pine or fir types that can provide open areas between the canopy and understory for foraging. Oak woodlands are identified as preferred winter habitat but not included in breeding habitat.	*Forest Type*	Calculated by algorithms using the plurality of stocking of all live trees tallied on the plot.
% Tree Canopy Cover	Percent of ground shaded by vertical projection of canopies of woody vegetation > 5 m in height.	*Crown Cover*	Measures crown intercept of live trees 2.54 cm or greater diameter at breast height (DBH) on every foot of four 7.62 meter long transects originating from the center of each subplot and running in each of the four cardinal directions.
% Shrub Crown Cover	Percent of ground shaded by vertical projection of canopies of woody vegetation < 5 m in height.	*Tree Cover-Aerial View* and *Shrub Cover-Aerial View*	Ocular estimate to the nearest 1% of the total canopy cover of trees < 2.54 cm DBH on each subplot. The same is done with shrub cover, and then the total cover of small trees and shrubs is added and averaged over the entire plot.
Number of Snags Per Hectare	Number of standing dead trees at least 30.48 cm DBH and 9.14 meters tall per hectare of otherwise suitable habitat.	*Status Code, Lean Code, Actual Height, and Diameter*	*Status Code* records whether the tree is alive or dead, Lean Code records whether the snag is standing at an angle <> 45%, Actual Height records the physical height of the tree, and Diameter records the diameter of the snag at "breast height" (1.37 meters).

The range of each EIT variable is divided into four "Indicator Ratings" ranging from "poor" to "very good". The breaks between each rating are based on available published and professional knowledge of the species. The Lewis's Woodpecker EIT documents some disagreement in the literature concerning the amount of woody understory that is needed to be considered optimal habitat. We report on both of the proposed ratings here. The EIT also acknowledges that although good crown cover for Lewis's Woodpecker is thought to be less than 53 percent, extremely low or the absence of crown cover may not be ideal for the species. Given this ambiguity and the fact that IW-FIA collects the majority of their data on plots with 5 percent or greater canopy cover, we defined a crown cover of 5-52 percent to represent suitable habitat. In consideration of the plant association indicator, those lands described as "non-stocked" by FIA standards were included in the area estimates as long as the field-derived forest type agreed with the EIT definition. Non-stocked lands are those that formerly supported forests but do not have the requisite cover or stocking to be classified to a forest type using the FIA algorithm. More often than not, these lands have been subject to fire, harvest, disease, or other disturbance(s) and likely will re-develop into the field-identified forest type at some point in the future through natural regeneration or plantings. While Figure 1 illustrates the estimated hectares of burned forests in Utah, all estimates provided hereafter will contain all non-stocked hectares regardless of disturbance type.

For this analysis, we considered variable measurements that fell within the two best ratings for each indicator to represent suitable habitat for Lewis's Woodpecker. These indicator rankings are shown in Table 2. An exception to this was snag densities, where IW-FIA methodologies only describe densities equal to or greater than about 2.4 snags per hectare. In order to account for the estimated area with snag densities from .2 - 2.3 snags per hectare, a model,

$$(y = 1/(a + bx^c)$$

was fit to a scatter plot of number of plots in a given forest type versus number of snags recorded on the plot. In this model y is the number of conditions that were estimated to have between .2 and 2.3 snags per hectare, a, b, and c were estimated parameters of the curve, and x was the number of snags tallied on the condition.

The extension of the curve below 2.4 snags per hectare gave a good fit (r^2 = 0.766) and resulted in an upward adjustment of 14 percent for all softwoods and a nine percent adjustment for the aspen forest type. These adjustments are reflected in the final estimates of area satisfying snag density thresholds.

The IW-FIA data for Utah collected in 2000-2006 were queried and plots identified that met each of the rating thresholds. Expansion factors were assigned to each plot and by summing the plots that met each criterion, a state-wide estimate of hectares was produced. All data compilations and summary statistics

were performed using SAS software (Ver. 8.02, SAS Institute, Cary, NC). All estimates expressed in percentages were rounded to the nearest whole percent.

TABLE 2: Indicators of "Key Ecological Attributes" and their qualitative ratings from the Utah Division of Wildlife Resources' Ecological Integrity Table for Lewis's Woodpecker. Oak woodlands and nut and fruit orchards are not considered important breeding habitat and were excluded from this analysis.

Indicator	"Good" and "Very Good" Rating
plant association	oak woodland, nut and fruit orchards, pine-fir, ponderosa pine, cottonwood or aspen riparian, logged or burned forest
% tree canopy cover	≤ 52%
% shrub crown cover < 5 m in height	≥ 25%
% shrub crown cover < 5 m in height	≥ 13%
number of snags ≥ 30.5 cm DBH and ≥ 9.1 m tall per ha	> 1.25

Results

Forest type

The amount of forest land in Utah potentially suitable for Lewis's Woodpecker was estimated at 1,504,111 hectares, or almost 19 percent of all forested lands in Utah (Figure 2). Of the forest types included in the highest indicator rating for this Key Ecological Attribute, aspen contributed the most acreage, with estimates exceeding 672,000 hectares. This accounts for approximately 45 percent of the total estimated suitable acreage in this indicator. The next highest contribution was from Douglas-fir, which approached 18 percent of the acreage at 266,253 hectares.

Crown cover

An estimated 867,343 hectares met both the requisite crown cover and stocked forest type criteria for Lewis's Woodpecker (Figure 3). This accounted for roughly 57 percent of the land estimated to be suitable for Lewis's Woodpecker in terms of stocked forest type. Aspen contributed more to this acreage than any two other forest types combined at approximately 323,543 hectares or roughly 38 percent of the total (Figure 4).

Woody Cover

The amount of suitable stocked forest with adequate crown cover that meets or exceeds the woody cover greater than 25 percent woody understory threshold was estimated to be 417,979 hectares (Figure 5). Aspen accounted for nearly half of this total with an estimated 206,018 hectares. Douglas-fir was the second-most abundant forest type meeting the 25 percent woody cover threshold at 76,987 hectares. The estimate of otherwise suitable habitat that only meets the woody cover greater than 13 percent threshold for woody understory was 686,905 hectares. Aspen forests again provided the bulk of this estimate at 269,442 hectares. Douglas-fir was the next biggest contributor when the woody understory threshold was dropped to 13 percent, providing approximately 123,536 hectares.

Snag Densities

An estimated 122,927 hectares in Utah have adequate snag densities while also meeting the requirements of forest type, crown cover, and woody cover greater than 25 percent (Figure 6). This estimate increased to 222,242 hectares when woody cover greater than13 percent was considered. The aspen forest type provided the majority of the acreage with 44,731 hectares at a woody cover greater than 25 percent threshold and 64,936 hectares at the woody cover greater than13 percent break. Lodgepole pine was the second greatest contributor when woody cover greater than25 percent was considered, while White fir got this distinction when the woody cover greater than13 percent criteria is applied.

Ownership

Figure 7 shows distribution of area that met forest type, crown cover, and woody cover greater than 13 percent thresholds for Lewis's Woodpecker by major ownership groups. Based on our assessment, the Forest Service manages roughly 71 percent of the potential breeding habitat of Lewis's Woodpecker in Utah. The bulk of remaining area is under private ownership, including almost 30 percent of the aspen predicted to be suitable habitat.

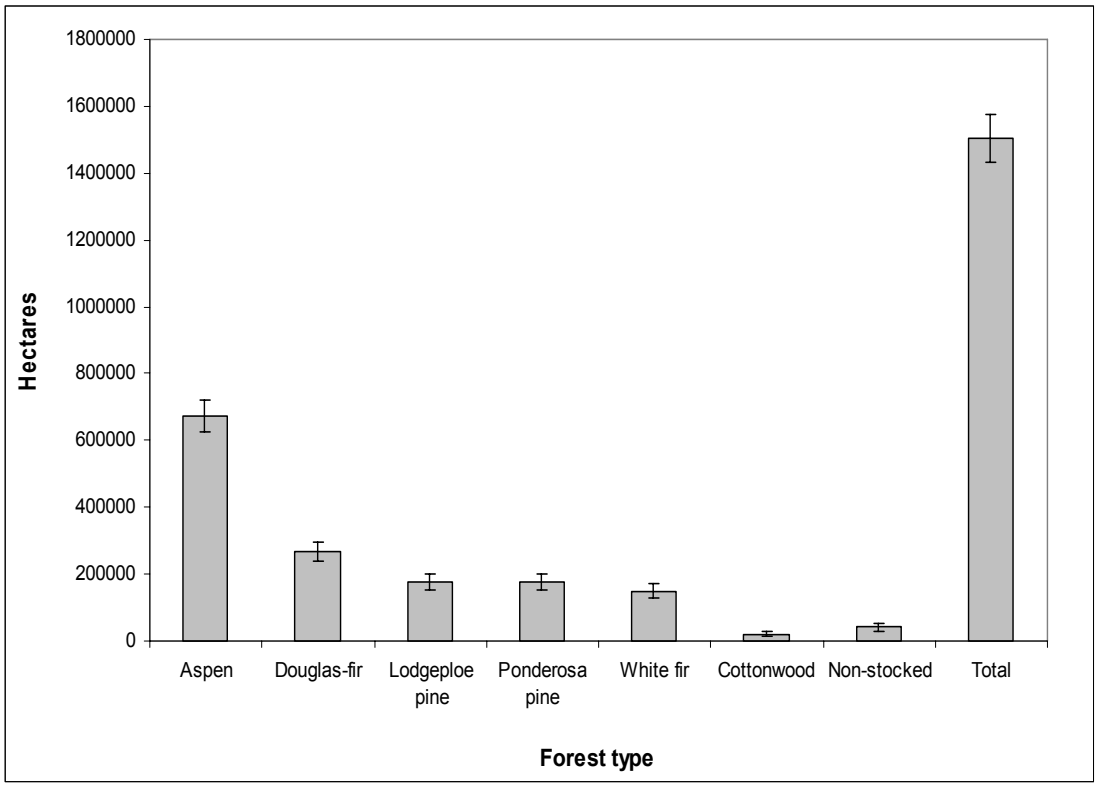


Figure 2: Estimated area of forest land in Utah by FIA forest types thought to be potentially useful to Lewis's Woodpecker for breeding and nesting purposes. Error bars represent the ninety-five percent confidence interval for the estimate.

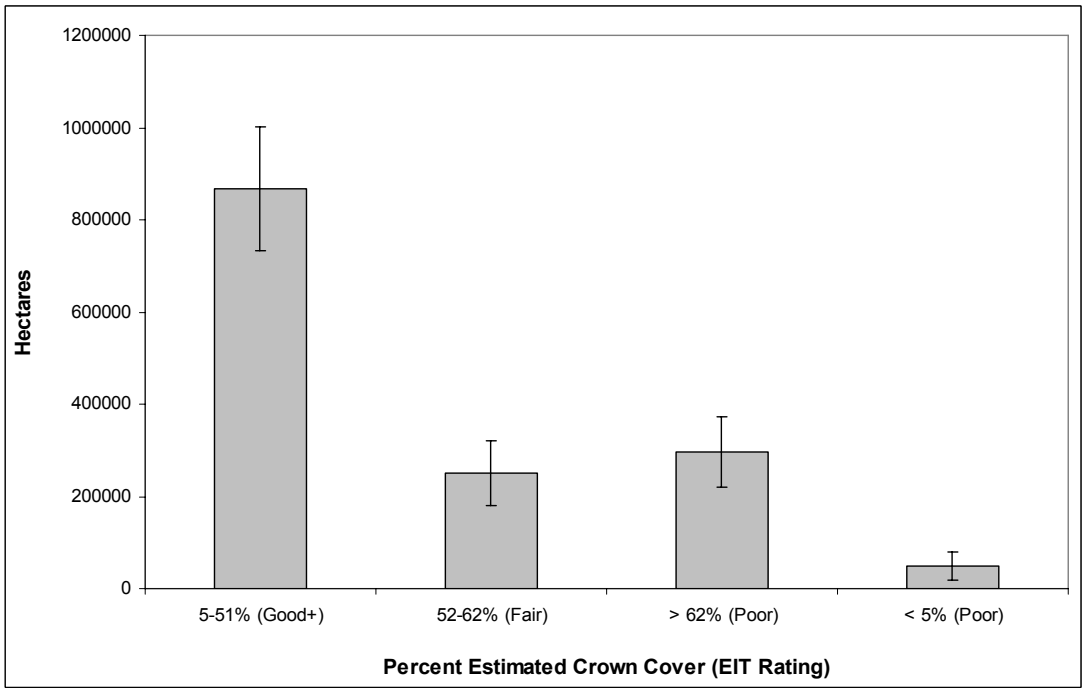


Figure 3: Estimated area of forest land in Utah by forest type and crown cover ratings. Error bars represent the ninety-five percent confidence interval for the estimate.

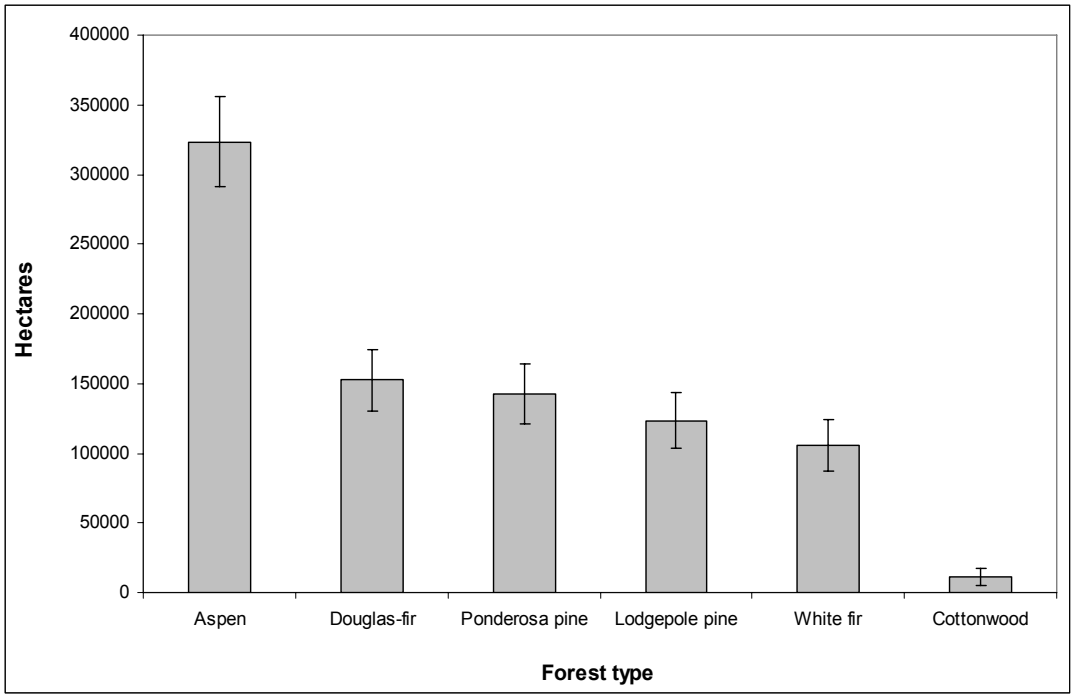


Figure 4: Estimated area of forest land in Utah that meets forest type and crown cover thresholds thought to be useful to Lewis's Woodpecker for breeding and nesting purposes. Error bars represent the ninety-five percent confidence interval for the estimate.

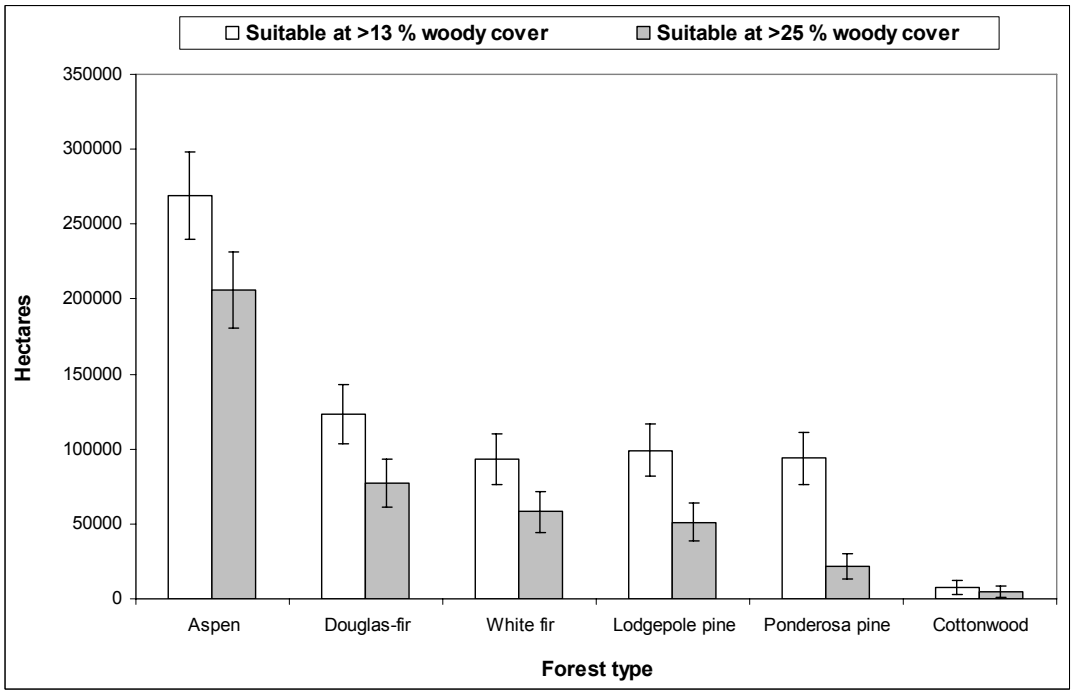


Figure 5: Estimated area of forest land in Utah that meets forest type, crown cover, and woody understory cover thresholds thought to be useful to Lewis's Woodpecker for breeding and nesting purposes. Error bars represent the ninety-five percent confidence interval for the estimate.

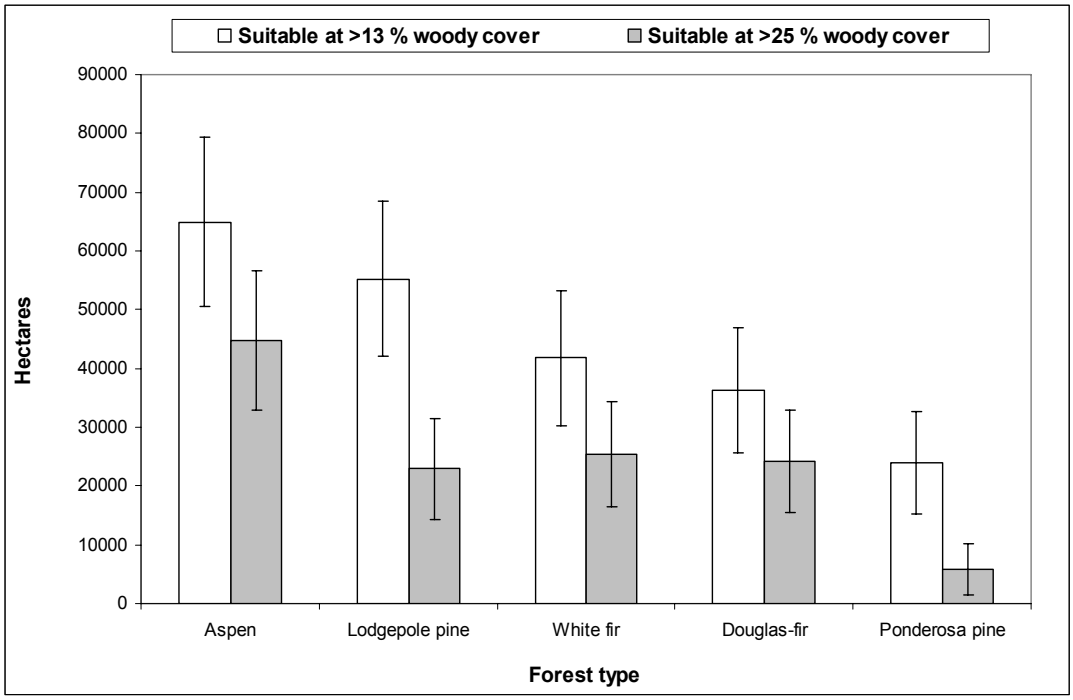


Figure 6: Estimated area of forest land in Utah that meets forest type, crown cover, woody understory cover, and snag density thresholds thought to be useful to Lewis's Woodpecker for breeding and nesting purposes. Error bars represent the ninety-five percent confidence interval for the estimate.

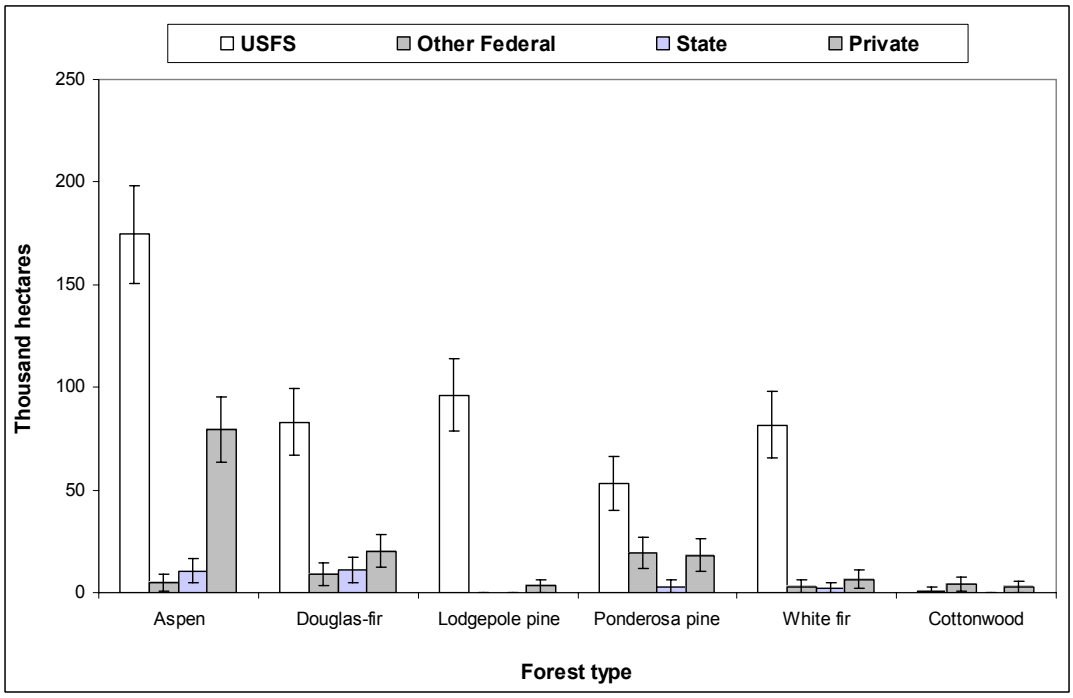


Figure 7: Estimated area of forest land in Utah that meets forest type, crown cover, woody understory cover, and snag density thresholds by land ownership. Error bars represent the ninety-five percent confidence interval for the estimate.

Discussion

The EIT for Lewis's Woodpecker was assembled using the best available published research and local knowledge of the species' ecology. However, research on the species in non-disturbed forests is relatively scarce, and peer-reviewed literature on the species in Utah is very sparse. Consequently, the EIT documents some disagreement in the literature concerning the amount of woody understory that is needed to be considered optimal habitat in Utah. It is not fully known whether Lewis's Woodpecker use Utah's landscape in the same manner as in other parts of their range where they have been observed or studied more intensely. The IW- FIA data presented here suggest that suitable habitat in Utah may be provided by forest types not strongly linked to Lewis's Woodpecker in the literature (Bock 1970, Diem and Zeveloff 1980, Sousa 1983, Vierling 1997, although see Medin and Clary 1991, Neel 1999 and Newlon 2005), although more study may be needed to verify this. In addition, how often Lewis's Woodpecker uses lodgepole pine stands for nesting sites is not well known. FIA data suggest the potential exists for this type to provide the structure these birds are thought to prefer. If they do utilize this resource, the lodgepole pine forests of the Uinta Mountains could provide approximately 99,134 hectares of potential habitat, giving it local importance if not statewide value. Additionally, there are several Douglas-fir stands that co-occur in the Uinta Mountains that scored high in the indicator ratings. These two forest types make up 64 percent of the potential habitat in the Uinta Mountains yet are not strongly associated with Lewis's Woodpecker in the literature.

The results indicate that when both canopy cover and woody understory cover are considered, aspen and Douglas-fir forest types provide the bulk of the preferred structure thought to be important for Lewis's Woodpecker for nesting and feeding activities outside of winter. Ponderosa pine, thought to be the most important forest type other than burned pine forests and cottonwood for these birds during the breeding season (Tobalske 1997), provides less than half the potential habitat that aspen does. When snag densities are considered much of the aspen acreage appears to be unsuitable. However, snag presence should not be as important in aspen types because of the ease of excavation of live aspen trees, the abundance of live trees with diseased portions of the bole, and the number of other excavating bird species that use aspen regularly and create new cavities annually (Dobkin et al, 1995, Daily 1993, Flack 1976). For these reasons, any aspen forest that meets the crown cover and woody cover requirements could be considered potentially suitable habitat.

Tobalske (1997) attributes declining Lewis's Woodpecker populations to the loss of burned stands and riparian cottonwood habitats. This conclusion is supported by the paucity of estimated burned forest land and cottonwood that has the present or future potential to support Lewis's Woodpecker in Utah. However, if the structural preferences of Lewis's Woodpecker are accurately reflected in the EIT, suitable habitat for this species may abound in Utah's unburned aspen

forests. Newlon (2005) showed high reproductive success and survivorship in riparian aspen communities in Idaho and suggested that these communities may provide better overall nesting habitat than do cottonwood forests. This is an important consideration as many land managers and researchers are concerned about the state of aspen throughout the Intermountain West. The same fire suppression activities that have been implicated in reducing availability of traditional habitats for this species may be the inadvertent architect of new habitats in aspen stands as lack of fire in aspen may result in dense canopies, less sub-canopy structure, and a shrubby understory where canopy breaks exist. Current and proposed efforts to convert aspen to younger stands/earlier seral stages in Utah could have both positive and negative effects to Lewis's Woodpecker breeding habitat. Treatments in stands with little or no understory could increase shrub and seedling cover, making it more productive for breeding birds. However, treatments in stands already providing sufficient habitat components would have negative consequences for Lewis's Woodpecker, especially if ponderosa pine and other traditional habitats are not managed in conjunction with aspen "restoration" efforts. It is also important to note that while aspen contributes relatively little to the total forested landscape in the Intermountain West, it is the fourth most common forest type in Utah, trailing only the pinyon-juniper, juniper woodland, and oak woodland types, respectively (USDA, unpublished data). These three woodland forest types are thought to provide very little breeding habitat for Lewis's Woodpecker because they lack the structure that provides good flycatching opportunities, and in the case of the pinyon-juniper and juniper woodland types, they often lack the desired woody understory component (Bock 1970).

There is some uncertainty regarding how much understory vegetation is required to meet the needs of Lewis's Woodpecker (Block and Brennan 1987, Tobalske 1997). It has been suggested that a certain amount of woody cover is required to support the invertebrate communities these birds depend on during breeding season (Bock 1970, Sousa 1983). However, the amount of woody understory necessary may vary by the general forest type selected and its association with riparian systems, agricultural fields, and oak stands. One could also argue that because of the ephemeral nature of insect availability in general, the understory structure directly surrounding the nest tree may not be as important as the nest's proximity to the aforementioned resources. The amount of understory the species requires is an important consideration in Utah, especially as it pertains to ponderosa pine. If woody cover of at least 25 percent is indeed what is required, the estimated amount of predicted suitable ponderosa pine habitat is reduced by roughly 76 percent. Estimates for suitable lodgepole pine are reduced by approximately 59 percent.

No information could be found documenting the foraging range of Lewis's Woodpecker around its nest. Block and Brennan (1987) suggest unburned stands of Jeffrey pine and Douglas-fir in California adjacent to burned areas are used for foraging after nesting season (e.g. after August 1). No nests or individuals were

observed in unburned stands prior to that time period. In addition, while the mean shrub cover of unburned areas used for nesting was 13.4 percent, unburned sites averaged 26.5 percent. These results conflict with the recommendations in Sousa (1983) and agree more closely with other published reports of successful breeding in areas with woody understory cover ranging from as low as 6.7 percent up to 13 percent 13 percent (Medin and Clary 1991, Linder and Anderson 1998).

The extent of the effects of temporal availability of resources between burned and unburned sites on foraging behavior has not been studied. There also may be behavioral traits that restrict adults to foraging in close proximity to their nests until their young have fledged. If insects and other food resources were available in both burned and unburned areas concurrently, nests constructed near the edge of these cover types would have facilitated adult foraging in the unburned areas while maintaining a close proximity to their unfledged young. The fact that no individuals were observed in the unburned stands while the nesting season was in process suggests some behavioral reluctance to venture into these areas. However, distance from nest tree to unburned stands was not reported by Block and Brennan (1987).

Although beyond the scope of this analysis, several FIA variables such as quadratic mean diameter, stand density index, and basal area may prove useful in refining the physical and spatial description of a plot such that other important components could be identified and quantified. The EIT assumes that all "pine-fir" forests have potential to provide habitat to the species, but it is doubtful that all of the types that could be considered pine-fir would provide such structure. In Utah it appears to be rare in the subalpine fir and spruce-fir forest types if anecdotal information reflects the general condition. Regardless of the potential of certain types to exhibit the sub-canopy space needed for flycatching, having variables that actually reflect this spatial arrangement and allow for area estimates of this habitat component within each forest type would be valuable for management purposes and model refinement.

Management Implications

The land area identified as suitable for Lewis's Woodpecker in terms of forest type, crown cover, and woody understory greater than 13 percent is mostly managed by the federal government. An estimated 489,350 hectares of these lands are managed by the United States Forest Service. This represents about 71 percent of the total estimated acreage for Lewis's Woodpecker potential breeding habitat in Utah. As such, an opportunity exists for broad-scale management for the benefit of Lewis's Woodpecker and other birds with similar needs. Site occupancy information would go far to refine and/or reinforce the assumptions made in the EIT and this type of analysis.

Sampling efforts for Lewis's Woodpecker using protocols described in Welsh et al (2006) and Zielinski et al (2006) could increase the utility of FIA data in

modeling wildlife habitats as well. All of these methodologies accommodate the use of FIA data as an estimator of existing habitat and as a monitoring tool of the physical environments important to many forest-dwelling species. FIA data can be useful in tracking trends in one or many important attributes, showing predicted future resource availability with and without a given management prescription, a continuing management policy, or as a result of natural disturbances. The nuances of how FIA data are collected and complied need to be fully understood before being applied to a particular question, but the potential for increased data quality and quantity along with substantial reduction in data collection costs should encourage researchers to explore the utility of using FIA data in their wildlife habitat research.

Conclusion

The information gleaned from this analysis is useful in several ways. The land area identified as suitable for Lewis's Woodpecker in terms of forest type, crown cover, and woody understory > 13 percent is mostly managed by the federal government. An estimated 489,350 hectares of these lands are managed by the United States Forest Service. This represents about 71 percent of the total estimated acreage for Lewis's Woodpecker potential breeding habitat in Utah. As such, an opportunity exists for broad-scale management for the benefit of Lewis's Woodpecker and other birds with similar needs. Site occupancy information would go far to refine and/or reinforce the assumptions made in the EIT and this type of analysis.

Literature Cited

Behle, William H., Sorensen, Ella D., White, Clayton M. 1985. Utah Birds: A revised checklist. Utah Museum of Natural History. University of Utah, Occas Publ 4.

Block, William M., Brennan, Leonard A. 1987. Characteristics of Lewis's Woodpecker habitat on the Modoc Plateau, California. Western Birds 18: 209-212.

Bock, Carl E. 1970. The ecology and behavior of the Lewis's Woodpecker (*Asyndesmus lewis*). University of California Publication of Zoology 92: 1-100.

Daily, Gretchen C. 1993. Heartwood decay and vertical distribution of Red-naped sapsucker nest cavities. Wilson Bulletin 89: 122-129.

Diem, Kenneth L., Zeveloff, Samuel I. 1980. Ponderosa pine bird communities. In: DeGraff, Richard M., tech coord. Proceedings on the workshop on management of Western forests and grasslands for nongame birds; 1980 February 11-14; Salt Lake City, UT. Gen Tech Rep. INT-86. Ogden, UT. US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station: 170-197.

Dobkin, David S., Rich, Adam C., Pretare, Jennifer A., Pyle, William H. 1995. Nest-site relationships among cavity-nesting birds of riparian and snowpocket aspen woodlands in the Northwestern Great Basin. The Condor 97: 694-707.

Flack, J. A. D. 1976. Bird population s of aspen forests in western North America. Ornithological Monographs 19. American Ornithologist's Union, Washington, D.C.
Gillespie, Andrew Jr. 1999. Rationale for a national annual inventory program. Journal of Forestry. 97: 16-20.
Linder, Kathleen A., Anderson, Stanley H. 1998. Nesting habitat of Lewis's Woodpecker in southeastern Wyoming. Journal of Field Ornithology 69 (1): 109-116.
Medin, Dean E., Clary, Warren P. 1991. Breeding bird populations in a grazed and ungrazed riparian habitat in Nevada. Research paper INT-441. Ogden UT. US Department of Agriculture, Forest Service, Intermountain Research Station: 7pp.
Neel, Larry A. 1999. Nevada Bird Conservation Plan. Nevada Partners in Flight, Nevada Department of Wildlife.
Newlon, Karen R. 2005. Demography of Lewis's Woodpecker, breeding bird densities, and riparian aspen integrity in a grazed landscape. M. S. thesis, Montana State University, Bozeman, Montana. 114pp.
Oliver, George V. 2009. Ecological integrity tables for the Utah animals of conservation concern. UDWR Publication No. 09-01, Utah Division of Wildlife Resources, 1594 West North Temple, Salt Lake City, UT 84114-6301.
Parrish, Jeffrey D., Braun, David P., Unnasch, Robert S. 2003. Are we conserving what we say we are? Measuring ecological integrity within protected areas. Bioscience 53 (9): 851-860.
Parrish Jimmie R., Howe, Frank P., Norvell, Russell E. 1999. Utah Partners in Flight Avian Conservation Strategy. Publ. No. 99-40. Utah Partners in Flight Program, Utah Division of Wildlife Resources, Salt Lake City, Utah.
Raphael, M.G. and White, M. 1984. Use of snags by cavity-nesting birds in the Sierra Nevada. Wildlife Monograph 86: 1-66.
Russell, Robin E., Saab, Victoria A., Dudley, Jonathan G. 2007. Habitat-suitability models for cavity-nesting birds in a postfire landscape. Journal of Wildlife Management 71 (8): 2600-2611.
Saab, Victoria A., Russell, Robin E., Dudley, Jonathan G. 2007. Nest densities of cavity-nesting birds in relation to postfire salvage logging and time since wildfire. Condor 109: 97-108.
Saab, Victoria A., Vierling, Kerri T. 2001. Reproductive success of Lewis's Woodpecker in burned pine and cottonwood riparian forests. Condor 103 (3): 491-501.
Sorensen, Ella. 1986. A precipitous decline in Lewis's Woodpecker in Salt Lake and Davis counties. Utah Birds 2: 45-54.
Sousa, Patrick J. 1983. Habitat Suitability Index models: Lewis's Woodpecker. FWS/OBS-82/10.32. U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C. 14pp.
Sutter, Janet V., Anderson, M. E., Bunnell, K. D., Canning, M. F., Clark, A.G., Dolsen, D.E., Howe, F. P. 2005. Utah Comprehensive Wildlife Conservation Strategy (CWCS). UDWR Publication No. 05-19. Utah Division of Wildlife Resources (UDRW), 1594 West North Temple, Salt Lake City, UT 84114-6301. http://wildlife.utah.gov/cwcs/
Tobalske, Brett. W. 1997. Lewis's Woodpecker (*Melanerpes lewis*). In: Poole, A. and Gill, F. Eds., The birds of North America, No. 284. The Academy of Natural Sciences, Philadelphia, and The American Ornithologists' Union, Washington DC.
USDA 2007. Interior West Forest Inventory and Analysis field procedures. Version 3.01. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 420 p.
Utah Division of Wildlife Resources 2007. (2007, December 14- last update) Utah Sensitive Species List [Homepage of Utah Division of Wildlife Resources], [Online].

Available: http://dwrcdc.nr.utah.gov/ucdc/ViewReports/ SSL&Appendices121407/ [2007, December 14].

Vierling, Kerri T. 1997. Habitat selection of Lewis's Woodpeckers in southeastern Colorado. Wilson Bulletin 109 (1): 121-130.

Welch Jr., Hartwell H., Dunk, Jeffery R., Zielinski, William J. 2006. Developing and applying habitat models using Forest Inventory and Analysis data: An example using a terrestrial salamander. Journal of Wildlife Management 70 (3): 671-681.

Zielinski, William J., Truex, Richard L., Dunk, Jeffery R., Gaman, Tom. 2006. Using Forest Inventory data to assess fisher resting habitat suitability in California. Ecological Applications 16 (3): 1010-1025.

Remote Sensing to Support Local to Regional Monitoring and Management Needs



Integrating Landsat-Derived Disturbance Maps with FIA Inventory Data: Applications for State-Level Forest Resource Assessments

Sonja N. Oswalt[1]
Chengquan Huang[2]
Hua Shi[3]
James Vogelmann[3]
Zhiliang Zhu[4]
Samuel N. Goward[2]
John Coulston[1]

Abstract: *Landsat images have been widely used for assessing forest characteristics and dynamics. Recently, significant progress has been made towards indepth exploration of the rich Landsat archive kept by the U.S. Geological Survey to improve our understanding of forest disturbance and recovery processes. In this study, we used Landsat images to map forest disturbances at biennial intervals from 1984 to 2007 for the State of Mississippi. Forest Inventory and Analysis plot data were used to characterize and validate the mapped disturbances. These products were used to produce multi-county forest assessments, which were compared with those derived from inventory data. We conducted our study, in part, to support a larger conterminous U.S.-level LANDFIRE effort.*

Keywords: Disturbance, FIA, LANDFIRE, Landsat, Mississippi.

Introduction

Foresters have long combined remotely sensed data, such as aerial photography, with field-collected data to aid in estimating population means and totals, develop regression equations, and study forest characteristics over varying spatial scales. Remotely sensed data provide a low-cost means for evaluating large areas when compared with increasing costs of plot-based surveys. This is particularly true now that the U.S. Geological Survey has begun providing free user access to historical Landsat data via the Internet.

---

[1] United States Forest Service; Southern Research Station; Forest Inventory and Analysis; 4700 Old Kingston Pike; Knoxville, TN 37919 USA; soswalt@fs.fed.us and jcoulston@fs.fed.us

[2] Department of Geography; University of Maryland; College Park, MD 20742 USA; cqhuang@geog.umd.edu and sgoward@umd.edu

[3] ARTS; Contractor at the USGS Center for EROS; Sioux Falls, SD 57198 USA; hua@usgs.gov and vogel@usgs.gov

[4] United States Forest Service; 1601 North Kent Street; Arlington, VA 22209, USA; zzhu@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

The Landsat program has been in operation since 1972, and imagery from the program has been used for multiple purposes in the fields of forestry, geology, agriculture, and even sociology, among others. The number of scientific articles referencing the Landsat program and its data exceeds 3,200, according to NASA (NASA http://landsat.gsfc.nasa.gov/about/appl.html). This unique time-series imagery, coupled with increasingly sophisticated imagery analysis software, provides scientists with a wide array of information related to vegetation change over time, which previously was possible only through the collection of field data (Cohen and Goward 2004).

The USDA Forest Service, Forest Inventory and Analysis (FIA) program has been collecting data in the field since the mid-1930s. Data collected by FIA includes plot and tree characteristics that may be used to evaluate changes in vegetation characteristics, including data specifically related to plot-level and tree-level disturbances. Field crews attempt to revisit the same plots through time so that trend analysis is possible. While some changes to the FIA plot design have occurred, and some changes in plot locations have been unavoidable, the FIA program still provides one of the best records of field-collected forest inventory data, worldwide.

In this study, we used Landsat images to map forest disturbances at biennial intervals from 1984 to 2007 in the State of Mississippi. FIA plot data were used to characterize and validate the mapped disturbances in a five-county area for disturbance year 2006. These products were used to produce multi-county forest assessments, which were compared with those derived from inventory data. We conducted our study in part to support a larger, conterminous U.S.-level LANDFIRE effort.

Methods

Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) images constituting a nominal temporal interval of 2 years per acquisition were used to map forest disturbances using the vegetation change tracker (VCT) method (Huang and others in press). The VCT process creates two attributes per detected disturbance: disturbance year and disturbance magnitude. For our study's purposes, we focused solely on disturbance year. Disturbance year was defined by the acquisition year of the earliest available Landsat image acquired after the disturbance. For each acquisition year, the VCT algorithm creates a map consisting of 7 categories: 0 = background (nonsampled); 1 = persisting nonforest (nonforest); 2 = persisting forest (forest); 4 = persisting water; 5 = previously disturbed (forest); 6 = current disturbed (nonforest); 7 = prior disturbance (nonforest). In this study, the map acquisition year assessed was 2006.

Five counties were selected for analysis based on the availability of 2006 Landsat thematic imagery for counties in southern Mississippi and those analyses

were based on a prior knowledge of disturbances that occurred in 2005 as a result of Hurricane Katrina. The five counties considered were Hancock, Pearl River, Lamar, Marion, and Forrest Counties along the Mississippi River (figure 1).

Output from the VCT from 2006 was converted from ENVI format to ESRI GRID format then tiled using tools in ArcMap™ Version 9.2. The raster mosaic was then clipped by the five-county polygon layer using Hawth's Tools. We imported Forest Service FIA plots into ArcMap™ and projected them to match the raster mosaic. Forest Inventory plots were located using actual (not fuzzed and swapped) latitudes and longitudes. Hawth's Analysis Tools 3.2 was used to assign a raster category to each forest inventory plot. We then exported FIA plot data and the corresponding disturbance category into comma-delimited files and imported them into SAS® (version 8) for summarization with additional FIA characteristics (e.g. condition and tree records on each plot).

We examined the agreement between FIA field-based classification and the Landsat-based classification using a standard confusion matrix (a tabulation of agreement and disagreement). For this assessment, four categories were considered: forest (VCT categories 2 and 5), nonforest (VCT categories 1, 6, and 7), water (VCT category 4), and nonsampled. In addition to examining agreement,

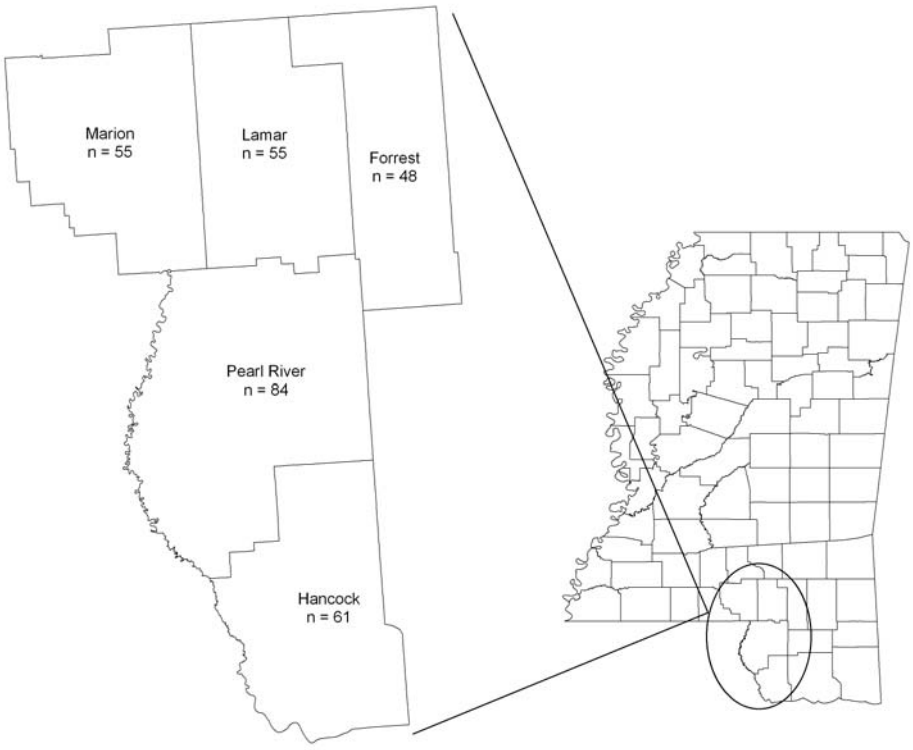


Figure 1: Five-county study area in Mississippi with number of sample plots and county name.

commission errors, and omission errors, we also performed a Kappa analysis. Kappa analysis is a multivariate technique that takes into account both diagonal and off-diagonal elements of the confusion matrix and provides an index of the agreement improvement between the classified image and reference data over that obtained by chance (Congalton and Mead 1983).

We also examined forested plots within each disturbance category for differences in forest structure. The disturbance categories used were persisting forest, post disturbance, and current year (2006) disturbance (categories 2, 5, and 6). Generally, we would expect tree density to be higher in post-disturbance forest and lowest in current year. We used mixed models analysis of variance and least square means to test the null hypothesis of no difference in the average number of trees per acre among VCT disturbance categories, controlling for diameter class, and to test the null hypothesis of no differences in average trees per acre by hardwood and softwood among disturbance categories.

Traditional FIA protocols for recording condition-level disturbances specify minimum impacted areas of at least 1 acre, and mortality and/or damage to at least 25 percent of all trees in a stand or 50 percent of an individual species' count (U.S. Department of Agriculture 2005). The Southern Research Station FIA program added regional protocols for collecting wind-related damage on plots in Mississippi following Hurricane Katrina in 2005 (U.S. Department of Agriculture 2005). The wind damage protocols do specify a minimum area of impact or minimum number of trees. Therefore, if any tree in any condition has been damaged, that condition receives a designation of damaged. Additional variables allow cruisers to indicate the degree of damage that occurred to each tree. We also compared plot level disturbance rates with the VCT-derived disturbance categories.

Results and Discussion

Forest-Nonforest Classification Agreement

We evaluated the agreement of 300 FIA plots in five counties with VCT classification in four categories: forest, nonforest, water, and nonsampled. The overall agreement across categories was 62.7 percent (table 1). With respect to the forest category, 91.9 percent of the pixels classified as forest by VCT also were identified as forest based on the FIA plots. However, the VCT classification had an omission error of 43.7 percent for the forest category. Based on the confusion matrix, 97 plots identified as forest by FIA were classified as nonforest based on the VCT classification. This led to a relatively high commission error rate for the nonforest category (63.3 percent). Disagreements between forest and nonforest classification led to a moderately low Kappa (Khat) value of 31.5. However, about 80 percent of the commission errors likely are due to differences in the definition of forest between FIA and the VCT algorithm (table 1).

Table 1: Forest Inventory and Analysis versus vegetation change tracker forest/nonforest confusion matrix with Khat value

	Vegetation change tracker classification						
Field-based classification	**Forest**	**Nonforest**	**Water**	**Non-sampled**	**Total**	**Agreement**	**Commission error**
						- - - - percent - - - -	
Forest	125	97			222	56.3	43.7
Nonforest	10	58			68	85.3	14.7
Water		2	5	1	8	62.5	37.5
Nonsampled	1	1			2	0.0	100.0
Total	136	158	5	1	300		
Agreement	91.9%	36.7%	1	0	Overall agreement	62.7	
Omission error	8.1%	63.3%	0	1	Khat	31.5	

The national FIA program defines forest land as "land at least 10 percent stocked with forest trees of any size, or formerly having had such tree cover, and not currently developed for nonforest use," (Bechtold and Patterson 2005). Although the area studied is not a naturally sparse forest ecosystem, regenerating forests typically have sparse canopy cover during the first years following disturbance. VCT does not use a specific definition of forest, but it likely maps young, regenerating forest land as post-disturbance nonforest (class 7), even though it meets the FIA definition of forest land and is classified as forest in the FIA database. Remaining commission errors may be attributable to the effects of forest edges on the VCT sample. Because VCT uses tens of images to produce disturbance maps, forest edges are more likely to be classified as nonforest due to residual registration errors in the satellite images.

Characteristics of Landsat Categories based on FIA Plot Data

The average density of trees (trees per acre) was similar among all VCT categories when diameter class was controlled ($p = 0.27$), and there were no notable interactions between diameter and VCT category ($p = 0.37$). Hardwood trees had the highest density among all classifications in the five-county area studied ($p = 0.0083$). When species group was controlled, trees per acre did not differ among VCT category ($p = 0.9$).

Average per-acre basal area was similar between persisting forest and post-disturbance forest classifications ($p = 0.99$). Average per-acre basal area was also similar between the persisting nonforest and nonforest disturbed in the year of measurement ($p = 0.99$). Persisting forest basal area (mean = 90.4) was higher than post-disturbance nonforest ($p = 0.006$; mean = 61.9). Post-disturbance forest basal area (mean = 87.0) was also higher than post-disturbance nonforest ($p = 0.03$). Category 7 post-disturbance nonforest was least like any of the groups, with the lowest average per-acre basal area. Category 6 post-disturbance

nonforest, though not significantly different from the other categories, had the next lowest average per-acre basal area.

Recorded Hurricane Katrina Disturbances

Ninety percent of FIA plots classified as persisting forest had experienced wind damage from Hurricane Katrina in 2005. Ninety-one percent of FIA plots classified as post-disturbance forest had experienced damage from Hurricane Katrina.

Although most plots showed some degree of damage, the proportion of trees on FIA plots that received damage was low, as shown in previous studies (Oswalt and Oswalt 2008). The proportion of trees damaged on plots was similar across all Landsat classifications. Sixteen percent of trees on plots classified as persisting forest had recorded damage, and 22 percent of trees on plots classified as post-disturbance forest were recorded as having had damage, while 20 percent of trees on plots classified as either post-disturbance nonforest or persisting nonforest were recorded as damaged.

Traditional FIA Disturbance Variables

Major disturbances were recorded on 125 of 222 forested FIA plots in the five-county area we studied. VCT detected about 58 percent of the disturbances noted by FIA cruisers (table 2). Weather-related events accounted for 36 percent of the disturbances noted. The majority (89 percent) of those weather events were wind-related, presumably a result of Hurricane Katrina. Sixty percent of FIA plots classified by VCT as persisting forest experienced major disturbances (primarily weather-related, but some fire and disease damage was recorded, as well). Forty-seven percent of plots classified as post-disturbance forest experienced major disturbance, while 49 percent of plots classified as post-disturbance nonforest and 12 percent of plots classified as persisting nonforest experienced major disturbances.

Table 2: Forest Inventory and Analysis versus vegetation change tracker disturbance confusion matrix with Khat value

	Vegetation change tracker classification				
Field-based classification	**No disturbance**	**Disturbance**	**Total**	**Agreement**	**Commission error**
				- - - - - percent - - - - -	
Disturbance	52	73	125	58.4	41.6
No disturbance	38	59	97	39.2	60.8
Total	90	132	222		
Agreement	42.2%	55.3%	Overall agreement	50.0	
Omission error	57.8%	44.7%	Khat	-2.4	

Conclusions

Landsat TM images analyzed using VCT algorithms require further study before they will be useful to the FIA program in a production mode. Several things preclude the ability to incorporate the VCT data into typical FIA outlets at this time. First, the file storage space and processor memory required to join and analyze FIA data in the spatial context of the imagery is larger than many laptops or standard personal computer towers may be able to handle without overloading systems and causing delays, particularly if users attempt to study State-level or larger areas. For example, analyses repeatedly caused system errors while conducting this study using a USDA Forest Service Lenovo model T7400 laptop equipped with an Intel®Core™2 processor, 100GB hard drive, and 2GB of random access memory running on Windows XP Service Pack 2. Landsat VCT images were stored remotely on an external 250GB hard drive. To avoid system errors, we chose to subset the State-level datasets to a more manageable five-county area.

A second potential challenge to using the VCT product in FIA production reports is the learning curve necessary to understand the imagery, merge the imagery with FIA data, and analyze the two datasets simultaneously. Many of the computer programs used to analyze remotely sensed data (e.g. ERDAS, ENVI, ARCGIS) require advanced knowledge of the software that extend beyond the expertise of scientists who do not work with geospatial data on a daily basis. As a result, the process of learning the programs while concurrently learning the product capabilities takes a long time, something that is not suitable for reports that require a quick turnaround.

More work will also be necessary to study the agreement rates between FIA field data and the VCT algorithms. In our study, images analyzed using VCT algorithms detected a little over one-half of the disturbances recorded by cruisers on FIA plots. Some of the disagreement between the image detection and the FIA field call probably can be explained by further examination of change magnitude maps in the context of FIA definitions of disturbance. For example, the VCT algorithm does a better job of detecting stand-clearing disturbances than relatively minor disturbances. Therefore, some disturbances that result in minor defoliation or crown dieback on a large enough scale to meet the FIA definition of disturbance may not cause enough of a change in spectral signature for algorithm detection. Thus, the VCT algorithm may be more useful for detecting harvests, land-use change (reversions and diversions), or massive natural disasters than for detecting other disturbance components.

In addition to disagreement between field-recorded disturbance and VCT-detected disturbance, there was disagreement in terms of field classifications of forest versus nonforest and VCT algorithm forest versus nonforest. Some of the disagreement can be explained by a known weakness in the imagery analysis,

whereby forest edge is assigned to the nonforest category. For example, the fact that basal area values are significantly lower for post-disturbance nonforest VCT categories suggests that either forest definitions differ (e.g. VCT is missing regeneration plots—a known phenomenon) or many of the plots were forest edge and were therefore classified by VCT as nonforest. Therefore, either the FIA data or the VCT algorithm could be further analyzed to try and isolate cases where forest and nonforest conditions exist on FIA plots; and/or pixels obviously are on a forest edge; to evaluate how much forest edge contributed to the agreement issue.

Additionally, forest/nonforest values were extracted from VCT imagery at FIA plot center, while FIA plots actually cover a 3 by 3 pixel square. In the future, it may be best to work with an average of the 3 by 3 area because of pixel-level spatial variation. Another potential explanation for disagreements is simply whether or not forest, as defined by FIA, is equivalent to forest as defined by the VCT algorithm. Finally, Global Positioning System receivers are prone to some error, particularly under a closed canopy. Therefore, FIA plot locations are generally, but not always, within 15m of a plot, introducing another, unavoidable, source of error.

The FIA program strives to quantify both the current status and the change in forest conditions in the United States. Estimators used by FIA require the stratification of forest inventory plots into approximately 4-6 strata. The purpose of stratification is to group like plots together, in order to reduce the variance of estimates. Currently, stratification schemes vary among FIA regions. However, they are all based on geospatial data at any single time. For example, the Southern FIA unit uses the National Land Cover Database to group inventory plots into the following stratum: forest, forest edge, nonforest, and nonforest edge. This stratification scheme does reduce the variance of current estimates of, for example, forest area but it likely does not decrease the variance of change estimates. The VCT data provides an excellent opportunity to stratify field inventory plots for estimating the components of change. We recommend that the VCT data be examined for its use in stratified estimation.

The change detection imagery produced using the VCT algorithm has the potential to be a tremendously useful product when coupled with FIA data, particularly for long-term land-use change analysis, in depth looks at harvest levels, and ecosystem-level analyses. The imagery has the potential to be particularly useful in cases where FIA data may not be sufficient (e.g. fine scales) or where data collection is particularly difficult (e.g. in Alaska). VCT products need to be improved and better characterized before they can be used as production-level products. More studies comparing and contrasting FIA field data with the VCT algorithm results are necessary, as well, to refine the VCT product and to understand how the two relate.

References

Bechtold, W.A.; Patterson, P.L., Editors. 2005. The enhanced forest inventory and analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS–80. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 85 p.

Cohen, W.B.; Goward, S.N. 2004. Landsat's role in ecological applications of remote sensing. Bioscience. 54(6): 535–545.

Congalton, R.G.; Mead, R.A. 1983. A quantitative method to test for consistency and correctness in photointerpretation. Photogrammetric Engineering and Remote Sensing. 49(1): 69–74.

Huang, C.G.; Schleeweis, S.N.; Thomas, K. [and others]. [In press]. Dynamics of national forests assessed using the Landsat record: case studies in Eastern U.S. remote sensing of environment. Remote Sensing of Environment.

Oswalt, S.N.; Oswalt, C.M. 2008. Relationships between common forest metrics and realized impacts of Hurrican Katrina on forest resources in Mississippi. Forest Ecology and Management. 255: 1692–1700.

U.S. Department of Agriculture Forest Service. 2005. Forest inventory and analysis national core field guide. Volume 1: Field data collection for phase 2 plots, version 3.0. [Online]. Arlington, VA: U.S. Department of Agriculture Forest Service, Forest Inventory and Analysis Program. http://fia.fs.fed.us/library/field-guides-methods-proc/docs/2006/core_ver_3-0_10_2005.pdf [July 7, 2008].

Analyzing Landsat Time-series Data Across Adjacent Path/Rows and Across Multiple Cycles of FIA: Lessons Learned in Southern Missouri

Mark D. Nelson[1], Sean P. Healey[2], W. Keith Moser[1], Mark H. Hansen[1], Warren Cohen[3], Mark A. Hatfield[1], Nancy Thomas[4], and Jeff G. Masek[5]

ABSTRACT: *The North American Forest Dynamics (NAFD) Program is assessing disturbance and regrowth in the forests of the continent. These forest dynamics are interpreted from per-pixel estimates of forest biomass, which are produced for a time series of Landsat 5 Thematic Mapper (TM) and Landsat 7 Enhanced TM Plus images. Image data are combined with sample plot data from the Forest Inventory and Analysis (FIA) program using Random Forests, a tree-based estimation method implemented here in the R statistical environment. The NAFD approach is based on a sample of image Path/Rows, resulting in most images being disjunct from and independent of other images in the sample. Increases in sample intensity and needs for assessing forest dynamics over geographic extents larger than a single image are leading to increased frequency of adjacent, overlapping images in the sample. We assessed the consistency of estimates of forest biomass and classification of forest/nonforest in southern Missouri, USA, across space and time, for adjacent images in Path 25/Row 34 and Path 24/Row 34, and for coincident images in Path 25/Row 34 acquired in 2000 and 2007. Results were consistent across space and time, implying consistency of both Landsat and FIA data, and supporting the NAFD image sample strategy and subsequent augmentation with overlapping images.*

KEYWORDS: North American Forest Dynamics, NAFD, Landsat, consistency analyses, forest inventory, FIA

[1] U.S. Forest Service Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108
[2] U.S. Forest Service Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401
[3] U.S. Forest Service Pacific Northwest Research Station, Forestry Sciences Laboratory, 3200 SW Jefferson Way, Corvallis, OR 97331
[4] Department of Geography, 2181 LeFrak Hall, University of Maryland, College Park, MD 20742
[5] Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, MD 20771

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

The North American Forest Dynamics (NAFD) project is an effort to map forest disturbance and regrowth across the continent (Goward et al. 2008). The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture Forest Service recently entered into close collaboration with NAFD, which is funded through the North American Carbon Program.

For NAFD, satellite image time-series data are carefully cross-normalised via state-of-the-art radiometric and geometric image processing procedures (Canty et al. 2004, Masek et al. 2008). The normalised image data are combined with forest inventory data, and a "state model differencing" approach is used to create a spectral model of a biophysical variable, e.g., forest biomass. The model is date-independent because images are selected for near anniversary dates and are cross-normalised; thus, relationships observed between the inventory data and contemporaneous imagery are assumed to hold throughout the time series. A single model is applied to every image in the time series, and estimates of change are obtained as differences in predictions between years. Relatively low prediction errors for state model differencing have been reported in Healey et al. (2006).

NAFD was designed for producing estimates from a sample of Landsat 5 Thematic Mapper (TM) and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) Path/Rows. This approach results in many images multi-temporal image stacks being disjunct and independent from other images in the sample. However, augmentation of the sample and needs for assessing forest dynamics over geographic extents larger than a single image are leading to more and more situations where newly added Path/Rows overlap pre-existing images in the sample.

Classification consistency, determined by comparing overlapping portions of individual Landsat Path/Row scenes, can be used as an indicator of classification quality (Cihlar et al. 2003). Relative accuracy assessments, or 'confidence overlays' complement conventional accuracy assessments that use "ground truth" data (Guindon and Edmonds 2002). In addition to assessing classification consistency, image overlap regions have been used to characterize the accuracy of landscape metrics (Brown et al.2000) and systematic surface reflectance and leaf area index (Butson and Fernandes 2004).

The question is, how consistent are independently modelled NAFD predictions of forest biomass within overlapping portions of adjacent Path/Rows, or between years within the same Path/Row? This study attempts to answer this question for an area of interest in southern Missouri, USA. We created predictive models using NAFD image data and assessed their performance using as prediction and validation input, imagery from: 1) The same image as the original model; 2) An adjacent, overlapping Path/Row image from the same year, e.g., performance

across space; and 3) The same Path/Row, but from a different year, having a similar anniversary date, e.g., performance across time.

Methods

Study Area

We evaluated relationships in three areas of interest within south-central Missouri, USA, generally encompassing the Mark Twain National Forest and surrounding forest land (Fig. 1): 1) The western, nonoverlapping portion of Landsat World Reference System (WRS2) Path 25, Row 34 (P25R34), labeled "A"; 2) The eastern, nonoverlapping portion of Path 24, Row 34 (P24R34), labeled "B"; and 3) The area of overlap between P25R34 and P24R34, labeled "O" (Fig. 2). The geographic extent of overlap in O contains image pixels from both P25R34 and P24R34, labeled “AO” and “BO”, respectively. For FIA plots AO = O = BO.

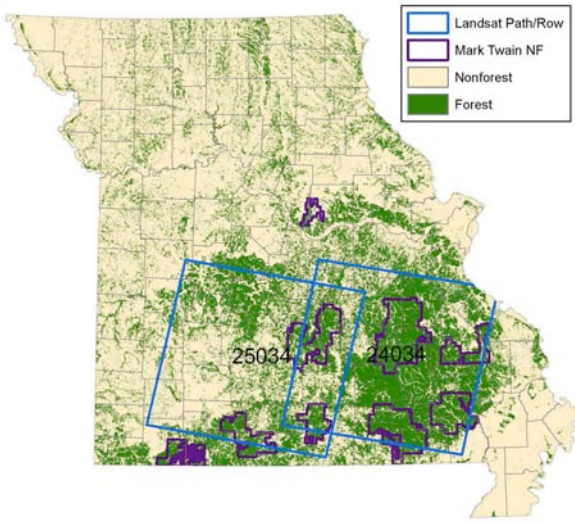


Figure 1: Landsat Path 25 / Row 34 and Path 24 / Row 34, southern Missouri, USA.

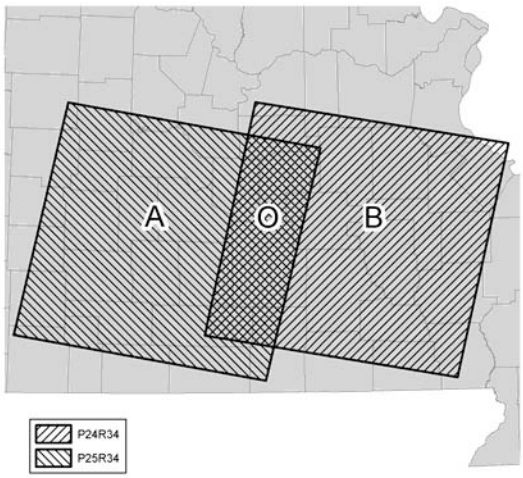


Figure 2: Study areas: A – western, nonoverlapping portion of P25R34, B – eastern, non-overlapping portion of P24R34, and O – extent of overlap between P25R34 and P24R34, southern Missouri, USA.

Data

FIA: FIA defines forest land as lands currently or formerly supporting a minimum level of tree stocking (10 percent) and not developed for a nonforest use such as agriculture, residential, or industrial use. Forest land includes commercial timberland, some pastured land with trees, forest plantations, unproductive forested land, and reserved, noncommercial forested land. FIA's definition of forest land also requires a minimum area of 1 acre and minimum continuous canopy width of 120 ft (U.S. Forest Service 2003). FIA sample plots follow a nationally consistent design comprised of four fixed-radius circular subplots, selected from a nationally consistent hexagonal sampling frame with at least one plot selected for each 6,000-acre hexagon (Bechtold and Scott 2005, Reams et al. 2005). On each FIA plot, land use (e.g., proportion forest cover), tree (e.g., species, height, and diameter) and other site variables are collected.

The FIA database was queried to obtain inventory field plot data collected between 1999 and 2007 within all Missouri counties intersecting P25R34 and P24R34. Geographic information system (GIS) data layers of inventory plot center locations were created based on global positioning system (GPS) coordinates obtained during field data collection. GPS coordinates were collected and maintained in North American Datum of 1983. The sample of FIA plots was further constrained to retain only those plots located within the geographic extent of P25R34 or P24R34. FIA plots measured during Missouri cycle 5 (1999-2003) were used for analyses across space (n = 2320), with 751 plots within A, 1278 plots within B and 291 plots within O (Figure 2). P25R34 plots measured during cycle 5 and remeasured during cycle 6 (2004-2007) were used for analyses across time (n=735). Per-plot estimates were produced for 1) Proportion forest land; and 2) Total gross biomass oven dry weight (pounds per acre) on forest land, based on trees 1.0 inch or larger, including all tops and limbs, but excluding foliage (DRYBIOT in FIA database, hereinafter: 'biomass').

Satellite Imagery: For comparisons across space, a TM image was obtained for P25R34, dated 29 August 2000; and an ETM+ image was obtained for P24R34, dated 30 August 2000 (Fig. 3). Comparisons across time were conducted within P25R34, using the 29 August 2000 TM image and a TM image from 2 September 2007. All three images were converted to surface reflectance by the NAFD Program (Goward et al. 2008) using NASA's Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al. 2008). However, the cross-normalization procedure was omitted for this particular study. Resulting images had 28.5-m spatial resolution and UTM projection, with North

American Datum of 1983. Clouds and shadows, which covered a minimal fraction of the study area, were excluded from further analyses.

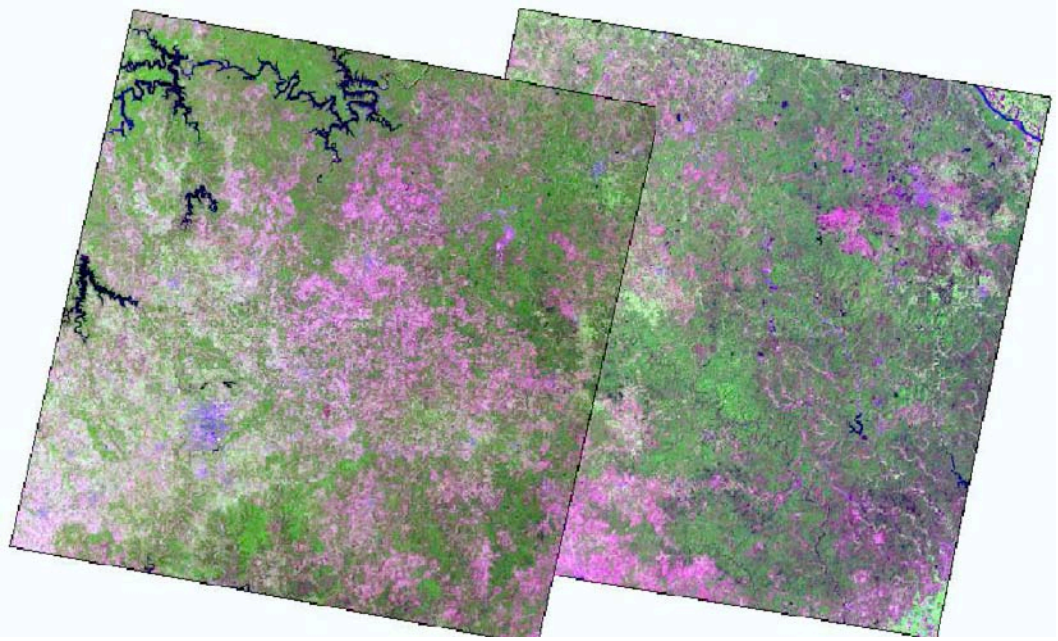


Figure 3: Landsat images of P25R34 (left) and P24R34 (right), late August 2000, southern Missouri, USA.

Modeling

NAFD protocols include methods for making spatially explicit (map-based) estimates of changes in aboveground forest biomass (Healey et al. 2007). Per-pixel predictions of biomass and forest probability were modeled using Random Forests. This work was done with R statistical software (R Development Core Team 2008) using an implementation of Random Forests developed by Liaw and Wiener (2002) and adapted by Freeman and Frescino (2008). For this approach, 2000 trees were created, with each tree using a different bootstrap sample of the data, and the best among a random subset of predictors selected for splitting each node. These trees were assembled into a 'forest' and each tree provides a 'vote' on the final, composite tree. Pixels with predictions of forest probability of 0.5 or greater were labeled forest class; all other pixels were labeled nonforest.

Across Space (within year 2000): To ensure an adequate number of validation plots in O, which is smaller in area than A or B, we chose a lower density of training plots and thus a higher density of validation plots. This ratio was reversed for A and B, where geographic extent and numbers of FIA plots were larger. Sixty percent of the FIA plots within A and 40 percent of the FIA plots within O were selected at random as model data for training P25R34. The remaining 40 percent of plots within A and 60 percent of plots within O were retained for validation

analyses. A model for P25R34 was applied to: 1) Spectral data from AO (i.e., image pixels from the same image for which the model was developed); and 2) Spectral data from BO (i.e., image pixels from the adjacent image - P24R34). The same set of FIA validation plots from O were used to assess both applications of the model.

The procedure was duplicated exactly with data from scene B to create predicted maps of biomass and forest proportion in area O using imagery from both A and B, thus allowing for comparison of predictions over the same area made with 2 different imagery sources. Within O, the same model and validation plots were used for both P25R34 and P24R34 models and validation tests.

Trading Space for Time (within P25R34): For the year 2000 model, 60 percent of the FIA plots from Missouri cycle 5 (1999-2003) were selected at random for model development and the remaining 40 percent of plots (n ~ 300) were retained for model validation. Similarly, for the year 2007 model, 60 percent of the FIA plots from Missouri cycle 6 (2004-2007) were selected at random for model development and the remaining 40 percent of plots (n ~ 300) were retained for model validation.

A model based on 2000 imagery was applied to: 1) Spectral data from 2000 (i.e., image pixels from the same image for which the model was developed); and 2) Spectral data from 2007. Similarly, a model based on 2007 imagery was applied to: 1) Spectral data from 2007 (i.e., image pixels from the same image for which the model was developed); and 2) Spectral data from 2000.

Validation

Site-specific (per-pixel) validation tests were conducted to assess overall classification accuracy of forest/nonforest (f/nf) classes, and root mean square error (RMSE) and coefficient of determination (R-squared) for predictions of biomass. Results of these tests were compared to determine effects on model performance of substituting model image input across space and across time.

When comparing models across space and across time, a naming convention is employed whereby the first term refers to the image source for model development, and the second term refers to the imagery for which models are implemented. Across-space comparisons are termed A-B or B-A, and across-time comparisons are termed 2000-2007, or 2007-2000. Model validation is termed A-A, B-B, 2000-2000, or 2007-2007 when using the same image source for model development and testing.

Results

Across Space

Overall accuracy of the f/nf models ranged from 92 to 94 percent (Fig. 4). Biomass predictions had R-squared values of 0.72 to 0.74, and RMSE (percent of mean) values of 0.56 to 0.60 (Fig. 5). In some cases, substituting imagery across space resulted in slightly higher accuracies (Figs. 4 and 5).

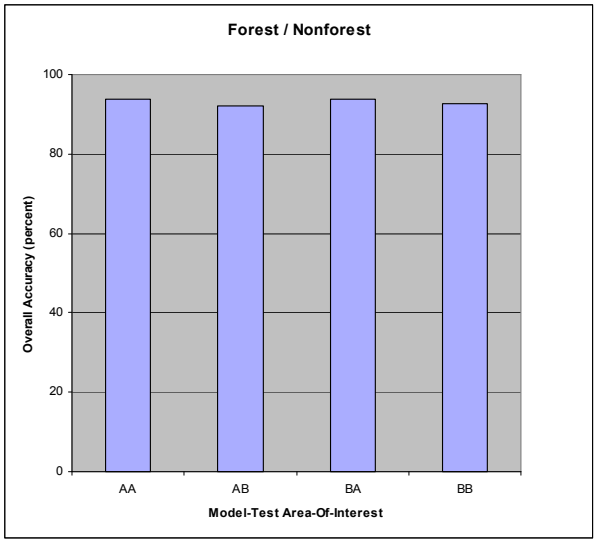


Figure 4: Overall classification accuracy of forest/nonforest classifications, 2000, southern Missouri, USA.

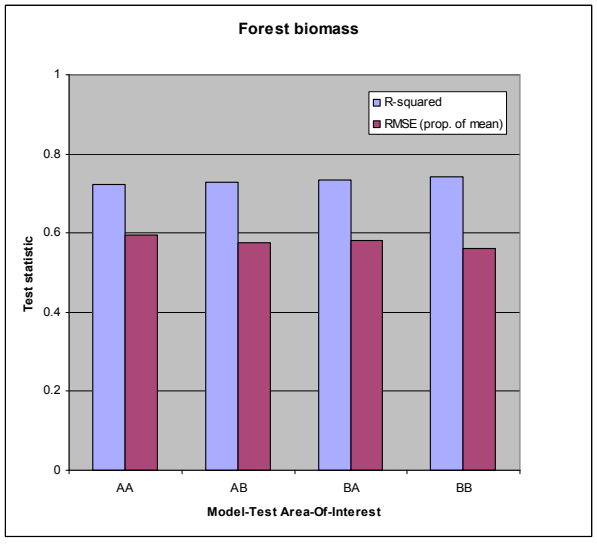


Figure 5: Error metrics for predictions of forest biomass, 2000, southern Missouri, USA.

Across Time

Overall accuracies of forest/nonforest classifications ranged from 88 to 90 percent (Figure 6). Biomass predictions had R-squared values of 0.74 to 0.79, and RMSE (percent of mean) values of 0.64 to 0.72 (Figure 7). Differences between the models built in 2000 and 2007 were negligible when it came to predicting the 2000 and 2007 test data. In some cases, the "off year" gave fractionally better predictions than models produced in the same year (Figs. 6 and 7).

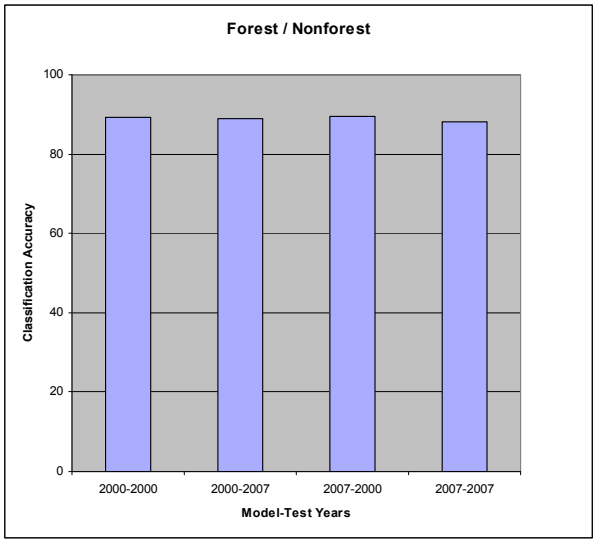


Figure 6: Overall classification accuracy of forest/nonforest classifications, P25R34, southern Missouri, USA.

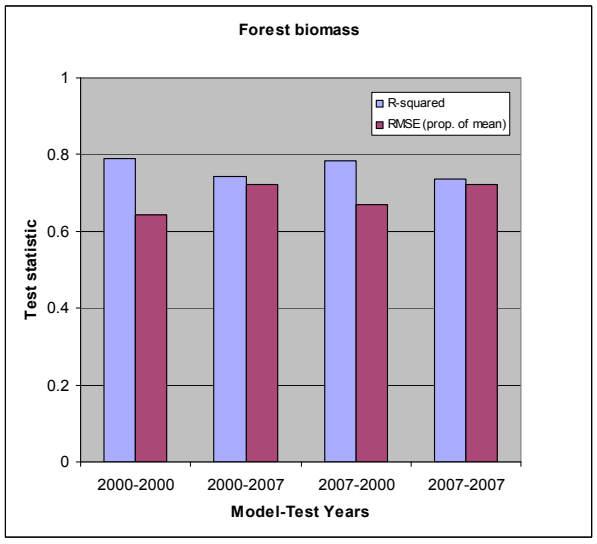


Figure 7: Error metrics for predictions of forest biomass, P25R34, southern Missouri, USA.

Discussion

Using NAFD procedures, forest/nonforest classifications and biomass predictions exhibited minimal effects across space (i.e., WRS Path/Row [Path 25 vs. Path 24]), or across time (i.e., FIA inventory cycle [cycle 5 vs. cycle 6]). While satellite sensor effects were not tested explicitly, results from Landsat 5 TM and Landsat 7 ETM+ images were consistent with each other. Consistency of results implies a consistency of both Landsat and FIA data across space and time.

The images used in this study had almost exact anniversary dates (within 2 days). Thus, sun angle effects, which usually introduce a major source of error, were minimized. Phenology, which can vary from year to year, was assumed to be similar for image anniversary dates from 2000 and 2007. Image cross-normalization, which typically is conducted during LEDAPS processing, was not performed for this study. The minimal differences across space and time observed in this study suggest that cross-normalization may be unnecessary under some circumstances.

At least a portion of the small differences observed in this study may be explainable as mostly random, since the Random Forests modeling process makes a certain number of trees (2000, in this case) from random bootstrap samples and random subsets of predictors.

A companion paper in these proceedings (Moser et al. 2008) discusses utility of NAFD products for assessing increasing eastern redcedar forest area in southern Missouri between 1985 and 2007. Assumptions of across-time consistency in Moser et al. (2008) are supported by this study.

References

Bechtold, W.A.; Scott, C.T. 2005. The forest inventory and analysis plot design. In: Patterson, Paul L., ed. The enhanced forest inventory and analysis program - national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture Forest Service: 27-42.

Brown, D.G.; Duh, J.; Drzyzga, S.A. 2000. Estimating error in an analysis of forest fragmentation change using North American Landscape Characterization (NALC) data. Remote Sensing of Environment. 71: 106-117.

Butson, C. R.; Fernandes, R. A. 2004. A consistency analysis of surface reflectance and leaf area index retrieval from overlapping clear-sky Landsat ETM+ imagery. Remote Sensing of Environment. 89: 369-380.

Canty, M.J.; Nielsen, A.A..; Schmidt, M. 2004. Automatic radiometric normalization of multitemporal satellite imagery. Remote Sensing of Environment. 91: 441-451.

Cihlar, J.; Latifovic, R.; Beaubien, J.; Trishchenko, A.; Chen, J.; Fedosejevs, G. 2003. National scale forest information extraction from coarse resolution satellite data, part 1. In: Wulder, M.A.; Franklin, S.E., eds. Remote sensing of forest environments: concepts and case studies. Boston: Kluwer: 337-357.

Freeman, E.; Frescino, T.A. 2008. Modelmap: creates random forest and stochastic gradient boosting models, and applies them to GIS '.img' files to build detailed prediction maps. Unpublished package (version 1.1) for R statistical software. Ogden, Utah.

Goward, S.N.; Masek, J. G.; Cohen, W. B.; Moisen, G. G.; Collatz, G. J.; Healey, S.; Houghton, R.A.; Huang, C.; Kennedy, R.; Law, B.; Powell, S.; Turner, D.; Wulder, M. A. 2008. Forest disturbance and North American carbon flux. Eos, Transactions American Geophysical Union. 89: 105-116.

Guindon, B.; Edmonds, C.M. 2002. Large-area land-cover mapping through scene-based classification compositing. Photogrammetric Engineering & Remote Sensing. 68: 589-596.

Healey, S.P.; Yang, Z.; Cohen, W.B.; Pierce, D.J. 2006. Application of two regression-based methods to estimate the effects of partial harvest on forest structure using Landsat data. Remote Sensing of Environment. 101: 115-126.

Healey, S.P.; Moisen, G.; Masek, J.; Cohen, W.; Goward, S.; Powell, S.; Nelson, M.; Jacobs, D.; Lister, A.; Kennedy, R.; Shaw, J. 2007. Measurement of forest disturbance and re-growth with Landsat and FIA data: anticipated benefits from Forest Inventory and Analysis' collaboration with the National Aeronautics and Space Administration and university partners. In: McRoberts, R.E.; Reams, G. A.; Van Deusen, P.C.; McWilliams, W. H., eds., Proceedings of the seventh annual forest inventory and analysis symposium; 2005 October 3-6; Portland, ME. Gen. Tech. Report WO-77. Washington, DC: U.S. Department of Agriculture, Forest Service: 171-178.

Liaw, A.; Wiener, M.. 2002. Classification and regression by randomForest. R News. 2/3: 18-22.

Masek, J.G.;Huang, C.; Wolfe, R.; Cohen, W.; Hall, F.; Kutler, J.; Nelson, P.. 2008. North American forest disturbance mapped from a decadal Landsat record. Remote Sensing of Environment. 112: 2914-2926.

Moser, W.K. ; Nelson, M.D.; Healey, S.P.; Hansen, M.H.; Cohen, W.B. 2008. Estimation of biomass growth of eastern redcedar in southern Missouri: trends in afforestation. In: McWilliams, W.; Moisen, G.; Czaplewski, R., eds. Forest inventory and analysis symposium; 2008 October 21-23; Park City, UT. Proc. RMRS-P-56CD, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

R Development Core Team. 2008. R: A language and environment for statistical computing. Computer program, URL http://www.R-project.org. Vienna, Austria.

Reams, G.A.; Smith, W.D.; Hansen, M.H.; Bechtold, W.A.; Roesch, F.A.; Moisen, G.G. 2005. The forest inventory and analysis sampling frame. In: Patterson, Paul L., ed.

The enhanced Forest Inventory and Analysis Program - national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture Forest Service: 11-26.

U.S. Forest Service. 2003. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 2.0. Washington, DC: U.S. Department of Agriculture, Forest Service. 281 p.

Trends in afforestation in southern Missouri

W. Keith Moser[1], Mark D. Nelson[1], Sean Healey[2], Mark H. Hansen[1], Warren Cohen[3]

***Abstract**: Past studies of forest disturbance traditionally have focused on biomass loss, e.g., blowdown in the Boundary Waters Canoe Area Wilderness, gypsy moth infestation, the impacts of Hurricanes Hugo and Katrina. Using FIA data and satellite imagery, this study examined a region of the country that is simultaneously experiencing biomass loss due to oak decline and biomass gain from afforestation of agricultural lands by eastern redcedar, following change to nonagricultural land use. This paper examines the increase in eastern redcedar in southwestern Missouri between 1985 and 2007. After converting stacks of Landsat imagery to delineations of forest and nonforest categories, we looked at changes between succeeding images. We observed a small increase in forest land area, particularly in the first 10 years of the analysis. Image-based estimates of forest land area did not differ significantly from estimates derived from FIA sample plot data. Differences in definition of forest may explain any real variations in the estimates.*

Keywords: Forest, inventory, FIA, remote sensing, eastern redcedar, agricultural abandonment, Landsat.

[1] USDA Forest Service; Northern Research Station; Forest Inventory and Analysis Program; 1992 Folwell Ave.; St. Paul, MN 55108
[2] USDA Forest Service; Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401
[3] USDA Forest Service Pacific Northwest Research Station, Forestry Sciences Laboratory, 3200 SW Jefferson Way Corvallis, OR 97331

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

Past forest disturbance studies traditionally have focused on biomass loss, e.g., impacts of hurricane and blowdown, insect infestation, wildfire, and harvest activities. Little attention has been paid to biomass accretion in disturbance-dependent ecosystems, such as the ecotone between the grasslands and forests in the Upper Midwest. A combination of changing economics of agriculture and human-influenced disturbance patterns in this region have resulted in a substantially different landscape than in past decades. One tree species taking advantage of the reduction in disturbance events in Missouri woodlands and rangelands is eastern redcedar (*Juniperus virginiana* L.). The suppression of fire and the reduction in grazing on pastureland have resulted in an unprecedented expansion of eastern redcedar in Missouri. The overall study, of which this paper reports some early analyses, intends to document the pattern and extent of redcedar expansion.

Eastern redcedar

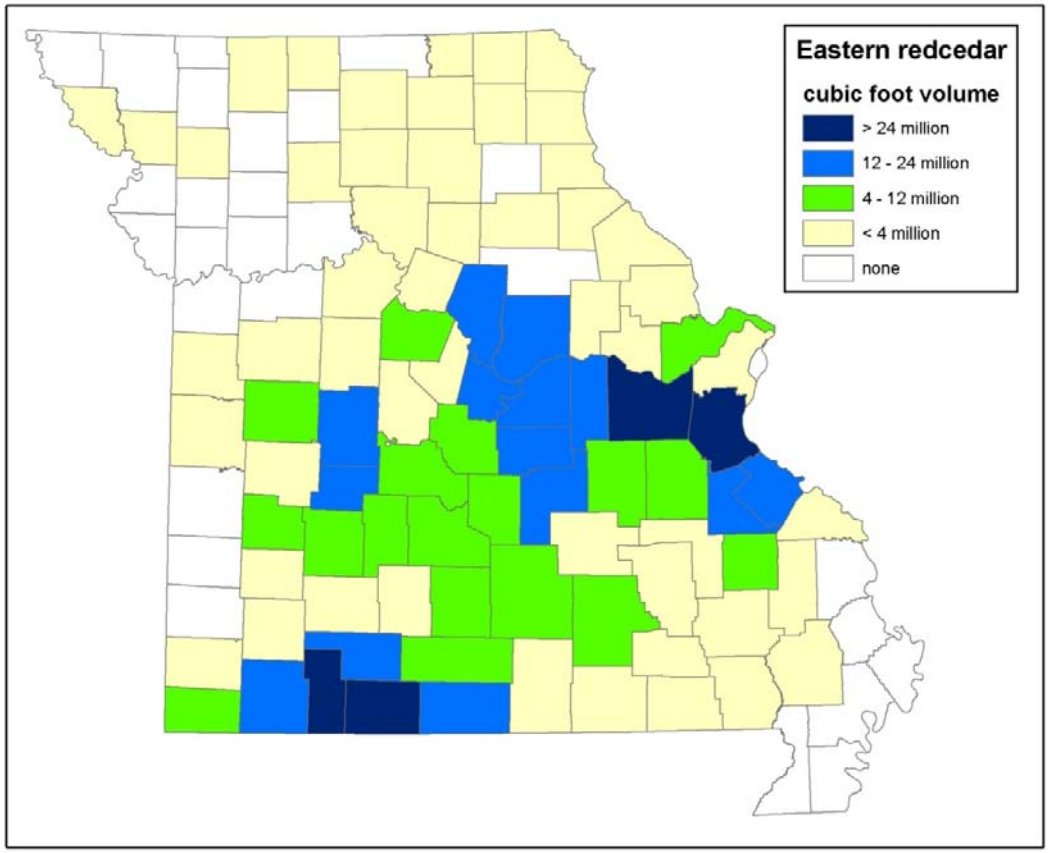


Figure 1: Eastern redcedar cubic foot volume on Missouri forest land, by county, 2007.

Eastern redcedar is a coniferous tree species common to the eastern United States (Lawson 1990). Historically limited to areas with infrequent fires, the species has long been known as a vigorous invader of old fields (Arend 1950, Lawson 1985). The two chief disturbances that suppress redcedar are grazing and fires – natural or prescribed (Beilmann and Brenner 1951, Buehring et al. 1970, Briggs and Gibson 1992). The species exhibits classic invasive behavior by producing prolific seed crops every 2 or 3 years. The seed are not light enough to disperse by wind, so dispersion depends heavily on birds and small mammals, which ingest and later defecate the seeds (Arend 1950, Parker 1951). The species can begin producing seeds at 10 years old, although it is most prolific as a seed producer between the ages of 25 to 75 years. Most seeds germinate in the spring of the second year after dispersal; those that germinated during the first year likely passed through an animal's digestive tract (Schopmeyer 1974). Seedling establishment is improved by exposure to bare soil, but intensive site preparation is not necessary (Ferguson et al. 1968). Since it is very shade intolerant (Baker 1949), eastern redcedar has higher survival rates under more open canopies (Parker 1952). Stands formed through invasion of old fields may start to break up

at around 60 years of age as hardwoods or other competing species become established.

Eastern redcedar has the potential to dramatically alter the structural characteristics of an ecosystem it enters. Dense redcedar stands not only change the character of pasture, range, or woodland to forest, but can also severely impact ground flora biomass and diversity, simultaneously creating a monolayer canopy and a monoculture. In native prairie ecosystems, encroaching redcedar is viewed as an undesirable invasive species. As previously mentioned, redcedar is suppressed by fire, as the stands are highly flammable and any established fires are extremely intense and usually stand-replacing.

This project provides several benefits to state agencies, resource managers, and others interested in trends in ecological succession along the prairie-woodland ecotone of the Upper Midwest. Gaining a better understanding of the relationship between satellite imagery and forest inventory data in this rapidly changing landscape will allow us to make better estimates of the resource. With an improved understanding of landscape change, we have the possibility to refine our post-sampling stratification. Our analysis of resource trends will improve as we can go back to times between periodic inventories and "fill in the gaps" on resource estimation. By obtaining a relatively intensive set of imagery (at 2-year intervals), we can simulate the effects of various image-sampling intensities to identify the most efficient and least expensive strategy for meeting the production needs of the U.S. Department of Agriculture, Forest Service's Forest Inventory and Analysis (FIA) program. This is accomplished by comparing the value of intensive data for growth (a relatively continuous process) vs. mortality and harvesting (examples of discrete events). Finally, this study will provide baseline data for analysis of carbon storage in the forest-woodland-prairie ecotone. To achieve these goals, this study had two objectives: 1) Assess the relationship between satellite image data and FIA field sample data, specific to forest biomass change processes; and 2) Document the timing and amount of forest biomass change over time using a combination of remote sensing imagery and FIA data.

Methodology

In 2007, we published an analysis of the Boundary Waters Canoe Area Wilderness (BWCAW) after the 1999 windstorm (Moser et al. 2007a). One of the challenges of that study was estimating the intensity and extent of the damage from the storm. A sketch map of forest blowdown was produced by U.S. Forest Service Forest Health Protection staff, who flew over the area in fixed-wing aircraft following the blowdown event, 'sketched' damage site locations on maps, and visually interpreted severity of blowdown damage. Reasonable estimates of the blowdown damage area were produced from sketch maps, but the coarse resolution of sketch map polygons limited their utility for site-specific assessments or linkage to inventory plot data and satellite image pixels. Therefore, we developed and employed alternative methodologies to estimate damage extent (Nelson et al. 2007).

The North American Forest Dynamics Program (NAFD) utilizes stacks of imagery from Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper plus (ETM+) satellite sensors to estimate loss and regrowth of forest biomass following fire, windstorms, harvesting, and other disturbances (Goward et al. 2008). Image dates are constrained to the growing season and have similar anniversary dates. A sample of Landsat Path/Row images was selected to produce estimates. NAFD is a cooperative effort between the University of Maryland, FIA, NASA, and the Forest Service's Pacific Northwest Research Station. Because NAFD addresses recovery and regrowth following disturbance, it has potential for estimating biomass increase resulting from afforestation of former agricultural lands.

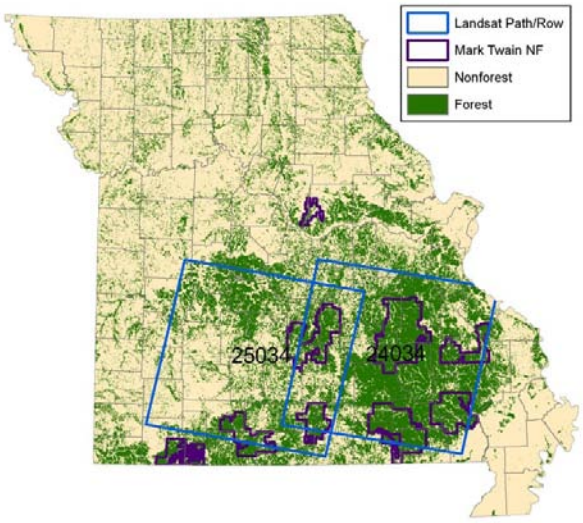


Figure 2: Map of Missouri with location of the two Landsat scenes 25034 and 24034, county boundaries, and the Mark Twain National Forest.

For this study, we selected a series of images acquired between 1985 and 2007, covering Path 25, Row 34 (25034) and Path 24, Row 34 (24034) (Fig. 2). This imagery was not part of the existing NAFD sample, but was purchased by the Forest Service's Northern Research Station, FIA program. For this paper, we will confine our analysis to Scene 25034. NAFD provided much of the processing supporting this work, which included the following steps (Nelson et al. 2008):

- Atmospheric correction of imagery, conducted by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) at NASA
- Download of additional free imagery and cloud masking at UMD
- Modeling of forest probability with Random Forest models and FIA data at FIA. These models were applied to all images in the time series.
- Application of an algorithm to reduce inter-annual spectral noise and to identify pixels showing steady increases in probability of forest.

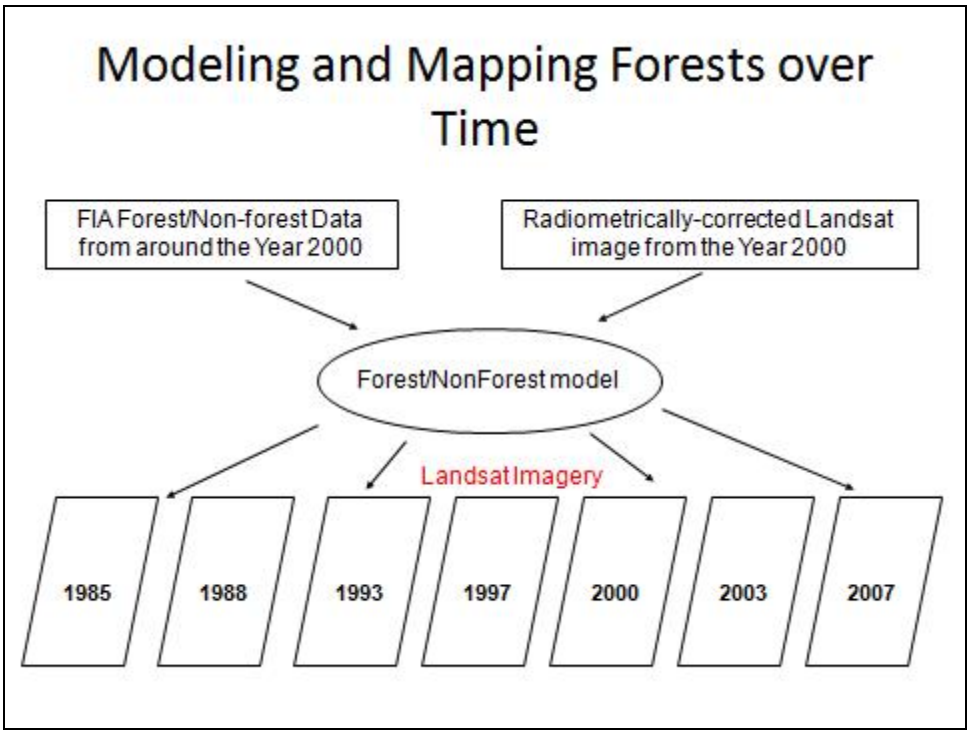


Figure 3: Schematic of the image processing effort.

A model predicting probability of forest (0-100%) was built based upon radiometrically corrected Landsat satellite imagery (Fig. 3) (Nelson et al. 2008). Because all of the imagery in the time series is similarly corrected, the model can be applied to all images to give an idea of how forest cover is changing over both space and time.

The algorithm to reduce inter-annual noise and identify trends has two primary steps: segmenting a pixel's temporal trajectory into discrete periods (usually punctuated by disturbance or, in our case, field abandonment), and fitting lines to the values in each period (Kennedy et al. 2007). The Y-axis in Figure 4 (Band 5 reflectance) is different from our Y-axis (probability of forest), but the process is the same. Our analysis is based on these fitted trajectories (i.e., the black lines).

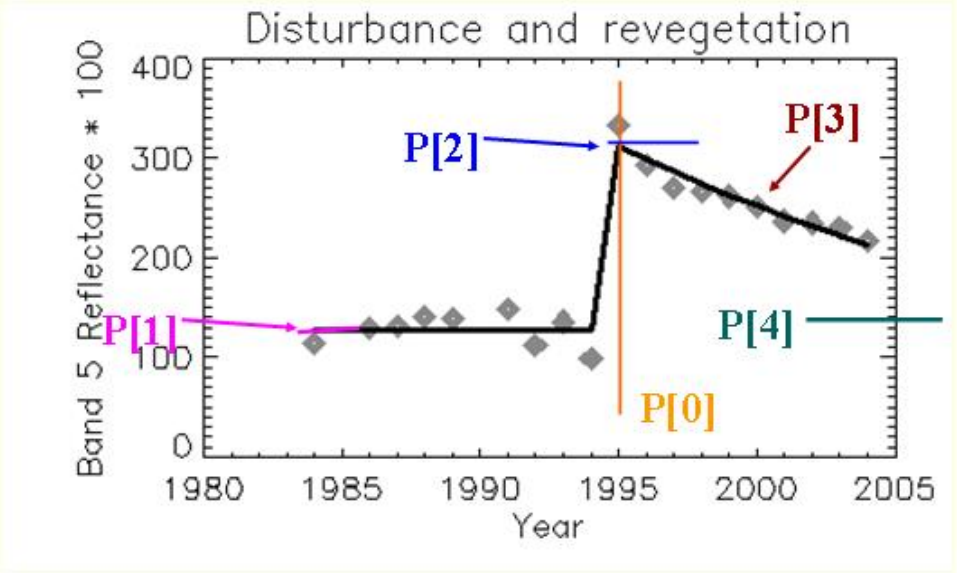


Figure 4: Methodology for reducing inter-annual noise. P(0) year of disturbance, P(1) pre-disturbance state, P(2) intensity of disturbance, P(3) period of recovery, and P(4) projected future state.

Figure 5 presents an example of an agricultural field (pink) gradually becoming forested (green) during a span of two decades.

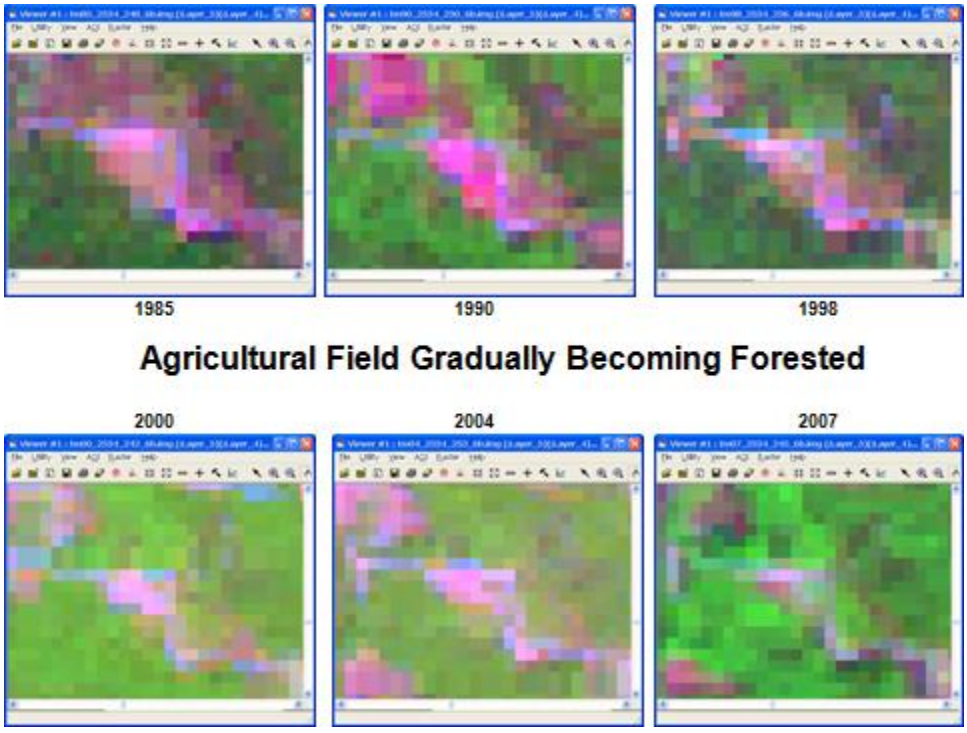


Figure 5: Example of an agricultural field gradually becoming forested.

Figure 6 is a graphical presentation of the process visualized in Figure 5. The series of blue symbols represents a pixel's change over time. For our study, we chose a 50 percent threshold. This minimum threshold was chosen as a reasonable cutoff to characterize a pixel as "forest" versus "nonforest". The series of pink

symbols represents a pixel that, while possibly revegating, did not cross the 50 percent threshold and thus is not considered (re)forested at any time. In a more open-canopy situation, such as a longleaf pine forest or a more woodland condition, pixels evaluated at a higher threshold (line A in the diagram) might better correspond with a definition of forest land. In situations with a vigorous shrub understory or a more dense canopy with a high leaf area index (LAI), a lower threshold (line B) might be appropriate.

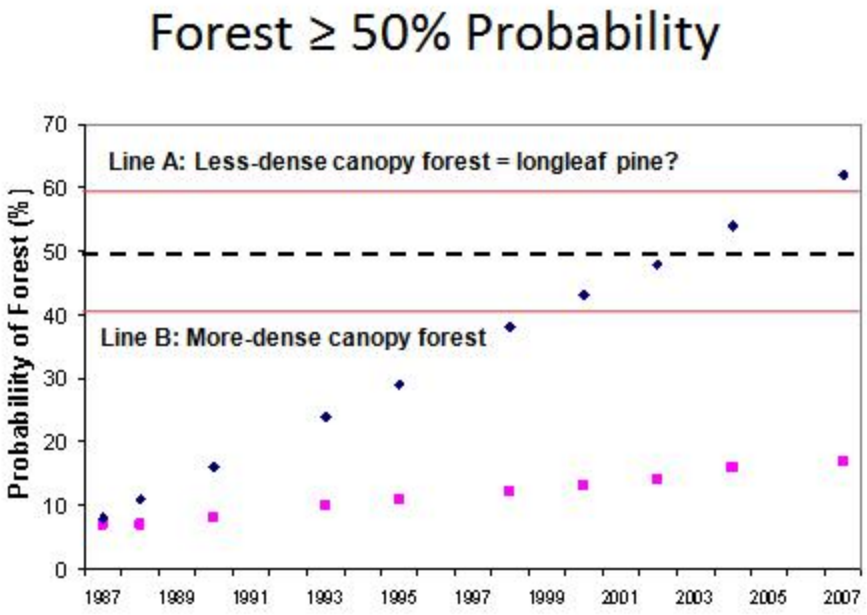


Figure 6: Example trajectories of increasing probability of a pixel representing a forested state and the selection of threshold values. The blue diamonds represent a pixel transitioning from open land to forest. The pink squares represent a pixel that remains classified as open land.

Pixel-based estimates of "probability of forest" (used here as an index of forest cover) can be tracked over time using NAFD data. Once a pixel exceeds the 50 percent forest-probability minimum threshold, it gets counted as forest area, e.g., for county-by-county analysis. However, definitional differences between land use and land cover mean that more trees do not always mean more forest. The algorithm predicts increases in forest probability in housing developments as trees are planted and irrigated. Obviously, new developments initially cause forest probability to go down.

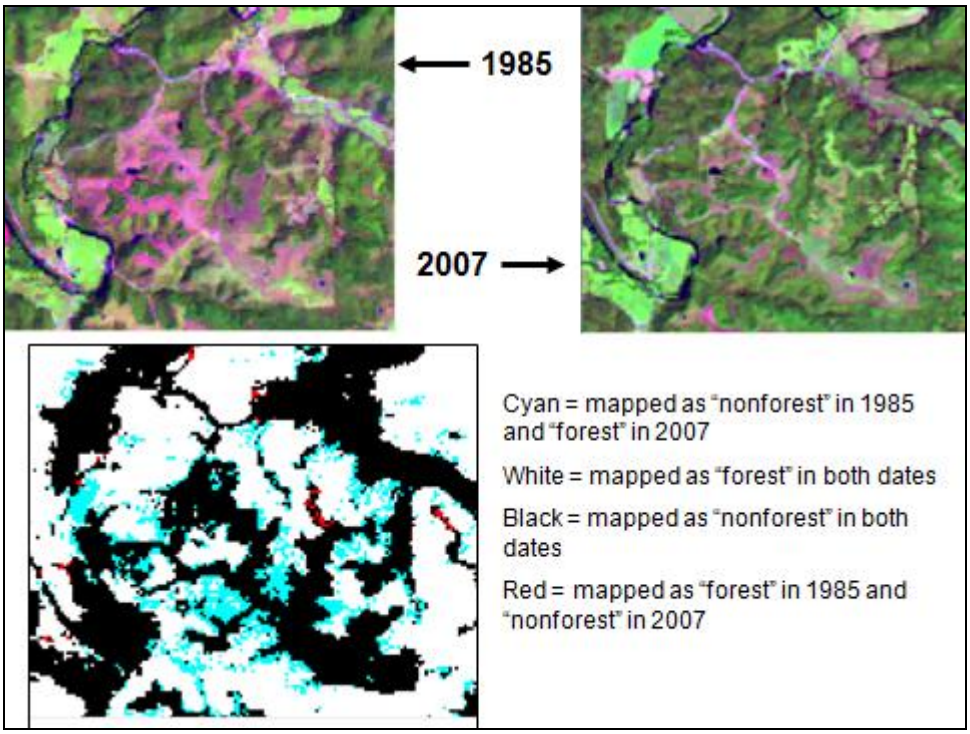


Figure 7: Example views from 1985 and 2007, and a transitional graphic illustrating forest vs. nonforest progression.

Using image reflectance data and the threshold level illustrated in Figure 6, we categorized each pixel as "forest" or "nonforest". By comparing scenes at time 1 and time 2, we then constructed a transition image dataset that predicted whether a pixel category remained constant, or changed to another category (Fig. 7). The temporal increment between time 1 and time 2 typically represents the number of years between successive images in the NAFD image stack – 2 years on average. However multiple increments can be combined such that intervals between time 1 and time 2 can span more than two decades.

We tracked four possible scenarios between succeeding images: Nonforest at both time 1 and time 2 (NFOR – NC), forest at both time 1 and time 2 (FOR-NC), change from nonforest at time 1 to forest at time 2 (toFOR), and change from forest at time 1 to nonforest at time 2 (toNFOR). We then produced estimates for each of the categories by summing image pixel areas for each year and county, for the 11 counties that are wholly within scene 25034.

FIA estimates

Estimates of forest land area were produced from FIA field sample data, using FIA data from the 1989 periodic inventory (Spencer et al. 1992), and the 1999-2003 (Moser et al. 2007c) and 2004-2006 (Moser 2007b) annual inventories. These three estimates were compared to image-based estimates of forest land area from similar time periods.

Results

Area estimates of FOR-NC, NFOR-NC, toFOR, and toNFOR are reported for Scene 25034, for each pair of successive image dates (Fig. 8). Most land within the study area exhibited almost no change between successive periods, both for forest and nonforest classes. A small but gradual decrease in nonforest and a corresponding increase in forest area were revealed by examining trends over multiple time increments. The two change categories, toFOR and toNFOR, portray more variability between successive time increments and reveal shorter term trends. A notable increase in area of toFOR occurred during earlier years of the study, peaking in the year 1993, and declining markedly thereafter. A corresponding decrease in toNFOR occurred during the earlier years, reaching a low in year 1988, then showing a gradual increase thereafter.

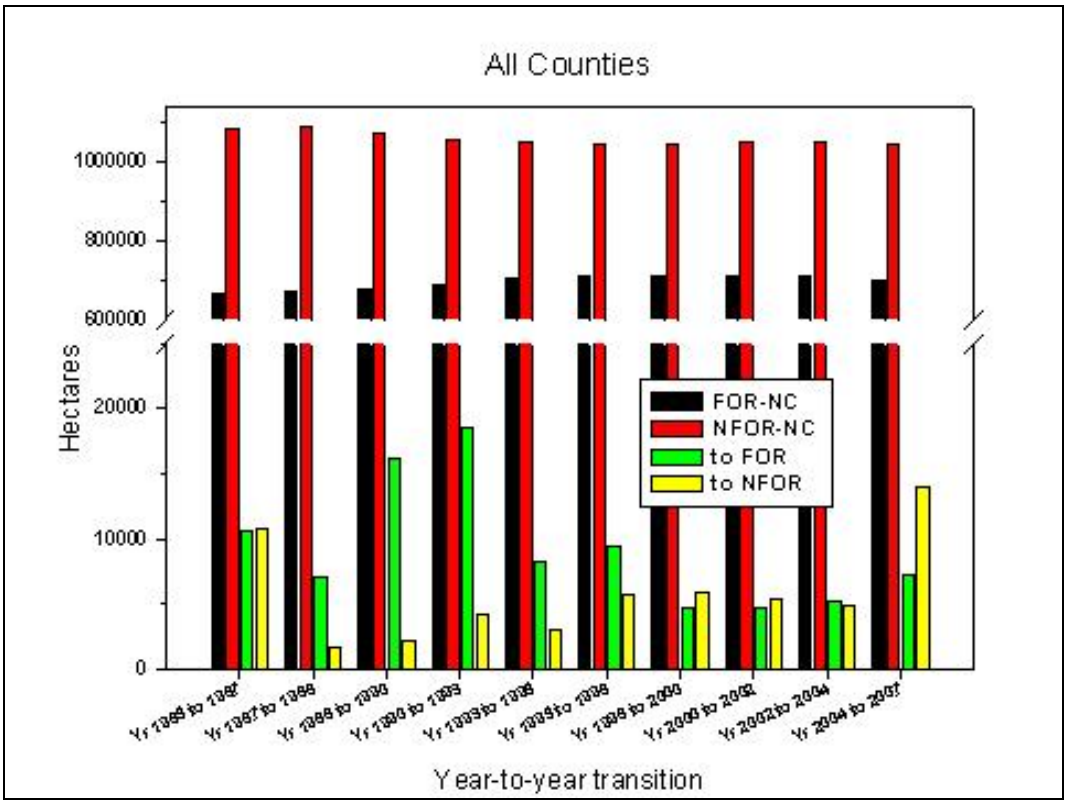


Figure 8: Image to image category transitions, in hectares of land in all counties of scene 25034, Missouri, 1985 to 2007.

This study examined trends in the total scene and in each of the 11 counties within the scene. Two counties – Pulaski (Fig. 9) and Camden (Fig. 10) – were chosen to represent typical predominantly forested and nonforested counties, respectively. Figure 11 portrays an example image of modeling results for Camden County, Missouri, over a time frame of 22 years. In Pulaski County (Fig. 9), there was a net increase in forested area through 1996, then a slight decline afterwards. In Camden County (Fig. 10), a county with a higher overall

proportion of nonforest, there was a more dramatic change to forested land through 1993; subsequent changes to forested land were smaller but always exceeded changes to nonforested land until the final interval (2004 to 2007). In Figure 11, a map comparing forest and nonforest landscapes in Camden County shows the likely urban-influenced change to nonforest in the southern and eastern parts of the county and changes to forest occurring more in the north and west.

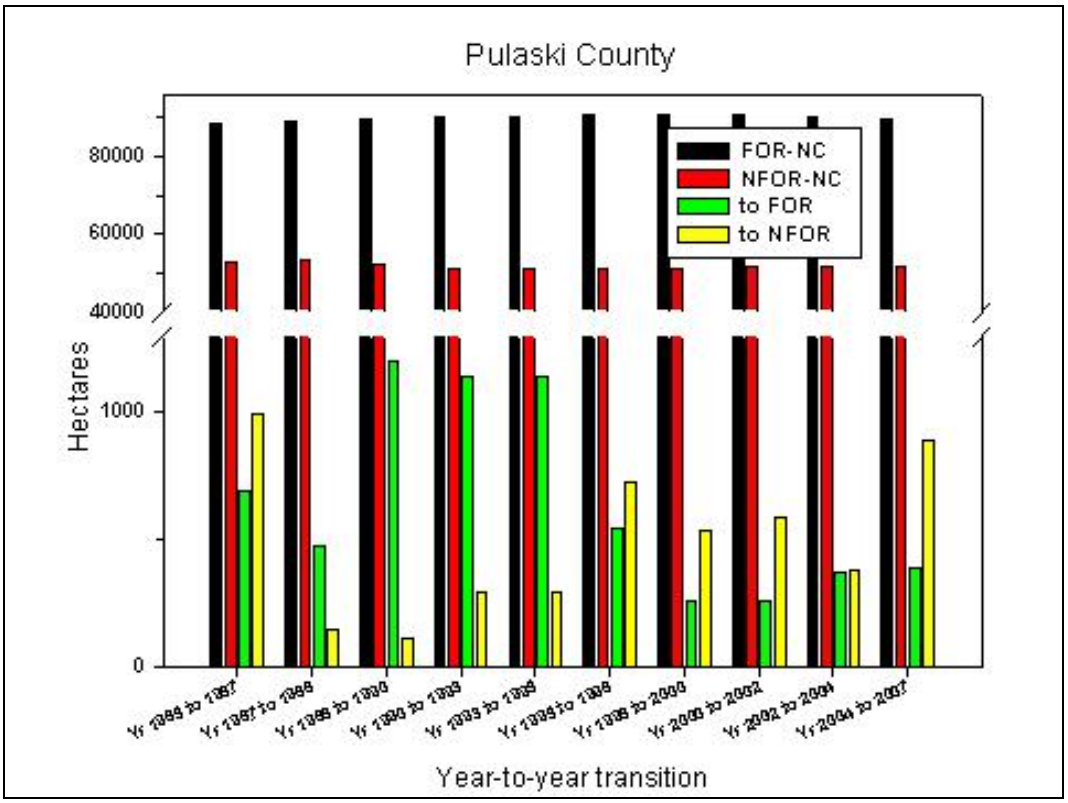


Figure 9: Image-to-image category transitions, in hectares of land, Pulaski County, Missouri, 1985 to 2007.

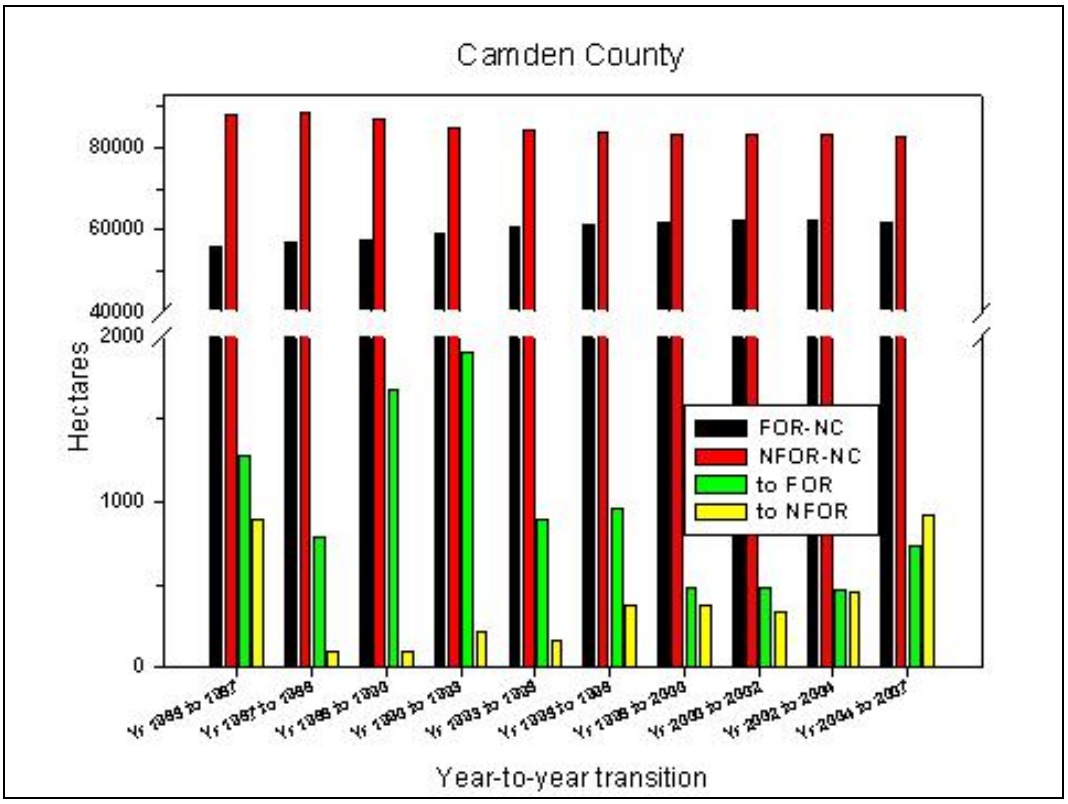


Figure 10: Image-to-image category transitions, in hectares of land, Camden County, Missouri, 1985 to 2007.

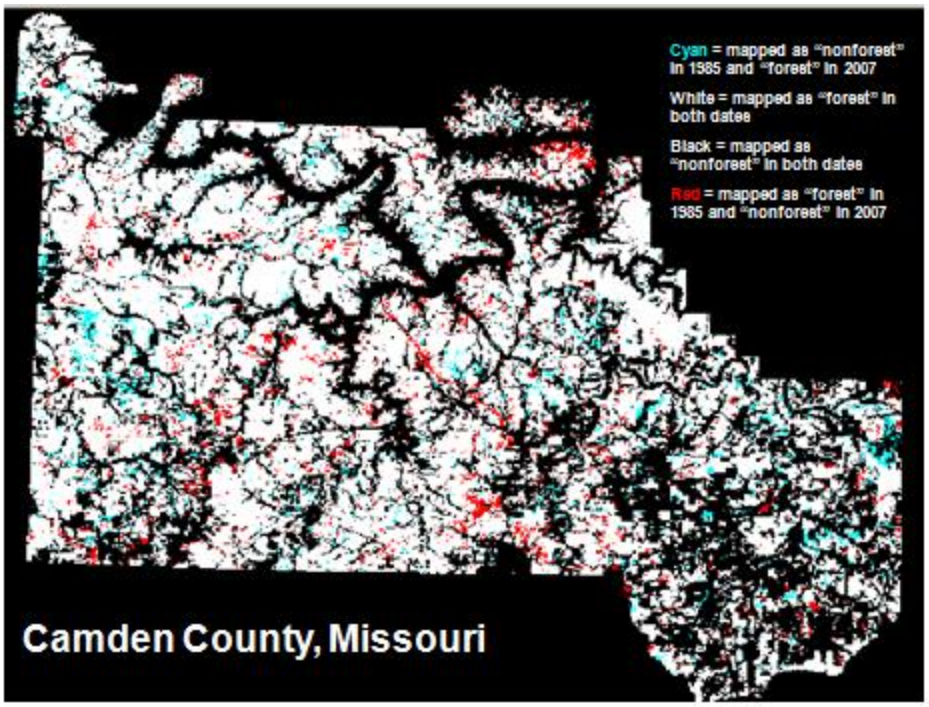


Figure 11: Example transition graphic for forest and nonforest conditions, Camden County, Missouri, 1985 to 2007.

Imagery analysis vs. FIA data

Because of FIA's forest definitional requirements, our initial expectations were that FIA data would have lower estimates of forest land than the estimates from the imagery. But FIA estimates were significantly higher, in some years markedly so (Fig. 12). There was a definitional change in forest land in the 1990s, which might explain the larger difference between the satellite estimates and the FIA estimates in the second half of our study period. There might have been sites where the forest was harvested, which FIA would have classified as forest land but the satellite imagery would not. Preliminary analysis (Moser, unpublished data) did not find any conclusive relationship between harvested area and the gap between the two estimates.

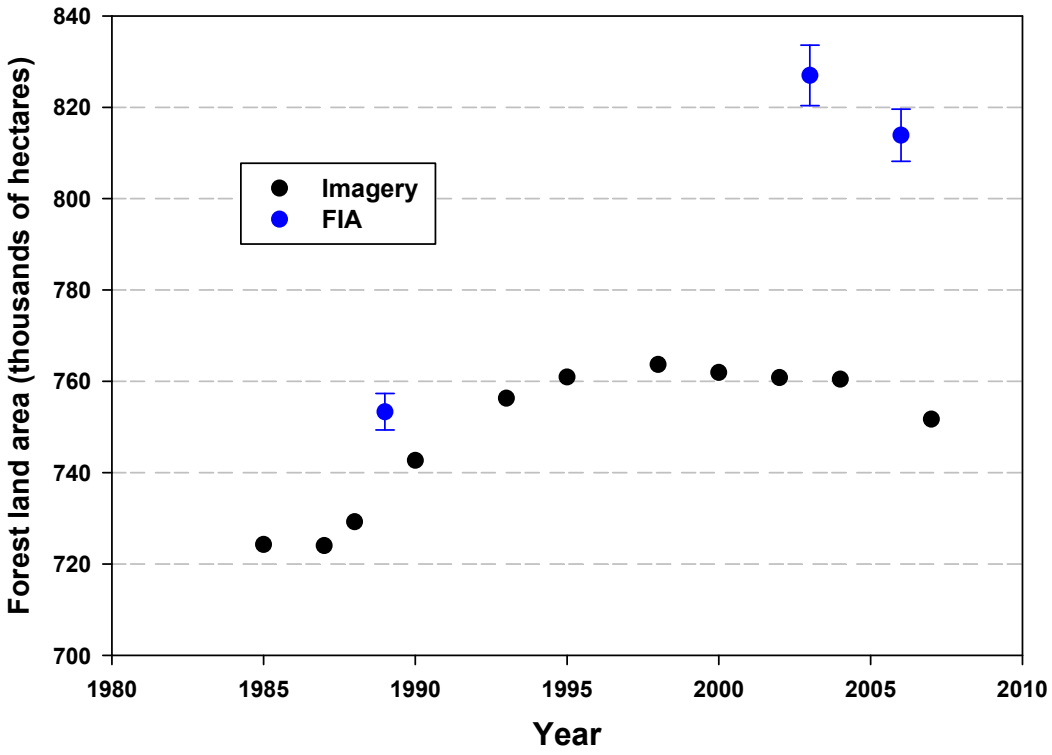


Figure 12: Estimates of forest land area using imagery (black) and FIA inventories (blue), 1985 through 2007. FIA inventories were conducted in 1989, 1999-2003, and 2004-2006.

Results were more variable at the county level (Fig. 13). In some counties, FIA estimates exceeded image-based estimates. In other counties, the opposite was true. Because the variance of FIA estimates at the county level was quite high, we did not observe any significant difference between the image-based estimates and the inventory-based estimates for any county, i.e., the image-based estimates of forest land area are within the confidence intervals of FIA plot-based estimates.

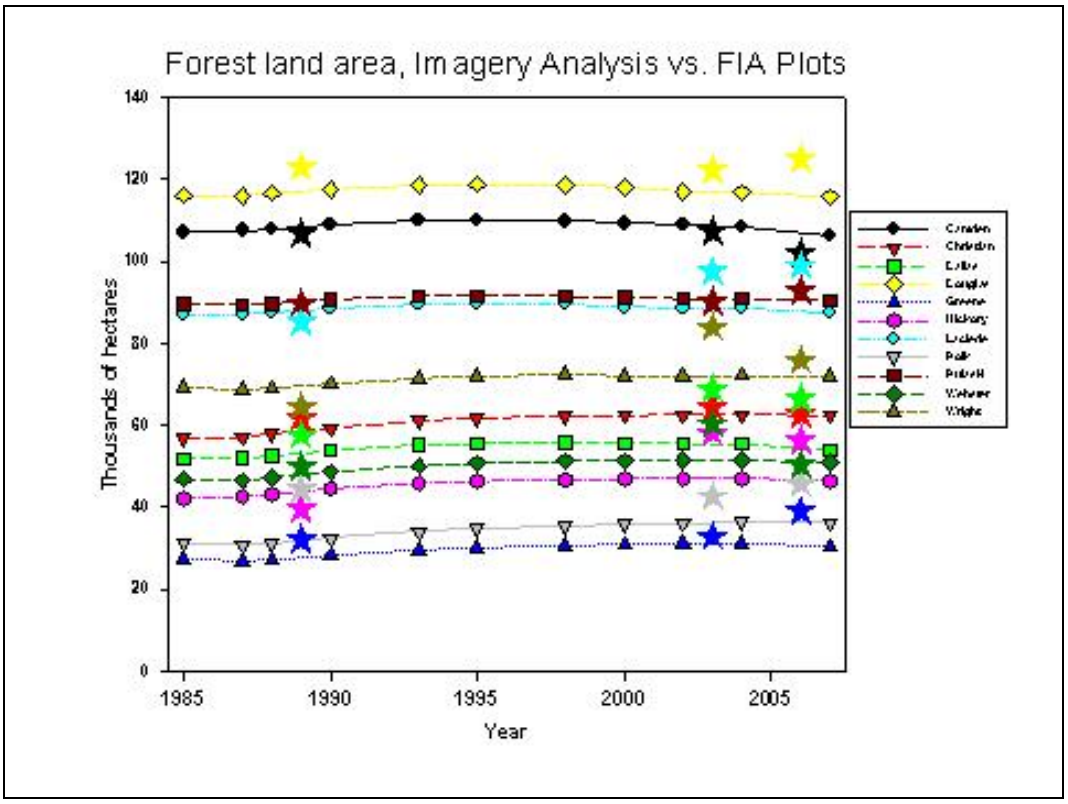


Figure 13: Estimates of forest land area, in hectares, of the 11 Missouri counties in Landsat scene 25034, 1985 though 2007 (symbols and lines), and estimates of forest land area using FIA data (stars).

Conclusion

This study examined patterns of afforestation in southwestern Missouri by processing a stack of Landsat TM imagery from Scene 25034 from 1985 through 2007, using NAFD procedures. Overall, we observed a slight but gradual net increase in forest land area over the period, with the bulk of the increase in the first half of the study period. We compared image-based estimates with estimates from FIA inventories conducted in 1989, 1999-2003, and 2004-2006. Although the FIA estimates of forest land area over the entire study area was significantly larger, the differences at the county level were not significant. While there were some counties with significant patterns of harvesting activity, we could not conclusively make the connection between harvested area and differences in the two forest land estimates at the county level. Continuing analyses from this study will extend to an adjacent image and further assess afforestation and resulting biomass changes resulting from eastern redcedar encroachment of abandoned agricultural lands.

Acknowledgements

The authors wish to acknowledge the contributions of the NRS-FIA program for purchasing the Landsat imagery, and to Robert Kennedy (Oregon State University), Nancy Thomas (University of Maryland), Jeff Masek (NASA Goddard), Chengquan Huang (University of Maryland), and Gretchen Moisen (IWFIA) for their contributions in processing the image data. Andrew Lister and Andrew Gray provided valuable comments on an earlier version of this manuscript.

References

Arend, J. L. 1950. Influence of fire and soil on distribution of eastern redcedar in the Ozarks. Journal of Forestry. 48:129–130.

Baker, F. S. 1949. A Revised Tolerance Table. Journal of Forestry. 47(3) 179-181(3).

Beilmann, A. P.; Brenner, L. G. 1951. The recent intrusion of forests in the Ozarks. Missouri Botanical Gardens Ann. 38:261 282.

Briggs, J. M.; Gibson, D. J. 1992. Effect of burning on tree spatial patterns in a tallgrass prairie landscape. Bulletin of the Torrey Botanical Club119:300–307.

Buehring, N.; Santelmann, P. W.; Elwell, H. M. 1971. Responses of eastern red cedar to various control procedures. Journal of Range Management. 24:378 38.

Ferguson, E. R.; Lawson, E. R.; Maple, W. R.; Mesavage, C. 1968. Managing eastern redcedar. Stat. Res. Pap. SO-37. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 15 p.

Goward, S.N.; Masek, J.G.; Cohen, W.; Moisen, G.; Collatz, G.J.; Healey, S.; Houghton, R.A.; Huang, C.; Kennedy, R.; Law, B.; Powell, S.; Turner, D.; Wulder, M.A. 2008. Forest disturbance and North American carbon flux. Eos, Transactions American Geophysical Union. 89: 105-116.

Kennedy, R. E.; Cohen, W. B.; Schroeder, T. A. 2007. Trajectory-based change detection of automated characterization of forest disturbance dynamics. Remote Sensing and the Environment. 110: 370-386.

Lawson, E. R. 1985. Eastern redcedar. An American wood. Washington DC: U.S. Department of Agriculture, Forest Service. FS-260. 7 p.

Lawson, E. R. 1990. Juniperus virginiana L. Eastern redcedar. In Burns, R.M. and Honkala, B.H., compilers. Silvics of North America, Vol. 1 Conifers. Agricultural Handbook 654. Washington, D.C.: U.S. Department of Agriculture, Forest Service,..

Moser W. K.; Hansen, M. H.; Nelson, M. D.; Crocker, S. J.; Perry, C. H.; Schulz, B.; Woodall, C. W.; Nagel, L.; Mielke, M. 2007a. The Boundary Waters and the Blowdown: A resource assessment of the Boundary Waters Canoe Area Wilderness, 1999-2003. Newtown Square, PA:U.S. Department of Agriculture, Forest Service, Northern Research Station. Gen. Tech. Rep. NRS-7. 54 p.

Moser, W. K.; Hansen, M. H.; Treiman, T. B. 2007b. Missouri's forest resources, 2006. Res. Note NRS-4. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 4 p.

Moser, W. K.; Hansen, M. H.; Treiman, T. B.; Jepson, E.; Leatherberry, E. C.; Liknes, G.; Perry, C. H.; Olson, C. L.; Piva, R. J.; Woodall, C. W.; Brand, G. J. 2007c. Missouri's forests 1999-2003, Part A. Resour. Bull. NRS-10. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 79 p.

Nelson, M. D.; Moser, W. K. 2007. Integrating remote sensing and forest inventory data for assessing the BWCAW blowdown. In: Greer, J. D., ed. New remote sensing technologies for resource managers; Proceedings of the 11th Forest Service Remote Sensing Applications Conference; 24-28 April 2006; Salt Lake City, UT. American Society for Photogrammetry and Remote Sensing: 8 p. CD-ROM.

Nelson, M. D.; Moser, W. K.; Healey, S.; Hansen, M. H.; Cohen, W. 2008. Analyzing Landsat time series data across adjacent Path/Rows and across multiple cycles of FIA: lessons learned in southern Missouri. In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2008. Forest Inventory and Analysis (FIA) Symposium 2008; 2008 October 21-23; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. CD-ROM.

Parker, J. 1951. Natural reproduction from red cedar. Journal of Forestry. 49:285.

Parker, J. 1952. Establishment of eastern red cedar by direct seeding. Journal of Forestry. 50:914–917.

Schopmeyer C.S. (ed). 1974. Seeds of woody plants in the United States. Agricultural Handbook 450. Washington, D.C.: U.S. Dept. Agric., For. Serv., 883 pp.

Spencer, J. S., Jr.; Roussopoulos, S. M.; Massengale, R. A. 1992. Missouri's forest resource, 1989: an analysis. Resour. Bull. NC-139. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 84 p.

National Inventory and Monitoring Applications



Overview of the National Inventory and Monitoring Applications Center (NIMAC)

Charles T. Scott [1]

Abstract: *The National Inventory and Monitoring Applications Center (NIMAC) was created by the Forest Inventory and Analysis (FIA) program in 2006. NIMAC addresses a growing need, expressed by FIA partners, for technical assistance in designing and implementing monitoring plans for forests at scales finer than that provided by the FIA standard inventory. NIMAC's goal is to develop leading-edge forest ecosystem monitoring methods and tools to help FIA and other organizations monitor forests, resulting in comparable results across the landscape. Methods and software tools are designed for use by land managers and across ownerships at landscape to national scales. To date, NIMAC has helped six states develop more specific inventories on their lands and is working with several National Forests on developing forest monitoring plans.*

NIMAC is developing a monitoring toolkit composed of a design tool, field data collection software, data storage and compilation tools, and spatial/tabular analysis tools. The toolkit will be applicable to a wide range of customers including the National Forest System, states, and FIA itself. Honduras is funding the development of a Spanish version of the toolkit. NIMAC is being funded to work with National Forest System monitoring coordinators and planners to develop the question-driven monitoring design and analysis tools to help ensure that data are collected and analyzed efficiently and in a scientifically defensible manner. NIMAC's software enhancements also will provide added flexibility to FIA customers.

Keywords: Forest, inventory, monitoring, FIA, sampling design, software.

Introduction

As a result of the 2006 combination of the North Central and Northeastern Research Stations of the U.S. Forest Service Research and Development branch, the Station's Forest Inventory and Analysis (FIA) units also were combined. This provided the opportunity to create a group dedicated to serving the forest inventory and monitoring needs of other organizations, such as states, the National Forest System, and other countries. The expertise of FIA has long been

[1] U.S. Forest Service, Northern Research Station, FIA, 11 Campus Blvd., Suite 200, Newtown Square, PA 19073 USA; ctscott@fs.fed.us; http://nrs.fs.fed.us/people/Scott

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

sought after, but FIA did not have the capacity to meet that need until the National Inventory and Monitoring Applications Center (NIMAC) was formed.

NIMAC's Mission

The mission of NIMAC is to develop leading-edge forest ecosystem inventory and monitoring methods and tools to help FIA and other organizations monitor forests, resulting in comparable results across the landscape. The emphasis is on providing expertise (consultation and training) and the tools needed for the partner organization to monitor its forests on its own in a way that is compatible with FIA estimates. NIMAC also develops new monitoring methods as problems are identified.

NIMAC's Assignments

To accomplish its mission, NIMAC works in several areas. We develop or enhance methods to make existing monitoring techniques more efficient, such as spatially balanced sampling methods (Lister and Scott 2008) and optimal plot design (Scott 1993). Similarly, we can develop new metrics or apply others' research and make their measures operational, such as for monitoring Montreal Criteria and Indicators. We also develop techniques to bridge the gap between the regional FIA survey and mid-scale planning needs, such as for developing forest plans, and even between mid-scale monitoring and stand-level inventory, such as small area estimation methods or ways to prioritize stand exams.

How NIMAC Can Help

NIMAC benefits customers in a number of ways. Forest managers benefit by having more efficient and effective forest monitoring plans. While these initially will focus on vegetation-related objectives, we hope to expand to other plot-based natural resource monitoring. The resulting methods are scientifically defensible in a court of law. The data are compatible and consistent with existing inventories, allowing comparisons with the forest matrix (context) and aggregation to ecoregion, regional, and national levels. Efficiency is also gained by building on existing FIA data and techniques. NIMAC provides the tools to implement the methods and to generate reports. NIMAC charges for its services – re-investing those funds in FIA's toolkit to speed up its development and to add new features, so everyone benefits.

NIMAC Projects

Much of the focus is on serving NIMAC customers. Most are traditional FIA partners: states, State & Private Forestry, and the National Forest System. We also work with other countries in cooperation with the International Forestry program of the U.S. Forest Service, and with nongovernmental organizations.

States

Wisconsin Continuous Forest Inventory: The Wisconsin Department of Natural Resources asked FIA to help conduct a continuous forest inventory (CFI) on 518,650 ac of State Forest lands so that they could monitor invasive plants and address the requirements for forest certification. NIMAC took this on as its first project. Wisconsin provided funds for NIMAC's technical guidance to help determine the objectives, monitoring questions, attributes to observe, precision, sampling and plot design, and plot locations. We modified the regional FIA field manual and data recorder program to meet their needs. Since the sampling intensity was increased to one plot per 166 ac, the number of 1/24-ac subplots was reduced to two so that multiple plots could be visited in a day. The proportion of Phase 3 (forest health) plots was increased from 1/16th to 1/3rd. Contractors were used to collect the data and to modify the FIA compilation program (NIMS) and the web-based analytical tool (FIDO). Some of the first season's data are available at: http://131.216.27.117/fido/mastf/index.html. The monitoring tools developed for Wisconsin are now being used to help other customers plan, conduct, process and analyze their own surveys, to the extent that they choose.

Indiana Continuous Forest Inventory: Based on the Wisconsin CFI experience, the Indiana Department of Natural Resources (DNR) asked NIMAC to help with a similar effort on their 157,211 ac of State Forests. Because the emphasis was on forest certification, only Phase 2 (tree) plots were selected. Another objective was to provide results at the analysis unit level – roughly 6,000 ac. Thus the sampling was increased to a single subplot every 40 ac. One concern with this design is the small number of sample trees on a plot with which to classify forest type. After considering different options, it was decided to rely only on the field crew classification. James Westfall conducted a study of the implications of reducing the number of subplots on the consistency of classification (Westfall 2008, this publication). Indiana DNR contracted the fieldwork to the same contractor currently measuring the statewide FIA plots. The first field season was completed in the fall of 2008. Indiana DNR staff will process the data using PC-based tools developed by NIMAC and funded by Indiana DNR.

Great Plains Initiative: The four states with the highest proportion of ash trees, relative to the total number of trees in the state, are North and South Dakota, Nebraska, and Kansas. Out of concern for the threat posed by the emerald ash borer, these States partnered with State & Private Forestry to fund a study of the potential ecological and economic impacts of this pest. As part of that study, they asked NIMAC for assistance in conducting a one-time inventory of trees outside of forests, including shelterbelts, riparian areas, scattered trees, and urban areas. In order to focus the field sampling effort, NIMAC developed a stratified two phase sampling system that used a percent tree cover layer from satellite imagery, a large subsample of aerial photographs, and ground samples. The allocation of samples to strata was optimized to efficiently sample plots with tree cover in areas

outside of what FIA calls forestland. Using PC-based tools, NIMAC will process the data, and provide the states with an analysis tool. The urban data will be processed by the urban forestry unit of the Northern Research Station.

National Forest System

NIMAC's assistance to the National Forest System started with a pilot project on the Mark Twain National Forest in 2006. We collaboratively identified the vegetation-related objectives, monitoring questions, metrics, and precision requirements. NIMAC then developed the sample size alternatives and resulting costs. The optimal solution was to continue doubling the intensity of the Phase 2 sample and to increase Phase 3 by seven times – much less than expected. Both Region 8 and 9 were encouraged by the pilot, so they asked NIMAC to work with several more National Forests. The ultimate objective is to apply NIMAC's 15 monitoring steps to all 30+ National Forests and Grasslands in the east over the next few years. The end result will be adequate forest-level information to monitor the vegetation-related aspects of their forest plans and to fulfill annual and 5-year reporting requirements under the 2008 Planning Rule. NIMAC is also providing assistance to the Spring Mountain National Recreation Area in Nevada (Region 4).

International

At the request of International Programs of the U.S. Forest Service, NIMAC is providing technical assistance to Honduras and Russia. For more than 75 years, Russia has been inventorying managed stands across the country. However, these stand data are difficult to roll up to national estimates and are representative of only managed forests. They have asked the United States and other countries for assistance in developing a national forest inventory (NFI). During several visits NIMAC has provided advice on NIMAC's monitoring steps and on the methods that FIA uses. The NFI will be critically important for sustainable forest management of these climate-sensitive areas which represent roughly 25 percent of the world's forests.

As a result of NIMAC's efforts, Honduras now has a monitoring system for its national forests in their highly productive pine regions. Initial work has begun on monitoring mahogany and other broadleaves. A high proportion of the harvests in these forests are illegal, so efforts are being made to foster sustainable forest management in cooperation with local communities. With funding from Honduras, NIMAC also is collaborating on the development of a Spanish version of the monitoring toolkit. These tools should then be useful in other Spanish-speaking countries.

Inventory and Monitoring Toolkit

The genesis of the development for a monitoring toolkit was the result of a meeting between FIA and The Nature Conservancy, NatureServe, the National

Park Service, and the state of New York, focused on how to more consistently monitor the forests of New York. The concept of a monitoring toolkit originated as a means of developing and fostering the use of consistent methods, based on a suite of tools that had been developed by the author. The vision was to first create a monitoring design tool for identifying consistent and efficient sampling methods. Then, develop data recorder, data compilation, and data analysis tools that all rely on those methods to integrate all steps and help foster consistency. Each of the tools would allow some flexibility, but would provide a common framework and a starting point that was consistent with FIA.

NIMAC has built upon that initial idea by adding functionality to the monitoring toolkit as needed for each new customer. The toolkit is applicable to a wide range of customers, including the National Forest System, states, and FIA itself. NIMAC is being funded to work with National Forest System monitoring coordinators and planners to develop question-driven monitoring design and analysis tools. The ultimate objective is to help ensure that data are collected and analyzed efficiently and in a scientifically defensible manner.

Design Tool

The Design Tool for Inventory and Monitoring (DTIM) is being developed to assist those who plan forest monitoring to work through the first 9 of the 15 monitoring steps. NIMAC has developed a working prototype to help:

1. Identify users and broad objectives of the monitoring.
2. Identify the monitoring questions to be asked.
3. Identify the tables of estimates (and their metrics and attributes) needed to answer the questions.
4. Assess whether existing data are sufficient to answer the questions, or to assess their variability if they are not adequate.
5. Set cost and/or precision constraints for collecting additional data
6. Develop an efficient plot and sampling design to balance cost and precision in order to address monitoring questions.
7. Select the plot locations in a spatially balanced way.
8. Generate field manual text for selected attributes
9. Generate training materials for selected attributes.

Based on a knowledge base from monitoring experts, users will be presented with lists of choices to guide the development of the monitoring system that best suits their needs. These steps are largely done within a spreadsheet. The version NIMAC is developing for the National Forest System will be more complete and will link with the other monitoring tools, especially the analytical tool (step 4 above).

Portable Data Recorder Tool

The use of portable data recorders with data collection software provides several important benefits. By not re-entering the data in the office, the data are available sooner and transcription errors are avoided. The data can be checked for completeness, for valid codes, and for cross checks between data elements in the field, where the data are most easily and accurately corrected.

Since late 2006, NIMAC customers have been using the Field Data Manager (FDM) system developed by the FIA unit in the North Central states. The new national system, Mobile Integrated Data Acquisition System (MIDAS), will soon replace FDM. It is highly flexible and more closely integrated with other FIA data processing software.

Data Storage and Compilation Tool

The data recorder software transfers the field data to servers via the Internet. These field data must then be loaded into the database and error checked again, including some checks that cannot be done on the data recorder. Once the data are determined to be clean and complete, then the data can be loaded into the main database and calculated fields added, such as volume per tree and forest-type class. In addition, the estimation methods used by FIA (Bechtold and Patterson 2005) require additional information about the stratification and population sizes. This information must be loaded into the various "population" tables.

FIA developed and uses the National Information Management System – Compilation System (NIMS-CS) to store and compile the data. NIMAC members have adapted the system for use in Wisconsin, since NIMAC is processing the data for this State. NIMAC has funded the development of the protocol sample design, which includes detailed metadata about each FIA or NIMAC inventory, the attributes collected, and their protocols. Once fully developed and integrated into NIMS, this will add the flexibility needed for NIMAC customers and for FIA to load past survey data into NIMS.

Since NIMS is designed for internal use, NIMAC chose not to try to transfer it to other NIMAC customers for their use yet. Instead, the Formatter program was written to transfer the data from the field tables (Oracle) to a Microsoft Access database that is an enhanced version of the FIADB (FIA Data Base) version available on the FIA website (http://fiatools.fs.fed.us/fiadb-downloads/fiadb3.html "Microsoft Access Database file ready for loading all of the FIADB data"). Currently, Formatter has no editing capabilities and limited ability to compute calculated fields. Enhancements will be added based on customer needs.

Spatial and Tabular Analysis Tools

Once the data are ready in the FIADB format (fia.fs.fed.us/library/database-documentation/FIADB_user_manual_v3-0.pdf, Forest Inventory and Analysis

2008), the data analysis can begin. FIA's primary analytical tool is Forest Inventory Data Online, or FIDO, (Wilson and Ibes 2005) and can be accessed at: http://www.fia.fs.fed.us/tools-data/. With it, the user can select the area of interest (state and counties), the survey years, the summary metric (e.g., area, volume, and biomass), the row and column categories, and apply data filters, if any. The results are presented as tables with associated sampling errors. Some results can also be displayed as maps.

NIMAC modified FIDO somewhat to analyze the Wisconsin inventory (as noted above). The Indiana DNR and Great Plains data will be analyzed using the EVALIDatorPC reporting tool (http://fiatools.fs.fed.us/fiadb-downloads/fiadb3.html). EVALIDatorPC uses the Microsoft Access database mentioned above to compute tables using the same estimation methods used by FIDO. NIMAC is enhancing EVALIDatorPC in order to provide some of the features, such as ratio estimation, included in NIMAC's TabGen analytical tool (Scott and Klopfer 1999).

NIMAC is also cooperating with Wisconsin on the development of a spatial intersection tool to select subsets of plots based on the mapped location. Analysts can use a geographic information system (GIS) to identify areas (polygons) of interest, such as broad buffer zones along streams or roads. These areas are then intersected with the plot locations to identify all the sample plots within those zones. Note that there are locational errors for both plots and map features, so this needs to be factored in when defining the polygon of interest. The plot list is then submitted along with the polygon area to FIDO for analysis. The tool will also be able to associate mapped categories with each plot so that these categories can be used in making the tables of estimates. If available, information on stratification for variance reduction can also be sent to FIDO.

The National Forest System has funded NIMAC to develop the Analytical Tool for Inventory and Monitoring (ATIM). ATIM will build on the strength of FIDO, focus on NFS inventory data, add GIS capabilities for input and output, and will add data query functionality. It will also give the user the flexibility to apply their own algorithms for calculated fields, such as volume equations.

Conclusions

NIMAC has been working on two primary areas: providing technical assistance to a variety of customers and developing inventory and monitoring software.

Technical Assistance

NIMAC continues to work with several states on intensifying their inventories on state-owned lands (IN and WI) or more targeted, specialized inventories (Great Plains). The work with the National Forest System is expanding, especially in the east. With the increasing emphasis on forest certification, carbon monitoring and on local estimates, the demand for NIMAC's assistance will likely increase both within and outside the United States.

Inventory and Monitoring Tools

While continuing to enhance the PC-based tools, Formater and EVALIDatorPC, much of NIMAC's focus will be on the development of the Design and Analytical Tools for Inventory and Monitoring. Although originally designed to meet the needs of National Forest System, many of the new features added will also be applicable to FIDO, thus enhancing FIDO's capabilities for all FIA customers.

It is the goal of NIMAC to assist customers (who are often already cooperators with FIA) in meeting their needs. This collaboration further strengthens the partnerships. In addition, the enhancements that NIMAC customers fund are then made available for use by FIA and other NIMAC customers. NIMAC's goal is to make more efficient and comparable estimates across the landscape, contributing to the sustainable use of our forest resources.

Literature Cited

Bechtold, W.A.; Patterson, P.L. 2005. The enhanced Forest Inventory and Analysis program--national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p. http://www.srs.fs.usda.gov/pubs/20371

Forest Inventory and Analysis Program. 2008. The forest inventory and analysis database: database description and users manual version 3.0 for Phase 2, revision 1. Washington, DC: U.S. Department of Agriculture, Forest Service. http://www.fia.fs.fed.us/library/database-documentation/FIADB_user_manual_v3-0.pdf.

Lister, A.J.; Scott, C.T. 2008. Use of space-filling curves to select sample locations in natural resource monitoring studies. Environmental Monitoring and Assessment. http://www.nrs.fs.fed.us/pubs/3672

Scott, C.T. 1993. Optimal design of a plot cluster for monitoring. In: Rennolls, K.; Gertner. G., eds. The optimal design of forest experiments and forest surveys; Proceedings, IUFRO S.4.11 conference. London, UK: University of Greenwich: 233-242.

Scott, C.T.; Klopfer, S.D. 1999. Standard forest sampling designs and their analysis using TabGen. In: Toward a unified framework for inventorying and monitoring forest ecosystem resources. Proc. RMRS-P-12. Ft. Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 87-89.

Wilson, B.T.; Ibes, W.S. 2005. Forest inventory and analysis information delivery architecture. In: Andersen, Kim V.; Debenham, John K. ; Wagner, Roland R., eds. Proceedings of 16th international workshop on database and expert systems applications (DEXA'05). [Place of publication unknown]: Springer: 706-710.

Design Tool for Inventory and Monitoring

Charles T. Scott [1]
Renate Bush[2]
Ken Brewer[3]

***Abstract**: Forest survey planning typically begins by determining the area to be sampled and the attributes to be measured. All too often the data are collected but underutilized because they did not address the critical management questions. The Design Tool for Inventory and Monitoring (DTIM) is being developed by the National Inventory and Monitoring Applications Center in collaboration with the National Forest System (NFS) and the Remote Sensing Applications Center. DTIM will assist inventory planners by stepping them through the monitoring process: 1) identifying key customers and broad objectives, 2) identifying monitoring questions, 3) identifying attributes needed to answer the questions, 4) evaluating existing data; then if additional data are needed, 5) setting precision and cost constraints, 6) selecting the sampling and plot designs and sample sizes; and, 7) selecting the sample locations. A prototype was successfully used on the Mark Twain National Forest and is being applied on several other National Forests. A team of NFS representatives is identifying the requirements for developing DTIM to help establish monitoring plans on National Forests as well as for use by other Forest Inventory and Analysis partners.*

Keywords: FIA, sampling design, plot design, software.

Introduction

Information on natural resources is important to any land management organization. The U.S. Forest Service spends at least $200 million annually on resource monitoring, and the expenditure could be more than $500 million. There are concerns that some of the monitoring is not statistically sound, defensible in court, efficient, consistent over time and space, well integrated across resources, or well utilized once data are collected. The 2008 National Forest Management Act Planning Rule requires that forest plans have an associated monitoring program to ensure that each national forest is progressing toward its desired

[1] United States Forest Service Northern Research Station, FIA, 11 Campus Blvd., Suite 200, Newtown Square, PA 19073 USA; ctscott@fs.fed.us; http://nrs.fs.fed.us/people/Scott

[2] United States Forest Service, Northern Region, 200 E Broadway, P.O. Box 7669, Missoula, MT 59807

[3] United States Forest Service, Remote Sensing Applications Center, 2222 W 2300 S, Salt Lake City, UT 84119

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

conditions and objectives. An inventory and monitoring toolkit is being developed to address these needs. A brief overview of the toolkit is presented followed by a more detailed description of the Design Tool for Inventory and Monitoring (DTIM).

Overview of Toolkit

Inventory and monitoring of natural resources is an involved and expensive process. The National Inventory and Monitoring Applications Center (NIMAC), part of Forest Inventory and Analysis (FIA) Program within the U.S. Forest Service, is focused on this issue. NIMAC has identified 15 inventory and monitoring steps from the initial identification of the information needs through the analysis and decision process. While the data collection process is familiar to many, most of the steps are not. Initial steps to identify measurable objectives and questions that can be answered are often overlooked or given little attention. Existing inventories are often used to answer questions after the fact without consideration for changes or managing costs in inventories. While analysis and reporting steps will be repeated after each remeasurement, the planning (design) steps are done once, thus land managers have little or no experience at initiating an inventory and monitoring system. Based upon experience gained with FIA, NIMAC provides technical assistance to others so that they can develop inventory and monitoring systems that are compatible with FIA data which are collected on all lands across the United States.

An inventory and monitoring toolkit is being built to help with most of the 15 steps. The toolkit has four main tools. The Design Tool for Inventory and Monitoring (DTIM) is used to identify monitoring needs, to evaluate existing data, and, if necessary, to specify the sampling design to balance cost and precision in order to address the monitoring questions. The Portable Data Recorder (PDR) Tool is software on PDRs and personal computers to collect, validate, and transfer field data. The Database and Compilation Tool is used to store the data and to compute calculated fields, such as tree volumes, biomass, and species richness. NIMAC and FIA are developing several spatial and tabular Analytical Tools for Inventory and Monitoring (ATIM) to estimate tables and produce maps. The tabular tools produce estimates with associated sampling errors based on the data. The spatial tool selects subsets of the data for estimation based on spatial layers, such as ownership, cover-type maps, and ecological classification maps. The vision for the completed set of tools, the relationships between the tools, and their roles are shown in Figure 1.

The Design Tool will play a major role in linking the tools. DTIM will provide the analytical tool, ATIM, with a list of tables to estimate. If existing data are insufficient, it also will provide metadata to the protocol sample design on data collection methods, the attributes to measure and variables to compute. DTIM computes the sample size and can create a list of plots which is then input to the data recorder software. DTIM also could be used to create portions of forest plan monitoring guides for national forests.

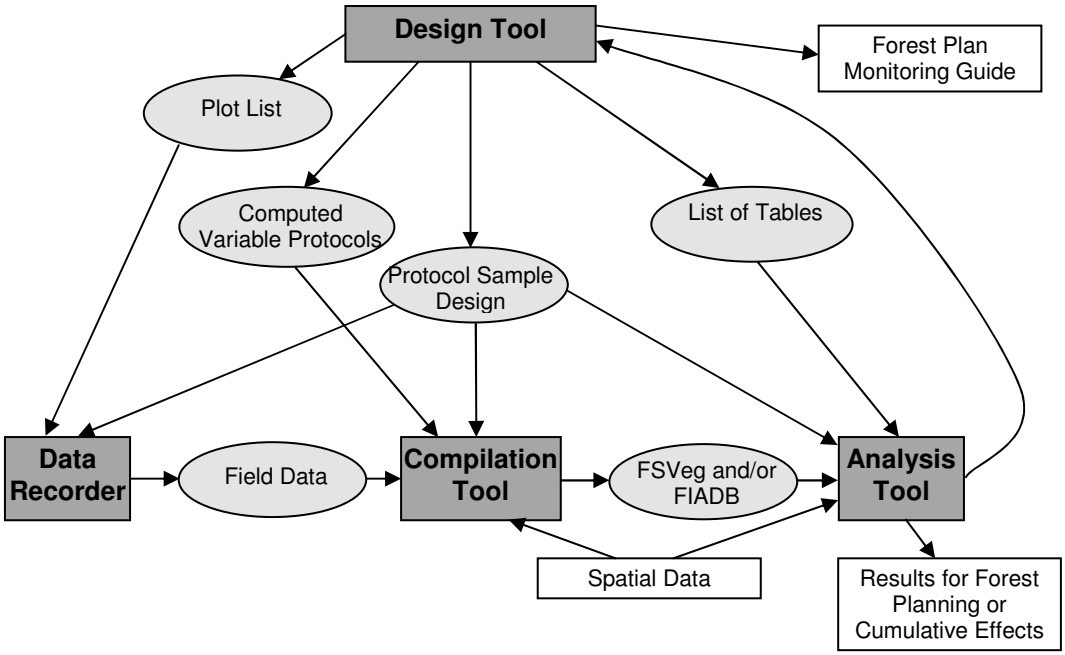


Figure 1: Diagram of the inventory and monitoring toolkit, its linkages, and its outputs.

Design Tool for Inventory and Monitoring (DTIM)

The Design Tool for Inventory and Monitoring is being developed to assist with the first 9 of the 15 monitoring steps. Using a knowledge base from monitoring experts, users will be presented with lists of choices to guide the development of the monitoring system that best suits their needs. These steps are done largely within a Microsoft Excel spreadsheet.

1. Identify customers and set broad objectives. – Pick from a list of broad objectives based on desired conditions or outcomes. Examples of broad objectives are: forest production, biodiversity, and ecosystem restoration.
2. Select monitoring questions. – Pick from list of generic questions based on the chosen objectives. Many of the questions relate to more than one broad objective.
3. Select attributes. – Pick from list of metrics for the questions chosen, then select from list of categorical attributes used to make tables to answer the questions. For example, to answer the question, "Are forests replacing themselves?", the chosen metric could be the number of seedlings and saplings by species (rows) and forest type (columns). Once the list of all attributes needed for tables is developed, then the attributes that were not selected should be evaluated to see if any should be included.
4. Assess existing data. – Use the Analytical Tool on FIA or other existing data to determine whether they are sufficient to answer the questions, in terms of precision of the estimates or whether the attributes have been

collected (data gaps). If so, then stop. If not, then use the ATIM to assess their variability for the next steps.

5. Set time/cost and precision constraints. – If existing data are imprecise or have gaps, then identify the costs of each component of monitoring. Specify the upper limit of funding. Specify precision requirements and the scale of analysis (e.g., forest or ecoregions) for key metrics, such as forest area, volume, or carbon.
6. Select sampling and plot designs. – Based on the selected attributes, specify which portions of the FIA plot need to be measured, or what components need to be added. DTIM can then be used to evaluate plot design options: FIA's Phase 2 (trees only), Phase 3 (forest health), or simplifications of them, including 1, 2 or 3 subplots or eliminating some of the components such as lichens, soils, downed woody material, or crowns. Based on the plot design chosen, determine sample size.
7. Select plots. – Based on the sample size chosen, select plot locations in a spatially balanced manner across the population.
8. Plan field work. – Since most attributes are chosen from the list provided in DTIM (so that they are compatible across the landscape), existing FIA field guides could supply the text for a field guide for the selected attributes.
9. Train field crews. – Since most attributes are chosen from the list provided in DTIM, existing FIA training materials could be accessed to provide training materials for the selected attributes.

This provides a quick overview of the nine steps that precede data collection, processing, analysis, and decisionmaking. The sample selection steps (steps 6 and 7) merit additional discussion.

Sampling Design Alternatives

Considerable research on forest inventory has resulted in a wealth of efficient alternatives for one-time surveys. However, when monitoring over time, many of these alternatives result in changing probabilities of selection over time, which can be difficult to handle. Hence, the standard sampling design alternative that DTIM evaluates is an intensification of FIA Phase 2 and/or Phase 3 plots (e.g., 2X for Phase 2 and 7X for Phase 3). This makes maximum use of the existing FIA plots which are permanent, fixed-area plots. If a higher intensity is required for certain areas, such as rare habitat, old growth, or other key management areas, then additional samples can be located within them. It is strongly recommended that this two-step process be used rather than partitioning the target areas first, then allocating the samples to each. This could create estimation problems for the whole population over time as those boundaries change. For this reason, pre-stratification with other than proportional allocation should only be used if it is a one-time inventory or when the strata boundaries are "permanent" (e.g., political boundaries rather than forest cover).

In general, NIMAC recommends intensifying using the standard FIA plot design (Bechtold and Patterson 2005). However, when high levels of intensification are used, then other plot designs may prove more efficient (e.g., using fewer or larger subplots). This is because the FIA design was chosen to be efficient when travel to and from the plot requires about half a day, thus the on-plot time should take much of the other half. If plots are located close to one another (within a mile or so), then sampling smaller plots allows the crew to sample more than one per day. However, smaller plots are more variable, so there is a tradeoff. Scott (1993) describes these tradeoffs by first identifying the different subplot types: overstory, understory, soils, etc. For each type of subplot, the survey planner can specify the size, the number of subplots, and their spatial arrangement (distance). By knowing the time it takes to do each and the resulting precision of the key metrics, an optimal plot design can be chosen. For FIA, the optimal design chosen has four subplots (Fig. 2) but for the much higher sampling intensities (less than 1 km apart), a single set of nested subplots can be more efficient (i.e., using only the center set of subplots).

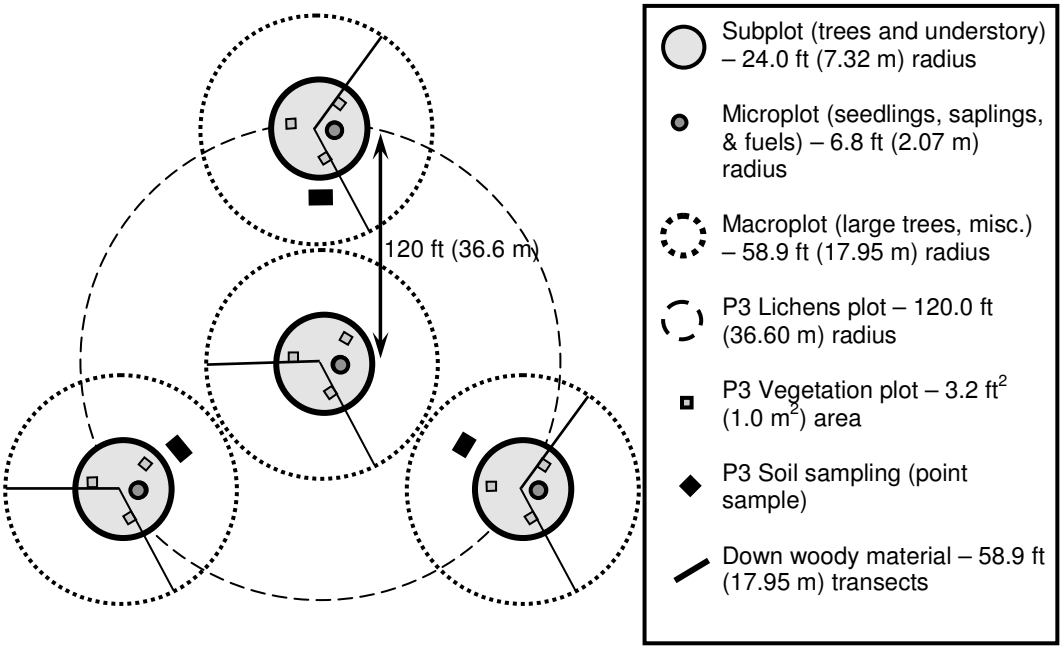


Figure 2: FIA plot design including both Phase 2 and Phase 3 subplots.

Optimization of the plot design features attempts to avoid over-sampling one attribute while under-sampling another. For example, if the regeneration sample is too variable relative to the overstory sample, then the regeneration subplot size or the number of regeneration subplots could be increased. Since the time required to sample the plot would thus increase, it might be necessary to decrease the overstory subplot size. DTIM will allow the user to evaluate these alternatives. Another alternative is to consider using a subsample for precise, expensive, or seasonal attributes, such as Phase 3 plots. In this case, not all attributes are measured on all plots, thus two (or more) plot kinds could be identified. DTIM helps evaluate tradeoffs between time on plot versus traveling

(increasing plot size versus increasing sample size). DTIM also attempts to efficiently use the available time in a day (no partial plots that require a second visit). Once the cost and precision relationships are developed, then the optimization can be performed in one of two ways: optimize precision for fixed cost, or minimize cost for fixed precision.

Select Plot Locations

FIA uses a hexagonal grid across the nation to spatially balance the plot locations (Bechtold and Patterson 2005). While this meets FIA's needs, it is difficult to achieve a fixed sample size when intensifying for a given area. NIMAC developed an alternative method which also is spatially balanced but results in the desired sample size (n) for the target population (Lister and Scott 2008). The approach is to use a geographic information system (GIS) to divide the population area into pixels, and use the GIS to aggregate the pixels into n equal area groups (clumps). Then randomly sample a point (or pixel) within each clump (Fig. 3).

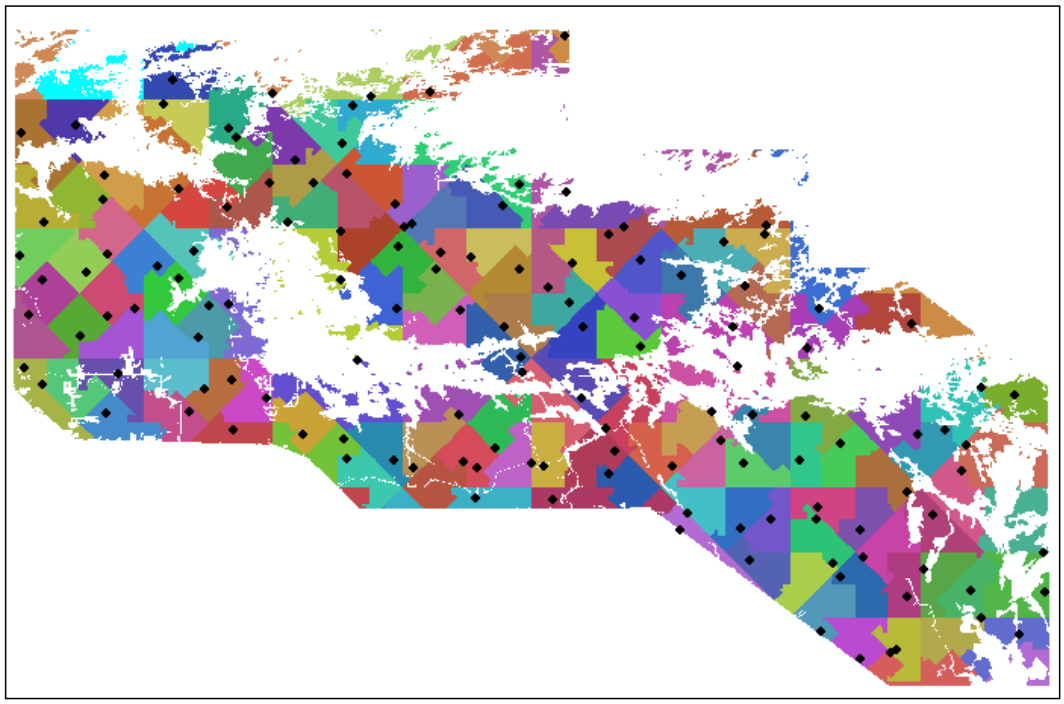


Figure 3: Example of spatially balanced sample using pixel groups from Lister and Scott (2008).

Space-filling curves are a mathematical means of creating a string of plots by connecting adjoining plots until all sample plots are included. A systematic sample of this string (e.g., every fifth plot) also is spatially balanced, thus the sample can be divided into annual panels (one for each measurement year) or into subsamples, such as Phase 3. Another approach to creating annual panels is to first group plots within compartments or other subpopulation of interest for which it would be helpful to have all plots completed in a single year. These groups can

then be formed into a space-filling curve and then divided into annual panels. This approach was used to monitor state forests in Indiana. It also could be applied to wilderness areas or other remote areas. The groups would be sized so that a crew could do an entire group during a single entry into the area (e.g., a 2-week sampling trip). For estimation purposes, the groups become the primary sampling units, thus variances would be computed using the group means.

Current Status and Path Forward

Early versions of DTIM were applied on the Mark Twain, Cherokee, Mississippi, Superior, Texas and Humboldt-Toiyabe National Forests, as well as to state forests in Wisconsin and Indiana. The Mark Twain collected the first season of data based on the intensified design in 2008. The other forests are in various stages of planning based on DTIM results.

The current version of this spreadsheet tool assists with steps 1-3, 5, and part of 6 (only evaluates sample size based on one to four subplots). A team of National Forest System (NFS) representatives is identifying the requirements for developing the Design Tool to help establish monitoring plans on National Forests as well as for use by other Forest Inventory and Analysis partners. Iterative use of DTIM has assisted in developing desired condition and objectives statements that are both meaningful and measurable. We will refine and complete DTIM so that it handles all 9 steps (with the help of ATIM). We also will include more questions and attributes with hope of extending beyond vegetation sampling to other resources that can be sampled with plots. We plan to expand DTIM's ability to evaluate sampling design alternatives and to optimize plot designs for customers who need to do so. DTIM will include a module to generate a spatially balanced list of plot locations. While designed to meet the needs of NFS, many of the new features also will be applicable to other NIMAC customers, such as states. NIMAC's goal is to make more efficient and comparable estimates across the landscape resulting in the sustainable use of our forest resources.

Literature Cited

Bechtold, W.A.; Patterson, P.L. 2005. The enhanced Forest Inventory and Analysis program--national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p. http://www.srs.fs.usda.gov/pubs/20371

Lister, A.J.; Scott, C.T. 2008. Use of space-filling curves to select sample locations in natural resource monitoring studies. Environmental Monitoring and Assessment. http://www.nrs.fs.fed.us/pubs/3672

Scott, C.T. 1993. Optimal design of a plot cluster for monitoring. In: Rennolls, K.; Gertner. G., eds. The optimal design of forest experiments and forest surveys; Proceedings, IUFRO S.4.11 conference. London, UK: University of Greenwich: 233-242.

Inventory of Trees in Nonforest Areas in the Great Plains States

Andrew Lister[1]
Charles Scott[2]
Steven Rasmussen[3]

***ABSTRACT**: The U.S. Forest Service's Forest Inventory and Analysis (FIA) program collects information on trees in areas that meet its definition of forest. However, the inventory excludes trees in areas that do not meet this definition, such as those found in isolated patches, in areas with sparse or predominantly herbaceous vegetation, in narrow strips (e.g., shelterbelts), or in riparian areas. In the Plains States, little is known about the tree resource in these noninventoried, nonforest areas, and there is a great deal of concern about the potential impact that invasive pests, such as the emerald ash borer, might have. To address this knowledge gap, the National Inventory and Monitoring Applications Center (NIMAC) has partnered with state cooperators to design and implement an inventory of trees in nonforest areas. The goal of the inventory is to characterize the tree resource using methods compatible with those of FIA so a holistic understanding of the resource can be obtained by integrating the two surveys. The goal of this paper is to give an overview of the goals and objectives of the inventory and to describe the plot and sample designs. Key findings related to the planning and establishment of the inventory are also provided.*

KEYWORDS: Trees outside of forest, nonforest tree inventory, emerald ash borer, isolated trees, Great Plains forest inventory, multiphase sample.

Introduction

The U.S. Forest Service's Forest Inventory and Analysis (FIA) program conducts an inventory of trees in areas meeting its definition of forested land use. This definition includes areas that are at least 1 acre in size, with certain geometric properties (e.g., at least 120 feet wide), of a current or former stocking level of at least 10 percent, and that are not subject to activities like mowing or understory clearing that would prevent normal regeneration (U.S. Forest Service 2007). FIA produces estimates of several forest parameters and creates statistical and analytical reports that are used by many customers including local, state, national

[1] U.S. Forest Service, Northern Research Station, FIA, 11 Campus Blvd., Suite 200, Newtown Square, PA 19073 USA; alister@fs.fed.us; http://nrs.fs.fed.us/people/alister
[2] U.S. Forest Service, Northern Research Station, FIA, 11 Campus Blvd., Suite 200, Newtown Square, PA 19073 USA
[3] Nebraska Forest Service, 510 North Pearl ST, Wayne, NE 68787 USA

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

and international scientists, land managers and other decision makers (Gillespie 1999).

North Dakota, South Dakota, Nebraska, and Kansas (hereafter referred to as the Plains States) are approximately 97 percent nonforest (Smith et al. 2004), and consist mostly of agricultural and grassland vegetation communities. Plains State resource agencies have recognized the lack of available information on the nonforest tree (NFT) resource, and how this knowledge gap might hinder wise management of these areas. Of particular concern is the spread of the emerald ash borer (EAB) (*Agrilus planipennis* Fairmaire), which, since being identified in 2002 near Detroit, MI, has been found in nine other Midwest and eastern states and into Canada. Although EAB has yet to be confirmed in the Plains States (as of January, 2009), this region has some of the highest relative density of EAB hosts (*Fraxinus spp.*), with mostly green ash (*Fraxinus pennsylvanica*) found along riparian areas, in conservation tree plantings and as a common tree in the communities across the plains (Figure 1). State forestry agencies in the Plains States, with funding assistance from the U.S. Forest Service's State & Private Forestry division, have undertaken an assessment of the potential economic and ecological impacts of EAB-induced ash mortality. The National Inventory and Monitoring Applications Center (NIMAC) was asked to assist with an inventory of the NFT resource to supplement the information that FIA collects on the tree resource in forested areas.

The goals of the inventory included obtaining state-level estimates of NFT parameters including area of land with different classes of NFT land use and estimates of total amounts of several continuous variables (e.g., basal area, volume, and biomass by species). An additional component of the inventory, not reported here, focused on NFTs in urban areas, with results serving as input to the Urban Forests Effects (UFORE) model, which, among other things, assigns estimates of value to urban forest components (Nowak and Crane 2000). Results from both inventories will be used by Plains States' resource agencies to gauge the economic and ecological impacts of potential ash species mortality, explore NFT utilization potential, develop educational materials, and to help formulate EAB mitigation strategies[4].

NIMAC helped incorporate FIA methodology into the design of the inventory to make NFT parameter estimates compatible with those obtained by the FIA program. In addition, the FIA data collection and processing infrastructure lends itself to being used in other, similar resource inventories. This paper describes how NIMAC extended traditional FIA plot and sample design methodology to the nontraditional Plains States NFT inventory.

[4] Nebraska Forest Service. 2007. Great Plains Tree and Forest Invasives Initiative; A multi-state cooperative effort for education, mitigation and utilization. Grant proposal. Unpublished document, on file at the Nebraska Forest Service, Lincoln, Nebraska.

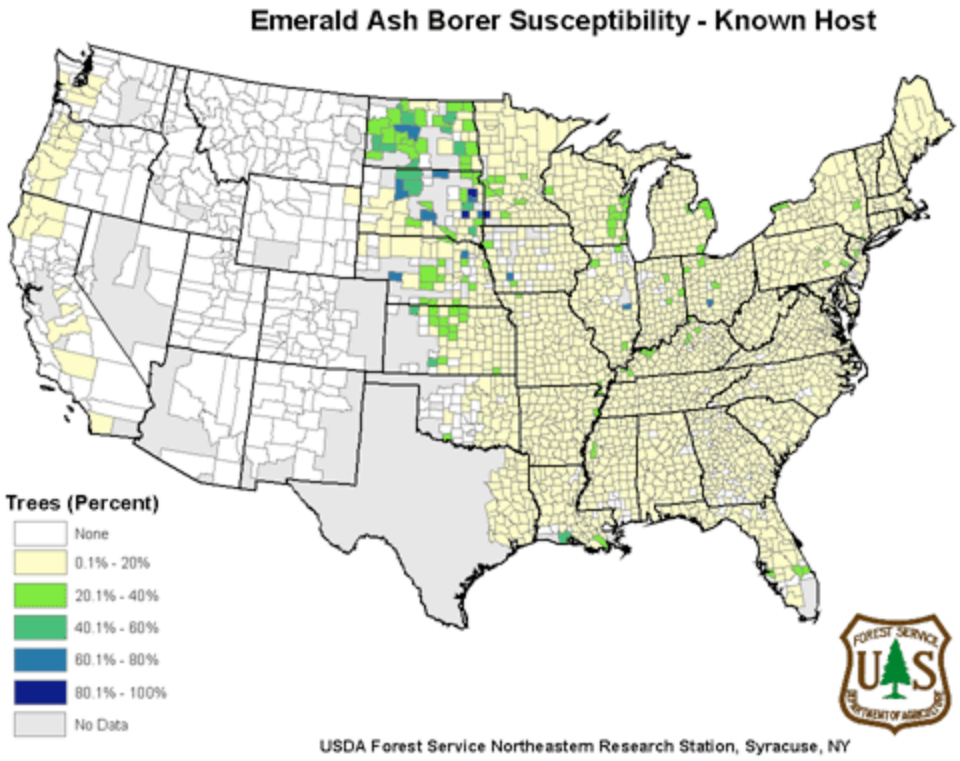


Figure 1. Map of relative density of species susceptible to the Emerald Ash Borer (from http://www.fs.fed.us/ne/syracuse/Data/Nation/data_list_eab.htm, accessed 11/15/2008).

Methods

NIMAC and Plains States forestry officials undertook a planning process that identified desired outcomes, precision requirements of NFT parameter estimates, existing FIA data sources, and new variables that were required to meet goals. The result of this process was the choice of a plot design that represents a tradeoff between a desire for compatibility with FIA methodology and cost effectiveness in the field. The field plot consisted of a single, 1/6-acre circular plot on which a variety of FIA, UFORE, and other site and tree-level attributes were recorded[5]. The FIA field guide, data recorder software program, and database storage system were adapted to accommodate the Plains States variables.

An assessment of the field data collection budget for the summer months of 2008 (the first year of data collection) revealed that 100 plots per state could be measured with the existing funding, which was divided equally among the states. This small number of plots compounded the concern that the attributes of interest,

[5] Great Plains Initiative Inventory Project, Data Collection Procedures. Version 1.0. Unpublished report on file at U.S. Forest Service, FIA Program, 11 Campus Blvd, Ste. 200, Newtown Square, PA 19073.

which are associated with NFT land use, occurred in only approximately 1 percent of the overall land area of the four states based on FIA estimates. In situations where there is potential to collect less costly information on a large number of elements in the population and to collect more costly, direct measurements of the attribute of interest on a subset, multi-phase sampling is suggested (Cochran 1977). For example, Holmgren et al. (1994) performed a study in which multi-phase sampling was found to be effective at characterizing the NFT resource in Africa. We chose a stratified, two-phase sample design for the NFT survey of the Plains States.

The first step was to stratify the four-state area into two strata (canopy and no canopy) using a derivative of the National Land Cover Dataset (NLCD). NLCD is a set of satellite image-based products produced by a consortium of federal agencies, led by the U.S. Geological Survey (Homer et al. 2007). These products are comprised of 30-m pixels, each labeled with a land cover category, percent impervious surface, and percent canopy cover estimates. To create the strata, a spatial filtering approach was applied to the percent canopy cover map. For each 3 x 3 block of pixels in the image (the focal window), the count of pixels with any estimated canopy cover in them was assigned to the center pixel of the block. The focal window was then shifted over one pixel, and the process was repeated for each pixel in the image. This resulted in an image containing pixel values of 0 (no canopy cover in the focal window) to 9 (all 9 input pixels contained canopy cover). This image was then recoded into the final stratum map: values of 0 were assigned to stratum 1, and all other values were assigned to stratum 2. For the four-state area, approximately 90 percent of the area fell into stratum 1, which we considered more likely to be devoid of trees.

The next step was to select elements within each stratum for the first phase of the two-phase sample. Phase 1 consisted of a large number of photo-interpretation (PI) plots overlaid on circa 2006 National Aerial Imagery Program (NAIP) imagery. Each plot consisted of 21 uniformly spaced points located within the footprint of potential ground plots – a 1/6-ac circle. Twenty-one points were chosen based on prior experience and a tradeoff between time cost per plot and completeness of area coverage for NFT assessment. The land use (using FIA definitions [U.S. Forest Service 2007]) of each of the 21 points was assessed and the count of points falling in the NFT land use category was recorded for each PI plot.

Based on consultation with PI specialists and a pilot assessment of the PI methodology, it was determined that the project budget allowed for 18,000 PI plots to be completed for each state. The number of PI plots per stratum was determined by optimal allocation because it was relatively inexpensive to perform the PI, and the population size and variability within some of these strata were large (Cochran 1977). To determine these optima, existing FIA ground plots were first assigned a stratum by intersecting them with the stratum map in a GIS. The variance of the binary-coded land-use category (NFT/non-NFT) of the center of

the center FIA subplots was then calculated for each stratum. These variances were used to proportionally allocate PI points to each stratum.

Phase 2 of the two-phase sample was established by assigning ground plots into substrata created from the PI results. Within each stratum, subsamples of PI plots were randomly chosen in a spatially balanced manner (Lister and Scott 2008) for ground visits based on a second stratification of the PI plots using NFT land use counts. The 0-21 range of potential NFT land-use count values was collapsed into three substrata within each primary stratum using information from FIA plot data as a guide. FIA plots all have a "percent forest" value assigned to them based on a ground mapping of land-use categories found on the plot. We assumed that the percent forest estimate on FIA plots is analogous to the proportion (out of 21) NFT land use estimate found on a PI plot. By making that assumption, we heuristically assessed how collapsing the FIA percent forest value into various configurations of three strata served to lower the variance of estimates of total number of trees and cubic foot volume from the FIA plots. We chose a count of three strata subjectively, guided by our desire to have an adequate number of sample plots in each stratum.

Once we arrived at a configuration of the FIA percent forest strata generally yielding the lowest variance for both tree density and total cubic foot volume (given an acceptable number of plots per stratum), we translated the stratum boundaries from the FIA percent forest scale (0-100 percent forest land use) to PI plot scale (0-21 NFT points/plot). We then used the resulting PI plot stratum boundaries and associated stratum areas to optimally allocate ground plots (phase 2) into each substratum. No ground plots were sampled in the first substratum of each stratum (the substrata with 0 NFT points). These substrata were assumed to have no NFT because of the high quality of the imagery and the cost of sampling the substratum. For example, Table 1 lists the area percentage in each primary stratum and the number of PI and ground plots by substrata for Kansas. The optimal allocation clearly shifts PI plots into stratum 2, eliminates substratum 0, and emphasizes substratum 2 versus 1.

TABLE 1: Example of allocation of photo-interpretation plots and ground plots by stratum and substratum for Kansas.

KANSAS	0-21 splits	Area (%)	Photo-interpretation Plots	Ground Plots
Stratum 1:		86.8%	10156	63
Substratum 0	0		9658	0
Substratum 1	1-6		300	25
Substratum 2	7-21		198	38
Stratum 2:		13.2%	7874	37
Substratum 0	0		6798	0
Substratum 1	1-8		630	16
Substratum 2	9-21		446	21

Conclusions

We learned much from the process of planning and establishing the Plains States NFT inventory. The process of setting inventory goals and choosing variables to meet these goals was a time-consuming, iterative process. Frequent meetings with all interested parties, as well as establishing expectations in light of the available budget, helped to ensure that the inventory would efficiently provide the answers to management questions.

Another finding was that photo-interpretation of nearly 80,000 plots (1,680,000 individual points to assess) required a great deal of effort to develop an efficient procedure, construct a manual, and manage and train analysts. We discovered a GIS procedure that vastly increased our productivity level and lowered costs. Furthermore, photo-interpreting plots in the Plains States was less complex because the vast majority of the plots assessed were completely devoid of trees. The experience we gained in this work will speed up future photo studies we conduct.

Finally, we learned that some of the existing FIA infrastructure, including the field guide and data recorder software, is adaptable for use in other, FIA-like resource inventories. By going through the process of adapting the FIA methodology to fit the NFT inventory, the project not only benefits from using pre-existing infrastructure, but also from the potential for integration of NFT results with those from FIA. Since the 800 urban plots and the 400 rural plots were collected with nearly identical, FIA-compatible methods, we plan to generate state-level estimates of tree parameters across all lands. The FIA program will benefit from this study by gaining additional information on the tree resource in these states for use in analytical reporting. It also will acquire new institutional knowledge related to its own inventory methodology as well as new techniques that can be implemented to address emerging resource issues.

Ground data from the 400 plot subsample of the PI plots, and more than 800 urban plots, were collected from the rural portions of the Plains States in the summer of 2008. As of December 2008, we are beginning data analysis. Results will be reported in future work.

Acknowledgments

We would like to offer special thanks to Eunice Alexander, Aaron Clark, Doug Griffith, and Andrea Stoneman for their many hours of hard work at helping develop and conduct the PI. We'd also like to thank Tom Albright, Dan Kaisershot, Cassie Kurtz, and Greg Liknes for their advice on FIA PI methods and logistics. Finally, we'd like to thank James Blehm, Katherine Johnson, Mary

Miller, and Jay Solomakos for their efforts to develop a field guide, data recorder, and training program for the project's field crews.

References

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

Gillespie, A. 1999. Rationale for a national annual forest inventory program. Journal of Forestry. 97(12): 16-20.

Holmgren, P.; Masakha, E.J.; Sjöholm, H. 1994. Not all African land is being degraded: a recent survey of trees on farms in Kenya reveals rapidly increasing forest resources. Ambio. 23(7): 390-395.

Homer, C.; Dewitz, J.; Fry, J.; Coan, M.; Hossain, M.; Larson, C.; Herold, N.; McKerrow, A.; VanDriel, J.N.; Wickham, J. 2007. Completion of the 2001 National Land Cover Database for the Conterminous United States. Photogrammetric Engineering & Remote Sensing. 73: 337-341.

Lister, A.J.; Scott, C.T. 2008. Use of space-filling curves to select sample locations in natural resource monitoring studies. Environmental Monitoring and Assessment. 149(1-4): 71-80.

Nowak, D.J; Crane, D.E. 2000. The Urban Forest Effects (UFORE) model: quantifying urban forest structure and functions. pp. 714-720. In: Integrated tools for natural resources inventories in the 21st century. Hansen, M.; Burk, T., eds. Gen. Tech. Rep. NC-212. St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station. 744 p.

Smith, W.B.; Miles, P.D.; Vissage, J.S.; Pugh, S.A. 2004. Forest resources of the United States, 2002. Gen. Tech. Rep. NC-241. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 146 p.

U.S. Forest Service. 2007. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 4.0. www.nrs.fs.fed.us/fia/data-collection/. Accessed 11/15/2008.

Effects of Plot Size on Forest-Type Algorithm Accuracy

James A. Westfall[1]

Abstract: *The Forest Inventory and Analysis (FIA) program utilizes an algorithm to consistently determine the forest type for forested conditions on sample plots. Forest type is determined from tree size and species information. Thus, the accuracy of results is often dependent on the number of trees present, which is highly correlated with plot area. This research examines the sensitivity of a forest-type algorithm to changes in amounts and types of input data that result from altering the sample plot area. Logistic regression was used to determine which plot metrics have the most influence on algorithm output. Relationships between plot area and key variables such as number of species, number of trees, and total basal area were established and applied to the regression models. The results allow for assessment of algorithm accuracy over a range of plot sizes. The algorithm was generally robust to changes in area for loblolly/shortleaf, oak/hickory, and oak/gum/cypress type groups. Algorithm accuracy was mediocre for other type groups, with oak/pine having the poorest performance. A comparison between field-observed forest type and algorithm output showed average agreement rates of near 90 percent when computed types were conifer. However, agreement rates were lower for hardwood groups, especially when the computed type was aspen/birch. Better alignment between the field- and algorithm-based determinations may be achieved by providing real-time algorithm output to field crews.*

Keywords: forest inventory, logistic regression, species diversity, classification

[1] Research Forester, USDA Forest Service, Northern Research Station, Newtown Square, PA.

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

Eyre (1980) describes forest type as a "descriptive classification of forestland based on present occupancy of an area by tree species". The contributions to site occupancy are often determined via the numbers and sizes (e.g., diameter at breast height [dbh]) of trees for each species (Hansen and Hahn 1992). The relative occupancies among species (or groups of species) are used to establish the forest-type classification. Due to the relatively large number of described forest types and pronounced similarities among a number of types, forest-type groups are often created. This allows a number of related forest types to be classified under a single designation, which is often useful for broader analytical summarizations.

In many forest inventories, the forest type may be assessed by the field crew at the time the sample data are collected, determined at a later time by applying a computer algorithm to the sample plot condition data, or both. The Forest Inventory and Analysis (FIA) program of the U.S. Forest Service uses both field- (USDA 2007) and algorithm-based (Arner et al. 2003) forest-type information. Generally, the algorithm-based forest type is used in estimation. However, if a forested condition is less than one subplot in area (~0.0415 ac) the field-based forest type is used. It is assumed the algorithm cannot accurately determine the forest type when the area is relatively small, because often few trees are present on which to make a determination. As area and numbers of trees are highly correlated, the question that arises is what affect does sampled-area size have on an algorithm-based determination of forest-type. An understanding of the accuracy of algorithm-determined forest type in relation to area sampled will allow forest managers to make informed decisions regarding the appropriate method of forest-type classification for particular forest inventory designs.

Data

Evaluation of algorithm classifications at various sampled-area sizes was accomplished using FIA data from Indiana (1999-2003), South Carolina (2002-2006), and Maine (1999-2003). The states were chosen so that many of the forest types encountered in the eastern United States. would be represented. The data were collected under the annual inventory design outlined by Bechtold and Patterson (2005). Sample plots are composed of four subplots, each having a 24-ft radius. Within each subplot is a microplot having 6.8-ft radius. Trees having 5.0 in. or larger dbh were tallied on the subplots. Sapling (1.0-4.9 in. dbh) and seedling (< 1.0 in. dbh with minimum height criteria) data were recorded on the microplots. To facilitate the analysis, only single-condition plots were retained. In order to have a large number of possible plot combinations for repeated simulations, only forest-type groups having more than 100 plots were evaluated. There were 3,712 plots in the study data representing 55 forest types within eight forest-type groups. Table 1 provides a summary of the data by forest-type group and forest type.

Methods

The forest-type algorithm used in this study is described by Arner et al. (2003). This algorithm uses relative stocking to assess the site occupancy of sample trees. Individual-tree stocking values are computed from species-specific equations using tree dbh. Further adjustments (e.g., weighting) may be made based on tree-size classification and social position. The individual-tree values are aggregated into initial type assemblages and the stocking totals of these initial groups are evaluated via decision rules to determine the final forest type. Forest types are hierarchically assigned to a more generic forest-type group, so forest-type group determination is straightforward once the forest type is established.

The accuracy of the algorithm-based classifications was examined for forest-type groups, which are assemblages of similar forest types. The analysis consisted of two phases: 1) combining a number of plots with the same forest-type group and then systematically reducing the area of the combined plots and re-evaluating the forest-type group to see if the classification changes; and 2) using the results of (1), perform logistic regression to evaluate which plot attributes are correlated with the classification changes and predict probabilities of correct classification.

In the first phase, a Monte-Carlo simulation (Metropolis and Ulam 1949) was performed by combining 30 randomly selected plots (without replacement) having identical forest-type group classification into a 'population' of 5 acres in size (30 x 1/6 ac = 5 ac). Forest-type group was determined for this combination of plots. The area was then reduced by 1/24 ac by removing a randomly selected subplot and the forest-type group was re-evaluated. This area reduction method was carried out until only a single subplot remained (1/24 ac). This allowed for evaluation of potential forest inventory plot sizes ranging from 1/24 ac to 5 ac. The resultant output for the 120 different plot sizes included a binary variable that indicated whether the classification had changed from the original type and also summary variables such as numbers of species and numbers of stems for seedlings, saplings (1.0-4.9 in. dbh), and trees (5.0+ in. dbh), and basal area for saplings and trees. This process was repeated 500 times for each forest-type group; results were quite stable after 300 iterations.

These data were then used in a logistic regression analysis where the binary response variable was whether or not the type classification had changed at any given reduced area. Independent model variables considered were the summary variables described above (with two-way and three-way interactions). A stepwise variable-selection procedure was used to identify variables having significant (α = 0.10) predictive ability. The α level of 0.10 was chosen to promote inclusion of more variables that may help explain the classification changes. These logistic regression models provided the basis for predicting the probability that forest-type group would be correctly identified at a specified plot size.

Regression models relating the summary variables to plot area were developed to describe average plot attributes at the various plot sizes. The relationships in the data suggested linear relationships between plot area and numbers of stems as well as plot area and basal area. Nonlinear relationships existed between area and numbers of species. The model forms were:

$$S_{jk} = \beta_{1jk} A_j + \varepsilon_{jk} \quad [1]$$

$$SPP_{jk} = \beta_{2jk} \times A_j^{\beta_{3jk}} + \varepsilon_{jk} \quad [2]$$

$$BA_{jk} = \beta_{4jk} A_j + \varepsilon_{jk} \quad [3]$$

where: j = tree size class (seedling, sapling, and tree)
k = forest-type group
S_{jk} = number of stems tallied for tree size class *j*, forest type *k*
SPP_{jk} = number of species tallied for tree size class *j*, forest type *k*
BA_{jk} = basal area of stems tallied for tree size class *j*, forest type *k*
A_j = sampled plot area (ac) for tree size class *j*
ε_{jk} = random error component for tree size class *j*, forest type *k*
$\beta_{1jk} - \beta_{4jk}$ = estimated coefficients for tree size class *j*, forest type *k*

The estimated coefficients are presented in Table 2. The predicted values from models [1] through [3] were used as inputs into the logistic regression model to predict the probability of misclassification for a given plot area. This analytical approach was carried out separately for each forest-type group.

Results

The logistic regression analyses were conducted for the eight forest-type groups. The general form of the model was:

$$P_k(\text{Correct}) = f(S_{jk}, BA_{jk}, SPP_{jk}, \times 2, \times 3) + \varepsilon_k \quad [4]$$

where: P_k(Correct) = Probability of correct classification for forest-type group *k*
×2 = all two-way interactions of the predictor variables
×3 = all three-way interactions of the predictor variables
ε_k = random error component for forest-type group *k*
all others as defined above

The variables chosen by the stepwise selection procedure varied considerably among the groups. Across all eight type groups analyzed, there were 34 different significant predictor variables related to the probability of correct classification of forest-type group (the detailed information is not provided here due to size limits). The models fit the data reasonably well with R^2 values ranging from 0.43 to 0.64 (Table 3). The AIC (Akaike 1974) statistics also showed that the addition of

covariates to the model substantially improved the prediction when compared to an intercept-only model.

The probability of correct classification of the white/red/jack pine and spruce/fir groups was influenced primarily by numbers of stems, numbers of different species, and basal area for saplings and trees. The classification accuracy of the loblolly/shortleaf pine group was affected mostly by numbers of stems, numbers of different species, and basal area for trees only. Conversely, the hardwood-type groups were more complex due to increased numbers of significant predictor variables, such numbers of stems and numbers of species for seedlings and various two-way interactions between these variables and the sapling and tree covariates. The oak/pine group had the most intricate model, with numerous three-way interactions being significant explanatory variables.

Inputs into the logistic regression model for each forest-type group were generated using models [1] through [3] for plot sizes ranging from 1/24 to 5 ac. The sensitivity of the algorithm to changes in forest parameters due to sample plot size was dependent upon the type group of interest. Within conifer types, the loblolly/shortleaf pine forest-type group was the most robust, as the probability of classification error was only 0.15 for 1/24 ac plot size (Figure 1c). The white/red/jack pine and spruce/fir type groups were more sensitive to area sampled, with the probability of misclassification being 0.3 - 0.4 at a plot area of only 1/24 ac (Figure 1a, 1b). A sampled area of roughly 0.2 ac. was needed to attain a nearly zero misclassification probability for loblolly/shortleaf, while the other two conifer types required about 0.5 ac.

For hardwood forest-type groups, the most stable classifications across the various plot sizes were in the oak/hickory and oak/gum/cypress groups (Figure 1e, 1f). For these groups, the probability of misclassification was near 0.1 at the smallest plot size evaluated (1/24 ac). Near-zero probabilities were achieved at a plot size of roughly 0.25 ac for oak/gum/cypress and nearly 0.5 ac for oak/hickory. The oak/pine group required plot sizes of over 2.5 ac to attain near-zero misclassification rates (Figure 1d). At a 1/24 ac plot size, the oak/pine group had misclassification probability of 0.62 and was 0.24 for the maple/beech/birch group. The maple/beech/birch group required a plot size of about 0.45 ac to obtain a misclassification probability less than 0.001 (Figure 1g). For the aspen/birch group, the maximum misclassification probability was near 0.29 (at $1/24^{th}$ ac plot size) and near-zero probabilities occurred at about 0.9 ac (Figure 1h).

The forest-type algorithm always provides the same forest-type group for a given set of input data from the sample plot. However, the field crews have the advantage of viewing the entire area – their determination is not limited to only trees within the sample plot. Also, a certain amount of subjectivity is introduced based on the field crew's perception of the area. These factors can result in differing outcomes between the field-based and algorithm-based forest-type group. Table 4 quantifies the agreement/disagreement proportions for the forest-

type groups analyzed in this study. Agreement was relatively high for softwoods, with red/white/jack pine having ~81 percent agreement and both spruce/fir and loblolly/shortleaf having agreement rates exceeding 90 percent. The conformity for hardwoods was poorer, as both aspen/birch and oak/pine had agreement rates less than 50 percent. When the algorithm determined the type was aspen/birch, the field call was spruce/fir for nearly 40 percent of the plots. The best agreement between algorithm and field hardwood type groups was for oak/gum/cypress, which had identical results for roughly 88 percent of the plots. Overall, agreement between field crew and algorithm occurred for ~ 75 percent of the plots.

Discussion/Conclusion

For the red/white/jack pine, spruce/fir, and maple/beech/birch groups, the algorithm classification accuracies decreased relatively quickly at plot sizes below 1/4 ac. This outcome is a reflection of the algorithm threshold for information needed to accurately classify these type groups. A review of the description for each type group indicates a wide range of species occur within these type groups (Eyre 1980). For example, spruce and fir species occur in areas where aspen, birch, and maple are also present. As plot size is reduced below 1/4 ac, the dominance of the spruce/fir species becomes more ambiguous, and the decision rules employed in the algorithm may produce a classification outside the spruce/fir group. The most common classification error for both red/white/jack pine and spruce/fir groups was maple/beech/birch. Similarly, a common misclassification of maple/beech/birch was spruce/fir type.

In contrast, there should be much less concern regarding misclassification of the loblolly/shortleaf pine group. These plots often come from planted areas where other species (primarily hardwoods) occur in the understory, which makes the preeminence of the primary species more apparent for smaller plots. In cases where loblolly/shortleaf was misclassified, oak/pine was by far the most common outcome.

A notable characteristic for the oak/pine and (to a lesser extent) aspen/birch groups was a relatively slow improvement in classification accuracy as plot sizes increased. For oak/pine, numbers of species, numbers of stems, and basal area among the three tree size classes all contributed to the misclassification rate. The confusion within aspen/birch was due primarily to species, stems, and basal area of trees having dbh 5.0 in. or larger. The oak/pine group required over 2.5 ac plot size to attain near-zero misclassification probabilities, while the aspen/birch group needed slightly less than 1 ac. In addition, the oak/pine group had the worst classification accuracy of all groups evaluated, with a probability of misclassification exceeding 0.6 when plot size was 1/24 ac. This gives further support to the argument given above related to species mixes. On plots where there is a wide range of species, it is difficult to determine the dominant type and relatively small shifts in the tree list can sway the classification in a different

direction. Common misclassifications of oak/pine were loblolly/shortleaf and oak/hickory groups. The aspen/birch group was most often mistaken with spruce/fir and while maple/beech/birch, owing to the primary species of this group often being replaced by more shade-tolerant species, resulting in relatively high numbers of species and differing tree sizes.

The relationships between area sampled and misclassification probability for the oak/hickory and oak/gum/cypress groups were similar to those for loblolly/shortleaf pine. This is presumably attributable to the tendency for these species groups to be fairly well defined, such that the dominant species are likely to survive and flourish relative to species that are primary to other type groups. The oak/gum/cypress sites also tend to be undisturbed and have large diameter trees. These large trees provide high stocking values that are very influential in the computations, especially at the smaller plot sizes. Misclassifications were due primarily to confusion with the oak/pine and either maple/beech/birch or elm/ash/red maple groups.

There are two primary differences between field observation and algorithm-based forest-type group determination. The field crews have the advantage of viewing the broader area, not just the area within the plot. However, there is also an element of subjectivity such that different crews may resolve different forest types when assessing the same area. A feature of the algorithm is that the same forest type will be computed for a given tree list, removing any subjectivity. The drawback of the algorithm is that performance is suspect when there are not many trees. These differences can result in conflicting determinations of forest-type group. It is shown in Table 4 that when a computed type group is either oak/hickory or oak/pine, a wide range of different types are recorded by the field crew. It is also shown that a computed aspen/birch type is seen as spruce/fir for almost 40 percent of the plots and is judged to be maple/beech/birch for 14 percent of the plots. This suggests that 1) the tree species and size composition over the broader area differs somewhat from that within the sample plot area only; and/or 2) the relative importance afforded to the various tree sizes and species differ between the field crew and the algorithm.

This leads to another point regarding species composition. One would expect that increases in species diversity occur in transition zones near the edges of stands of differing type groups and more generally near the indistinct boundaries of natural ranges of type groups. In these zones, the increased diversity may lead to higher levels of classification error, as well as additional disparity between the field determination and algorithm output. Such analyses are beyond the scope of this paper, but the concept is worth highlighting as a future research topic.

A dilemma for analysts is whether to use an algorithm or the field-observed forest-type group. This choice could result in large shifts in estimated area for certain forest-type groups. There is a need to better align the field forest-type group with that computed by the algorithm. Given that crews collect data with

electronic data recorders, improved consistency may be obtained by having the algorithm provide real-time feedback on the computed forest-type group. This would allow the field crews to see when there is disagreement. This may 1) allow the field crews to calibrate their observations to be more consistent with algorithm output; and 2) provide feedback that sheds light on needed modifications to improve algorithm accuracy.

In summary, the algorithm was generally robust to changes in plot size for loblolly/shortleaf, oak/hickory, and oak/gum/cypress groups. For classification of other forest-type groups, the recommended plot size should reflect the relative proportions of occurring type groups and be consistent with levels of misclassification that are considered tolerable. For example, if the area is composed primarily of aspen/birch then a larger plot size should be considered than if the area is mostly oak/gum/cypress. Ultimately, it would be desirable to refine the algorithm such that all forest-type groups had similar (small) misclassification probabilities. This paper provides an analytical framework for evaluating whether changes to the algorithm provide improved classification consistency.

Literature Cited

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 19(6): 716–723.

Arner, S.L.; Woudenberg, S.; Waters, S.; Vissage, J.; MacLean, C.; Thompson, M.; Hansen, M. 2003. National algorithms for stocking class, stand size class, and forest type. Available: http://www.fia.fs.fed.us/library/field-guides-methods-proc/docs/National%20algorithms.doc

Bechtold, W.A.; Patterson, P.L., eds. 2005. The enhanced forest inventory and analysis program – national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

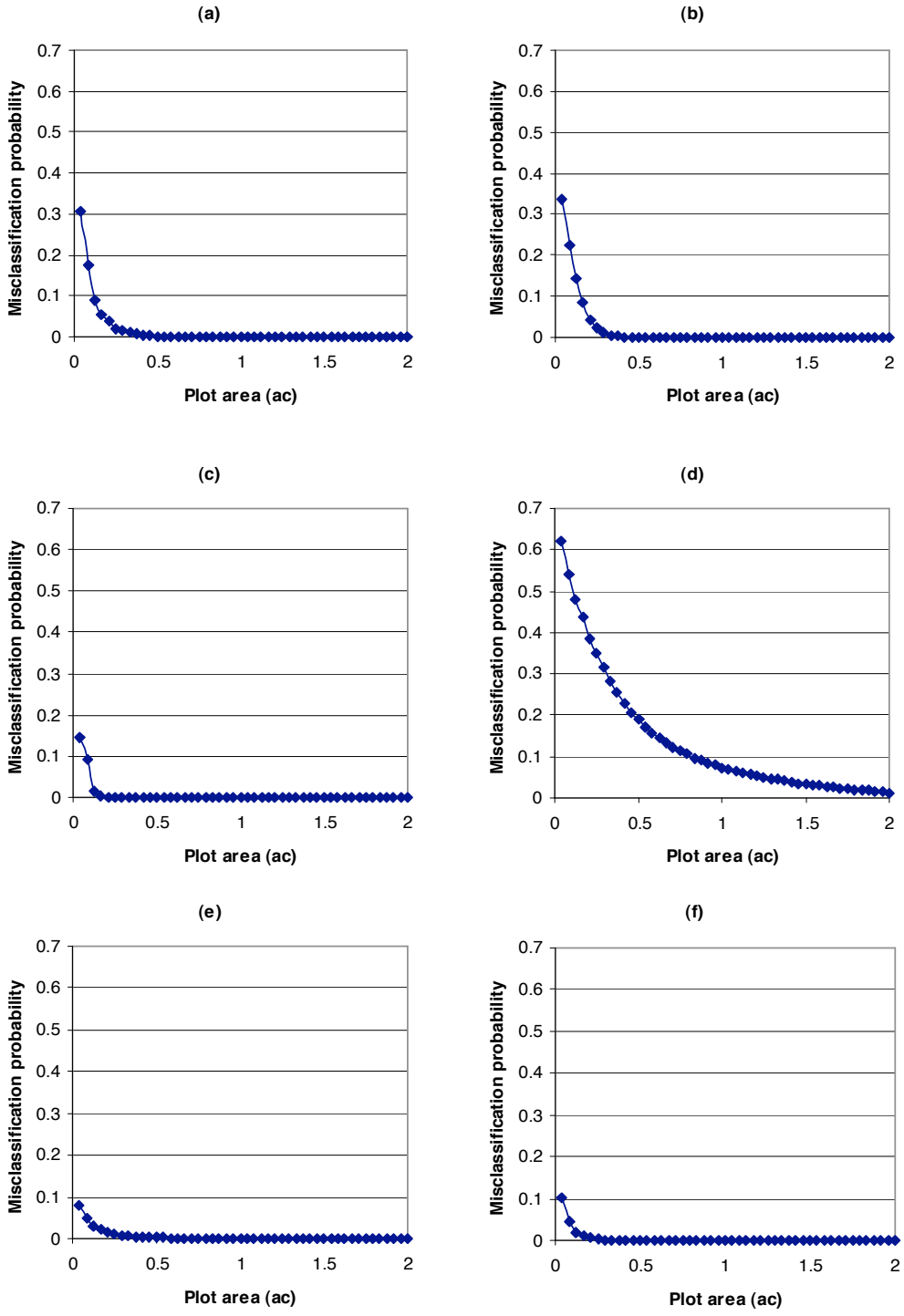
Eyre, F.H., ed. 1980. Forest cover types of the United States and Canada. Washington, DC: Society of American Foresters. 148 p.

Hansen, M.H.; Hahn, J.T. 1992. Determining stocking, forest type, and stand-size class from forest inventory data. Northern Journal of Applied Forestry. 9(3): 82-89.

Metropolis, N.; Ulam, S. 1949. The Monte Carlo method. Journal of American Statistical Association. 44: 335-341.

U.S. Department of Agriculture. 2007. Forest inventory and analysis national core field guide: volume I: field data collection procedures for phase 2 plots, version 4.0. Washington, DC: U.S. Department of Agriculture, Forest Service, Forest Inventory

and Analysis. Available: http://www.fia.fs.fed.us/library/field-guides-methods-proc/docs/core_ver_4-0_10_2007_p2.pdf


(a)
(b)
(c)
(d)
(e)
(f)
Misclassification probability
Plot area (ac)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.5
1
1.5
2

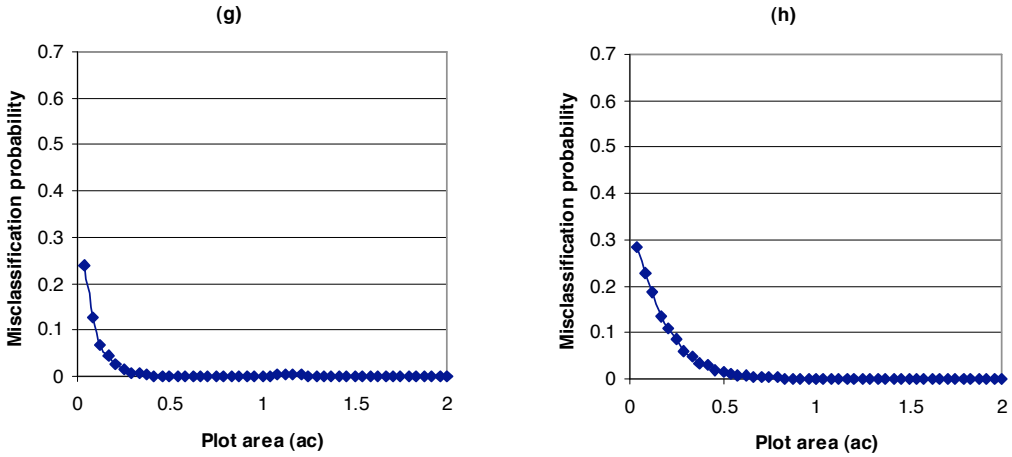


Figure 1: Misclassification probability vs. plot area for a) red/white/jack pine; b) spruce/fir; c) loblolly/shortleaf; d) oak/pine; e) oak/hickory; f) oak/gum/cypress; g) maple/beech/birch; and h) aspen/birch type groups.

Table 1: Data summary statistics by forest type and forest-type group.

Forest type group	Forest type	# plots	No. stems/plot[a]			No. species/plot		
			Min.	Mean	Max.	Min.	Mean	Max.
White/red/jack pine	Jack pine	1	55	55	55	5	5	5
White/red/jack pine	Red pine	7	41	86	113	6	10	14
White/red/jack pine	Eastern white pine	58	30	74	154	3	8	15
White/red/jack pine	White pine/hemlock	28	36	74	119	6	8	14
White/red/jack pine	Eastern hemlock	30	43	93	135	4	9	13
		124	**30**	**79**	**154**	**3**	**9**	**15**
Spruce/fir	Balsam fir	269	7	95	194	2	8	15
Spruce/fir	White spruce	14	12	53	103	1	5	9
Spruce/fir	Red spruce	151	32	104	187	1	8	13
Spruce/fir	Red spruce/balsam fir	137	25	97	177	3	8	13
Spruce/fir	Black spruce	73	4	73	141	1	5	13
Spruce/fir	Tamarack	5	31	72	122	5	7	10
Spruce/fir	Northern white-cedar	119	38	116	177	3	9	16
		768	**4**	**97**	**194**	**1**	**8**	**16**
Loblolly/shortleaf pine	Loblolly pine	512	4	65	308	1	7	17
Loblolly/shortleaf pine	Shortleaf pine	7	3	89	167	1	13	19
Loblolly/shortleaf pine	Virginia pine	9	32	76	114	10	14	19
Loblolly/shortleaf pine	Pond pine	9	33	59	106	2	7	10
		537	**3**	**66**	**308**	**1**	**7**	**19**
Oak/pine	White pine/red oak/white ash	37	12	66	103	3	9	14
Oak/pine	Eastern redcedar/hardwood	14	7	68	124	6	13	20
Oak/pine	Longleaf pine/oak	13	20	41	70	4	7	12
Oak/pine	Shortleaf pine/oak	9	31	72	97	6	11	18
Oak/pine	Virginia pine/southern red oak	6	28	60	92	3	13	22
Oak/pine	Loblolly pine/hardwood	80	12	62	147	3	9	19
Oak/pine	Slash pine/hardwood	4	31	44	60	9	10	12
Oak/pine	Other pine/hardwood	6	42	65	92	3	6	8
		169	**7**	**62**	**147**	**3**	**10**	**22**
Oak/hickory	Post oak/blackjack oak	12	20	68	116	8	12	18
Oak/hickory	Chestnut oak	12	24	55	115	4	8	15
Oak/hickory	White oak/red oak/hickory	190	15	63	283	5	12	24
Oak/hickory	White oak	30	20	66	138	4	12	18
Oak/hickory	Northern red oak	19	34	65	84	5	8	14
Oak/hickory	Yellow-poplar/white oak/red oak	35	21	67	158	6	14	26
Oak/hickory	Sassafras/persimmon	19	1	54	101	1	9	18
Oak/hickory	Sweetgum/yellow-poplar	50	27	59	122	3	10	18
Oak/hickory	Bur oak	1	24	24	24	3	3	3
Oak/hickory	Scarlet oak	3	57	63	72	8	11	14
Oak/hickory	Yellow-poplar	9	33	71	152	9	12	16
Oak/hickory	Black walnut	2	27	33	38	7	8	9
Oak/hickory	Black locust	1	68	68	68	10	10	10
Oak/hickory	Southern scrub oak	10	19	35	73	1	6	11
Oak/hickory	Chestnut oak/black oak/scarlet oak	15	18	47	87	3	11	21
Oak/hickory	Red maple/oak	11	2	59	143	1	8	12
Oak/hickory	Mixed upland hardwoods	103	2	62	380	2	10	21
		522	**1**	**62**	**380**	**1**	**11**	**26**
Oak/gum/cypress	Swamp chestnut oak/cherrybark oak	7	33	46	64	8	11	15
Oak/gum/cypress	Sweetgum/Nuttall oak/willow oak	84	3	47	97	1	9	17
Oak/gum/cypress	Overcup oak/water hickory	4	13	27	46	6	10	14
Oak/gum/cypress	Baldcypress/water tupelo	33	25	54	92	1	7	15
Oak/gum/cypress	Sweetbay/swamp tupelo/red maple	77	19	56	145	3	8	21
		205	**3**	**51**	**145**	**1**	**8**	**21**
Maple/beech/birch	Sugar maple/beech/yellow birch	978	20	92	172	4	9	19
Maple/beech/birch	Black cherry	2	17	32	47	4	6	7
Maple/beech/birch	Cherry/ash/yellow-poplar	33	22	72	156	4	10	20
Maple/beech/birch	Hard maple/basswood	9	33	60	119	5	10	14
Maple/beech/birch	Elm/ash/locust	1	22	22	22	6	6	6
Maple/beech/birch	Red maple/upland	66	10	87	147	2	9	14
		1089	**10**	**91**	**172**	**2**	**9**	**20**
Aspen/birch	Aspen	114	3	90	151	1	9	17
Aspen/birch	Paper birch	173	4	89	185	1	8	17
Aspen/birch	Balsam poplar	11	57	97	162	6	10	15
		298	**3**	**90**	**185**	**1**	**9**	**17**

[a] Includes all tallied seedlings, saplings, and trees.

Table 2: Estimated coefficients for numbers of stems, numbers of species, and basal area models [1] through [3] by tree size class.

		Number stems [1]			Number spp [2]			Basal area [3]	
Forest type group	**Tree size class**	β_1	**Pr > |t|**	β_2	**Pr > |t|**	β_3	**Pr > |t|**	β_4	**Pr > |t|**
Red/white/jack pine	Seedling	2718.5473	<.0001	47.4815	<.0001	0.3425	<.0001	-	-
Spruce/fir	Seedling	3725.9365	<.0001	35.6997	<.0001	0.2836	<.0001	-	-
Loblolly/shortleaf	Seedling	1618.1861	<.0001	60.6337	<.0001	0.4410	<.0001	-	-
Oak/pine	Seedling	2172.1193	<.0001	101.0000	<.0001	0.4567	<.0001	-	-
Oak/hickory	Seedling	2504.3036	<.0001	92.2299	<.0001	0.4079	<.0001	-	-
Oak/gum/cypress	Seedling	996.9864	<.0001	62.8282	<.0001	0.5128	<.0001	-	-
Maple/beech/birch	Seedling	3993.0822	<.0001	50.4331	<.0001	0.3389	<.0001	-	-
Aspen/birch	Seedling	3694.8156	<.0001	41.9937	<.0001	0.2634	<.0001	-	-
Red/white/jack pine	Sapling	548.4302	<.0001	35.0012	<.0001	0.4799	<.0001	16.8095	<.0001
Spruce/fir	Sapling	1468.3205	<.0001	29.8807	<.0001	0.3948	<.0001	36.6782	<.0001
Loblolly/shortleaf	Sapling	623.1041	<.0001	45.3810	<.0001	0.5509	<.0001	21.1109	<.0001
Oak/pine	Sapling	648.6880	<.0001	83.9266	<.0001	0.6005	<.0001	19.4474	<.0001
Oak/hickory	Sapling	517.2282	<.0001	76.0099	<.0001	0.5895	<.0001	15.9016	<.0001
Oak/gum/cypress	Sapling	647.3272	<.0001	51.7741	<.0001	0.5139	<.0001	19.9905	<.0001
Maple/beech/birch	Sapling	857.7573	<.0001	36.6037	<.0001	0.4043	<.0001	23.9603	<.0001
Aspen/birch	Sapling	1277.4473	<.0001	35.1636	<.0001	0.3859	<.0001	32.3657	<.0001
Red/white/jack pine	Tree	215.0946	<.0001	18.1165	<.0001	0.3987	<.0001	112.2513	<.0001
Spruce/fir	Tree	175.9560	<.0001	12.8520	<.0001	0.2973	<.0001	69.5593	<.0001
Loblolly/shortleaf	Tree	220.1037	<.0001	13.0503	<.0001	0.5233	<.0001	84.2581	<.0001
Oak/pine	Tree	147.5297	<.0001	29.4719	<.0001	0.4790	<.0001	67.6585	<.0001
Oak/hickory	Tree	132.9210	<.0001	31.7232	<.0001	0.4219	<.0001	79.1111	<.0001
Oak/gum/cypress	Tree	177.5039	<.0001	25.3210	<.0001	0.4436	<.0001	113.7109	<.0001
Maple/beech/birch	Tree	161.3880	<.0001	18.1601	<.0001	0.3663	<.0001	74.5506	<.0001
Aspen/birch	Tree	152.9498	<.0001	15.5264	<.0001	0.3395	<.0001	54.5049	<.0001

Table 3: Fit statistics by forest-type group for model [4].

		AIC		
Forest type group	$R^{2\,a}$	Intercept only	Intercept + covariates	% reduction (covariates)
White/red/jack pine	0.57	4419.4	1975.7	55.3%
Spruce/fir	0.50	2952.0	1525.5	48.3%
Loblolly/shortleaf pine	0.64	1389.9	519.6	62.6%
Oak/pine	0.46	25862.0	15441.7	40.3%
Oak/hickory	0.43	2657.6	1536.3	42.2%
Oak/gum/cypress	0.51	2025.0	1010.3	50.1%
Maple/beech/birch	0.58	4854.6	2090.3	56.9%
Aspen/birch	0.49	7717.3	4054.7	47.5%

[a] Max. rescaled R^2

Table 4: Frequency of agreement between field forest-type group and computed forest-type group for 3,712 FIA plots.

Algorithm Forest Type Group	Field Forest Type Group[a]												
Frequency Row Pct Col Pct	AB	EAR	ES	LLS	LS	MBB	OGC	OH	OP	PJ	WRJ	SF	Total
AB	133 44.63 74.72	2 0.67 3.70	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	42 14.09 5.17	0 0.00 0.00	1 0.34 0.24	2 0.67 0.90	0 0.00 0.00	1 0.34 0.55	117 39.26 10.75	298
EAR	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0
ES	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0
LLS	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0
LS	0 0.00 0.00	1 0.19 1.85	0 0.00 0.00	6 1.12 40.00	492 91.62 90.61	0 0.00 0.00	0 0.00 0.00	4 0.74 0.95	34 6.33 15.25	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	537
MBB	34 3.12 19.10	15 1.38 27.78	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	693 63.64 85.34	0 0.00 0.00	43 3.95 10.17	25 2.30 11.21	0 0.00 0.00	54 4.96 29.67	225 20.66 20.68	1089
OGC	0 0.00 0.00	19 9.27 35.19	0 0.00 0.00	0 0.00 0.00	1 0.49 0.18	0 0.00 0.00	180 87.80 93.75	4 1.95 0.95	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	1 0.49 0.09	205
OH	2 0.38 1.12	12 2.30 22.22	1 0.19 50.00	3 0.57 20.00	16 3.07 2.96	55 10.54 6.77	4 0.77 2.08	355 68.01 83.92	70 13.41 31.39	1 0.19 25.00	2 0.38 1.10	1 0.19 0.09	522
OP	1 0.59 0.56	2 1.18 3.70	0 0.00 0.00	5 2.96 33.33	34 20.12 6.26	1 0.59 0.12	8 4.73 4.17	16 9.47 3.78	82 48.52 36.77	1 0.59 25.00	17 10.06 9.34	2 1.18 0.18	169
PJ	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0
WRJ	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	1 0.81 0.12	0 0.00 0.00	0 0.00 0.00	10 8.06 4.48	0 0.00 0.00	101 81.45 55.49	12 9.68 1.10	124
SF	8 1.04 4.49	2 0.26 3.70	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	20 2.60 2.46	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	1 0.13 25.00	7 0.91 3.85	730 95.05 67.10	768
Total	178	53	1	14	543	812	192	423	223	3	182	1088	3712

[a] AB = Aspen/Birch, EAR = Elm/Ash/Red Maple, ES = Exotic Softwood, LLS = Longleaf/Slash Pine, LS = Loblolly/Shortleaf Pine, MBB = Maple/Beech/Birch, OGC = Oak/Gum/Cypress, OH = Oak/Hickory, OP = Oak/Pine, PJ = Pinyon/Juniper, WRJ = White/Red/Jack Pine, SF = Spruce/fir.

Wisconsin State Forests Continuous Forest Inventory: A Look at the First Year

Randall S. Morin[1]
Teague Prichard[2]
Vern Everson[3]
Jim Westfall[4]
Charles Scott[5]

Abstract: *The demand for timely, consistent, and reliable forest inventory and monitoring information for Wisconsin's state forests has increased significantly. A wide range of publics and partners, including businesses, organizations, and citizens alike are well aware of the benefits of sustainable forestry and are working together to increase knowledge through an annual comprehensive forest inventory and monitoring program. In response, the Wisconsin Department of Natural Resources (WDNR) and the National Inventory and Monitoring Applications Center (NIMAC) of the U.S. Forest Service, Forest Inventory and Analysis (FIA) program have developed a Wisconsin State Forest Continuous Forest Inventory (WisCFI) program for the 518,680 acres of land distributed across 10 state forests. To accomplish the monitoring objectives of the WDNR, NIMAC designed a survey of approximately 3,145 FIA-like plots to be measured over 5 years (1 plot per 165 acres). A limited suite of summer only variables (soils, understory vegetation, down woody materials, and crowns) will be measured on one-third of the plots. The first panel of data was collected in 2007. We used this first panel to describe the benefits of the intensification on the precision of forest-level estimates for state forest land. Additionally, estimates of species richness were compared for plots where introduced species were present versus those where they were absent.*

Keywords: continuous forest inventory, intensification, forest health, introduced species, species richness

Introduction

[1] Research Forester, U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA, 610-557-4054, 610-557-4250 FAX, rsmorin@fs.fed.us
[2] Wisconsin Department of Natural Resources, Bureau of Forest Management, Madison, WI, 608-264-8883
[3] Wisconsin Department of Natural Resources; Bureau of Forest Management, Madison, WI, 608-264-8883
[4] Research Forester, U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA, 610-557-4043, 610-557-4250 FAX
[5] Program Manager, U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA, 610-557-4020, 610-557-4250 FAX

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

The demand for timely, consistent, and reliable forest inventory and monitoring information for Wisconsin's state forests has increased significantly. A wide range of publics and partners, including businesses, organizations, and citizens alike are well aware of the benefits of sustainable forestry and are working together to increase knowledge through an annual comprehensive forest inventory and monitory program. The Wisconsin Department of Natural Resources (WDNR) is directed to protect and sustainably manage state forests to provide ecological, economic, and social benefits for current and future benefits (Wisconsin State Stature 28.04). Forest certification on state forest lands makes a continuous forest inventory an even more critical need. Additionally, managers face an increasing number and type of issues.

The WDNR and the National Inventory and Monitoring Applications Center (NIMAC) of the U.S. Department of Agriculture, Forest Service's Forest Inventory and Analysis (FIA) program have developed a Wisconsin State Forest Continuous Forest Inventory (WisCFI) program for the 518,680 acres of land distributed across 10 state forests (Table 1 and Fig. 1). The primary purpose of the WisCFI is to collect and report on the condition of the forest in a statistically sound manner on an annual basis for each state forest. The information will be used to track the status and trends in forest extent, cover, growth, mortality, habitat, and overall health. The WisCFI will provide unbiased, reliable information at the property level that will assist in planning, management, and monitoring. Because the method is parallel to FIA methods, the inventory will be able to address regional trends. Other goals of the WisCFI include providing information on the condition and health of the forest and tracking changes over time; integrating data, methods, and tools in the planning and decision-making processes; developing and maintaining data input, models, and methods for forestry analysis and planning; and developing up-to-date and easy-to-use information products and services for land managers and the public.

Table 1: Name and area of each Wisconsin state forest

State Forest Property	Acres
Black River State Forest	67,809
Brule River State Forest	40,881
Coulee Experimental Forest	2,972
Flambeau River State Forest	90,281
Governor Knowles State Forest	19,910
Northern Highland American Legion State Forest	223,294
Kettle Moraine State Forest - Northern Unit	29,275
Kettle Moraine State Forest - Southern Unit	20,939
Peshtigo River State Forest	5,871
Point Beach State Forest	2,902

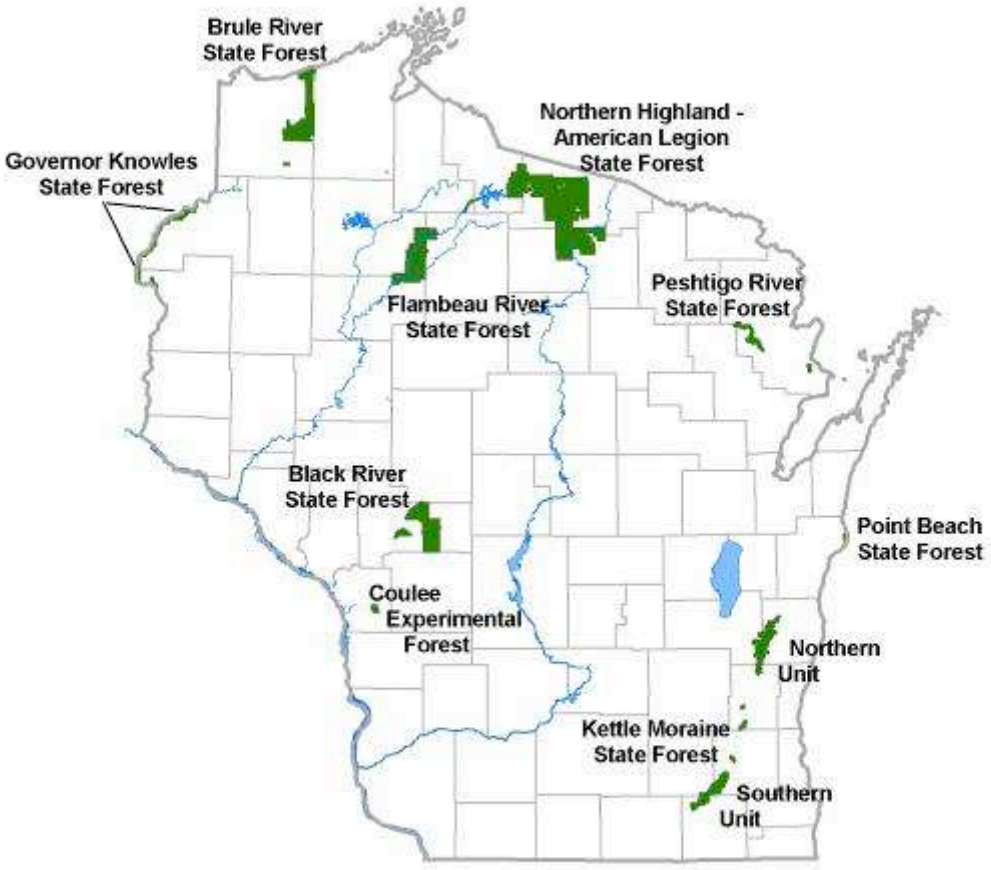


Figure 1: Map of Wisconsin's state forests.

The purpose of this study was to analyze the first year of data collected in the WisCFI to describe the benefits of the increased plot intensity (as compared to the FIA plot sample) on the precision of selected coarse forest-level estimates for state forest land. Additionally, estimates of understory species richness were compared among plots stratified by the presence of introduced vegetation species. Because these analyses are based on only one year of field data, the results are preliminary and incomplete. They are meant to illustrate the potential utility of the WisCFI.

Methods

The WDNR is using methods similar to the Forest Inventory and Analysis (FIA) program to conduct the WisCFI (U.S. Forest Service 2006 a, b). The FIA program of the U.S. Forest Service, the only congressionally mandated national inventory of U.S. forests, conducts a three-phase inventory of forest attributes of the country (Bechtold and Patterson 2005). The FIA sampling design is based on a tessellation of the United States into hexagons approximately 2,428 ha in size with at least one permanent plot established in each hexagon. In phase 1, the

population of interest is stratified and plots are assigned to strata to increase the precision of estimates. In phase 2 (P2), tree and site attributes are measured for forested plots established in each hexagon. Phase 2 plots consist of four 24-ft fixed-radius subplots on which standing trees are inventoried, and phase 3 (P3) plots include an extended suite of forest health measurements. The WDNR elected to reduce the number of subplots from four to two to maximize field efficiency and precision of estimates.

To accomplish the monitoring objectives of the WDNR, NIMAC designed a survey of approximately 3,145 FIA-like plots to be measured over 5 years (1 plot per 165 acres). A limited suite of summer only P3 variables (soils, understory vegetation, down woody materials, and crowns) will be measured on one-third of the plots. This results in a 36x intensification of P2 plots and a 193x intensification of P3 plots.

The WisCFI surveyed 628 P2 plots in 2007. Estimates and sampling errors were generated separately for forest area (acres), net volume of all live trees on forest land (ft^3), and net volume of white pine on forest land (ft^3) using both the WisCFI plots (measured in 2007), the standard FIA P2 plots (measured in 2007 only), and the standard FIA P2 plots (measured in 2003-2007). The resulting estimates and sampling errors were evaluated.

The origin of each understory species sampled in the 200 P3 plots measured in 2007 was identified using the Natural Resources Conversation Service (NRCS) PLANTS database (USDA 2008). Total species richness, native species richness, and introduced species richness per plot were calculated for each state forest in Wisconsin, and the mean richness values were examined for spatial trends. Additionally, the mean native species richness values were compared for plots where introduced species were present versus those where they were absent for all state forest land.

Results

Forest Area and Volume Estimation

The state forests in Wisconsin cover 518,650 acres. In 2007, the WisCFI surveyed 628 P2 plots (of 3,145 projected for the complete 5-year inventory cycle), and the FIA program visited 19 plots on state forest-owned land as a part of the standard annual inventory system. When the entire 5-year annual inventory of Wisconsin (2003-2007) is included, the number of plots increases to 189. The single-year FIA estimate of forest land is within 10 percent of the WisCFI estimate, but the sampling error is nine times higher (Table 2). When the full cycle of FIA plots is used the estimate rises to within five percent of the WisCFI 1-year estimate, and the sampling error is reduced to 2.4 times the sampling error on the WisCFI 1-year estimate (Table 2).

The single-year FIA estimate of net volume of all live trees is only 1 percent higher than the WisCFI 1-year estimate, but the sampling error is more than five times higher (Table 2). When the full cycle of FIA plots is used the estimate decreases to 3.5 percent below the WisCFI 1-year estimate, but the sampling error

is reduced to only two times the sampling error associated with the WisCFI 1-year estimate (Table 2).

The single-year FIA estimate of net volume of white pine is 83 percent lower than the WisCFI 1-year estimate, and the sampling error is nearly six times higher (Table 2). When the full cycle of FIA plots is used the estimate matches the WisCFI 1-year estimate, and the sampling error is only 1.3 times larger than the FIA sampling error (Table 2).

Table 2: Estimates and sampling errors of forest characteristics on State Forests in Wisconsin from WisCFI and FIA surveys

Variable	Survey	Estimate	Sampling Error
Acres of forest land	WisCFI (2007)	471,668 acres	1.1%
	FIA (2007)	422,730 acres	10.3%
	FIA (2003-2007)	450,238 acres	2.6%
Net volume of all live	WisCFI (2007)	654 million $ft.^3$	3.6%
	FIA (2007)	662 million $ft.^3$	18.6%
	FIA (2003-2007)	631 million $ft.^3$	6.3%
Net volume of white pine	WisCFI (2007)	98 million $ft.^3$	13.1%
	FIA (2007)	17 million $ft.^3$	75.9%
	FIA (2003-2007)	98 million $ft.^3$	17.5%

Understory Vegetation

The mean total species richness of vegetation plots is similar among most of Wisconsin's state forests (Fig. 2). A spatial trend is not evident, but there were some differences among the forests. The Black River, Point Beach, and Northern Highland-American Legion State Forests are the only forests that average fewer than 50 total species per plot.

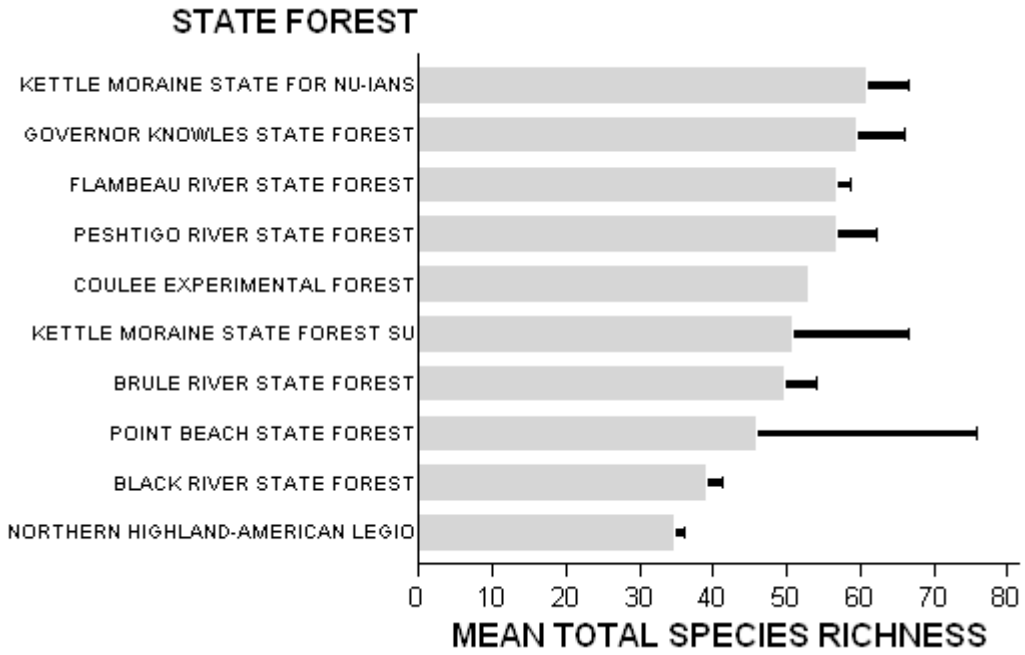


Figure 2: Mean total species richness per plot of understory vegetation by Wisconsin state forest, 2007.

Similarly, the mean native species richness of vegetation plots is similar among most of Wisconsin's state forests (Fig. 3). A spatial trend is not apparent, but the Black River, Point Beach, Kettle Moraine Southern Unit, and Northern Highland-American Legion State Forests are the only forests that average fewer than 42 native species per plot.

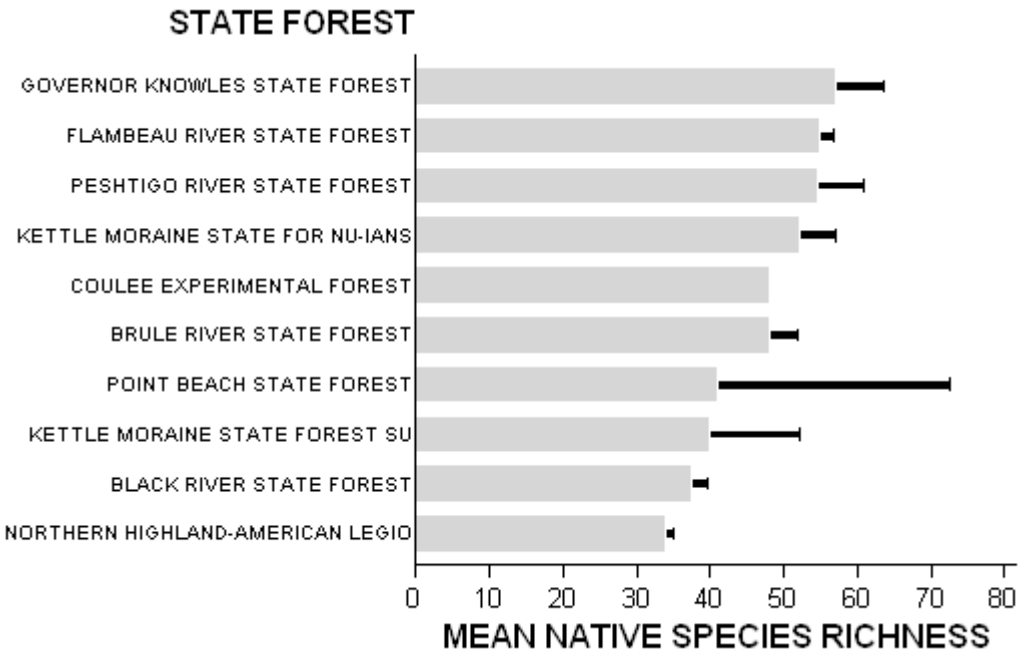


Figure 3: Mean native species richness per plot of understory vegetation by Wisconsin state forest, 2007.

By contrast, there are larger differences in the mean introduced species richness of vegetation plots in Wisconsin's state forests (Fig. 4). The mean introduced species richness on the northern and southern units of the Kettle Moraine State Forest is nearly double that on any of the other forests. There appears to be a spatial trend where state forests in the southern half of Wisconsin generally have higher mean introduced species richness values. The most common introduced species encountered were Japanese honeysuckle (*Lonicera japonica*), garlic mustard (*Alliaria petiolata*), Japanese barberry (*Berberis thunbergii*), and several buckthorn species (*Rhamnus* spp. and *Frangula* spp.). Native species richness was higher on average when introduced species were present on plots (Fig. 5).

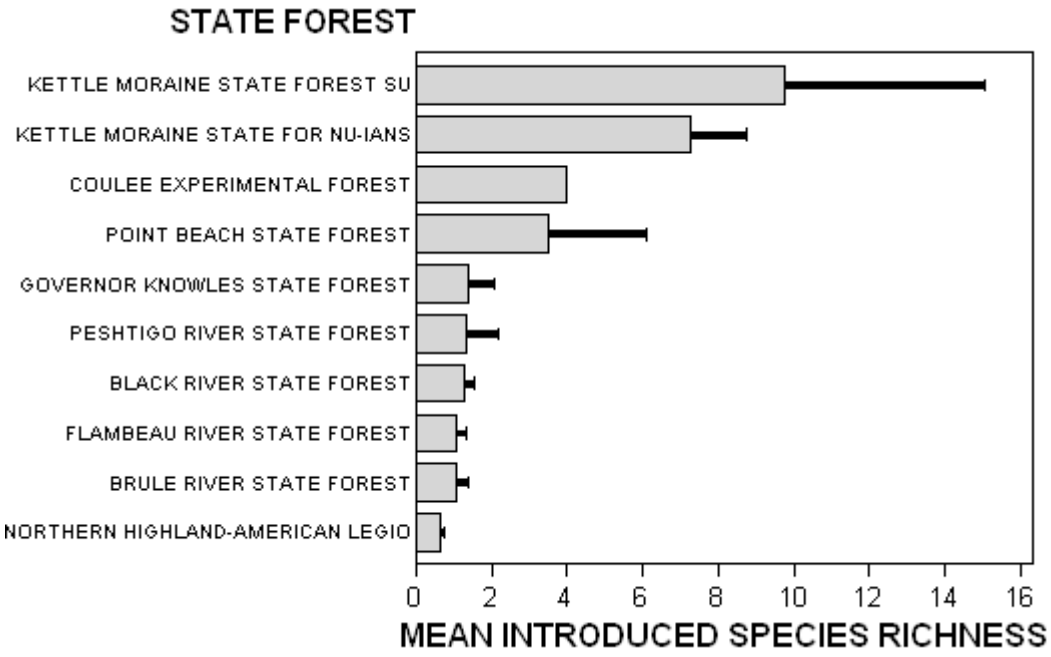


Figure 4: Mean introduced species richness per plot of understory vegetation by Wisconsin state forest, 2007.

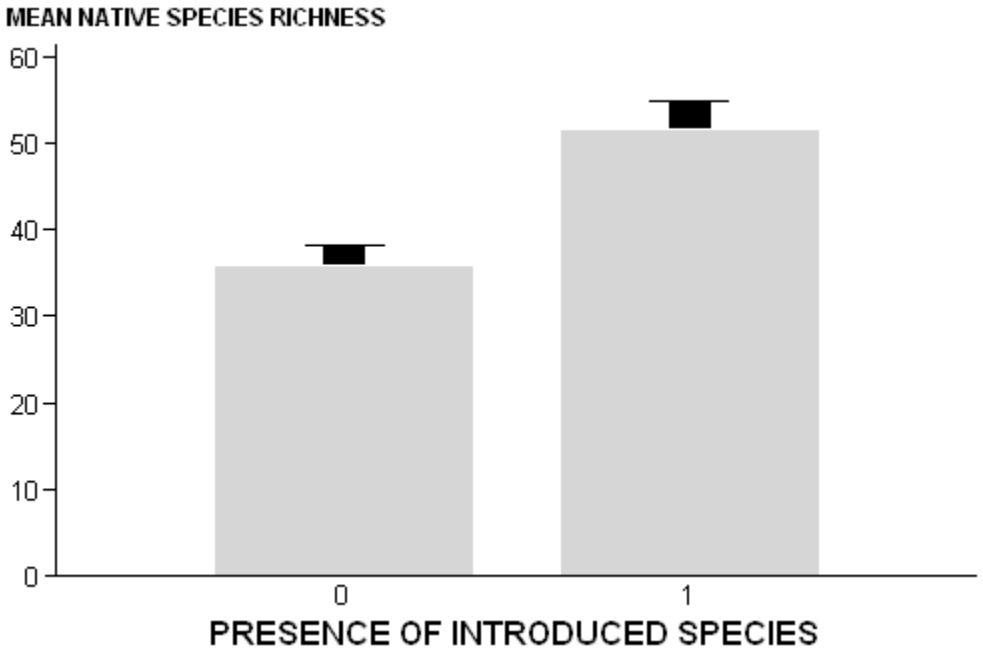


Figure 5: Mean native species richness per plot of understory vegetation by presence/absence of introduced species, Wisconsin state forests, 2007.

Discussion

The results of the estimation portion of the study indicate that a single year of WisCFI-intensified data yields more precise estimates that those generated from 1-year or a full 5-year cycle of FIA data (Table 2). Because the WisCFI is also on a 5-year cycle, year-to-year comparisons have the same sampling intensity as the two full 5-year cycles that will be achieved when the WisCFI is complete. Differences in sampling errors ranged from five times for net volume of live trees to nine times for forest land area. The precision of these estimates will improve as more data become available. In time, the WisCFI will allow for estimates of forest land, volume, and other important forest characteristics at the individual state forest level that would not be reasonable using the standard FIA grid because the sample is too limited.

Many introduced plant species have escaped and now exist in natural ecosystems in the United States, and some of those have displaced native plant species (Morse et al. 1995, Pimentel et al. 2000). To determine if the presence of introduced species is correlated with native species richness in the state forests of Wisconsin, we compared the average number of native species per plot. The plots were stratified by the presence and absence of introduced species. Total species richness and native species richness are similar among the Wisconsin state forests (Figs. 3, 4) and no spatial patterns are apparent. By contrast, there are differences in the mean introduced species richness of vegetation plots in Wisconsin's state forests (Fig. 4). There appears to be a spatial trend where state forests in the southern half of Wisconsin generally have higher mean introduced species richness values.

A widely recognized hypothesis that exotic species might more easily invade areas of low species diversity than areas of high species diversity has been supported by several studies (Case 1990, Stohlgren et al. 1999, Tilman 1997, Vitousek et al. 1996). Somewhat surprisingly, native species richness in our study was higher on average when introduced species were present on plots (Fig. 5). Other studies have yielded similar results where introduced species were more prevalent in areas of high native species richness (Huston and DeAngelis 1994, Stohlgren et al. 1999). In fact, Huebner et al. (2009) found a similar relationship between native species richness and presence of exotic species on the Allegheny National Forest in Pennsylvania using intensive FIA P3 plot data. However, disturbance may be a more important factor in invasibility than native species richness (Robinson et al. 1995; Ross et al. 2002).

Once completed, the WisCFI will provide more than 1,000 understory vegetation plots for analysis. Due to its high spatial intensity, this dataset will be statistically powerful and could be used for many purposes. A few suggestions are to re-visit the native species richness analysis with the full cycle of data (e.g., by forest type, stand size, disturbance, site quality, and/or forest fragmentation metrics), to study the effect of non-native invasive species on the distribution of native species, and to develop distribution maps of important, rare, and/or invasive plants for each state forest.

Literature Cited

Bechtold, W.A.; Patterson, P.L., eds. 2005. Forest Inventory and Analysis National Sample Design and Estimation Procedures. Gen. Tech. Rep. SRS-GTR-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Case, T.J. 1990. Invasion resistance arises in strongly interacting species-rich model competition communities. Proceedings of the National Academy of Science. 87: 9610–9614.

Huebner, C.D.; Morin, R.S.; Zurbriggen, A.; [et al.]. 2009. Patterns of exotic plant invasions in Pennsylvania's Allegheny National Forest using intensive Forest Inventory and Analysis plots. Forest Ecology and Management. 257: 258–270.

Huston, M. A.; DeAngelis, D. L. 1994. Competition and coexistence: the effects of resource transport and supply rates. American Naturalist. 144: 954–977.

Morse, L. E.; Kartesz, J. T.; Kutner, L. S. 1995. Native vascular plants. In: LaRoe, E. T.; Farris, G. S.; Puckett, C. E.; **[**et al.**],** eds. Our living resources: a report to the Nation on the distribution, abundance, and health of US plants, animals, and ecosystems. Washington, DC: U.S. Department of the Interior, National Biological Service: 205–209.

Pimentel, D.; Lach, L.; Zuniga, R.; [et al.]. 2000. Environmental and economic costs of nonindigenous species in the United States. BioScience. 50(1): 53–65.

Robinson, G.R.; Quinn, J.F.; Stanton, M.L. 1995. Invasibility of experimental habitat islands in a California winter annual grassland. Ecology. 76: 786–794.

Ross, K. A.; Fox, B.J.; Fox, M.D. 2002. Changes to plant species richness in forest fragments: fragment age, disturbance and fire history may be as important as area. Journal of Biogeography. 29: 749–765.

Stohlgren, T.J.; Binkley, D.; Chong, G.W.; [et al.]. 1999. Exotic plant species invade hot spots of native plant diversity. Ecological Monographs. 69(1): 25–46.

Tilman, D. 1997. Community invasibility, recruitment limitation, and grassland biodiversity. Ecology. 78: 81–92.

USDA NRCS. 2008. The PLANTS Database [online]. Available online at: http://plants.usda.gov [November 25, 2008].

U.S. Forest Service. 2006a. Forest Inventory and Analysis National Core Field Guide Volume I: Field Data and Analysis National Core Field Guide for Phase 2 Plots. Version 3.1. Northern Edition (modified for WisCFI). Unpublished paper on file at: U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA. 211 p.

U.S. Forest Service. 2006b. Forest Inventory and Analysis National Core Field Guide Volume I: Field Data and Analysis National Core Field Guide for Phase 3 Plots. Version 3.1. Northern Edition (modified for WisCFI). Unpublished paper on file at: U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA. 89 p.

Vitousek, P. M.; D'Antonio, C. M.; Loope, L. L.; [et al.]. 1996. Biological invasions as global environmental change. American Scientist. 84: 468–478.

FIA Map Products: Production and Assessment



Combining Forest Inventory, Satellite Remote Sensing, and Geospatial Data for Mapping Forest Attributes of the Conterminous United States

Mark Nelson[1], Greg Liknes[2], and Charles Perry[3]

***Abstract**: Analysis and display of forest composition, structure, and pattern provides information for a variety of assessments and management decision support. The objective of this study was to produce geospatial datasets and maps of conterminous United States forest land ownership, forest site productivity, timberland, and reserved forest land. Satellite image-based maps of forest land cover, geospatial datasets of land protection and ownership, and data from the U.S. Department of Agriculture Forest Service's Forest Inventory and Analysis (FIA) program were integrated to produce forest land maps. Forest land cover was derived from the U.S. Forest Service Forest Types Map; land ownership and protection is derived from the Conservation Biology Institute's Protected Areas Database; and forest attributes are derived from FIA, summarized by a hexagon sampling array from the Environmental Protection Agency Environmental Monitoring and Assessment Program. This approach provides a technique for efficiently producing forest resource maps over large geographic extents.*

Key words: Forest, inventory, FIA, mapping, CONUS

Introduction

The U.S. Department of Agriculture Forest Service's Forest Inventory and Analysis (FIA) program produces estimates of United States forest land composition and structure across all classes of land ownership. These estimates provide information about the amount, condition, health, and change in forest resources. As a sample-based inventory, FIA reports estimates for defined estimation units, typically states and counties.

---

[1] U.S. Forest Service; Northern Research Station; Forest Inventory and Analysis Program; 1992 Folwell Avenue; St. Paul, MN 55108 USA; mdnelson@fs.fed.us

[2] U.S. Forest Service; Northern Research Station; Forest Inventory and Analysis Program; 1992 Folwell Avenue; St. Paul, MN 55108 USA; gliknes@fs.fed.us

[3] U.S. Forest Service; Northern Research Station; Forest Inventory and Analysis Program; 1992 Folwell Avenue; St. Paul, MN 55108 USA; charleshperry@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Spatial analysis and map display of forest information provides support for a variety of assessments and management decisions of natural resources. Displaying FIA attributes spatially across typical estimation units is a straightforward procedure, however, many natural resource practitioners desire maps with more spatial specificity and better integration with supporting geospatial datasets than are available from traditional FIA estimates.

This paper describes one approach to combine data from FIA, satellite remote sensing, and other geospatial products for mapping forest attributes across the conterminous United States (CONUS). We produced geospatial datasets and maps of CONUS forest land ownership, forest site productivity, timberland, and reserved forest land by integrating satellite image-based maps of forest land cover, geospatial datasets of land protection and ownership, and FIA data. Previous mapping efforts (Stein et al. 2006, Nelson and Liknes 2007, Nelson and Vissage 2007) provide the basis for this approach.

Data

FIA data

FIA conducts inventories across land of all ownership categories, using a nationally consistent plot with four fixed-radius subplots, and a systematic national sampling design based on a hexagonal grid. At least one FIA plot is established within each hexagon, representing approximately 6,000 acres each.

Forest land is defined by FIA as "land that is at least 10 percent stocked by forest trees of any size, or land formerly having such tree cover, and not currently developed for a nonforest use." (Bechtold and Patterson 2005, glossary). Additional requirements include minimum forest patch size (1 acre) and minimum forest patch width (120 ft).

Three subcategories of FIA forest land include reserved forest land, timberland, and other forest land. Reserved forest land is defined as "land permanently reserved from wood products utilization through statute or administrative designation." Timberland is defined as "forest land that is capable of producing in excess of 20 cubic feet/acre/year of wood at culmination of mean annual increment (MAI)", excluding reserved forest lands. Other forest land comprises the remainder of forest land, being neither timberland nor reserved forest land (Bechtold and Patterson 2005, glossary).

FIA plot data from each state are aggregated to form a nationwide database U.S. Forest Service 2008) for producing forest resource assessments of the United States as mandated by the Forest and Rangeland Renewable Resources Planning Act of 1974 (RPA), P.L. 93-378, 99 Stat. 4765.

RPA 2007 data (Smith et al., in press) on forest site productivity are assigned to one of seven categories, based on the culmination of mean annual increment of

fully stocked natural stands, recorded in units of cubic feet/acre/year. Plots were assigned the approximate midpoints of their productivity class, with one exception (Table 1).

Table 1: Forest site productivity.

SITECLCD	Class range (cubic feet/acre/year)	Midpoint
1	225+	225
2	165-224	195
3	120-164	142
4	85-119	102
5	50-84	68
6	20-49	35
7	0-19	10

RPA 2007 data on forest land ownership differentiates public vs. private, corporate vs. noncorporate, and various other ownership classes. Corporate owners include "forest industry and forest management companies, timber investment management organizations, and other companies that may or may not have forest management as a primary ownership objective" (Smith et al., in press).

EMAP hexagons

A hexagon sampling array from the Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP) was used for summarizing RPA plot data (White et al. 2005). The original EMAP grid of hexagons provide a CONUS sampling grid with each hexagon measuring approximately 648 square kilometers (160,000 acres) in area (U.S. Environmental Protection Agency 2002). Unlike states and counties, EMAP hexagons are equal-sized estimation units, and therefore are likely to contain similar numbers of plots. A number of natural resource monitoring activities are based on the EMAP sampling array and EMAP hexagon IDs have been assigned to each plot in the FIA database. Because the vast majority of EMAP hexagons in CONUS contain multiple RPA plots, per-hexagon estimates of forest attributes can be produced for each hexagon. These hexagons allow information to be analyzed and mapped at a finer spatial resolution than is available from typical FIA estimation units. RPA mean forest site productivity and percent corporate ownership of private forest land was calculated within each EMAP hexagon.

Land cover

We employed a raster geospatial dataset of 2001 CONUS and Alaska forest types for portraying forest and nonforest classes. This map of forest types was produced by the U.S. Forest Service at 250 m pixel resolution from Moderate-resolution Imaging Spectroradiometer satellite image data, ancillary geospatial data, and FIA plot data (Ruefenacht et al. 2008).

Land ownership

Land ownership data was derived from the Conservation Biology Institute's Protected Areas Database (PAD), version 4.6 (DellaSala et al. 2001), with revisions by FIA. To match FIA definitions, PAD polygons having International Union for Conservation of Nature (IUCN) protection classes I through V were defined as 'reserved' land, and areas not labeled as public were assumed to be in private ownership.

Mapping

A geographic information system (GIS) was used to prepare and manage geospatial datasets in Albers Conic Equal Area projection with standard U.S. Geologic Survey (USGS) parameters, North American Datum of 1983, and converted to raster format with 250 m pixel resolution. Figure 1 illustrates the spatial pattern of attributes from the four categories of data described above. For all maps, geographic base data are from the USDA National Agricultural Statistics Service (http://www.nass.usda.gov/census/census02/atlas02/CoGenAll_D02>) and the National Atlas of the United States (http://www.nationalatlas.gov/maplayers.html).

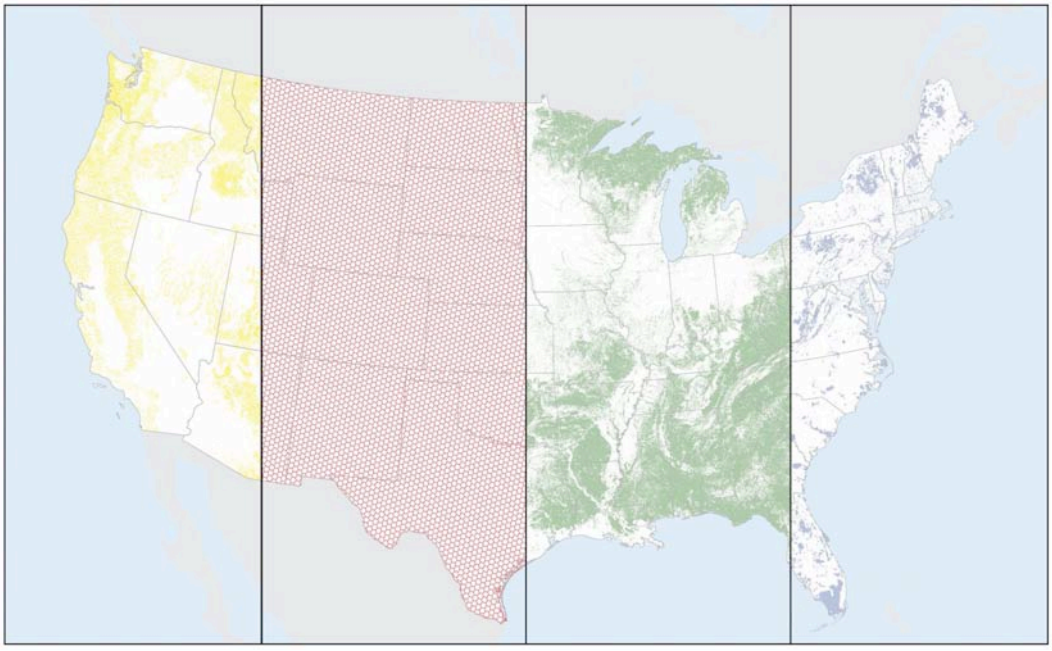


Figure 1: Geospatial datasets used for mapping forest attributes in CONUS, from left to right: approximate locations of forested FIA plots (U.S. Forest Service 2008), EMAP hexagon sampling array (U.S. Environmental Protection Agency 2002), forest land cover (Ruefenacht et al. 2008), and delineation of protected areas and land ownership (DellaSala et al. 2001).

Forest land ownership

Land ownership was mapped for all pixels labeled forest in the forest type map. Forest pixels were labeled as public or private ownership based on PAD. The percentage of private forest land in corporate ownership (FIA data) within each EMAP hexagon was assigned to each private forest pixel (Fig. 2).

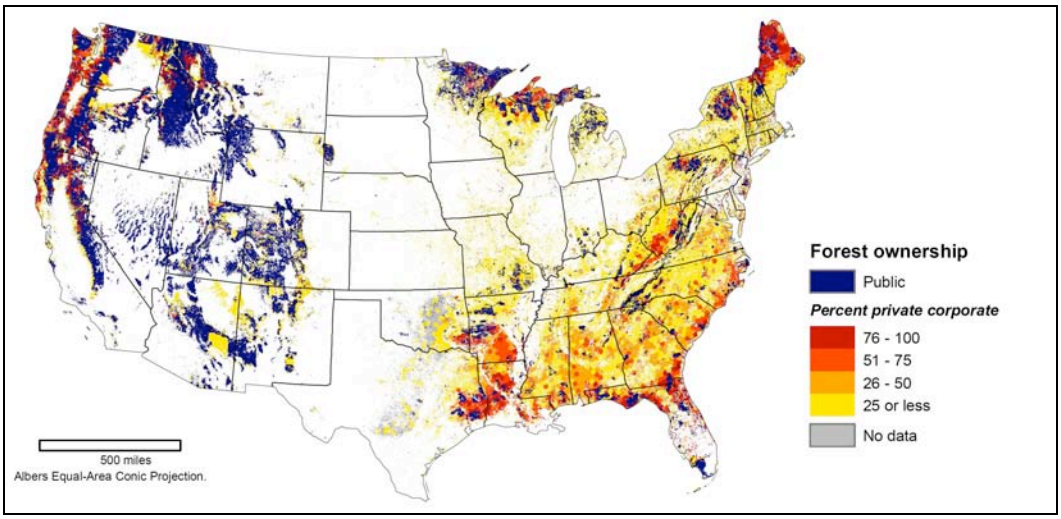


Figure 2: Public, private, and corporate forest land ownership in CONUS.

Forest site productivity

A map of net forest site productivity was produced by classifying mean site productivity within each hexagon into standard RPA classes, and assigning this class to all forest pixels (Fig. 3).

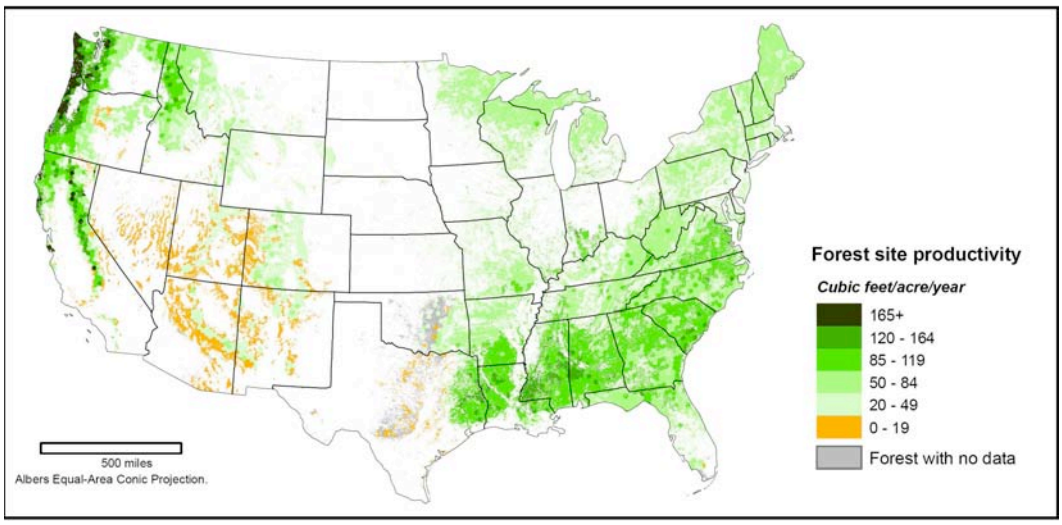


Figure 3: Net primary forest site productivity in CONUS.

Timberland, reserved forest land, and other forest land

Three subcategories of forest land were mapped: reserved forest land, timberland, and other forest land. Forest pixels in IUCN protection categories I through V were classified as reserved forest land. These lands are managed primarily for science, protection, conservation, or recreation. Nonreserved forest pixels were combined with the preceding forest site productivity map to identify pixels meeting the minimum timberland productivity threshold. All remaining forest pixels, i.e., not timberland and not reserved, were classified as other forest land (Fig. 4).

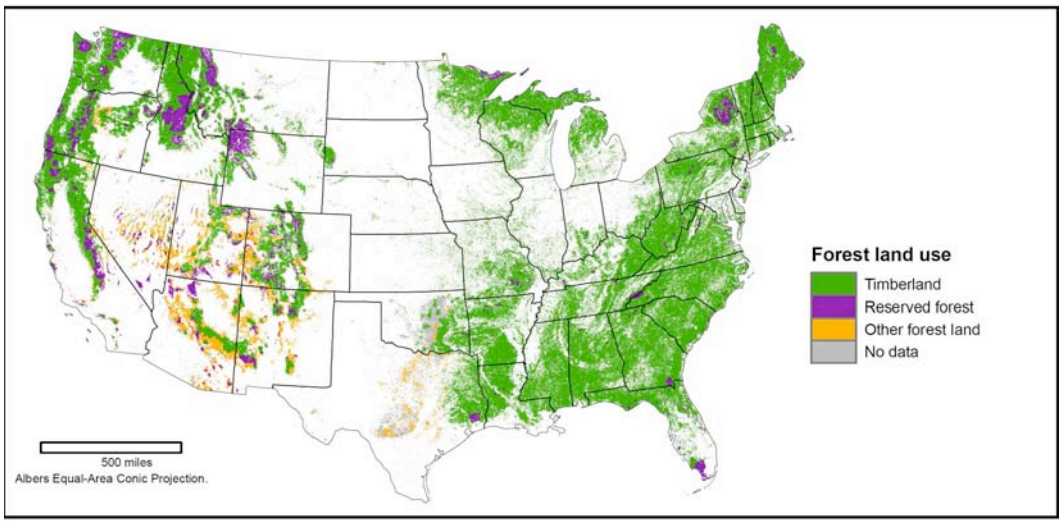


Figure 4: Timberland, reserved forest land, and other forest land in CONUS.

The maps depicted in Figures 3 and 4 were combined to further differentiate timberland into three general productivity classes (Fig. 5), selected to portray geographic variability:

- Low productivity — 20 to 50 cubic feet/acre/year,
- Medium productivity —50 to 85 cubic feet/acre/year,
- High productivity — 85+ cubic feet/acre /year

. This map (Fig. 5) reveals that timberland comprises most forest land in the East, with highest productivity occurring in the Pacific Northwest and in the South. Reserved forest land is most prevalent in the West where large national parks and wilderness areas are most common, but also is present in the East. Low productivity timberland and other forest land are most noticeable in the arid Southwest.

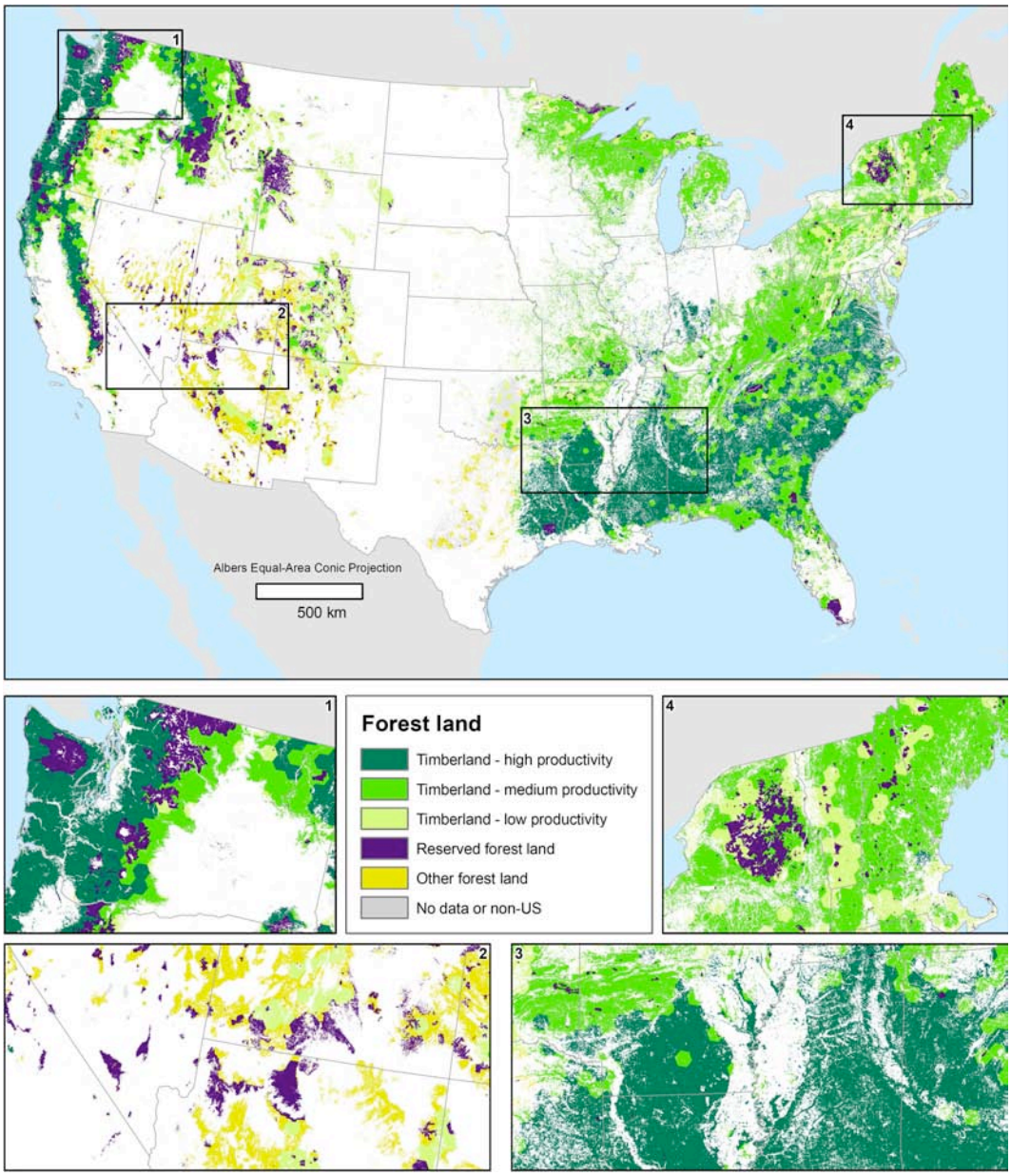


Figure 5: Reserved forest land, other forest land, and forest site productivity on timberland in CONUS.

Summary

We presented an approach for integrating satellite image-based maps of forest land cover, geospatial datasets of land protection and ownership, and forest inventory statistics to produce geospatial datasets and maps of CONUS forest land ownership, forest site productivity, timberland, and reserved forest land. This approach has been employed to produce and publish maps for the ESRI Mapbook, ESRI "Mapping Forestry" book, U.S. Forest Service's "Forests On The Edge" (FOTE) reports, RPA 2007 report, and other products.

Literature Cited

Bechtold, W.A.; Patterson, P.L., eds. 2005. Forest inventory and analysis 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.

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. 21:124-135.

Nelson, M.D.; Liknes, G.C. 2007. Forest land ownership in the conterminous United States. In: Law, M., ed. ESRI Map Book. Redlands, CA: ESRI Press: 76.

Nelson, M.D.; Vissage, J. 2007. Mapping forest inventory and analysis forest land use: timberland, reserved forest land, and other forest land. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; McWilliams, W.H., eds. Proceedings of the seventh annual forest inventory and analysis symposium. Gen. Tech. Report WO-77. Washington, DC: U.S. Department of Agriculture, Forest Service: 185-190.

Ruefenacht, B.; Finco, M.V.; Nelson, M.D.; Czaplewski, R.L.; Helmer, E.H.; Blackard, J.A.; Holden, G.R.; Lister, A.J.;Salajanu, D.; Weyermann, D.; Winterberger, K.. 2008. Conterminous U.S. and Alaska forest type mapping using Forest Inventory and Analysis data. Photogrammetric Engineering & Remote Sensing. 74:1379-1388.

Smith, W.B.; Miles, P.D.; Perry, C.H.; Pugh, S.A. In press. Forest resources of the United States, 2007. General Technical Report. Washington, DC: U.S. Department of Agriculture, Forest Service.

Stein, S.; McRoberts, R.; Nelson, M.; Theobald, D.; Eley, M.; Dechter, M. 2006. Forests on the edge: A GIS-based approach to projecting housing development on private forests In: Aguirre-Bravo, C.; Pellicane, P.J.; Burns, D.P.; Draggan, S., eds. 2006. Monitoring science and technology symposium: Unifying knowledge for sustainability in the western hemisphere; Proceedings RMRS-P-42CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 736-743.

U.S. Forest Service. 2008. The forest inventory and analysis database: database description and users manual version 3.0 for Phase 2, revision 1. Washington, DC: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis Progra,. http://www.fia.fs.fed.us/library/database-documentation/FIADB_user_manual_v3-0.pdf

U.S. Environmental Protection Agency. 2002. Original EMAP grid 648 square kilometer hexagons - electronic dataset. Corvallis, OR: U.S. Environmental Protection Agency, Western Ecology Division. http://www.epa.gov/wed/pages/staff/white/getgrid.htm

White, M.A.; Shaw, J.D.; Ramsey, R.D. 2005. Accuracy assessment of the vegetation continuous field tree cover product using 3954 ground plots in the south-western USA. International Journal of Remote Sensing. 26: 2699-2704.

Disturbance Impacts



An Assessment of the Relationship Between Emerald Ash Borer Presence and Landscape Pattern

Susan J. Crocker and Dacia M. Meneguzzo[1]

***ABSTRACT**: Six years after its 2002 detection near Detroit, MI, the emerald ash borer (EAB) (Agrilus planipennis* Fairmaire*) has spread hundreds of miles across the Upper Midwest and Mid-Atlantic regions of the United States. Human-assisted transportation of infested ash materials is the primary mechanism of EAB dispersal over long distances. Natural spread occurs locally and is influenced by factors, such as host availability, meteorological conditions, and landscape configuration. This study looks at the effects of ash density and landscape pattern on current EAB distribution. Forest Inventory and Analysis (FIA) data were used to calculate plot-level landscape metrics (total edge length, edge density, and forest proportion) in Ohio, Indiana, Illinois, and Wisconsin. Initial results indicate that while EAB is primarily found in areas with a low proportion of forest land, it is also found in areas with a relatively high concentration of ash. Counties containing EAB infestations were also found to have high relative edge densities.*

Keywords: Emerald ash borer, landscape metrics, forest inventory.

Introduction

The emerald ash borer (EAB) (*Agrilus planipennis* Fairmaire, Coleoptera: Buprestidae) is a wood-boring beetle that is believed to have been introduced to the United States from Asia during the early 1990s (McCullough and Katovich 2004, Siegert et al. 2008). Undetected until 2002, EAB was initially found near Detroit, MI, and in adjacent Windsor, Ontario, Canada. Since that time, EAB has killed tens of millions of ash trees (*Fraxinus* spp.) across the Upper Midwest and along the East Coast, causing it to become one of the leading threats to the nation's forest resource (de Groot et al. 2008). EAB is currently (as of October 2008) found in 10 U.S. states (Michigan, Ohio, Indiana, Maryland, Illinois, Pennsylvania, West Virginia, Virginia, Missouri, and Wisconsin) and two Canadian provinces (Ontario and Québec) (USDA APHIS 2008).

Natural spread of EAB has been facilitated by human transportation of infested ash materials, such as movement of firewood and nursery stock. This form of artificial spread has been the source of many localized, outlier infestations (Cappaert et al. 2005). Once an outlier infestation has been established, the gradual, natural spread of EAB becomes paramount. Natural dispersal is influenced by many factors, including host availability, meteorological conditions, insect flight ability, and the configuration of the landscape (Cappaert et al. 2005). Additional research on natural spread from outlier infestations is

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

needed to help identify specific conditions which influence the direction and rate of natural dispersal of EAB. This information will enhance detection methods and aid in making management decisions.

Alterations in landscape pattern affect the quantity and suitability of the habitat that is available to organisms. Human activities, such as urban sprawl, land-use conversion, and commercial and residential development are the leading cause of landscape change. The most important outcome of changing landscape pattern is that it often leads to increases in forest fragmentation, i.e., the break up of large, contiguous forest parcels into small, isolated patches (Forman 1995).

The purpose of this study is to investigate the effects of ash density and landscape pattern on current EAB distribution. Results will provide insight on specific landscape conditions that may influence dispersal, or increase habitat suitability and thereby increase the risk of EAB invasion.

Methods

Study Area

Data from Ohio, Indiana, Illinois, and Wisconsin were examined in this study. EAB has been positively identified in each of the four states within the study area; however, infestations have not been found in every county. The study area was chosen to reflect the variation in the length of time that a known EAB infestation has been present within a state. Ohio (2003) and Indiana (2004) represented long infestation periods, while Illinois (2006) and Wisconsin (2008) represented short infestation periods. Michigan was not included in the study as the Lower Peninsula is considered generally infested.

Mapping Current EAB Distribution

Using quarantine and available EAB-positive location data (Ohio Dept. of Ag. 2008, Indiana Dept. of Nat. Res. 2008, Illinois Dept. of Ag. 2008, Wisconsin Dept. of Nat. Res. 2008), a map of the current distribution of EAB (i.e., EAB presence/absence) and state quarantine boundaries was constructed using ArcMap 9.2 (Fig. 1).

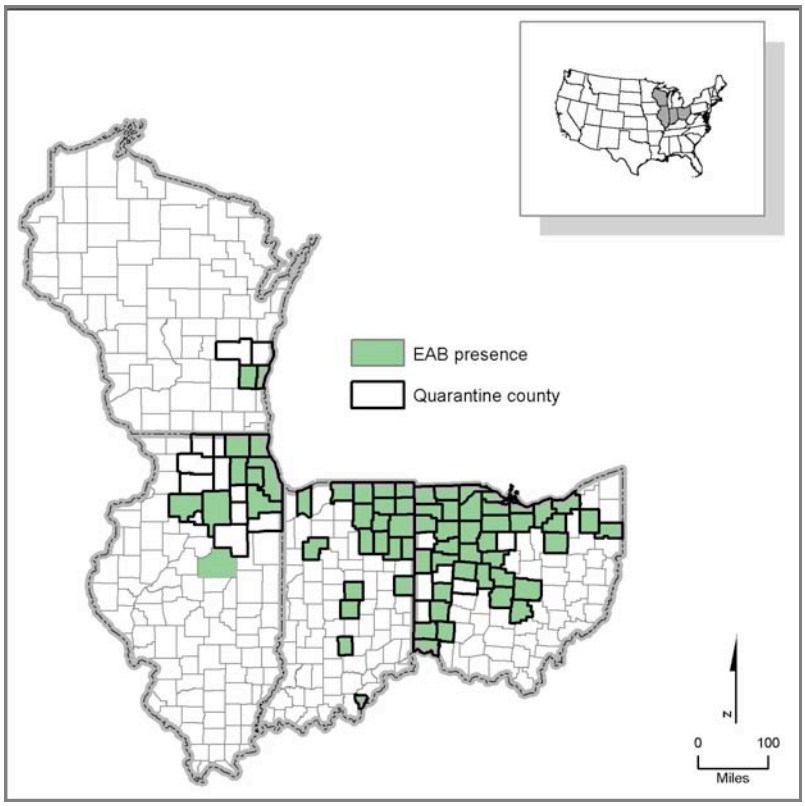


Figure 1: EAB presence and state quarantine boundaries in Ohio, Indiana, Illinois, and Wisconsin

Estimating Ash Density

Basal area was used as a measurement of ash density. Estimates of ash basal area (as a percentage of total live tree species basal area) were calculated using FIA plot data for all counties in the study area.

Landscape Metrics

Forest fragmentation that occurs as the result of natural disturbance events has a greater impact on the landscape than fragmentation that results from human activities, such as urban development. Human-caused fragmentation occurs more often, is less random, and has long-lasting effects that occur over a larger area (Haynes 2003). The process of fragmentation greatly alters the shape and pattern of continuous forest, leaving behind a collection of patches that vary in size and shape and that contain an increased amount of edge habitat. This increase in edge habitat has many negative effects on the remaining vegetation, including increased susceptibility to pest outbreaks. Therefore, it is important that the resulting landscape pattern be quantified for management and monitoring

purposes. Landscape pattern can be quantified using landscape metrics, i.e., "measurements designed to quantify and capture aspects of landscape pattern and include such measures as fragmentation indices" (Griffith et al. 2000). FIA plot data were used to calculate several landscape metrics in order to describe the landscape pattern within the study area. Calculated metrics included forest proportion, total edge length, and edge density. Total edge length was required to calculate edge density; however, measurements of total edge length are not reported.

Forest Proportion: Forest proportion represents the percentage of forest land area relative to the total land area, including noncensus water. These estimates were obtained by applying two algorithms that expanded plot condition level data (i.e., accessible forest, nonforest, and noncensus water) from the FIA database to county-level estimates for each state. The first algorithm was used to calculate area (acres) of accessible forest land, while the second was used to calculate the total area of all land and noncensus water. Forest land area was then divided by total land area to find the proportion (%) of forest land in each county.

Total Edge Length: The requirement to map different condition classes on FIA subplots allows the boundary, or edge length, between the conditions to be obtained. However, for this study, edge length was only obtained between forest and nonforest condition classes. Edge length may be measured as a straight line or as two lines that meet at a corner. When two different conditions are encountered on a subplot and the edge is a straight line, the right and left azimuths are recorded from subplot center, where the differing conditions intersect the subplot circumference. Given that the radius of the subplot is fixed at 24 feet, edge length has a maximum length of 48 feet. If the edge between the two conditions is not a straight line and contains a corner, a corner azimuth and a corner distance (from subplot center) are also recorded. When this occurs, the total edge length may be longer than 48 feet. Using azimuth, radius, and/or corner distance information, the total edge length for both types of edges was calculated using the law of cosines:

$$a^2 = b^2 + c^2 - 2bc * \cos A$$

where a is the edge length, and b and c are the length of the radii, or when a corner is present, the length of the segments from the corner to the subplot circumference. Angle A represents the corner angle, opposite the straight-line edge, at subplot center. If a corner was present, two angles were recorded and the law of cosines was used to find two edge lengths, which were then summed to find the total edge length. After edge lengths were obtained for each subplot, they were summed to find the plot-level edge length total (only if there was more than one subplot containing a forest/non-forest edge), which was then multiplied by the plot expansion factor. Finally, these expanded plot estimates were summed to find the total edge length for each county and state (as a whole) in the study area.

Edge Density: To make edge length estimates comparable across different levels of forestation, total edge length was converted to edge density equivalent to miles of edge per square mile of forest land. This was calculated by converting forest land area estimates from the FIA database from acres to square miles. The total edge length, in miles, was then divided by the forest land area, in square miles, to obtain edge density in miles per square mile. Similar to total edge length, edge density was calculated at the county and state level. Measures of relative edge density were then calculated for each county as the ratio of its estimate to the state-level estimate; this was done to determine how levels of edge density in each county compared to edge density in the state overall.

Results

Forest Proportion and Ash Basal Area

As of October 2008, EAB has been positively identified in 67 counties, or 18 percent of the study area (Fig. 1). EAB was predominantly found in counties where the land area was less than 25 percent forested (Fig. 2). While EAB-positive counties generally had a low percentage of forest, many of these counties also had a significant ash component (Fig. 2, Fig. 3). In these counties, the percentage of ash basal area ranged from zero to 60 percent of the total basal area of all live tree species (or the total live-tree basal area) (Fig. 3). On average, ash made up 13 percent of total live-tree basal area in counties containing EAB.

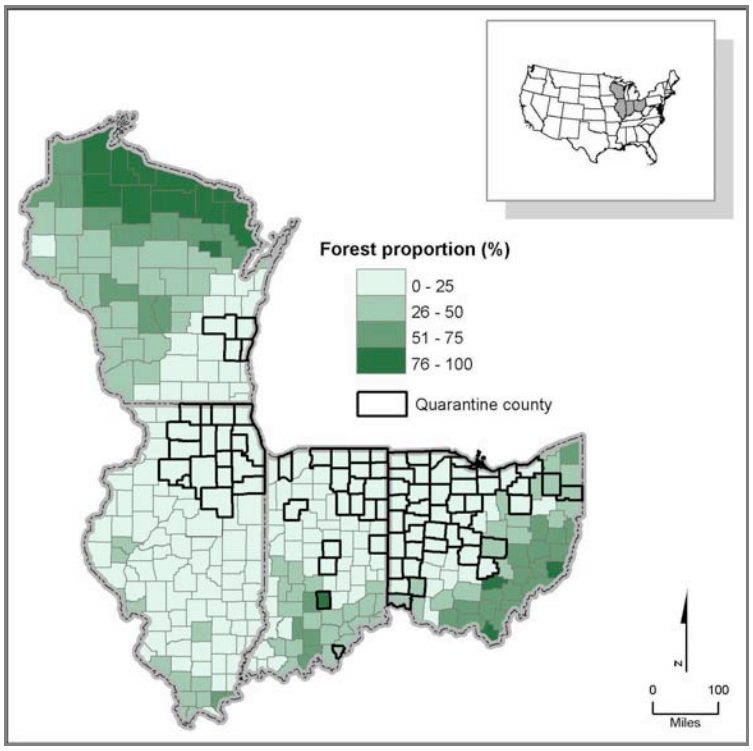


Figure 2: Forest proportion by county in Ohio, Indiana, Illinois, and Wisconsin

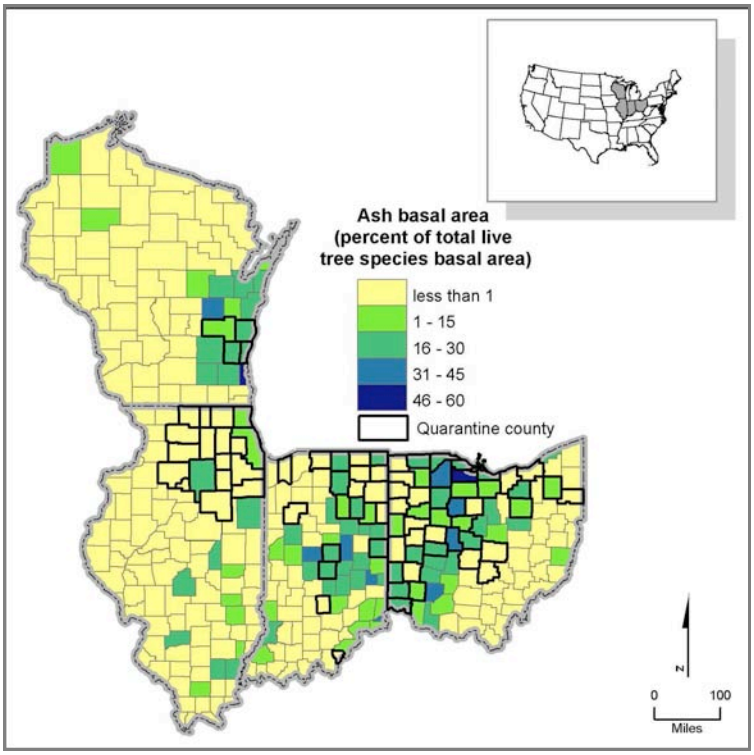


Figure 3: Ash basal area, as a percentage of total live tree species basal area, by county in Ohio, Indiana, Illinois, and Wisconsin

Edge Density

Counties with high proportions of forest land (Fig. 2) generally had low relative edge densities (Fig. 4). Not surprisingly, edge density was higher in counties containing EAB than in counties where EAB was absent (Fig. 5). Sixty-nine percent of EAB-positive counties had relative edge densities of 1.1 or greater. Conversely, 43 percent of counties where EAB has not been found had relative edge densities of 1.0 or less.

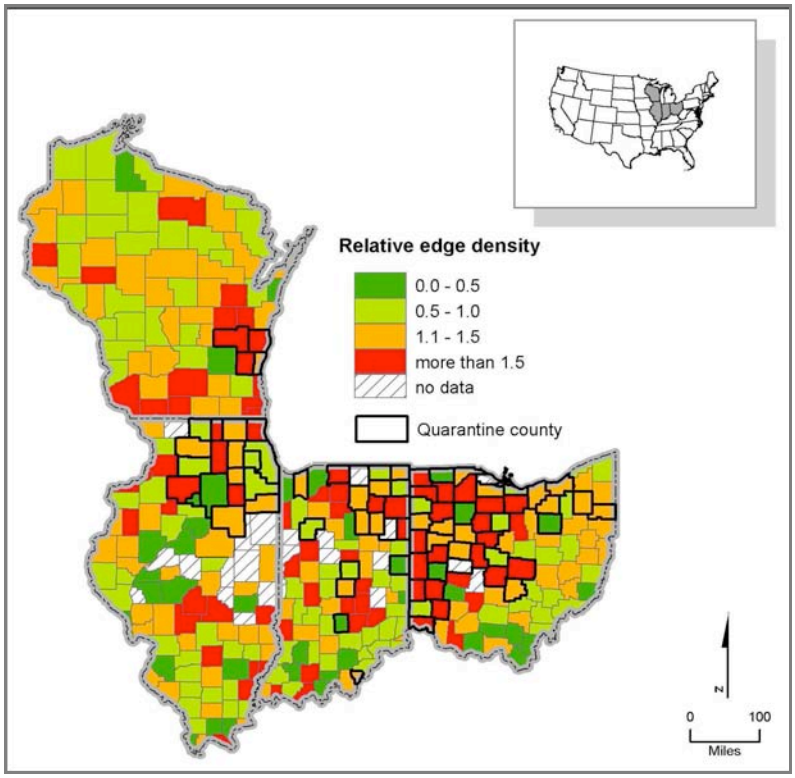


Figure 4: Relative edge density by county in Ohio, Indiana, Illinois, and Wisconsin (counties that did not contain accessible forest land and/or a forest/nonforest edge appear as 'no data')

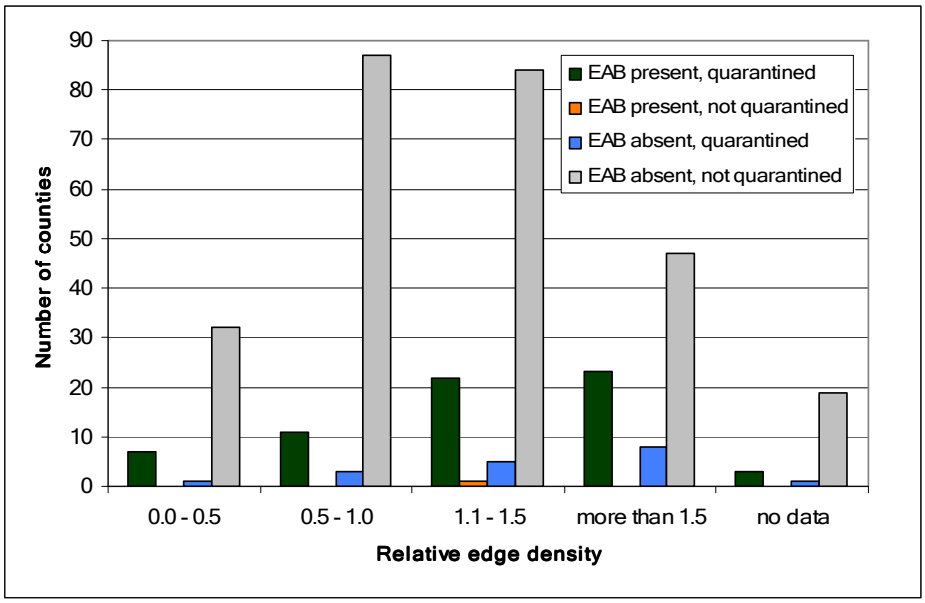


Figure 5: Relative edge density inside and outside of quarantine boundaries in Ohio, Indiana, Illinois, and Wisconsin

Discussion

In this study, we explored the relationship between EAB distribution, relative host abundance, and landscape pattern. We calculated county-level landscape metrics and compared them based on EAB presence or absence. Presence of EAB within a county was generally associated with low forest proportion, a high percentage of ash basal area, and high relative edge density. This is indicative of available habitats that consist of smaller parcels of fragmented forest that contain large amounts of ash.

Because EAB is found predominantly in areas with a relatively high percentage of ash, there is greater potential for damage and mortality of ash trees. A higher density of ash and therefore, more available phloem area increases EAB reproductive capacity. As a result of these factors, insect populations are more capable of rapid build up; therefore, smaller parcels of habitat may be quickly exhausted and spread of EAB may occur at a higher rate.

Initial results indicate a relationship between EAB presence and landscape pattern. Future work will involve the use of additional metrics, such as those derived from land cover imagery, to further characterize the landscape. These metrics will include road density, mean patch size, inter-patch distance, and connectivity. Furthermore, surrounding land uses will be examined in relation to EAB-positive locations and host abundance as a means of investigating the potential effects of changing landscape pattern on EAB dispersal.

Literature Cited

Cappaert, D.; McCullough, D.G.; Poland, T.M.; Siegert, N.W. 2005. Emerald ash borer in North America: a research and regulatory challenge. American Entomologist. 51: 152-165.

de Groot, P.; Grant, G.G.; Poland, T.M; et al. 2008. Electrophysiological response and attraction of emerald ash borer to green leaf volatiles (GLV) emitted by host foliage. Journal of Chemical Ecology. 34: 1170-1179.

Forman, R.T. 1995. Land mosaics: the ecology of landscapes and regions. Cambridge, UK: Cambridge University Press. 632 p.

Griffith, J.A.; Martinko, E.A.; Price, K.P. 2000. Landscape structure analysis of Kansas at three scales. Landscape and Urban Planning. 52: 45-61.

Haack, R.A.; Jendek, E.; Liu, H.; Marchant, K.R.; Petrice, T.R.; Poland, T.M.; Ye, H. 2002. The emerald ash borer: a new exotic pest in North America. Newsletter of the Michigan Entomological Society. 47: 1-5.

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

Illinois Department of Agriculture. 2008. [Online]. Emerald ash borer – Quarantine restrictions. http://www.agr.state.il.us/eab/index.php?pg=28 [Accessed December 15, 2008].

Indiana Department of Natural Resources. 2008. Overview map of all quarantined sites in Indiana (November 13, 2008). [Online]. http://www.state.in.us/dnr/files/ep-EAB-Quarantine-11132008.jpg [Accessed December 15, 2008].

McCullough, D.G.; Katovich, S.A. 2004. Emerald ash borer. NA-PR-02-04. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, State and Private Forestry Northeastern Area. 2 p.

Ohio Department of Agriculture. 2008. Ohio's EAB quarantine. [Online]. http://www.agri.ohio.gov/divs/plant/eab/eab-index.aspx#tog [Accessed December 15, 2008].

Poland, T.M.; McCullough, D.G. 2006. Emerald ash borer: invasion of the urban forest and the threat to North America's ash resource. Journal of Forestry. 104: 118-124.

Siegert, N.W.; McCullough, D.G.; Liebhold, A.M.; Telewski, F.W. 2008. Dendrochronological reconstruction of the establishment and spread of emerald ash borer. In: Mastro, V.; Lance, D.; Reardon, R.; Parra, G., comps. Emerald ash borer research and development meeting; 2007 October 23-24; Pittsburgh, PA. FHTET 2008-07. Morgantown, WV: U.S. Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team: 4-5.

U.S. Department of Agriculture, Animal and Plant Health Inspection Service. 2008. [Online]. Emerald ash borer. http://www.aphis.usda.gov/plant_health/plant_pest_info/emerald_ash_b/background.shtml [Accessed December 15, 2008].

Wisconsin Department of Natural Resources. 2008. Firewood quarantine in Wisconsin. [Online]. http://dnr.wi.gov/forestry/fh/pdf/FirewoodQuarantineWI.pdf [Accessed December 15, 2008].

Mountain Pine Beetle Infestations and Sudden Aspen Decline in Colorado: Can the Forest Inventory and Analysis Annual Inventory System Address the Issues?

Michael T. Thompson[1]

Abstract.: *There are two events occurring in Colorado that are concerning forest managers in Colorado. There is severe and widespread mortality of lodgepole pine due to the mountain pine beetle and aspen forests in some areas of the state have experienced widespread, severe, and rapid crown deterioration leading to mortality. Implementation of the Forest Inventory and Analysis annual inventory coincided with the two mortality events, providing an opportunity to test the utility of the annual inventory system. Preliminary analysis suggests that annual inventory data can quantify status and trends.*

Keywords: forest inventory, FIA, conifer, mortality, bark beetles

In the mid-1990's, the U.S. Department of Agriculture (USDA) Forest Service Forest Inventory and Analysis (FIA) program began a transition from periodic to annual inventories (Gillisepie 1999). Under the periodic system, the entire sample grid of plots was measured in a given State over a period of 1 to several years. Subsequent periodic inventories occurred anywhere from 5 to 20 years. Users of FIA data and reports became concerned about the time lag between periodic inventories and suggested ways be explored to produce more timely information. In response, the FIA program began implementation of an annual inventory system that would provide inventory estimates on a yearly basis (Gillisepie 1999).

[1] United States Forest Service; Rocky Mountain Research Station; Forest Inventory and Analysis Program; 507 25th Street; Ogden, UT 84404 USA; mtthompson@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

The same systematic sample grid that was used for periodic inventories was also used for the annual inventory grid. The difference in an annual inventory is that each plot in a state is assigned to one 6000-acre hexagon. These hexagons are then systematically formed into groups that are referred to as panels. Panels are selected for measurement on an annual rotating basis, and measurement of all accessible plots in one panel is completed before measurement of plots in a subsequent panel is initiated (Reams and Van Deusen 1999). Therefore, theoretically, annual panels of data are free from geographic bias. In the western United States, one panel represents approximately one-tenth of all plots in a State with the objective being that all plots will be measured in ten years. The advantage of the annual inventory system is that data are available every year which provides researchers with more timely data across the landscape. This system results in data being collected and compiled every year, which allows for more flexible analysis options and provides opportunities to monitor forest change in ways that were not possible using periodic inventory data.

The Interior West FIA (IW-FIA) program operates in Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming. The annual inventory system was implemented in Utah in 2000 and in most of the other IW-FIA states since then. The annual inventory began in Colorado in 2002, and as of 2007, six panels have been completed in the State. About the same time that annual inventories were implemented in the Interior West, forest managers began noticing significant increases in the incidence of insects and disease for certain tree species. Some of these incidents were attributed to drought that began in the late 1990's. Others are attributed to stand conditions, stand dynamics, and unusual weather patterns.

Two events, in particular, are considered significant issues in Colorado. There is severe and widespread mortality of lodgepole pine (*Pinus contorta*) due to mountain pine beetle (*Dendroctonus ponderosae*). The current mountain pine beetle epidemic is currently considered to be catastrophic and unprecedented. Aspen forests in some areas of Colorado and other regions have experienced widespread, severe, and rapid crown deterioration leading to mortality. This phenomenon is now being referred to as Sudden Aspen Decline (SAD). These two events can be considered an opportunistic test of the ability of the FIA annual inventory system for quantifying rapid change over a State. Analysis of the event may test some assumptions that have been made about the FIA annual inventory system.

This article describes mortality of lodgepole pine and aspen from the FIA annual inventory in Colorado, preliminary analysis of the results, and how the design of the annual inventory may influence the final analysis.

The FIA Annual Inventory Design

There are more analysis options for monitoring forest change under the annual inventory system than the periodic inventory system. One option is to simply compile the data as if it were a periodic inventory, ignoring the fact that some panels are several years older when compared to previous periodic inventories. This may be referred to as the traditional periodic option (Reams and Van Deusan 1999). Another analysis option considers the fact that the annual inventory is an interpenetrating design. This may be referred to as the independent panel option. Under this approach, each annual panel is analyzed independently and estimates are produced on a yearly basis. The independent panel option is a design where the n units (FIA sample plots) are divided into $k = 10$ panels, each panel containing $m = n/k$ units. Panel 1 plots are measured in year 1, panel 2 in year 2, etc., such that all plots have been visited by the end of year 10. The panel cycle is repeated into perpetuity (Reams and Van Deusan 1999).

When FIA began implementation of annual inventories in the mid 1990's, questions immediately arose regarding the inventory design. The most fundamental question was, are there enough plots in one annual panel to detect the progression of events such as tree mortality due to natural disturbance? A systematic grid of plots serves as the base for annual inventories where one field plot represents approximately 6000 acres. Therefore, under the annual inventory in the Interior West, one plot in an annual panel represents about 60,000 acres ($1/10^{th}$ of the periodic grid). Table 1 illustrates the number of plots that sample forest land in Colorado by measurement year and panel number. An average of about 394 plots is scheduled to be visited in any given year in Colorado.

Since each panel represents about $1/10^{th}$ of all plots in Colorado, the smaller number of plots results in a higher variance around the individual panel estimate. Mortality estimates, in particular, are subject to high variances since tree mortality is relatively infrequent. Tree mortality does not always occur on every plot. For situations like insect and disease outbreaks where mortality may be spatially spotty, the signature of the event may not be adequately captured with a limited sample size. Caution must also be exercised when screening data for single panels by certain attributes such as forest type or analyzing data for very small domains since the small number of plots may result in an unacceptably high variance. Therefore, there is a tradeoff between variance and temporal currency. Does having current inventory estimates with a high level of statistical variance outweigh having more dated data with lower variance?

Estimating Tree Mortality

The IWFIA program defines a mortality tree as one determined to have died within 5 years of the plot measurement date (USDA 2007). Therefore, field crews make a distinction between trees that have recently died and long-standing dead trees on all plots that sample forest land. For this analysis, population estimates of the number of all formerly live lodgepole pine and aspen trees with a minimum diameter at breast height of 5.0 inches that qualify as mortality on forest land are used for comparisons. All trees that qualify as mortality on FIA plots are assigned a cause of death code that estimates the primary agent that caused the tree's death. Only lodgepole pines that were assigned a cause of death of insects were used in mortality estimates. All aspen trees, regardless of cause of death, were used in mortality estimates. The annual number of mortality trees is the number of mortality trees expanded to a population level averaged over a 5-year period and reported as an average annual number. For example, the mortality tree estimate for the year 2002 represents the average number of trees that died each year during the period 1998-2002. This period is termed mortality period in this paper (Table 1). The reason behind this mortality estimation procedure in Colorado, along with several other IW regions, is the inventory data was limited to initial plot measurements. Complete remeasurement data for the State, where the status of the plot and all trees on the plot are known at two points in time, will not be available in Colorado until all ten panels of data are completed and remeasurement begins in the eleventh year.

Table 1: Number of Inventory Plots That Sampled Forest Land in Colorado by Mortality Period, Measurement Year, and Panel Number.

Measurement Year	Mortality Period	Panel Number	Number of Plots
2002	1997-2002	1	362
2003	1998-2003	2	387
2004	1999-2004	3	409
2005	2000-2005	4	407
2006	2001-2006	5	391
2007	2002-2007	6	411

The Mountain Pine Beetle Epidemic

The USDA Forest Service Region 2 Forest Health Management Group began statewide aerial surveys of mountain pine beetle infestations in 1996. This monitoring effort has recorded significant increases in the area of lodgepole pine forests affected by the mountain pine beetle since 1996 and as of 2007, the epidemic is believed to be catastrophic and unprecedented. Mountain pine beetle outbreaks are occurring at greater intensity and at locations where they have not previously occurred (Logan and Powell 2001, Caroll et al. 2004). These epidemics are not unique to Colorado—much of western North American and Canadian lodgepole pine forests are also suffering from mountain pine beetle infestations.

The mountain pine beetle is a native insect to western pine forests in North America and innocuous populations are almost always present in forests. Transition to epidemic populations is a function of the beetle's capacity to locate, colonize and reproduce within suitable host trees in a weather pattern conducive to overwintering survival, emergence, and dispersal (Caroll et al. 2004). The current epidemic in Colorado is of major concern because lodgepole pine, the primary host of the insect in the state, is an important tree for timber production, recreation, and wildlife.

Sudden Aspen Decline

Rapid mortality of trembling aspen has been reported from multiple locales in southwestern Colorado by varied observers since 2004 (Worrel et. al. 2008). The suddenness of the phenomenon is in striking contrast to the stand-level mortality processes typically observed in aspen stands. Adding to the concern, there appears to be a lack of regeneration occurring in stands were the overstory mortality is unusually high.

Evidence to date suggests that it is a decline disease incited by acute, warm drought. Predisposing factors include low elevation, south and southwest aspects, droughty soils, open stands, and physiological maturity. Contributing factors, which kill the stressed trees, include Cytospora canker, two bark beetle species, poplar borer, and bronze poplar borer.

What is happening to aspen in Colorado? Unlike mountain pine beetle infestations, an event that has been extensively researched, there is no recorded precedent for this phenomenon and the aspen mortality appears to be function of several agents. Aspen forests are dynamic and have always changed in response to climate, frequency and intensity of disturbance, and natural succession. There

is no conclusive evidence to date that indicates that this event will continue into the future nor what the eventual impact on the aspen resource will be.

What the Data Show

Annual inventory data indicate an upward progression of lodgepole pine mortality caused by bark beetles in Colorado. The average annual number of lodgepole pine trees killed by insects in the 2002-2007 mortality period is over 10 times that recorded in the 1997-2002 mortality period (Figure 1). The annual average of lodgepole pine mortality indicated an increase with each successive mortality period over the six years of measurement with the largest increase occurring after the 2000-2005 mortality period. Average annual mortality rates of beetle-killed lodgepole pine progressed upward from .22 percent during the 1997-2002 mortality period to 2.6 percent in the 2002-2007 mortality period.

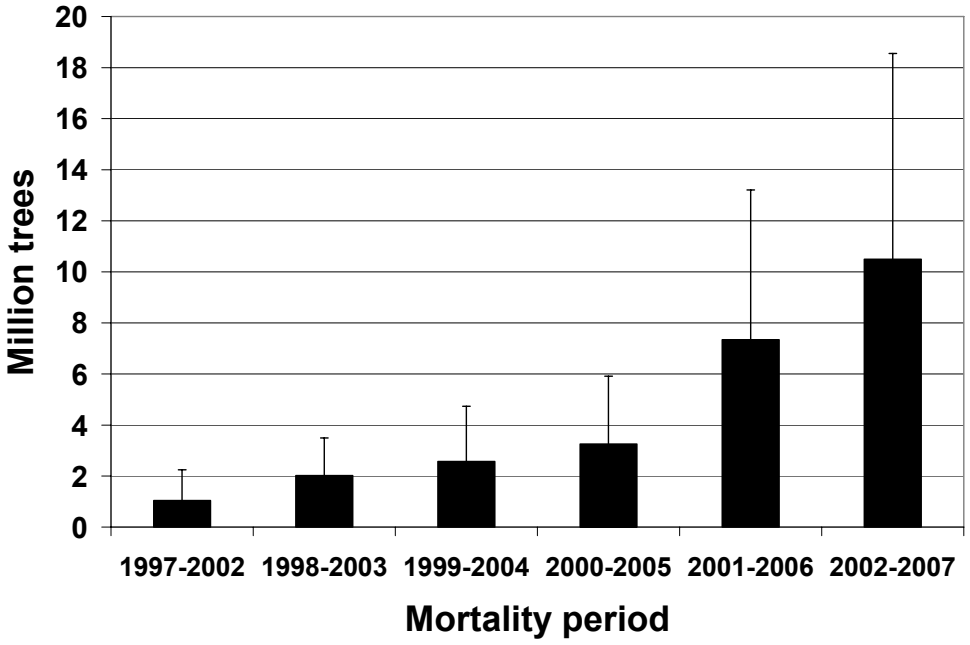


Figure 1: Average annual number of lodgepole pine trees killed by insects by mortality period in Colorado. Error bars represent 95% confidence interval.

Statistical analysis of the mortality data was performed in two steps. First, an analysis of variance was done to detect mortality period effect using the ANOVA procedure version 8.02, of the SAS system for Windows Version 5.0.2195 (SAS, 2001). The difference in the mean number of lodgepole pine mortality trees per acre among mortality periods panels was significant (P=0.0130). Second, all combinations of mortality periods were tested against each other using the Bonferroni correction factor for multiple tests. Comparisons among panels

identified two combinations of mortality periods significant at the 0.05 level; 1997-2002 compared to 2002-2007, and 1998-2003 compared to 2002-2007.

Trends in number of aspen mortality trees do not indicate the same distinctive upward trend as lodgepole pine beetle-killed mortality over the same period in Colorado (Figure 2). The highest level of aspen mortality occurred in the 2002-2007 period where the average annual number of trees was nearly 15 million trees. A test for mortality period effect indicated no significant difference (P=0.5387).

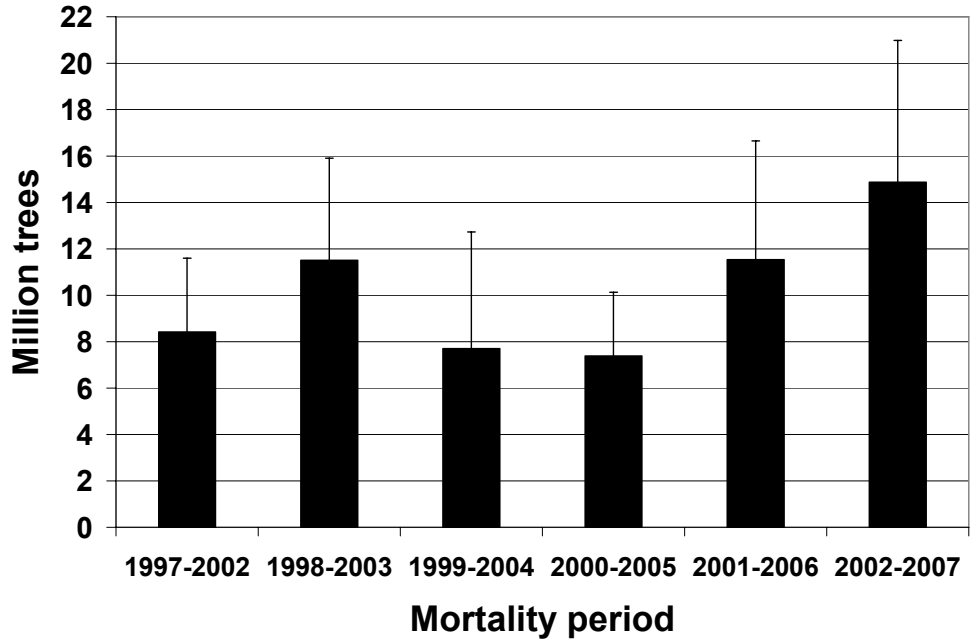


Figure 2: Average annual number of aspen trees classified as mortality by mortality period in Colorado. Error bars represent 95% confidence interval.

Discussion

The current high rate of lodgepole pine mortality due to the mountain pine beetle event has forest managers very concerned about the future of this species in Colorado. There is concern that the high rate of mortality will continue and could possibly deplete the supply of large-diameter trees in as little as five years. Trends revealed by annual data of other forest estimates agree with the progression of lodgepole pine mortality. For example, examination of the percentage of live lodgepole pine trees damaged by bark beetles by individual year indicates similar trends (Figure 3). Bark beetle damage rates jumped from 4 percent in 2005 to over 12 percent in 2006 and continued upward to 15 percent in 2007. If the 2007 mortality and damage rates continue to increase, assuming that most of the

damaged trees will die in the near future, the impact of the remaining living lodgepole pines in Colorado could be dramatic.

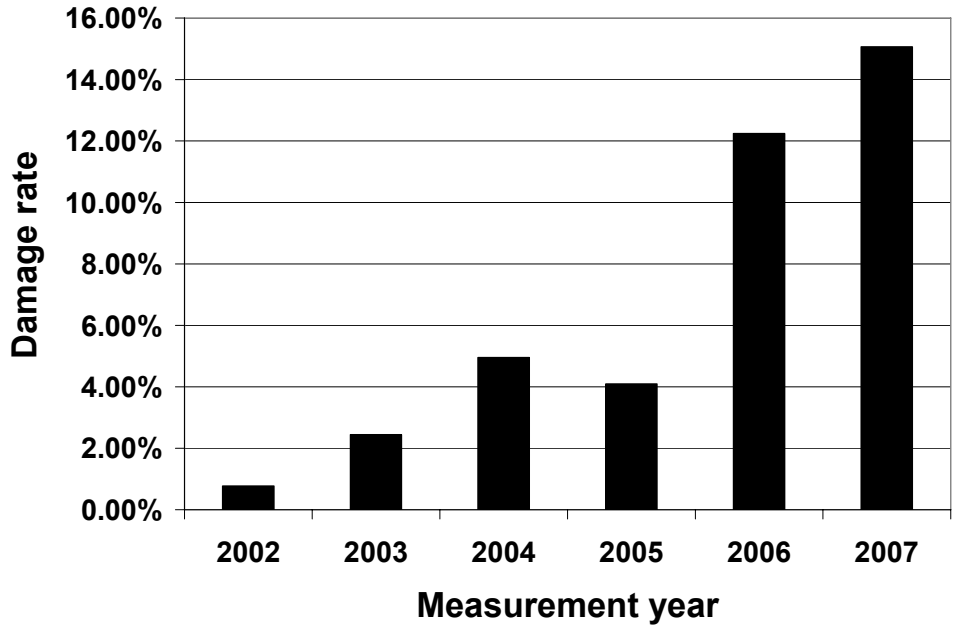


Figure 3: Percentage of live lodgepole pines damaged by bark beetles by measurement year in Colorado.

Unlike lodgepole pine, trends in aspen mortality over the annual inventory period do not indicate an immediate serious threat to the statewide inventory of aspen. There are several reasons why SAD is not reflected in Colorado's annual inventory. First, the rapid mortality associated with this phenomenon is relatively recent—it was first observed by researchers in 2004 (Worrel et. al. 2008). Second, it may be more of a localized event that will impact aspen in domains too small to be adequately captured in a broad-scale inventory. Third, the diverse factors associated with the decline make it difficult to assess cause of death and damaging agents that are specific to SAD. Fourth, aspen stand dynamics are complex. It is a relatively short lived species susceptible to a host of pathological organisms. Aspen is also a clonal tree species showing a variety of characteristics between individual clones.

Is there additional data from the annual inventory that might shed light on aspen decline? Figure 4 illustrates the percentage of live aspen trees that had evidence of damage. Disease was identified as predominate damaging agent of this species in Colorado and the highest disease damage rate of nearly 11 percent was recorded in 2007. As the annual inventory progresses in Colorado, and more temporal data is collected, the eventual impact that Sudden Aspen Decline may have on the aspen resource can be assessed with more confidence.

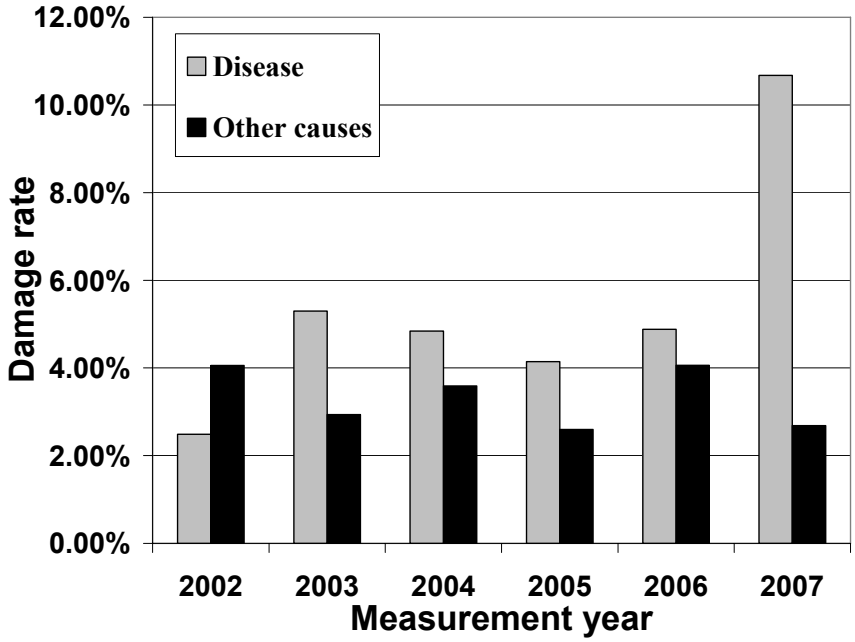


Figure 4: Percentage of live aspen trees that have evidence of damage by cause and measurement year in Colorado.

References

Carroll, A.L.; and , L Safranyik. 2004. The bionomics of the mountain pine beetle in lodgepole pine forests: establishing a context. Pages 19-30 *in* T.L. Shore, J.E. Brooks, and J.E. Stone, editors. Mountain pine beetle Symposium: Challenges and Solutions, Proceedings: Challenges and Solutions. October 30-31, 2003, Kelowna, British Columbia, Canada.

Gillespie, A.J.R. 1999. Rationale for a national annual forest inventory program. Journal of Forestry. 97(12): 16-20.

Logan, J. A., and J. A. Powell. Ecological consequences of climate change altered forest insect disturbance regimes. *In* F. H. Wagner (ed.), Climate change in western North America: evidence and environmental effects. Allen Press *(In review).*

Patterson, P.L., Reams, G.A. 2005. Combining panels for forest inventory and analysis estimation. In: Bechtold, W.A. and P.L. Patterson. 2005. The Enhanced Forest Inventory and Analysis Program—National Sampling Design and Estimation Procedures. General Technical Report SRS-80, USDA Forest Service, Washington, D.C.

Reams, G.A., and P.C. Van Deusen. 1999. The Southern Annual Forest Inventory System. Journal of Agricultural, Biological, and Environmental Statistics, 4, (4) 346-360.

SAS Institute. 1999-2001. The SAS system for Windows. Release 8.02. SAS Inst., Cary, NC

USDA Forest Service, 2007. Interior West Forest Inventory and Analysis Forest Survey Field Procedures, version 3.0. Available online at http://fsweb/manual/pdf/p2_manual_april_5_07_R.pdf. U.S.

Worrel, J.J., Egeland, L., Eager, T., Mask, R.A., Johnson, E.W., Kemp, P.A., Shepperd, W.D. 2008. Rapid mortality of Populus tremuloides in southwestern Colorado, USA. Forest Ecology and Management 255 686-696.

Large Area Mapping Applications I



Harmonizing Estimates of Forest Land Area from National-Level Forest Inventory and Satellite Imagery

Bonnie Ruefenacht[1], Mark D. Nelson[2], Mark Finco[1], Ken Brewer[1]

Abstract: *Estimates of forest land area are derived both from national-level forest inventories and satellite image-based map products. These estimates can differ substantially within subregional extents (e.g., states or provinces) primarily due to differences in definitions of forest land between inventory- and image-based approaches. We present a geospatial modeling approach for redefining satellite image-based pixels to meet inventory definitions. We compare resulting estimates of forest land area for six test states – Arizona, Minnesota, Montana, New York, North Carolina, and Oregon – using image estimates based on Moderate Resolution Imaging Spectroradiometer (MODIS) and inventory estimates from the U.S. Department of Agriculture Forest Service's Forest Inventory and Analysis Program (FIA). Our geospatial model utilizes several ancillary geospatial datasets to simulate conditions required by FIA's definition of forest land, including minimum forest patch area and width, minimum tree stocking or canopy cover, and exclusion of lands not used primarily as forest land.*

Keywords: forest, inventory, land use, land cover, dasymetric mapping, FIA

Introduction

Forest Inventory and Analysis

The mission of the USDA Forest Service Forest Inventory and Analysis Program (FIA) is to inventory the renewable forest and rangeland resources of the US. To inventory these resources, FIA has established field sample plots throughout the US at an intensity of approximately one plot per 2,400 ha (6,000 acres) (USDA Forest Inventory and Analysis 2007). FIA uses an annual rotating panel system whereby between 10 to 20 percent of each state's plots are sampled every year. From this plot data, FIA produces annual estimates of forest land area, reported in the form of tabular data at the county and state level. FIA defines forest land as land that meets at least one of the two following criteria:

[1] USDA Forest Service Remote Sensing Applications Center, 2222 West 2300 South, Salt Lake City, UT 84119
[2] USDA Forest Service Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

1. The land is at least ten percent stocked by trees of any size or has been at least ten percent stocked in the past. It is not subject to non-forest use(s) that prevent normal tree regeneration and succession such as regular mowing, grazing, or recreation activities. The area stocked by trees is at least one acre in size and at least 120 feet wide.
2. The land is a western woodland type where stocking cannot be determined, and has at least five percent crown cover by trees of any size, or has had at least five percent cover in the past. Additionally, the land is not subject to non-forest use(s) that prevent normal regeneration and succession such as regular mowing, grazing, or recreation activities. The area stocked by trees or with five percent crown cover is at least one acre in size and at least 120 feet wide.

Conversely, non-forest land is defined by FIA as land that does not support, or has never supported forests, and lands formerly forested where use for timber management is precluded by development for other uses. Non-forest land includes areas used for crops, improved pasture, residential areas, city parks, improved roads of any width and adjoining rights-of-way, power line clearings of any width, and non-census water. If intermingled in forest areas, unimproved roads and non-forest clearings must be more than 120 ft wide or more than one acre in size to qualify as non-forest land. Structures such as houses and cabins are considered non-forest land regardless of the size of the housing unit.

Although FIA forest land and non-forest land often are referred to in terms of "land use", FIA definitions reveal components of both land use and land cover. Requirements for minimum tree stocking or crown cover, minimum patch size, and minimum patch width all refer to land cover characteristics. Presence or absence of specific human activities (e.g., clearing of power line rights-of-way, or mowing of city parks) refers to land use characteristics. In addition, a temporal component is inherent within FIA's definitions: previously forested lands continue to be defined as forest following removal of tree canopy (e.g., from harvest, fire, or windstorm) if those lands have not been converted to another use and future tree regeneration is expected to achieve minimum thresholds of stocking or crown cover. FIA forest land is further differentiated into three sub-components: timberland, reserved forest land, and other forest land, which are defined by forest productivity and protection status.

Satellite Image-Based Land Cover Classification

Satellite imagery incorporates large geographic extents and commonly is used for mapping land cover. Land cover classifications typically include all landscape entities such as urban areas, forests, shrublands, grasslands, open water, etc. Usually these map products are defined in terms of land cover such as the National Land Cover Database (NLCD) of 2001 (Homer and others 2007). However, a seminal classification system for remotely sensed data (Anderson

and others 1976) refers to both land use and land cover, as do many current assessments of "land use/land cover" change and trends.

Nelson and others (2005) explored the efficacy of satellite image-derived maps for estimating forest land area by comparing estimates obtained from FIA, the USDA Natural Resources Conservation Service's (NRCS) National Resources Inventory (NRI), and four satellite image-derived data sets: 1991 Forest Cover Types, 1992–93 Land Cover Characteristics, 2001 Vegetation Continuous Fields, and the 1992 NLCD. The four satellite image-derived land cover maps, differing in date of image acquisition, classification scheme, and spatial resolution, showed varying degrees of similarity with inventory estimates of forest land across the conterminous United States (CONUS).

Forest Land Cover and Land Use

Both FIA and satellite image-based classifications may include components of land cover and land use but the importance of cover vs. use differs between these approaches. Therefore, a mapping approach for differentiating forest land use versus forest land cover would provide a more consistent basis for comparing classified satellite imagery with FIA estimates of forest land area. This topic comprises components of a broader research project being conducted by FIA and the Forest Service Remote Sensing Applications Center to produce a "four-in-one" geospatial dataset of pixel predictions for percent 1) tree canopy cover, 2) forest cover, 3) forest use, and 4) subcomponents of FIA forest land.

Nelson and others (2004) addressed the first component – tree canopy cover – by calibrating satellite image-based per-pixel predictions of percent tree canopy cover such that resulting estimates of forest land area were comparable to FIA estimates. Canopy cover is defined as "the proportion of the forest floor covered by the vertical projection of the tree crowns" and canopy closure is defined as "the proportion of the sky hemisphere obscured by vegetation when viewed from a single point" (Jennings and others 1999). Although terms such as crown closure, crown cover, and canopy closure sometimes are used interchangeably, (e.g., Avery and Burkhart 1994), we use the term canopy cover as defined in Jennings and others (1999). Although per-state estimates of forest land area derived from tree canopy cover data were strongly and positively correlated with FIA estimates, the two were not equivalent, due in part to definitional differences in land cover and land use.

The fourth component – differentiating timberland, reserved forest land, and other forest land – has been addressed by combining geospatial datasets of forest inventory attributes, forest land cover, and land ownership and protection (Nelson and Vissage 2007).

The focus of this study is to address the second and third components of the

"four-in-one" project – forest cover vs. forest use – by using existing geographic information systems (GIS) / remote sensing data layers to produce a map of forest land for the conterminous United States (CONUS), corresponding to FIA's definition of forest land. The initial map product is designed to have spatial resolution of 250 m, although the modeling procedure was designed to allow for production of forest maps having other spatial resolutions. Six states were chosen to develop and test the methodology: Arizona, Minnesota, Montana, New York, North Carolina, and Oregon. The subject of this paper is the development of the forest land use product for these six states.

Methods

An overview of the geospatial modeling process is presented below, and is diagrammed in the flowchart in figure 1. Following this overview is a more in-depth description of the methodology for specific land cover components.

FIA forest land is characterized by minimum tree stocking or crown cover, minimum patch area and width, and absence of non-forest uses. Non-forest land is defined in terms of agriculture lands, roads, urban areas, water, and small patch area or width. GIS and remote sensing data layers that characterize non-forest land cover classes already exist for CONUS. Therefore, forest land can be characterized and mapped by eliminating everything that is not non-forest land.

A raster tree canopy cover data layer was filtered to serve as a surrogate for FIA tree stocking/crown cover. Because the tree canopy dataset used in this study had coarser spatial resolution (250 m) than the other land cover datasets (30 m), a dasymetric mapping procedure was employed to spatially reallocate tree canopy cover to only areas with potential tree cover within the 250 m pixels so that accurate percent tree canopy cover estimates could be made.

The first step involved selection of land cover pixels where trees could potentially grow; trees cannot grow or are prohibited from growing on several land cover types (e.g., roads, water, and agricultural lands). GIS and remote sensing data layers of these land cover classes were combined and labeled as non-forest; everything else was labeled as potential forest.

Within each canopy cover pixel, non-forest cover pixels were labeled as non-tree cover; remaining areas were labeled as potential tree cover. The percent tree canopy cover attributed to the coarser pixel was reallocated to represent a more spatially explicit distribution of tree canopy cover within each pixel. This "true" "canopy cover layer was used to eliminate areas of potential forest that did not meet the minimum threshold. Initial GIS processing was conducted in a vector file format. This procedure was compared with a raster-based approach, which was used thereafter.

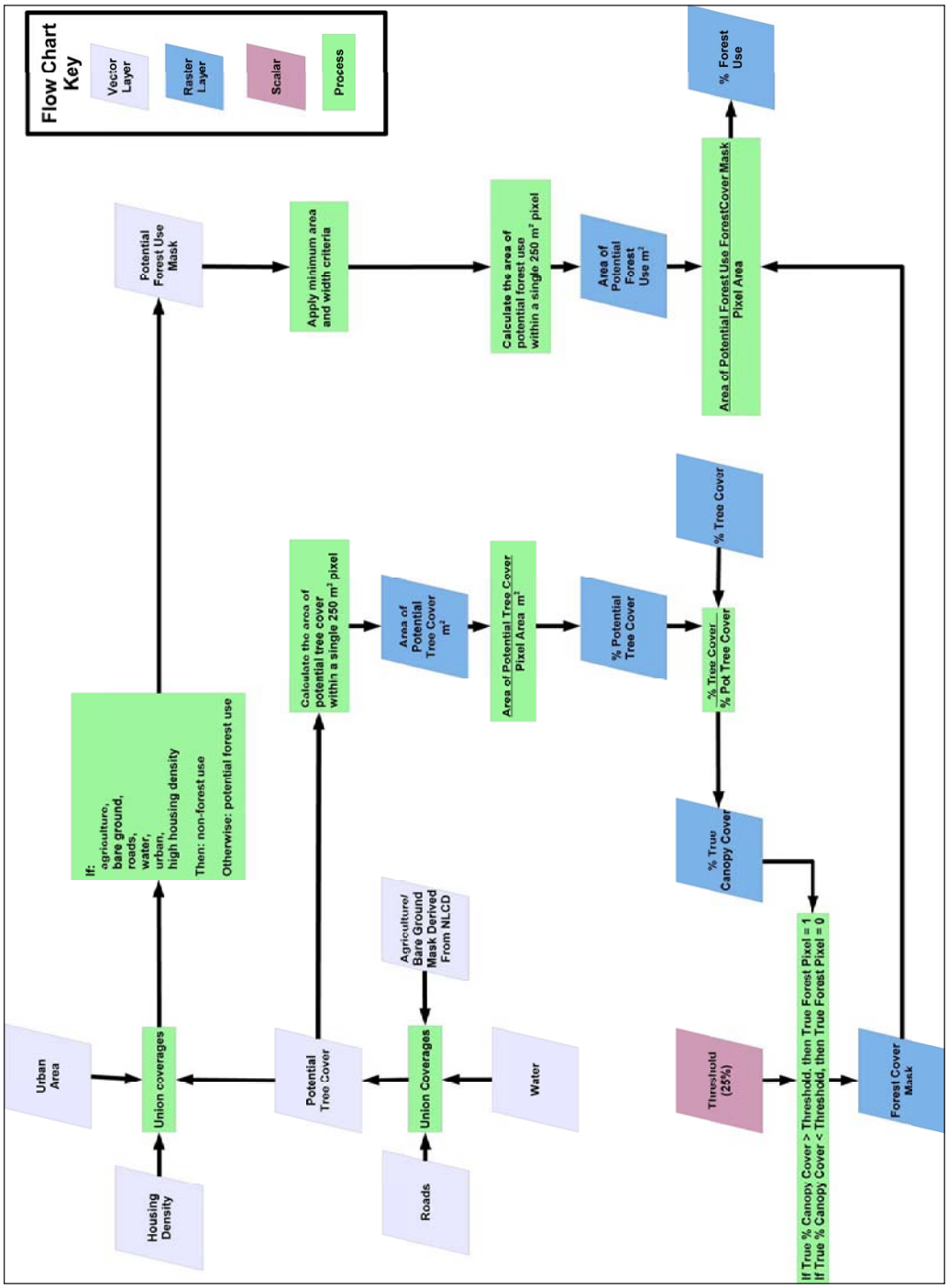


Figure 1: Flowchart depicting the process used to create the percent forest use product.

Non-Forest Land Use

Agriculture Lands: The database used to delineate agriculture lands was the 2001 NLCD (Homer and others 2007), which categorizes the entire U.S. at a spatial resolution of 30 m into 29 land cover classes (http://landcover.usgs.gov). The classes of interest to this project were 81 (pasture/hay), 82 (cultivated crops), and 31 (barren). Overall classification accuracies reported for 2001 NLCD ranged from 70 to 98 percent across mapping zones, with a nationwide average accuracy of 83.9 percent (Homer and others 2007). To create an agriculture/bare land mask, pixels with codes 31, 81, or 82 were recoded to a value of "1"; everything else was recoded to value of "0". The agriculture/bare land mask was converted to a polygon coverage.

Roads: The roads database, which is a line coverage at a scale of 1:100,000, was obtained from the Bureau of Transportation Statistics (http://www.bts.gov). Roads are attributed as indicated in the first two columns of table 1. According to FIA's non-forest land definition, unimproved roads that are intermingled in forest areas and less than 120 feet wide are considered forest land. Four-wheel-drive (4WD) trails generally are narrow and unimproved, meaning that they are not typically defined as non-forest. Therefore, these roads, with FCC codes A5 - A53, were deleted from the roads database. Since roads in the database are represented as centerlines that have no inherent width, the lines were buffered according to the road widths shown in table 1. These road widths were either derived from federal road standards or from general accepted standards of road widths.

Table 1: Road Attributes and assigned widths

FCC	Road Type	Total Road Width
A1 – A18	Interstate Highways	86 ft
A2 – A29	State Highways	44 ft
A3 – A38	City Main Thoroughfares	30 ft
A4 – A48	Local Roads	20 ft
A5 – A53	4WD Trails	10 ft
A61	Cul-de-Sac	120 ft
A62	Traffic Circle	120 ft
A60, A63	Access Ramps	26 ft
A64	Service Road	10 ft
A7 – A73	Other	10 ft

Water: The water data layer was extracted from the Wildland-Urban Interface (WUI) database (http://silvis.forest.wisc.edu/projects/WUI_Main.asp), which is based on the U.S. Census Bureau TIGER line files. The scale of the water data layer is 1:100,000.

Urban Areas/Structures: The smallest geographic entity for which the U.S. Census Bureau tabulates decennial data is the census block (http://www.census.gov). Aggregations of blocks are termed block groups. Blocks and block groups are irregular in shape and vary in size. Urbanized areas (UA) and urban clusters (UC) portray aggregations of core blocks or block groups that have a population density of at least 1,000 people per square mile, and surrounding blocks or block groups that have an overall density of at least 500 people per square mile. The UA and UC areas were considered to be non-forest. Census UA/UC data layers address population centers, but do not include rural areas. In addition, houses rarely are evenly distributed within census blocks or block groups. Normally, houses are located close to roads and are not located on steep slopes. Using the roads database described above, roads were buffered by 300 feet to capture a zone of proximity in which most housing structures are expected to be located.

A data layer of topographic slope from the USGS Elevation Derivatives for National Applications (EDNA) (http://edna.usgs.gov/) was used to produce a slope mask, where a value of "1" indicated areas less than 15 degrees of slope and a value of "2" indicated areas greater than or equal to 15 degrees of slope.

The slope mask and the road buffer data layers were joined together to predict locations of housing structures for areas outside the urban areas and urban clusters. Using a housing density attribute in the WUI database, the number of houses in each census block was redistributed to a geographic subset of each block occurring within the zone of road buffers, where slopes were less than 15 degrees. This dasymetric mapping approach was used to recalculate housing density within non-steep road buffers. Raster cells then were assigned their corresponding housing density value. If the resulting housing density was less than one house per ten acres, then the area was considered to be undeveloped. Otherwise, the area was considered to be developed and defined as non-forest.

The non-forest land use products described above are used in the development of the potential tree cover data layer discussed in the next section.

Tree Cover

Potential Tree Cover: A data layer representing non-treed lands was created by combining data layers for agriculture lands, road buffers, and water, described above. Remaining areas were assumed to have potential for growing trees. The potential tree cover data layer is used in the development of the potential forest data layer described in the next section.

Percent Potential Tree Cover: The next step was to calculate the areal extent of potential tree cover within each 250 m x 250 m cell, using the Moderate Resolution Imaging Spectroradiometer (MODIS) 250 m Vegetation Continuous Fields (VCF) percent tree canopy cover dataset. The VCF dataset consists of

per-pixel predictions of woody vegetation, herbaceous vegetation, and bare ground, which together sum to a total of 100 percent cover for each pixel. The VCF data layer of interest to this project is woody vegetation, i.e., tree canopy cover, which will be referred to simply as VCF for the remainder of this document.

Using VCF and the PIXELX and PIXELY functions in the ERDAS Imagine software package, a raster product was created where each pixel had a unique value based on the spatial location of the cell. This raster product was converted to a polygon coverage consisting of 250 m x 250 m squares. This polygon coverage was joined with the potential tree cover data layer making possible the calculation of area of potential tree cover occurring within each square polygon. This product was converted to a raster with a spatial resolution of 250 m. The area values were converted to percentages, thereby creating a percent potential tree cover data layer.

Area-Weighted Canopy Cover: VCF percent tree canopy cover values represent per-pixel predictions. However, canopy cover can vary substantially within each pixel because source MODIS pixels are coarse in spatial resolution relative to patterns of forest structure. For example, a VCF cell value might be 30% but this does not imply that woody vegetation is evenly distributed, at a density of 30%, throughout the cell.

To calculate the "true" canopy cover of the cell, the VCF per-pixel prediction of percent tree canopy cover was divided by the percent potential tree cover. If the percent potential tree cover was 100% and the VCF value was 30%, the woody vegetation was assumed to be evenly distributed throughout the cell and the "true" canopy cover is also 30%. If, however, only 50% of the cell has potential tree cover and the VCF value was 30%, then the VCF value is reallocated to the half of the cell with potential tree cover and the area-weighted prediction of canopy cover for that cell is 60% (i.e., 30/50).

Forest Cover Mask: A minimum threshold of tree stocking (or crown cover) must be present to meet FIA's definition of forest land. Because no geospatial dataset of tree stocking was available, the area-weighted canopy cover dataset was used as a surrogate. To satisfy the stocking requirement, a 25% minimum threshold of canopy cover was employed. If the area-weighted canopy cover for a pixel was below the threshold, the pixel was assigned a "0". Otherwise, the pixel was assigned a "1". This forest cover mask dataset was used to remove areas of non-forest land from the forest land in the next section.

Forest Land

Potential Forest: Potential forest was modeled by combining the potential tree cover data layer with the urban areas/structures data layer. Pixels were labeled as potential forest if they had potential tree cover and were not labeled as developed.

Minimum Acreage and Width: The next step was to eliminate areas from the potential forest data layer that were too small or too narrow, according to FIA's definition. All forest polygons smaller than one acre were relabeled as non-forest. Forested areas that were less than 120 feet wide were relabeled as non-forest. Because of coverage processing limitations, all procedures were performed on geographic subsets of data that measured 150 km x 150 km in extent.

FIA Forest Land: The resulting data layer produced from the steps above was joined with the 250 m x 250 m polygon grid created previously making possible the calculation of area of potential forest land occurring within each square polygon. This product was converted to a raster with a spatial resolution of 250 m and the area values were converted to percentages. The forest cover mask described above was used to remove pixels with less than 25% area-weighted canopy cover. These steps created a percent forest use or percent forest land data layer. The percent forest use data layers for Arizona, Minnesota, Montana, New York, North Carolina, and Oregon are displayed in Figures 2 – 7.

FIA produces tabular county summaries for various forest attributes which are estimated using sample plot measurements. These estimates are made available to the public as published reports and via web-based estimation tools such as Forest Inventory Data Online, EVALIDator, and FIA MapMaker (http://www.fia.fs.fed.us/tools-data/). FIA forest land area estimates were compiled for the states and counties for each of the six states included in this study. Corresponding estimates of sampling errors were used to construct 95 percent confidence intervals surrounding per-state estimates.

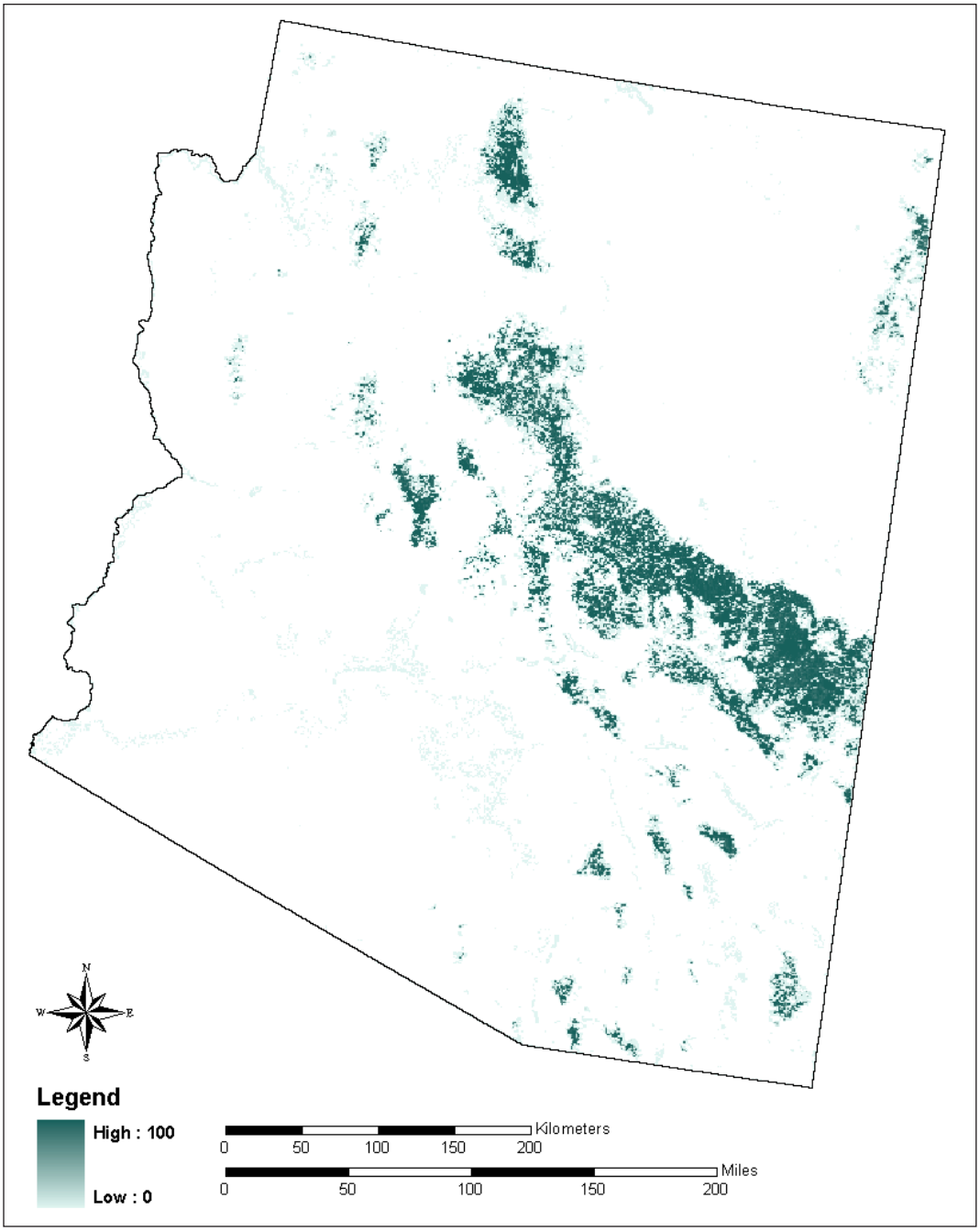


Figure 2: Percent forest use at 250 m spatial resolution of Arizona.

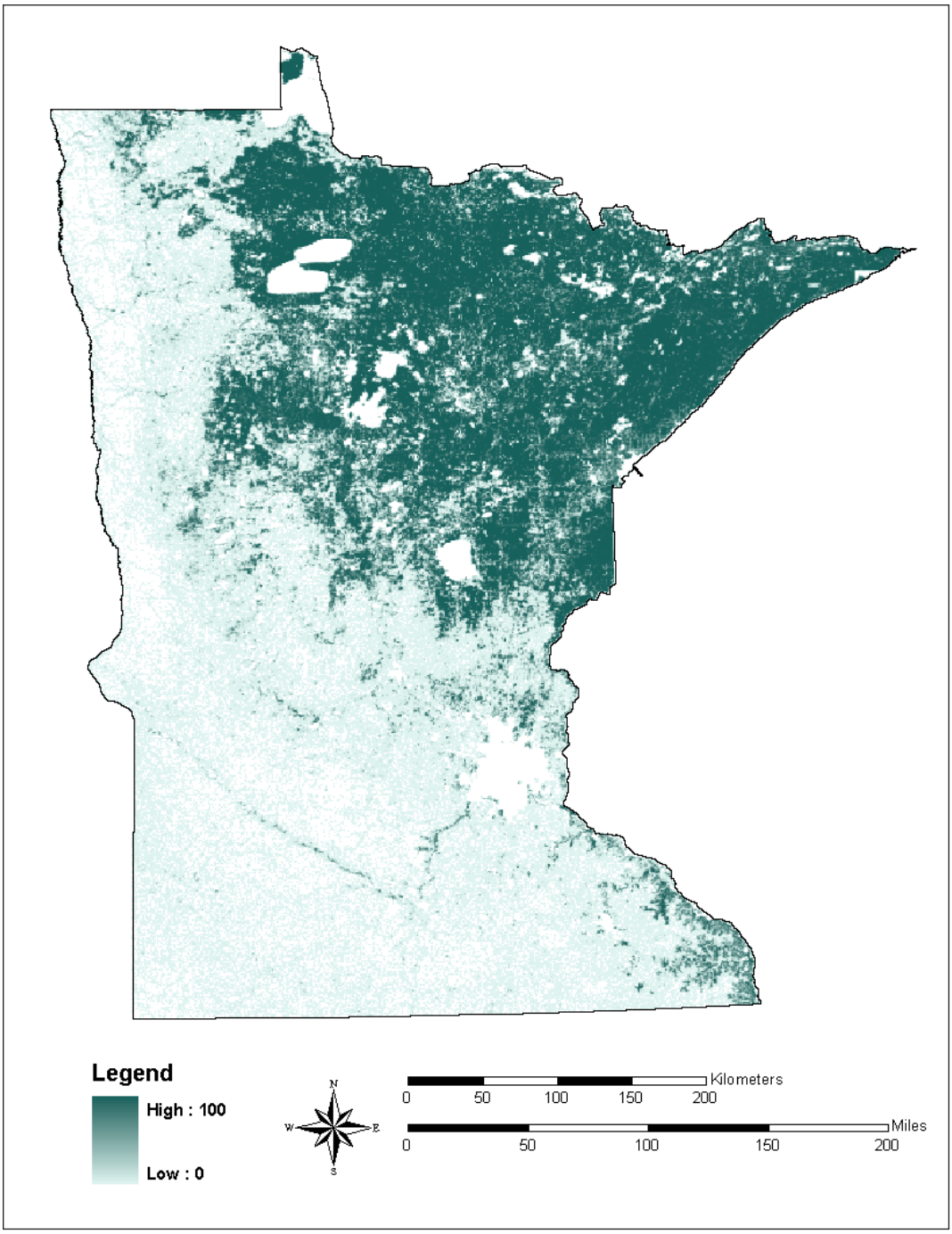


Figure 3: Percent forest use at 250 m spatial resolution of Minnesota.

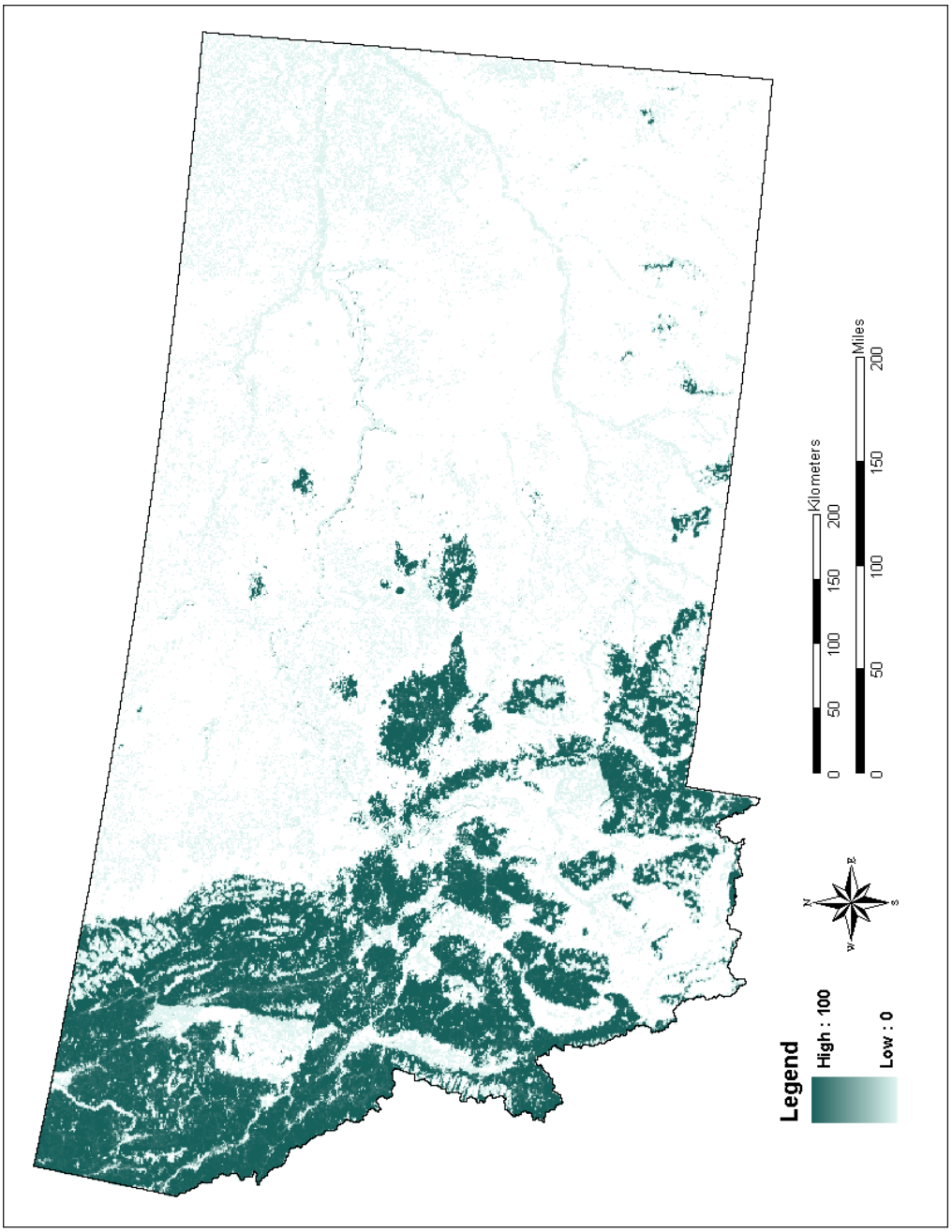


Figure 4: Percent forest use at 250 m spatial resolution of Montana

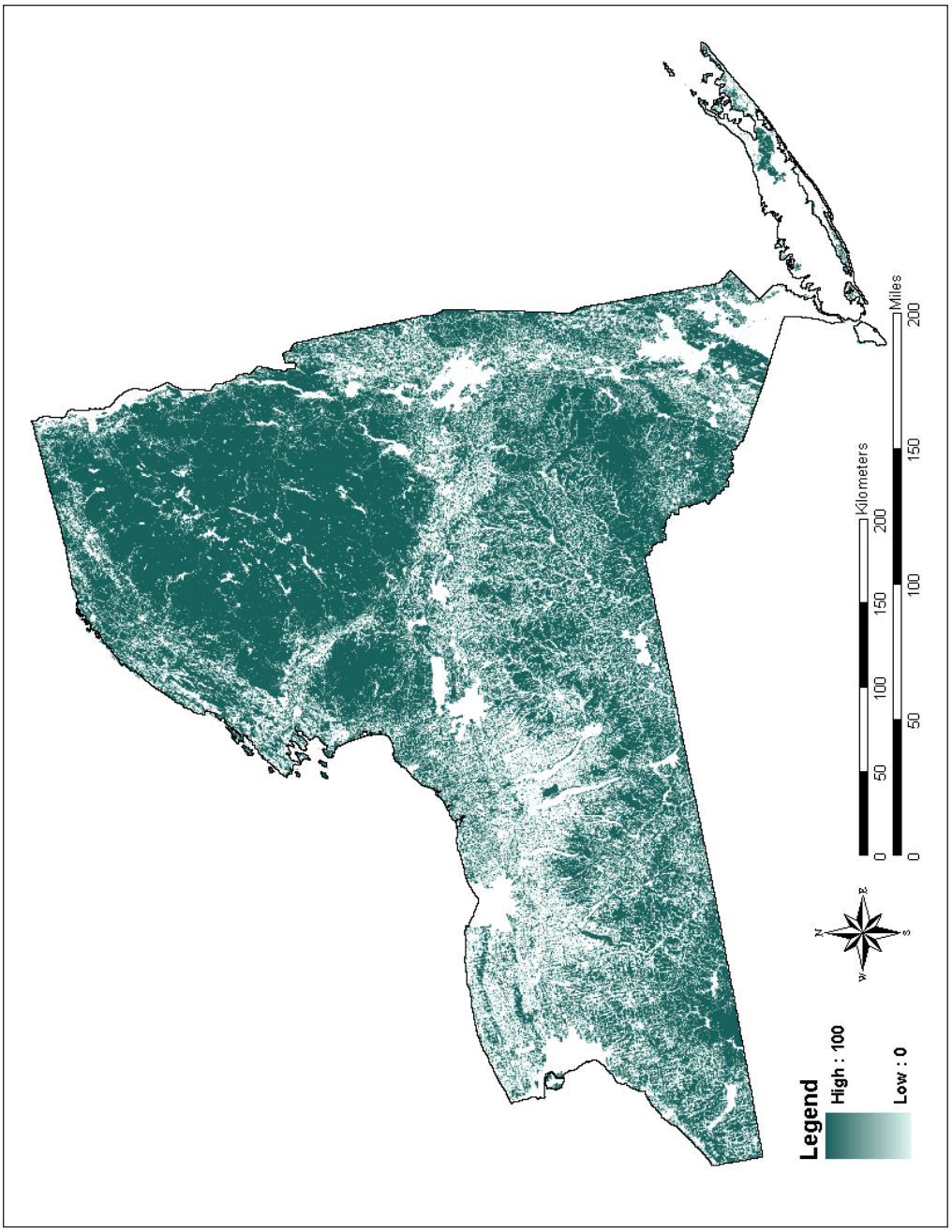


Figure 5: Percent forest use at 250 m spatial resolution of New York.

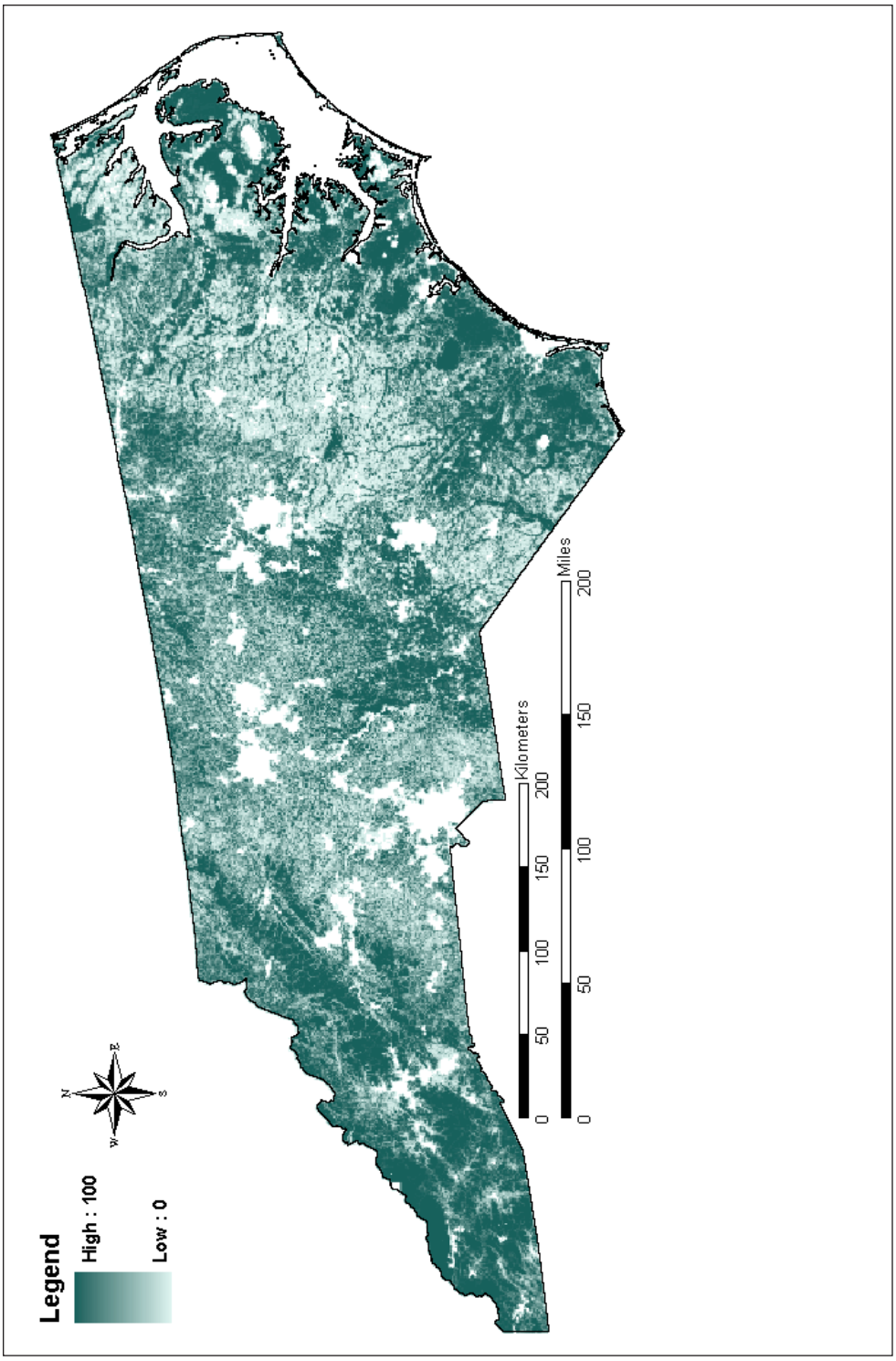


Figure 6: Percent forest use at 250 m spatial resolution of North Carolina.

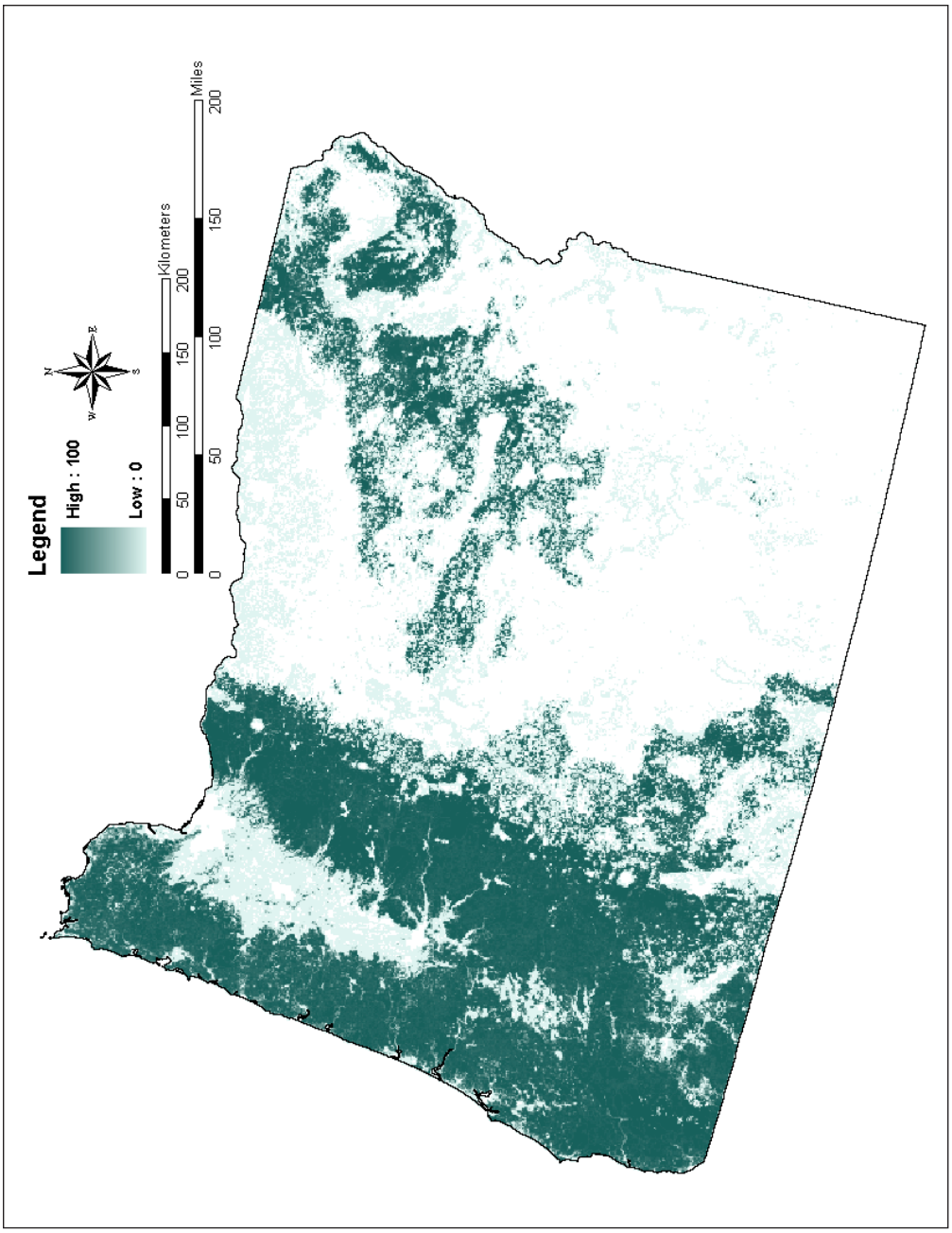


Figure 7: Percent forest use at 250 m spatial resolution of Oregon.

GIS Processing Effects

A major technical question is: can a geospatial forest map product be produced using a raster GIS file format, rather than the vector GIS file format employed in this study? Due to improved processing efficiency, working in a raster environment would eliminate the need to subset databases and could greatly reduce computer processing time. However, some vector processes cannot be replicated in a raster environment, or may produce different results (Wade and others 2003). For example, the reallocation of WUI housing density to different size polygons is a vector operation and raster data do not have polygons. Also, the elimination of areas less than 120 feet wide would be difficult to perform in a raster environment, unless pixel spatial resolution was small, which would negate some of the gains in processing efficiency. The end products of this project are at 250 m spatial resolution. The effects of simply eliminating the minimum width and area processes on the overall forest land area estimates were not investigated as part of this study.

For comparison, the vector methodology described in this paper was adapted for a raster environment, with one process altered and another process omitted. The process altered was the elimination of areas less than 120 feet wide. Instead, a minimum mapping size criterion was applied, which partially dealt with the 120 foot width requirement. The process omitted was the reallocation of housing density. All the data layers used in the vector-based approach were converted to a raster file format with spatial resolution of 30 m. The same basic procedures described above were followed, using the ERDAS Imagine software package.

Comparisons

Forest Land Area: Estimates of forest land area derived from this study's geospatial datasets were compared with FIA's plot-based estimates. Additional comparisons were made with estimates of forest land area derived from the 2001 NLCD (Homer and others 2007), the 2001 Forest Types Map (Ruefenacht and others 2008), and with FIA estimates of timberland – a major sub-component of FIA forest land.

For each state, countywide FIA estimates of forest land area were compared with modeled geospatial estimates to produce area weighted root mean square deviations (RMSD) using methods derived from Häme and others (2001):

$$RMSD_{rs} = \sqrt{\sum_i \frac{a_i}{A}\left(\hat{p}_{ir} - \hat{p}_{is}\right)^2} \qquad (1)$$

where a_i is the area of the ith county, A is the total area within a state (sum of a_is for all counties in that state), and $\hat{p}_{ir}$ and $\hat{p}_{is}$ denote the estimated proportion of forest land in the ith state obtained from the FIA sample plots (r) and modeled geospatial dataset (s) estimates.

Results

Differences in results derived from the raster and vector products were insignificant (table 2), but raster processing took only two days per state, as opposed to seven days for vector processing. These results support those of Wade and others (2003) who reported faster processing time with raster data, and ecologically insignificant differences in results. Subsequent results are reported only for modeling of raster geospatial file formats.

Table 2: Comparison of forest land area estimates (acres) derived using vector and raster GIS file formats.

State	Raster	Vector	Difference (Raster – Vector)	RMSD (Acres)	RMSD (Prop.)	RMSD (Percent)
Arizona	5,076,039	5,106,296	-30,257	11,209	0.0017	0.17
Minnesota	17,442,442	17,502,500	-60,058	21,082	0.0010	0.10
Montana	19,091,504	19,688,044	-596,540	33,000	0.0009	0.09
New York	18,309,622	16,839,980	1,469,642	33,880	0.0022	0.22
North Carolina	17,187,758	16,634,397	553,361	10,791	0.0006	0.06
Oregon	22,731,634	24,130,448	-1,398,814	79.867	0.0014	0.14

Figure 8 shows the per-state FIA estimates of forest land (FIA forest land) and timberland (FIA timberland) area compared to estimates derived from raster geospatial modeling of this study (GIS raster model), 2001 NLCD (NLCD01), and 2001 Forest Type Groups (ForTypGrp Map) datasets. Forest land area estimates from the raster model were substantially lower than FIA forest land area estimates for three western states (Arizona, Montana, Oregon), but were more similar for three eastern states (Minnesota, New York, North Carolina). However, all estimates were outside the lower and upper error bounds of the FIA forest land area estimate 95 percent confidence intervals. Modeled estimates were lower than 2001 NLCD estimates for all states, and lower than 2001 Forest Types estimates for all states except Minnesota, where the two were similar. By definition, timberland comprises a sub-component of FIA forest land area. For all states except North Carolina, FIA area estimates of timberland were significantly lower than estimates of forest land.

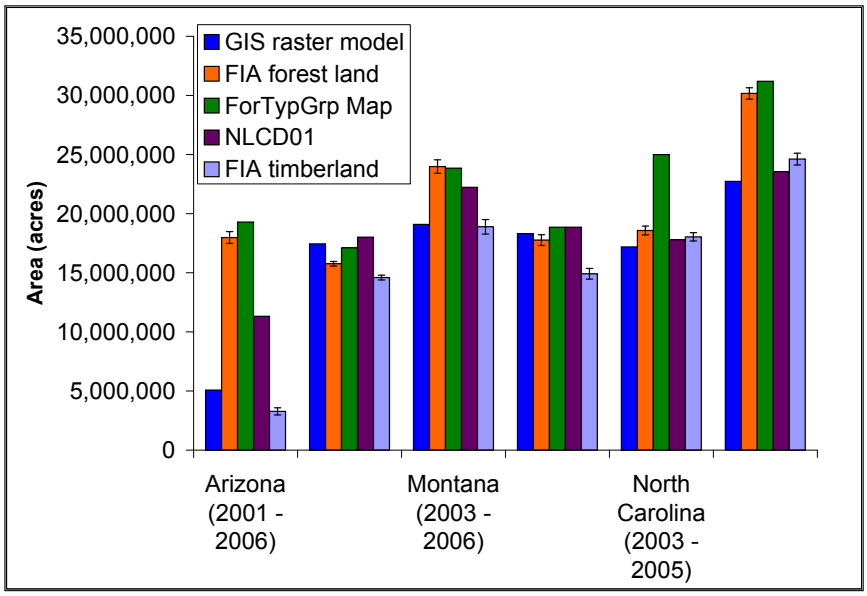


Figure 8: Estimates (in acres) of forest land area derived from the GIS raster model, FIA forest land, 2001 Forest Type Groups Map, 2001 NLCD; and FIA timberland area. Error bars for FIA forest land and FIA timberland show the 95 percent confidence intervals.

Figure 9 shows the RMSD between FIA estimates and raster GIS model-based estimates of countywide forest land area for each of six states included in this study. FIA forest land area estimates were in agreement with the forest land area estimates produced in this study. Arizona shows the largest RMSD, exceeding 3 percent, while Minnesota, New York, and North Carolina all had RMSD values below 0.5 percent. Montana and Oregon showed RMSD values between those of Arizona and the three eastern states.

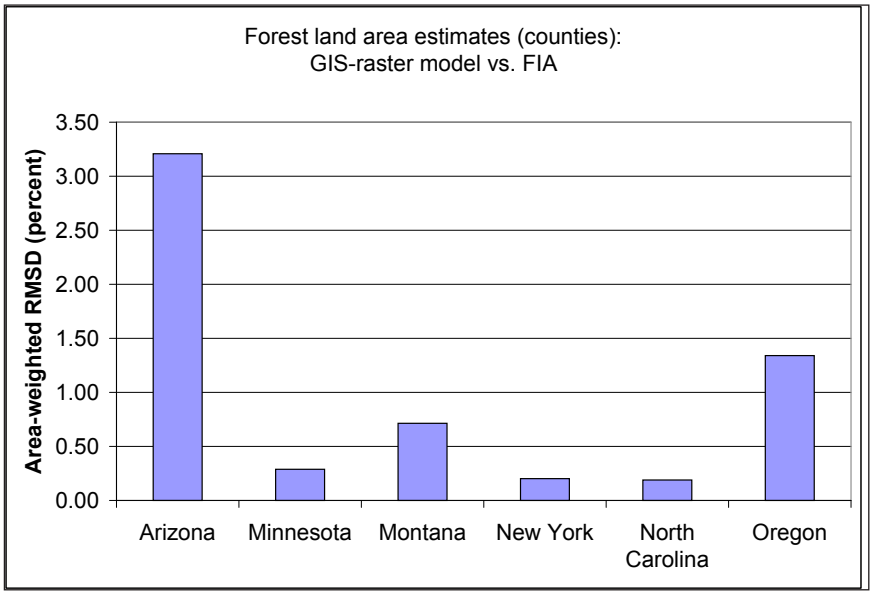


Figure 9: Area-weighted RMSD between FIA-based and GIS raster model-based countywide estimates of forest land area.

Discussion

Figure 8 reveals disappointing results in the raster GIS model-based estimates of forest land area for all states, with the possible exception of New York. One processing step in particular – applying a minimum canopy cover threshold of 25% – likely was a major cause of the discrepancy in estimates between FIA and the raster GIS model. A change in this threshold would greatly influence the resulting forest land area estimates. Nelson and others (2004) reported that mean VCF percent canopy cover varies substantially among states. Thus, considerable improvement in estimates is expected to result from optimizing the tree canopy cover dataset.

One of the goals of this project was to design an operational modeling framework where interchangeable pieces could be taken out and replaced with newer or more current data, or used for other spatial resolutions. Because the geospatial modeling procedure is now established and documented, it is possible to rerun the models with alternate data sources, and recalculate forest land area estimates, with resulting improvement in estimates of forest land area.

One of the major challenges with the methodology was the amount of time involved in processing geospatial databases in the GIS. For example, the computer processing time for each state spanned a minimum of seven days, which did not include analyst time required to obtain and transfer files, manage datasets, set up processes, etc. To process one state might require a total of two to three weeks. One of the reasons why the processing takes a considerable amount of time is due to the necessity of sub-setting the data layers into small subsets. This made it necessary to keep track of neighboring subsets, which was challenging. Extra processes needed to be incorporated to handle border issues, which significantly increased the processing time. Based upon our results, it is strongly recommended that future processing be performed in a raster environment, with modeling steps modified accordingly.

Acknowledgements

We would like to thank Jock Blackard, Rachel Riemann, and Nancy Solomon for reviewing and editing this manuscript. We greatly appreciate Gail Shaw for preparing this document for publication. We also thank the FIA Techniques Research Band for sponsoring this project.

References

Anderson, J.R.; Hardy, E.E.; Roach, J.T.; Witmer, R.E. 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964, Washington, D.C. 28 p.

Avery, T.E.; Burkhart, H.E. 1994. Forest Measurements, 4th edition. Series in forest resources. NY: McGraw-Hill. 408 p.

Häme, T.; Stenberg, P.; Andersson, K.; [and others]. 2001. AVHRR-based forest proportion map of the Pan-European area. Remote Sensing of Environment 77:76–91.

Homer, C.; Dewitz, J.; Fry, J.; [and others]. 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogrammetric Engineering & Remote Sensing. 73: 337–341.

Jennings, S. B.; Brown, N. D.; Sheil, D. 1999. Assessing forest canopies and understory illumination: canopy closure, canopy cover and other measures. Forestry. 72: 59-73.

Nelson, M.D.; McRoberts, R.E.; Hansen, M.C. 2004. Forest land area estimates from vegetation continuous fields. In: Greer, J.D., ed. Remote sensing for field users; proceedings of the tenth biennial Forest Service Remote Sensing Applications Conference; April 5–9, 2004; Salt Lake City, UT. Bethesda, MD: American Society for Photogrammetry and Remote Sensing. Unpaginated CD-ROM.

Nelson, M.D.; McRoberts, R.E.; Lessard, V.C. 2005. Comparison of USA forestland area estimates from Forest Inventory and Analysis, National Resources Inventory, and four satellite image-derived land cover datasets, pp. 129–137. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; [and others], eds. 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.

Nelson, M.D.; Vissage, J. 2007. Mapping Forest Inventory and Analysis forest land use: timberland, reserved forest land, and other forest land. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; [and others], eds., Proceedings of the seventh annual Forest Inventory and Analysis symposium; October 3–6, 2005; Portland, ME. Gen. Tech. Report WO-77, U.S. Department of Agriculture Forest Service: 185–190.

Ruefenacht, B., Finco, M.V.; Nelson, M.D.; [and others]. 2008. Conterminous U.S. and Alaska forest type mapping using Forest Inventory and Analysis data. Photogrammetric Engineering & Remote Sensing. 74:1379–1388.

U.S. Department of Agriculture, Forest Inventory and Analysis. 2007. Forest Inventory and Analysis national core field guide. Volume 1: field data collection procedures for phase 2 plots [online]. Available: www.fia.fed.us [December 8, 2008]

Wade, T.G.; Wickham, J.D.; Nash, M.S.; [and others]. 2003. A comparison of vector and raster GIS methods for calculating landscape metrics used in environmental assessments. Photogrammetric Engineering & Remote Sensing. 69: 1399-1405.

Enhancing FIA for Offset Monitoring



Investigation into Calculating Tree Biomass and Carbon in the FIADB Using a Biomass Expansion Factor Approach

Linda S. Heath[1], Mark H. Hansen[2], James E. Smith[1], W. Brad Smith[3], and Patrick D. Miles[2]

Abstract: *The official U.S. forest carbon inventories (U.S. EPA 2008) have relied on tree biomass estimates that utilize diameter based prediction equations from Jenkins and others (2003), coupled with U.S. Forest Service, Forest Inventory and Analysis (FIA) sample tree measurements and forest area estimates. However, these biomass prediction equations are not the equations used in the current public national FIA dataset (FIADB3), which utilizes regionally specific prediction equations, nor are they based on current FIA volume estimates. We describe and investigate an approach that is proposed for biomass estimates in the FIADB version 4 (FIADB4), due to be released in April, 2009, and that would produce national-level biomass and carbon estimates consistent with FIA volume estimates at the tree-level. The approach, called the component ratio method (CRM), is based on: 1) converting the sound volume of wood in the bole to biomass using a compiled set of wood specific gravities; 2) calculating the biomass of bark on the bole using a compiled set of percent bark and bark specific gravities; 3) calculating the biomass of tops and limbs as a proportion of the bole biomass based on component proportions from Jenkins and others (2003); 4) calculating the biomass of the stump based on equations in Raile (1982); and 5) summing the parts to obtain a total aboveground live biomass. Root biomass is also available as a proportion of the bole biomass based on component proportions from Jenkins and others (2003). The CRM approach is based on assumptions that the definition of bole in the volume prediction equations is equivalent to the bole in Jenkins and others (2003), and that the Jenkins and others (2003) component ratios accurately apply.*

We compare results between estimates calculated using equations in Jenkins and others (2003), current regional FIA equations, and this approach. The CRM approach is promising because the estimates are congruent with FIA volumes and compiled specific gravities. However, because FIA units currently use different volume equations the resulting estimates are not nationally consistent (that is, biomass of the same diameter and species tree will differ between regions). Because a number of volume equations are currently used by FIA, this approach can be complex for those wanting to take their own tree data and estimate biomass with FIA prediction equations especially when data cross regional boundaries. In the long-term, a planned and coordinated research study, as well as an accompanying operational implementation plan, for volume and biomass estimation methods would greatly add to the credibility of these estimates in the publicly available national FIA dataset.

Keywords: biomass equation, forest inventory, greenhouse gas inventory, FIA

[1] U.S. Forest Service, Northern Research Station, Durham NH 03824, USA; email: Lheath@fs.fed.us, tel: 603-868-7612
[2] U.S. Forest Service, Northern Research Station, St Paul, MN, USA
[3] U.S. Forest Service, Washington Office, Washington DC, USA

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

The U.S. Forest Service, Forest Inventory and Analysis (FIA) program is receiving an ever-increasing number of requests for forest biomass and carbon estimates, in addition to the traditional volume estimates that have been central to the FIA program. Because the carbon content of wood and bark is about 50 percent of dry biomass (Houghton and others 1997), carbon estimates are obtained by multiplying dry biomass estimates by 0.5, and all discussion of carbon estimation focuses on the estimation of dry biomass. Previous analysis of the data in FIA's national database revealed inconsistencies in the biomass estimation approaches and resulting estimates in the FIA regions (for example, see Hansen 2002) suggesting that FIA needs a national approach to biomass estimation. More importantly, the standard prediction equations used nationally in conjunction with FIA tree measurement data to produce the official forest greenhouse gas (GHG) inventories of the United States (U.S. EPA 2008) are the biomass prediction equations developed by Jenkins and others (2003). Note that we use the phrase "biomass prediction equation" to indicate the equations are fitted models; the word "equation" alone implies an equality.

In addition to a relationship between biomass and carbon, it is logical to assume there should be a relationship between volume and biomass. This relationship is implicit in the biomass expansion factor (BEF) approach (Brown and others 1989, Houghton and others 1997, Somogyi and others 2007). The FIA program for years has developed and maintained a statistically sound, sample based inventory of forests of the United States, including estimated volumes of individual sample trees based on tree measurements applied to volume prediction equations. These prediction equations used by FIA, such as those described in Hahn and Hansen (1991) and Flewelling and Rayner (1993) have been developed specifically to obtain the best possible estimates of individual tree volumes on a regional basis, have received scientific peer review, and are being used extensively for volume estimation purposes. FIA volume estimation procedures take into account major species, diameter, and height, or other factors that help predict a tree's volume, as well as taking deductions for atypical tree form. Traditionally, FIA has focused on the estimation and reporting of net volume of wood in the bole (net means deductions for nonmerchantable portions of the bole are made); however, in recent years the focus has shifted to the estimation of sound volume where only deductions for missing and rotten portions of the bole are made. Tying biomass to sound volume, and then multiplying biomass by a carbon conversion factor, provides not only consistent volume and biomass estimates, but also 'matching' carbon estimates.

One way for FIA to calculate carbon estimates at this time in the national FIA databases is to simply adopt the Jenkins and others (2003) equations, arguably the current standard for carbon estimates in the United States (for example, see Smith and others 2006, U.S. Dept. of Energy 2006, U.S. EPA 2008). This would make the FIA database consistent with past national carbon reporting. It would not

provide direct linkage between biomass and carbon estimates and the volume estimates.

The objective of this study is to conduct a preliminary analysis of a biomass expansion factor approach to investigate its potential to be a nationally consistent biomass computation procedure to calculate dry weight biomass in the publicly available national-level Forest Inventory & Analysis database, the FIADB (U.S. Forest Service 2008). Because "biomass expansion factor" is often used generically and consequently is ambiguous, we use the phrase "component ratio method" (CRM) to describe our BEF approach. We present the CRM approach in detail and apply it to a specific example as well as to all data from annualized surveys in the FIADB.

Background and Current Status

BEF and Forest GHG Inventories

FIA conducts statistically sound forest surveys over large areas (Bechtold and Patterson 2005). Measurements are taken, and prediction equations applied to calculate volume or biomass. Biomass may be calculated from measured tree attributes using biomass prediction equations, or calculated indirectly by multiplying the volume estimates by biomass factors that expand or convert the volume estimates to biomass. In the latter case, these factors are called "biomass expansion factors," originally applied only to expand stand-level volumes or volume growth. However, this phrase now has been applied generically at the tree level and has been used to mean a number of things, including converting units rather than factors that expand. See Somogyi and others (2007) for an extensive discussion of various definitions and facets of BEF approaches. A BEF approach is listed as the preferred method for some of the tiers in the Intergovernmental Panel on Climate Change guidance for national greenhouse gas inventories (Penman and others 2003). However, the higher tier methods call for greater specificity, such as country-level factors and factors specific to species. It is generally recognized that when individual tree data is available, biomass estimates based on individual trees are preferred.

A Standard Way to Develop Equations

The standard empirical way of developing credible biomass prediction equations is to collect data from a sample of trees across the range of sizes, from species and the area of interest. This approach was taken in Canada by Lambert and others (2005) and Ung and others (2008) using data from thousands of trees collected under the Energy from the FORest (ENFOR) project in the early 1980s. With these data, the authors could truly develop an internally consistent set of national allometric equations, including validation and testing. Two sets of equations were developed: one based on diameter at breast height (d.b.h) only and the other based on d.b.h and height. Such an approach provides not only

predictions of individual tree biomass but also estimates of the bias and random error associated with these prediction equations.

A second credible approach is to collect unrelated datasets for a wide-ranging, well studied species without having to collect additional samples, and reanalyze the data. For instance, Wirth and others (2004) studied Norway spruce using this approach. Out of 688 trees, only 78 were completely sampled for biomass, and young trees especially were under-represented. In spite of this limitation, this study features important points to consider when designing a study to derive biomass equations. The credibility of the equations and estimates are strengthened by setting evaluation criteria on the process and resulting equations. This is discussed below.

Evaluation Criteria

Quality, science-based information for land management, at the strategic to applied level, is needed (USDA FS 2007). "Science-based" typically implies some type of peer-review, either as peer-review in journal publications or as a review by a designated panel of experts. The most stringent criteria for choosing participants in peer-review are for highly influential scientific assessments (OMB 2004), but even this document notes that different types of peer review are appropriate for different information. The Forest Survey Handbook[4] calls for "high quality, consistent and reliable data" in FIA databases, but provides little guidance on how to do that. More detailed guidance is given within the Forest Service Research and Development quality assurance program. Acceptance and publication in a peer-reviewed research journal is often an acceptable standard, but for highly influential work or in which the turnaround time is critical, expert panels are often preferred.

In the past, for individual studies, evaluation criteria for carbon accounting studies focused on criteria of accuracy, precision, consistency over time, and transparency, yet would also be cost-effective and usable by other scientists and managers and. Having a consistent approach over time is absolutely critical because it is the change in carbon over time, not just carbon stocks, which is of most interest in the terms of the carbon issue. Inconsistent approaches over time can affect the amount claimed to be sequestered. In this preliminary investigation, we do not formally evaluate the equations, but note issues and results to consider.

As new information needs and science results become available, it is important to re-evaluate existing systems, in this instance, equations, and consider adopting new approaches. To maintain consistency over time when implementing an "improved prediction equation" it is important to be able to apply the new method to not only current data but also to all previous data that form the historical record, to recode all tools that use the method being updated, and to work with

[4] Forest Service Handbook 4809.11 Amendment No. 4809.11-2001-1, approved 12/28/2008.

users to again develop credibility in the methods. Thus, changing estimation techniques is costly in many ways, and therefore changes are not undertaken without clear benefits.

Carbon Estimation in Forest GHG Inventories of the United States

The Jenkins and others (2003) biomass prediction equations are one of the pillars of the carbon estimates used for forests in the official greenhouse gas inventories of the United States (U.S. EPA 2008), which have arguably served as the "gold standard" for carbon. Scientific studies (for example, Potter and others 2008) compare their carbon results to carbon estimates based on the application of Jenkins and others (2003) to the FIA data, or use the estimates to calibrate their models. Virtually all the policy-relevant carbon estimates and carbon tools, such as the U.S. Greenhouse Gas Inventories (for example, U.S. EPA 2008, USDA 2008), Heinz Center carbon storage indicators (Heinz Center 2002, 2005, 2008), carbon indicators for the 2010 Sustainable Forests, the updated 1605b Voluntary Reporting Program of the United States (Smith and others 2006, Pearson and others 2008, U.S. Dept of Energy 2006, NCASI 2008); Carbon Calculation Tool (Smith and others 2007); Hoover and Rebain 2008 (FVS-Carbon) in the United States are based to at least some degree on FIA data and the biomass equations from Jenkins and others (2003). There are a number of studies that have used FIA regional biomass estimates for carbon (such as Schroeder and others 1997), but these studies only covered only a part of the conterminous United States.

Carbon estimates for trees based on databases of older plot-level FIA data are based on Smith and others (2003), which were developed based on biomass from Jenkins and others (2003). The Jenkins and others (2003) equations were developed at the time specifically because 1) large differences in tree biomass carbon between FIA units[5] for the same species and size tree sometimes occurred; 2) documentation for existing equations was scattered and uneven in its quality so it was difficult to check the data or know the source of the estimate; and 3) databases at the time did not include mass for standing dead trees. Some of these items are still true. Perhaps most importantly, forest carbon inventories were still viewed with some suspicion by many communities as highly uncertain, and having a method based on a peer-reviewed publication provided credibility, especially with carbon becoming a commodity in the marketplace.

Jenkins and others (2003) features 10 equations covering all tree species in the conterminous United States, based on a meta-analysis of a thorough compilation of all biomass equations (Jenkins and others 2004) found in the literature. These equations are based on diameter only because the databases available at that time included tree diameter in all tree records, but only occasionally included measured height (that is, not estimated from diameter). A similar approach was also

[5] The FIA units are designated by Northern Research Station (NRS), Pacific Northwest Research Station (PNWRS), Rocky Mountain Research Station (RMRS), and Southern Research Station (SRS).

adopted by Muukkonen (2007) to develop generalized allometric volume and biomass equations for some species in Europe for regional analyses, also based on diameter. Moreover, Muukkonen (2007) included equations with height in the underlying compilation of equations by using diameter-height equations, which is usually a strong relationship.

Some users of Jenkins and others (2003) equations have reported the equations estimate greater biomass than they expect at large diameters because height is not included, that the form of the equation forces biomass to continue to increase as diameter increases (see Figure 1). Users expect the rate of increase at larger diameters to be smaller with total biomass in a tree approaching some maximum upper limit rather than continuing to increase at an increasing rate. This issue may be worth revisiting.

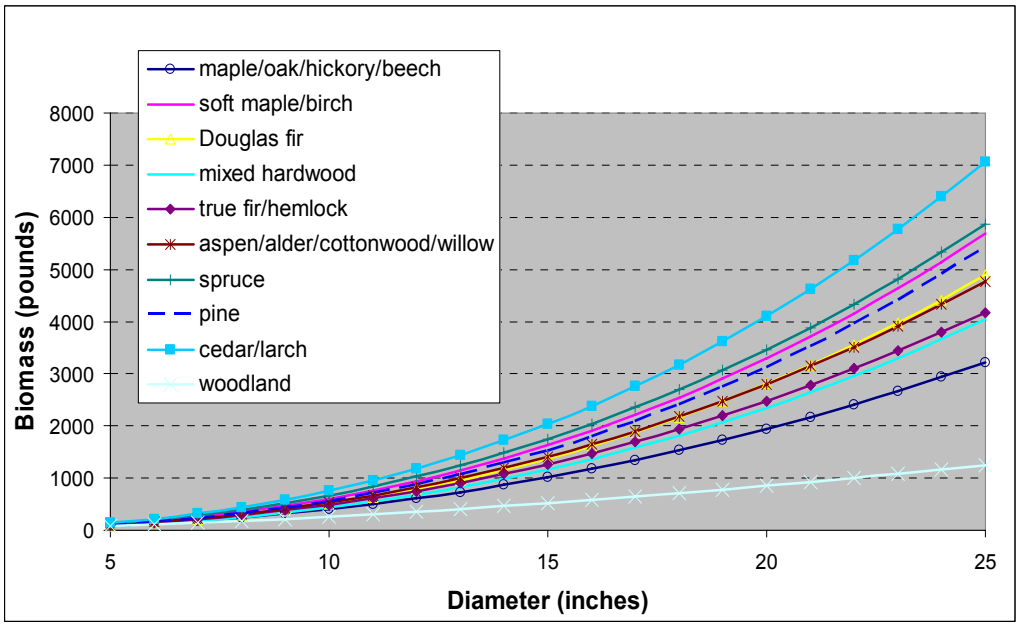


Figure 1: Total aboveground biomass from Jenkins and others (2003) estimated for each of the species groups. Note the diameter of woodland species may be measured at diameter root collar because some woodland species are multi-stemmed; this woodland equation is based on d.b.h.

FIA Biomass and Volume Estimation for Forests of the United States

There is currently no single publication that lists the tree biomass estimation approaches for all the FIA units. Current FIA biomass and volume equations have different forms for the regional FIA units, and were developed at different times from different datasets (Hansen 2002). Hansen (2002) documented the different volume and biomass estimation procedures in the eastern FIA units, which, due to historical reasons, included three sets of approaches for the current Northern Research Station (NRS) FIA alone. Methods were compared for 67 species that cross regional boundaries. Based on the results, Hansen cautioned users of FIA data from making regional comparisons of volume or biomass

estimates for small diameter trees. He suggested that FIA needs to move to a consistent method to estimate tree volume and biomass nationwide that uses common measurement data. However, he also noted that consistency over time is an important consideration in revising equations because these will affect the calculation of changes over time.

Methods

Our approach entails compiling species-level wood densities (dry mass per unit green volume) to multiply by green bole volume for a dry biomass estimate per bole. This approach makes the biomass estimate for the bole portion of the tree equivalent to the FIA volume of that portion; adopting the Jenkins and others (2003) equations would not. In addition to tree biomass prediction equations, Jenkins and others (2003) presents equations to predict the proportion of the biomass in foliage; tops, limbs, and stumps; bark of bole; bole wood; and coarse roots; to the total aboveground biomass, respectively, for hardwood and softwood species by d.b.h. We use ratios developed from the component equations in Jenkins and others (2003) for a consistent approach to predicting biomass in other components of the tree besides the bole. We calculate the component ratio estimates based on the equation sets from Jenkins and others (2003), and produce proportions of tops and limbs, and root components in terms of bole wood biomass. The calculation of stump and bark components that the ratios are built from are based on different methods described below. We multiply these ratios in terms of bole biomass by our calculated bole biomass to calculate the biomass in each component pool.

Understanding the concepts underlying the current approaches is necessary to devise a method to calculate biomass from volume. First we define types of volumes and biomass used by FIA. We then briefly discuss current regional biomass computation. Finally, we describe the steps we used to calculate biomass from sound volume.

Definitions

Definitions of the various volume, biomass, and carbon components are key. FIA volumes are green wood basis—that is, they represent the volume of wood as standing or freshly cut, not the volume of dry wood, and bark is not included. The unit of measure of interest for this study is cubic feet, although other units of volume, such as board foot, are available from FIA. Volume is defined for trees greater than 5 inches diameter, and only includes the central bole of the tree from 1 foot aboveground to the point where the central stem has a diameter outside bark of 4 inches (or the point where the tree forks into branches all of which are less than 4 inches). This is the standard volume that has been used historically in most volume studies in the United States for more than 100 years. Trees less than 5 inches d.b.h., called saplings, are assumed to have zero volume.

Users of biomass equations and estimates in general need to aware that these may be on a dry or green weight basis; may include or exclude bark, foliage, stump and root portions of trees; may include seedlings (trees < 1 inch diameter); or may include species that FIA considers to be shrubs rather then trees. Also the units of measure for reporting biomass include pounds, tons and kilograms. Biomass in this study is on an oven-dry basis and includes bark, but excludes foliage. The unit of measure is pounds unless otherwise noted and FIA biomass is defined only for trees greater than 1-inch diameter. FIA biomass estimates typically include only the aboveground portion, however, with the introduction of FIADB4 a prediction of the biomass in the coarse roots portion of all trees greater than 1-inch diameter has been added.

Volume

FIA defines and calculates gross, sound, and net bole volumes, of all live trees at least 5 inches d.b.h (USDA FS 2008b) (Fig. 2). These volumes are estimated for the central stem (bole) from a 1-foot stump to a minimum 4-inch top diameter outside bark, or to a point where the central stem breaks into limbs. The only time there are differences between these three volume estimates for a tree is when the estimated rotten or missing parts of the tree are nonzero, or when the tree has poor form. Gross volume is the total potential volume; rotten or missing parts of the tree and poor form effects on tree volume have not been deducted from gross volumes. Sound volume is gross volume with missing and rotten volumes of the tree deducted. Net volume is gross volume minus deductions for rot, roughness, and poor form. Depending on the FIA unit, either gross volume or net volume will be calculated first, as well as the missing and rotten, or volume affected by form, and then sound volume is calculated by adding or subtracting the appropriate portion. Many of the gross volume prediction equations used by the Pacific Northwest Research Station FIA unit are based on the integration of taper equations that predict the diameter of the bole at any height, such as those in Flewelling and Rayner (1993). The Northern Research Station FIA unit is in the process of converting its volume estimation to a taper equation-based system. These taper-based systems are capable of predicting the bole volume in any portion of the bole from the ground to the top of the tree.

Biomass

FIADB3 (version 3 of FIADB) includes two biomass variables: total gross (named DRYBIOT in the database) and merchantable stem (DRYBIOM) biomass. Total gross aboveground biomass includes main stem, bark, tops, limbs and stump of all live trees 1 inch in diameter or larger, but excludes foliage and roots. Merchantable stem biomass includes only trees greater than or equal to 5 inches d.b.h from a 1-foot stump to a minimum 4-inch top outside bark of the central stem. All trees less than 5 inches d.b.h have total biomass, but they have a merchantable biomass of zero. Gross biomass minus merchantable biomass produces the amount of biomass in tops, limbs, and stumps, as well as all the

biomass in trees less than 5 inches diameter. See Figure 3 for an illustration of the differences between these types of biomass.

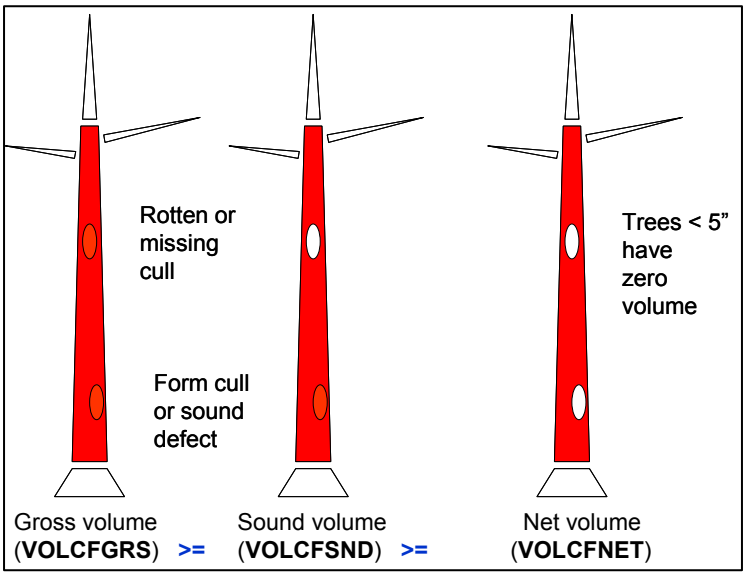


Figure 2: Illustration of gross, sound, and net volume at the tree-level.

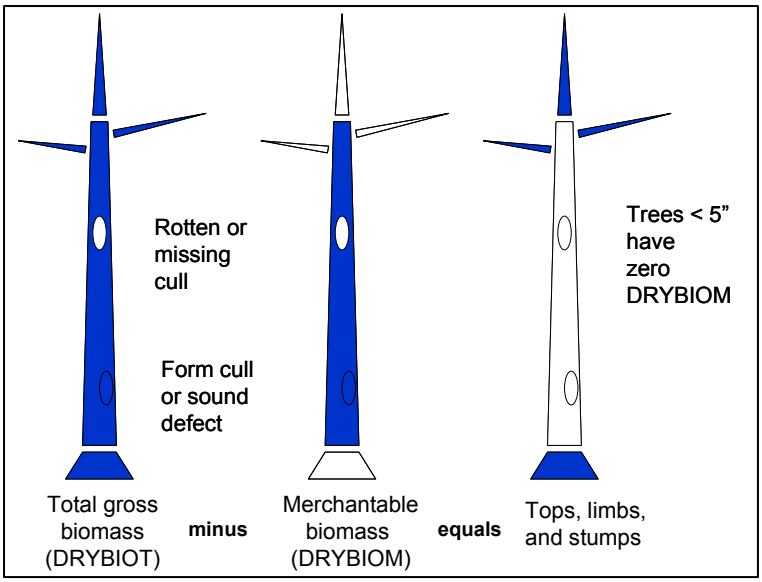


Figure 3: The difference between gross and merchantable biomass as stored in FIADB3.

The components of biomass in this approach are illustrated in Figure 4. The biomass components for top and limbs (labeled DRYBIO_TOP) are broken out from the stumps (DRYBIO_STUMP) and the merchantable (bole) biomass has been labeled DRYBIO_BOLE because it is not the same value as in previous FIA datasets. These three variables are computed for all species where FIA measured

d.b.h., and are defined to be zero for woodland tree species (because diameter is measured at root collar) and for trees less than 5 inches d.b.h. To avoid confusion, the attributes DRYBIO_SAPLING (total aboveground biomass in trees 1 to 5 inches d.b.h) and DRYBIO_WDLD_SPP (total aboveground biomass in woodland species) have been added. Belowground biomass estimates in coarse roots are not part of FIADB3; however, these estimates are of interest and are shown in the illustration as DRYBIO_BG.

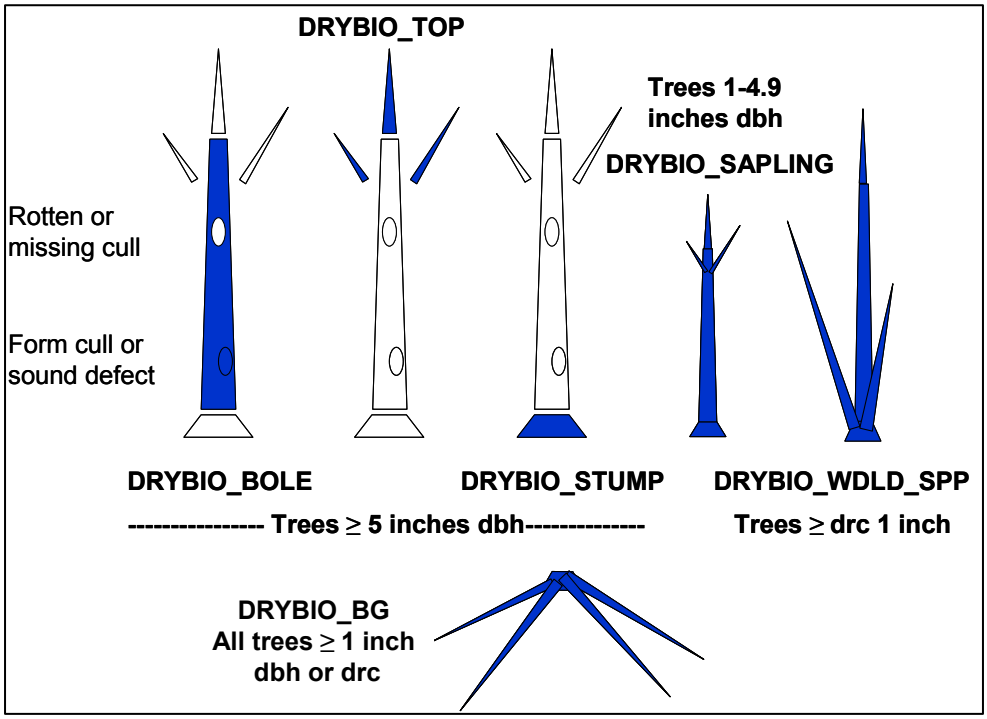


Figure 4: Biomass variables needed to implement this method in the FIADB.

Inconsistencies in estimation of component biomass

FIA biomass estimation procedures have shown unexpected differences in the average amount of biomass in tops, limbs, and stumps. Table 1 shows an example of the differences between eastern FIA units using select red oak trees of 10 inches d.b.h and 60-70 feet in height from all trees measured over the period 1999-2006, using regional equations and the Jenkins and others (2003) equations. The percent in tops, limbs, and stumps varies from approximately 10 to 30 percent in the eastern units, based on the current biomass calculation procedures. The pounds of wood per cubic foot of sound wood vary from 40 lbs/cu ft to almost 54 lbs/cu ft. The region with the higher lbs/cu ft has the lowest percentage in tops, limbs, and stumps. A revised approach should result in these components being similar between regions.

Table 1: Biomass statistics for all 10-inch diameter at breast height, select red oak growing stock trees, 60-70 feet tall, from all FIADB3 observations, 1999-2006.

FIA Region[a]	DRYBIOM/VOLCFSND	(DRYBIOT-DRYBIOM)/DRYBIOT
	--Pounds wood and bark per cu ft sound wood--	*--Percent of total biomass in tops, limbs, and stumps--*
Using regional equations:		
NRS-East	53.5	11.3
NRS-West	40.9	29.8
SRS	41.4	21.8
Using Jenkins and others (2003):		
NRS-East	40.7	26.5
NRS-West	47.7	26.5
SRS	38.3	26.5

[a] NRS-East: the eastern portion of the Northern Research Station which is the area covered by the former Northeastern Research Station, NRS-West: the western portion of the Northern Research Station which is the area covered by the former North Central Research Station, and SRS: Southern Research Station..

Component Ratio Method (CRM)

For trees 5 inches in diameter and greater, total aboveground biomass is computed as the sum of three components: bole of the tree, tops and limbs, and the stump. Bole biomass is the largest portion of aboveground biomass. Stumps, tops and limbs, and saplings are a significant yet much smaller portion of the total aboveground biomass in most forests. The biomass of saplings, that is, trees less than 5 inches in diameter but greater than or equal to 1-inch diameter, are based on an adjustment of Jenkins and others (2003) equations because they have zero volume. Belowground biomass, that is, biomass of coarse roots, is predicted as a ratio of aboveground tree biomass. We present the details of these calculations by section below.

Aboveground Biomass of Trees ≥ 5 inches d.b.h., Dry Weight: Total aboveground biomass (dry weight) of trees greater than or equal to 5 inches d.b.h. is calculated as the sum of three components of the tree:

$$AGBIOT5 = DRYBIO_BOLE + DRYBIO_STUMP + DRYBIO_TOP \quad [1]$$

where AGBIOT5 (lbs) = total aboveground biomass (dry weight), including bark but excluding foliage, of a tree ≥ 5 inches d.b.h.,

DRYBIO_BOLE (lbs) = biomass (dry weight, including wood and bark) of the main stem of tree that also defines sound volume,

DRYBIO_TOP (lbs) = biomass of top and limbs (dry weight, excluding foliage but including bark) of trees ≥ 5 inches d.b.h., and,

DRYBIO_STUMP (lbs) = biomass (dry weight) of wood and bark from ground level to 1 foot stump.

Biomass of the Bole: Biomass of the bole of a particular species is calculated by multiplying green volume (cu ft) by the weight of one cubic foot of water (62.4 lbs/cu ft) to convert to a weight basis, and then multiplying by the specific gravity

of the component, including wood and bark, separately, for that species. Specific gravities are obtained by laboratory studies, and the results compiled by species where available, or assigned to similar species. Here, specific gravities in terms of dry weight per unit of green volume are used. Because the specific gravity is different for bark and wood, these two components are calculated separately and then summed. The calculation for bark has an additional term, bark as a proportion of wood volume. Bark volume in terms of percentage is given in Table 2; proportions are equal to percent divided by 100. Specific gravities used in this study are from Miles and Smith,[6] which build on the compilations by Smith (1991) and Jenkins and others (2004).

DRYBIO_BOLE (lbs) = (VOLCFSND x 62.4 x SG_BARK x BRK_VOL_PROP)
+ (VOLCFSND x 62.4 X SG_WOOD) [2]

with DRYBIO_BOLE (lbs) = biomass (dry weight, including wood and bark) of the main stem of tree that also defines sound green volume (VOLCFSND),

VOLCFSND (cu ft) = sound green wood volume of a tree ≥ 5 inches d.b.h.,

SG_BARK = dry weight specific gravity of green bark volume of tree bole,

BRK_VOL_PROP = ratio of green volume of bark to green volume of sound wood (see Table 2), and,

SG_WOOD = dry weight specific gravity of green wood volume of tree bole.

Table 2: Bark as a percent of wood volume by Jenkins and others (2003) species groups.

Jenkins groups[a]	(Bark as % of wood volume)	Jenkins groups[a]	(Bark as % of wood volume)
Aspen/alder/ cottonwood/willow	20	Cedar/larch	15
Soft maple/birch	14	Douglas-fir	14
Mixed hardwood	18	Pine	18
Hard maple/oak/ hickory/beech	19	Spruce	12
Woodland	12	True fir/hemlock	15

[a]Exceptions to these species groups are coastal redwood, giant sequoia, baldcypress, eastern, western, and Carolina hemlock – 25%; beech, sycamore – 7%

Factors for Calculating Top & Limbs and Stump Biomass: The biomass in the stumps and tops and limbs of large trees make up the next largest components of aboveground biomass in most forests. Jenkins and others (2003) provide equations that calculate total aboveground biomass, and also a set of equations that estimate the proportion of biomass in the tops and limbs as well as other tree components. The CRM uses the component equations from Jenkins and others (2003) to compute the ratio of the component to Jenkins total aboveground biomass. These ratios are then multiplied by the bole biomass calculated in [2] using the CRM approach to produce the biomass in tops and limbs (DRYBIO_TOP), and biomass of the stump (DRYBIO_STUMP). We cannot apply those equations directly because the value of our biomass bole is not the same.

[6] Miles, Patrick, and W.B Smith. In review. Wood and bark specific gravity for tree species in the continental United States. USDA Forest Service, Northern Research Station. Research Note.

Applying these component ratios to biomass equations other than Jenkins and others (2003) results in different absolute biomass estimates than would be produced by Jenkins and others (2003). However, we assume that the proportions are the same, and thus we calculate these factors to multiply by our bole biomass. Equation 3 shows the formula for calculating the proportion in tops and limbs. Stump biomass in Jenkins and others (2003) is based on stump volume equations from Raile (1982). Equations 4 and 5, respectively, show the computations for factors related to stumps.

TPLMB_PROP = ((BIO_TOP_JENKINS)/(DRYBIO_BOLE_JENKINS)) [3]

where TPLMB_PROP = proportion of bole biomass that is biomass in top and limbs,
BIO_TOP_JENKINS (kg) = biomass in top and limbs using Jenkins, and,
DRYBIO_BOLE_JENKINS (kg) = biomass of the bole based on Jenkins.

DRYBIO_STUMP_RAILE (kg) = d.b.h (inch)*d.b.h (inch)*ParameterB [4]
where DRYBIO_STUMP_RAILE (kg) = stump biomass, and,
Parameter B = coefficient from Table 1 in Raile (1982).

STUMP_PROP = ((DRYBIO_STUMP_RAILE(kg))
/(DRYBIO_BOLE_JENKINS(kg))) [5]
where STUMP_PROP = proportion of bole that is stump biomass,
DRYBIO_STUMP_RAILE (kg) = biomass in stump, and,
DRYBIO_BOLE_JENKINS (kg) = biomass of the bole using Jenkins.

Top and Limb Biomass: Equation 6 shows the computation for estimating biomass in the top and limbs.

DRYBIO_TOP = DRYBIO_BOLE x TPLMB_PROP [6]

where DRYBIO_TOP (lbs) = biomass of top and limbs (dry weight, excluding foliage but including bark) of trees ≥ 5 inches d.b.h., and,
DRYBIO_BOLE (lbs) = biomass (dry weight, including wood and bark) of the main stem of tree that also defines sound volume (VOLCFSND), and,
TPLMB_PROP = proportion of bole biomass that is biomass in top-limbs.

Stump Biomass: Equation 7 shows the computation for stump biomass.

DRYBIO_STUMP = DRYBIO_BOLE x STUMP_PROP [7]
where DRYBIO_STUMP (lbs) = biomass (dry weight) of wood and bark from ground to 1 foot stump,
DRYBIO_BOLE (lbs) = biomass (dry weight, including wood and bark) of the main stem of tree that also defines sound volume, and,
STUMP_PROP = proportion of bole that is stump biomass, see above.

As FIA converts to a taper-based system to predict sound volume, it will be possible to directly calculate the sound wood volume in any section of the bole, including the stump. Thus, the separate stump calculation based on Raile (1982) will not be needed.

Aboveground Saplings: The biomass of saplings is based on biomass computed from Jenkins and others (2003) on the observed diameter multiplied by an adjustment factor. For the purposes of this preliminary investigation, the adjustment factor was computed as a national average ratio of the CRM total biomass divided by the Jenkins total biomass for all 5-inch trees, which is the size at which biomass based on volume begins. Each species group has an adjustment factor, which is given in Table 3. Computations are shown in Equation 8.

DRYBIO_SAPLING = (BIO_SAP_JENKINS - FOLIAGE) X (1-
JENKINS_SAPLING_ADJUSTMENT)) [8]

with DRYBIO_SAPLING (lbs) = aboveground biomass of trees < 5 inches d.b.h and ≥ 1.0 inch d.b.h., including wood, bark, and stump, but excluding foliage,
BIO_SAP_JENKINS (lbs) = aboveground biomass calculated using Jenkins and others (2003), converted to pounds
FOLIAGE (lbs) = dry weight of foliage from Jenkins and others (2003) converted to pounds, needed to subtract off foliage,
JENKINS_SAPLING_ADJUSTMENT = factor that adjusts Jenkins biomass for trees < 5 inches d.b.h for a smooth transition at 5-inch trees (see Table 3).

As with the stump biomass (DRYBIO_STUMP), when taper equations are available for volume estimation in all FIA units, it will be possible to calculate the central stem component of sapling biomass if the taper equations have been fit to datasets that include an adequate sample of smaller trees. This may prove to provide a better prediction of biomass in sapling size trees.

Belowground (Root) Biomass: Equation 9 shows the computation for coarse root biomass.

DRYBIO_BG = DRYBIO_BOLE x ROOT_PROP [9]

where DRYBIO_BG (lbs) = biomass of coarse roots,
DRYBIO_BOLE (lbs) = as above,
ROOT_PROP = ((ROOT _JENKINS)/(DRYBIO_BOLE_JENKINS)), which is the proportion of bole biomass to biomass in coarse roots,
ROOT_JENKINS (kg) = biomass in roots calculated using Jenkins, and,
DRYBIO_BOLE_JENKINS (kg) = biomass of the bole based on Jenkins.

Table 3: Adjustment factors applied to Jenkins and others (2003) sapling equations for the component ratio method.

FIA species codes	Jenkins sapling adjustment	Common name
58-60,62,63,65,66,69,106,133,134,140,141,143	0.352	Juniper, pinyon
745,747-749	0.378	Cottonwood
211	0.410	Redwood
116,122,135	0.434	Ponderosa pine
92,93,96	0.442	Spruce
41,42,50-55,64,72,101-104, 109,112-114,118, 120,124, 127,137-139,142,201,212, 231,251, 264, 299	0.458	Pines, other conifers
98	0.463	Sitka spruce
202	0.526	Douglas-fir
117	0.557	Sugar pine
119	0.574	Western white
81	0.588	Incense-cedar
11,14,15,17-22	0.602	Fir
10,12,90,94,95,97	0.608	Balsam fir, spruce
260-262	0.628	Eastern hemlocks
16,40,43,56,57,61,67,68,70,71,91,100,136,144, 200, 220, 230,232,240,241,252	0.631	Various conifers
531	0.632	American beech
105	0.643	Jack pine
300,321,322,475,755-758, 803,810,814,829,843,846, 847,850,81,902,990	0.651	Variety, woodland species
263	0.671	Western hemlock
950-953	0.672	Basswood
740-744, 746,752,753	0.691	Cottonwood
242	0.705	Western redcedar
822,832,835,836,838,840, 841,844	0.722	Oaks
125,129	0.729	Red pine, white pine
400-413,316,317	0.744	Hickory, misc. maples
611	0.749	Sweetgum
351	0.750	Red alder
110,131	0.763	Shortleaf, loblolly pine
802,804,808,823,825,826	0.770	White oaks
801,805,807,811,815,818, 821,839	0.774	Western oaks
806,809,812,817,820,824,827,828,830,831,837	0.780	Black oaks
371	0.789	Yellow birch
313,331,332,334,337,350,355,370,373,375,377, 379,422,452,460-463,555,580-583,600,601,605, 650-653, 655,657,658,712,729,731,762,911, 912,915,922,924,927,928,929,931,970-976,992	0.792	Mixed hardwoods
73	0.800	Western larch
813,833,834	0.811	Oaks
All other species not listed elsewhere	0.840	All other species
310,311,314,318,320,323, 690, 691,693, 694	0.841	Maples, tupelo
621	0.852	Yellow-poplar
602	0.872	Black walnut
108	0.883	Lodgepole pine
111,121	0.922	Slash, longleaf pine
372,450,491,510,513,521, 550, 551,552, 571,591,680- 683,800,858,901,977	0.932	Various hardwoods
541,543-546,548,549	0.936	Ash
221-223	0.952	Taxodium
312,330,333,352,353,361-363,374,378,431,492,511, 542,547,603,604,606,631, 661,730,732,768,981	0.964	Various hardwoods
107,115,123,126,128,130, 132	1.011	Various pines

Example

We present an example for using Jenkins and others (2003) and CRM for estimating biomass for a 4-inch tree and 25-inch tree, red oaks in the NRS-East region. For the 25-inch CRM, we use the volume calculated in the FIADB because we do not have the volume equations readily available. Mass is in terms of dry weight.

Information needed for the Jenkins estimates:

- The species and d.b.h for the individual tree.
- Red oak, species code 833, is in the "hard maple/oak/hickory/beech" group (mo). Therefore the paired coefficients (b0,b1) needed for estimates are: (-2.0127, 2.4342) for total aboveground biomass, (-4.0813, 5.8816) for foliage, and (-1.6911, 0.816) for the coarse root component.
- Metric-English conversions for length and mass: 1 inch equals 2.54 centimeters, and 1 kilogram equals 2.2046 pounds.

For a 25-inch d.b.h tree excluding foliage, aboveground biomass is based on deducting foliage from the aboveground-biomass equation:

- Total aboveground biomass = exp(b0 + b1 × ln(d.b.h)) = exp(-2.0127 + 2.4342 × ln(25×2.54)) = 3267.4 kg dry weight = 3267.7 × 2.2046 = 7203.4 pounds.
- Foliage component ratio = exp(b0 + b1/d.b.h) = exp(-4.0813 + 5.8816/(25×2.54)) = 0.01852. Thus, foliage biomass (total × component) = 7203.4 × 0.01852 = 133.4 pounds.
- Therefore, the aboveground biomass excluding foliage for a 25-inch d.b.h tree is the difference: 7203.4 - 133.4 = 7070 pounds.

Similarly, for a 4-inch d.b.h tree

- Total aboveground biomass = exp(b0 + b1 × ln(d.b.h)) = exp(-2.0127 + 2.4342 × ln(4×2.54)) = 37.75 kg = 37.75 × 2.2046 = 83.22 pounds.
- Foliage component ratio = exp(b0 + b1/d.b.h) = exp(-4.0813 + 5.8816/(4×2.54)) = 0.03013. Thus, foliage biomass (total × component) = 83.22 × 0. 03013 = 2.507 pounds.
- Therefore, the aboveground biomass excluding foliage for a 4-inch d.b.h tree is the difference: 83.22 - 2.507 = 80.71 pounds.

Belowground, or coarse root, biomass for the 25- and 4-inch d.b.h trees are based on total aboveground biomass and the coarse root component:

- Coarse root component = exp(b0 + b1/d.b.h) = exp(-1.6911 + 0.816/(25×2.54) = 0.1867 for a 25-inch d.b.h tree and exp(-1.6911 + 0.816/(4×2.54) = 0.1997 for a 4-inch d.b.h tree.
- Therefore, the belowground biomass for a 25-inch d.b.h tree is the product: 7203.4 × 0.1867 = 1345 pounds.
- Therefore, the belowground dry weight for a 4-inch d.b.h tree is the product: 83.22 × 0.1997 = 16.62 pounds.

Information needed for the CRM estimates:

- The species and volume of sound wood (VOLCFSND) from the FIADB.
- The set of Jenkins biomass and component coefficients for red oak, as provided above plus the additional paired coefficients (b0,b1): (-2.0129, -1.6805) for the stem bark component, (-0.3065, -5.424) for the stem wood component, and (-1.6911, 0.816) for the coarse root component.
- Five additional species-specific factors: (1) the ratio of volume of bark to volume of wood (aka BARK_VOLUME_PROP); (2) specific gravity of wood (aka SG_WOOD); (3) specific gravity of bark (aka SG_BARK); (4) a factor for estimating stump biomass based on Raile (1982, aka RAILE_STUMP_B1); (5) an adjustment factor applicable to trees less than 5 inches d.b.h (JENKINS_SAPLING_ADJUSTMENT).
- Most calculations are in English units. However, the metric-to-English conversion of 1 kg equals 2.2046 pounds may be necessary if stump biomass units are in pounds. The density of water is 62.4 pounds per cubic foot.

For a 25-inch d.b.h tree excluding foliage, aboveground biomass is based on determining merchantable biomass and then expanding according to the top and stump component ratios:

- Merchantable biomass, or biomass of the bole = VOLCFSND × (BARK_VOLUME_PROP × SG_BARK × density of water) + VOLCFSND × (SG_WOOD × density of water) = 103.04 × (0.19 × 0.65 × 62.4) + 103.04 × (0.56 × 62.4) = 794.07 + 3600.63 = 4394.8 pounds.
- Note that volume equations are needed, as well as information on defects and form to calculate sound volume; we do not include that here as this information is not readily available.
- Top component ratio = (Jenkins total aboveground – Jenkins merchantable – Raile stump – Jenkins foliage) / (Jenkins merchantable) = (7203.4 - 5805.0 – 176.3 - 133.4) / (5805.0) = 0.1875.
  - Where Jenkins merchantable = (Jenkins total aboveground) × (Jenkins stem bark component + Jenkins wood component) = (7203.4) × (exp(-2.0129 + -1.6805/(25×2.54)) + exp(-0.3065 + -5.424/(25×2.54))) = 7203.4 × (0.1301 + 0.6758) = 7203.4 × 0.8059 = 5805.0 pounds.
  - Where Raile stump = DIA × DIA × RAILE_STUMP_B1 = 25×25 × 0.12798 = 79.99 kg x 2.2046 lbs/kg = 176.3 pounds.
- Stump component ratio = Raile stump / Jenkins merchantable = 80.0 / 5805.0 = 0.03038.
- Therefore, the aboveground biomass excluding foliage for a 25-inch d.b.h tree is = 4394.8 × (1 + top component ratio + stump component ratio) = 4394.8 × (1 + 0.1875 + 0.03038) = 5352.3 pounds.

Aboveground biomass, excluding foliage, for a 4-inch d.b.h tree = Jenkins aboveground biomass without foliage × JENKINS_SAPLING_ADJUSTMENT = 80.71 × 0.81068 = 65.43 pounds.

Belowground, or coarse root, biomass for the 25-inch d.b.h tree is based on merchantable biomass and the coarse root component ratio:

- Coarse root component ratio = (Jenkins coarse root) / (Jenkins merchantable) = 1344.9 / 5805.0 = 0.2317.
- Therefore, the belowground biomass for a 25-inch d.b.h tree is the product: 4394.8 × 0.2317 = 1018 pounds.

Belowground, or coarse root, biomass for a 4-inch d.b.h tree = Jenkins belowground biomass × JENKINS_SAPLING_ADJUSTMENT = 16.62 × 0.81068 = 13.47 pounds.

Results and Discussion

This approach was applied to all trees in all the annualized surveys in the FIABD3, using a preliminary set of specific gravities by species. In terms of merchantable biomass and percent in tops, limbs, and stumps, results in Table 4 indicate that the larger differences in Table 1 have been resolved. Figure 5 also indicates that other problems with tops, limbs, and stump can be resolved using the CRM approach. In particular, the regional percentages for PNWRS-softwoods and NRS-East hardwoods are quite different compared to the other regions, but the CRM-based results are more similar among units. The range in average top, limbs, and stump by unit is about 10 to 33 percent using the regional approaches, but the range is 16 (PNWRS-softwoods) to 27 (RMRS – hardwoods) percent using CRM. These results conform to what is expected given the nature of the predicted tops and limbs proportions (TPLMB_PROP) from Jenkins and others (2003) which predict the smallest proportions for large diameter softwoods and the largest proportions for small diameter hardwoods. The PNWRS has the largest average diameter of softwoods and the RMRS region has the smallest average diameter of hardwoods.

Table 4: Select red oak, 10-inch d.b.h., 60-70' height, growing stock trees,—CRM equations.

FIA region[a]	DRYBIOM/VOLCFSND	DRYBIOT-DRYBIOM/DRYBIOT
	--Pounds wood and bark per cubic feet of sound wood--	*--Percent of total biomass in tops, limbs, and stumps--*
NRS-East	42.7	26.5
NRS-West	42.7	26.5
SRS	43.6	26.5

[a] NRS-East: the eastern portion of the Northern Research Station which is the area covered by the former Northeastern Research Station, NRS-West: the western portion of the Northern Research Station which is the area covered by the former North Central Research Station, and SRS: Southern Research Station.

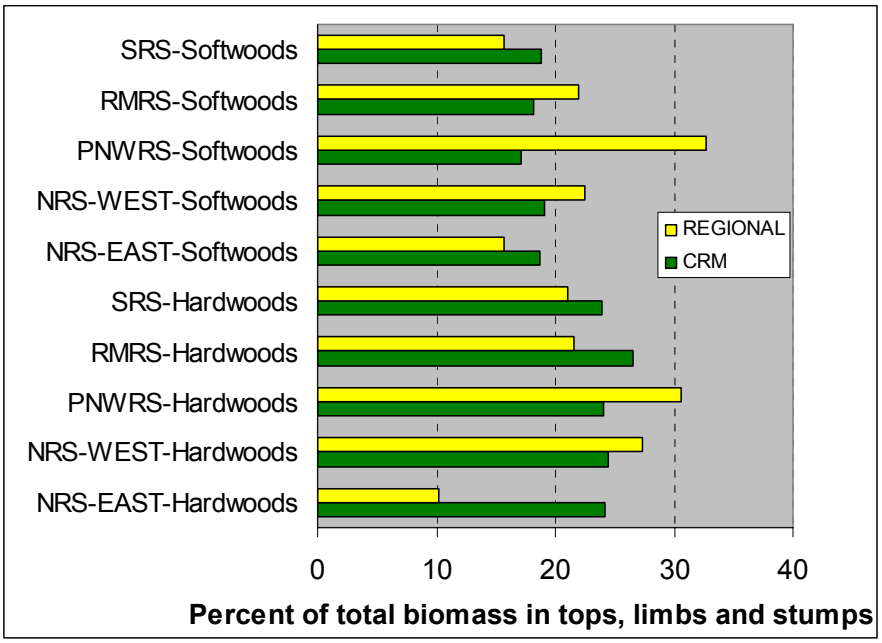


Figure 5: Percent of live tree biomass of all forest land trees ≥ 5 inches d.b.h. in tops, limbs, and stumps for the current regional approach and the proposed CRM, by softwood/hardwood and FIA unit. Abbreviations for unit designations are: SRS = Southern Research Station, RMRS = Rocky Mountain Research Station, PNWRS = Pacific Northwest Research Station, NRS = Northern Research Station where NRS-East = the former Northeastern Research Station and NRS-West = the former North Central Research Station. Note: This figure currently includes component ratios of woodland species; however, woodland species were not included in the Jenkins and others (2003) calculation of component ratios.

Figure 6 shows the overall effect that implementing either the CRM or Jenkins and others (2003) on a national basis would have on the total estimated biomass in trees 5 inches diameter and larger within each of the regions. Aboveground biomass estimates for trees ≥ 5 inches d.b.h on a tons per acre basis from Jenkins and others (2003) tend to be greater than estimates from the other approaches. The current regional approach for NRS-East is a compilation of biomass equations, including some that were used in the development of Jenkins and others (2003). The biomass prediction methods in NRS-West, RMRS and for many species in PNWRS[7] are based on volume predictions and are therefore very similar to the CRM. Thus, the CRM estimates in these regions are quite close to the regional estimates. The CRM approach reduces biomass densities even further compared to the current regional equations, with the exception of

[7] For example, for PNWRS, bole wood volume is predicted based on species, diameter, and height measurements, and a library of volume equations specific to the species or species groups and portion of region. These equations take on different forms, but many are based on the integration of taper equations such as Flewelling and Raynes (1993). Bole bark volume is computed from a variety of sources including methods that calculate inside bark and outside bark volume using equations from Pillsbury and Kirkley (1984). These are then multiplied by wood weight to get bark mass. These bole wood and bole bark volume predictions are converted to biomass using specific gravity estimates, many of which are in Table 3. Branch, top, and stump mass calculations come from a variety of sources, such as Snell and others (1983), Gholz and others (1979), and Cochran and others (1994). These components are added to obtain predictions of total aboveground biomass.

estimates for RMRS which are only slightly greater on a per acre basis. Note we are not presenting any validation evidence to prove that the CRM-based estimates are any more accurate than the regional estimates or Jenkins and others (2003).

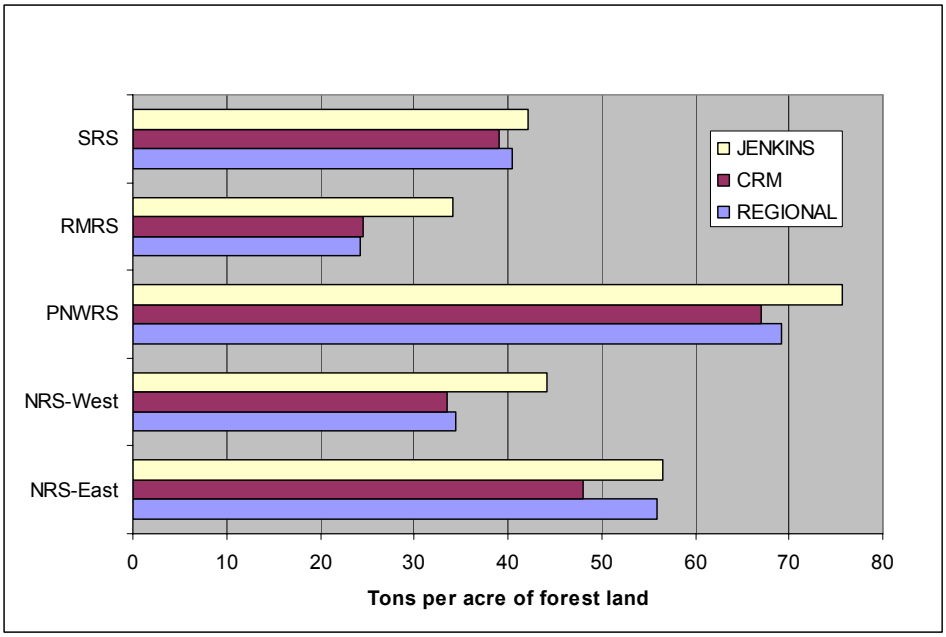


Figure 6: Aboveground live tree biomass per acre by approach by FIA unit, all trees ≥ 5 inch, conterminous United States.

In this preliminary investigation, we have not examined how the different equations affect change in biomass over time. If the older estimates are not updated, we will be comparing a smaller estimate based on the CRM method to a larger estimate from an older method. Even if the tree did not change in size, the trees would show a notable artificial loss in biomass if a new method were adopted before the older estimates could be updated.

The measurement of change in volume over time and the breakdown of this change into various components of change such as growth, removals and mortality has been a central part of FIA estimation. To produce these estimates it is vital that both old and new observations are based on the same prediction methods. Whenever a new volume estimation procedure is implemented in FIA, there has been a need to recalculate previous inventory methods. A volume based system such as CRM facilitates recalculations of biomass and the computation of biomass change into the standard FIA data processing system. Similarly, carbon change has been central in the GHG inventory estimates and the use of Jenkins and others (2003) equations. It is absolutely crucial to recalculate all the biomass data going back in time to ensure the change over time for biomass and carbon sequestration calculations is not simply due to a change in equations.

Figure 7 shows a comparison of Jenkins and others (2003) in terms of average total aboveground biomass by d.b.h. with both the FIA regional and CRM

methods for only one species group, alder/aspen/willow/cottonwood. This is probably the most wide-ranging group, growing throughout the conterminous United States. In the NRS-West, PNWRS, and RMRS regions the CRM and regional methods produce very similar results to the CRM. In those regions the regional method is based on bole volume, rather than on independent biomass equations as it is in NRS-East and SRS. The results from various FIA units bound the results from the Jenkins and others (2003) equations, illustrating how those equations effectively yield an average estimate composited from published equations across the United States. Note that differences in these results could be due to different equations used in the different regions, tree size, number of trees, and different species mix.

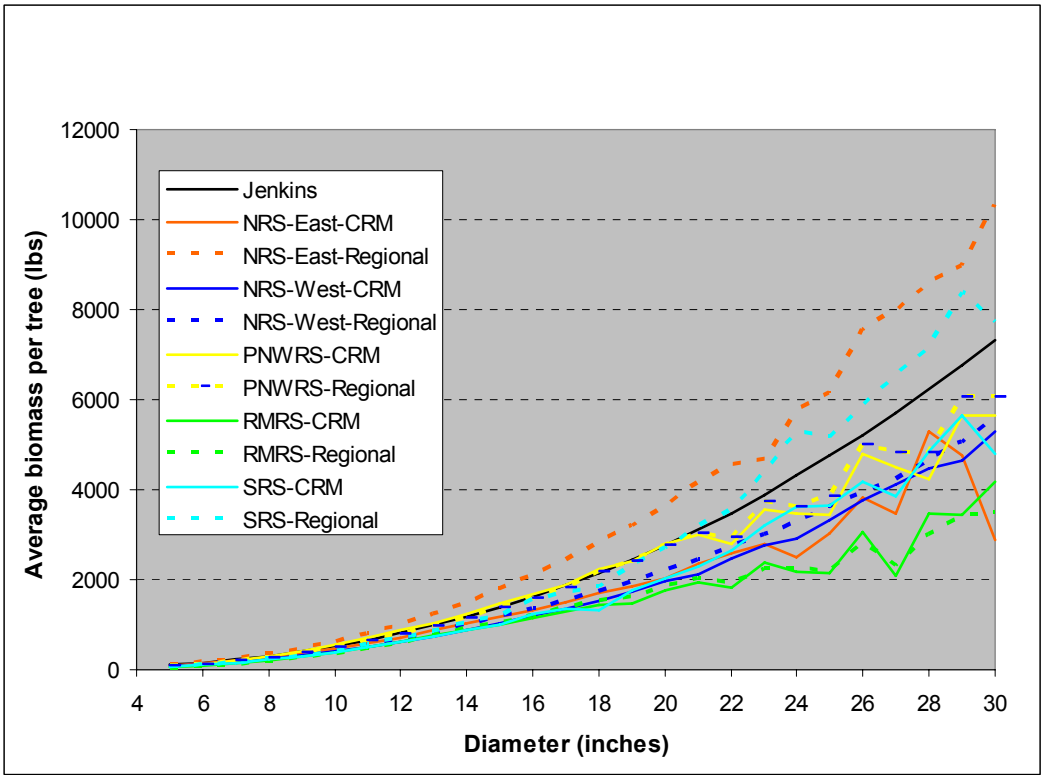


Figure 7: Average aboveground tree biomass for each FIA region based on both the current FIA regional and CRM methods by diameter, and the aboveground tree biomass from the Jenkins and others (2003) equation for the alder/aspen/willow/cottonwood species group. For the regional and CRM methods, all tree data of alder/aspen/willow cottonwood species from 137,701 trees (NRS-EAST 10,507 trees, NRS-WEST 81,812 trees, PNWRS 9,558, RMRS 31,498, and SRS 4,326) in FIADB3 from annualized surveys are used (biomass in lbs; diameter in inches)

Comparing Jenkins and others (2003) with the CRM approach, as shown in Figure 7, estimates from all the units are less than those from Jenkins and others (2003). Note that differences in these results could be due to different volume equations, which are the basis for the CRM, used in the different regions, tree size, number of trees, and different species mix. Although densities of wood and bark do not change at FIA unit boundaries, volume equations do (Hansen 2002),

and since biomass based on CRM is based on volume equations, biomass estimation still changes at FIA unit boundaries.

Several items in the CRM were identified in this preliminary investigation as needing further work to meet the Forest Survey Handbook standard of "high quality, consistent and reliable data." Component ratios for woodland species need to be derived. Stump equations need to be reworked to match specific gravity and bark estimates from the tables. The adjustment process to the Jenkins and others (2003) biomass predictions for use in CRM for diameters less than 5 inches needs further consideration. The implementation of a well designed national system for bole volume prediction that is based on taper equations would address these issues. Datasets for validation should be compiled, at least for major species to test the accuracy of the equations. For transparency, a full well-documented compilation of volume equations is needed for all FIA units and species. A complete set of specific gravities by species, well-documented and consistent with existing estimates in the published literature is needed. Such documentation will also meet the needs of users of FIA data, who sometimes collect their own inventory data, and would like to apply the same compilation procedures as FIA so they can compare their biomass and carbon results to FIA-plot results either for planning, double-checking, or verification. This is especially important if FIA data are used for carbon monitoring.

The CRM is based on the assumption that component ratios calculated in Jenkins and others (2003) can be accurately adopted and applied to the predictions of sound bole volume. That is, it is assumed the merchantability standards for a tree bole in Jenkins and others (2003) are the same merchantability standards for a bole measured for FIA volume. (For instance, top height is a standard, such as height to a 4-inch top.) This method allows the user to plug in any volume and convert it to a biomass estimate. An interesting hypothesis to test is whether it would be more accurate to predict aboveground biomass and then estimate volume as a proportion of biomass. Merchantability standards of volume have continued to change, but the definition of total biomass is has always included all biomass of the tree.

The urban tree biomass scientific community also has biomass equations for their estimates, and these equations and estimates should also be taken into consideration when adopting methods for forest biomass estimates. Biomass estimates for forest land that recently converted to urban land without loss of trees should be similar to urban forest biomass estimates. Otherwise, artificial changes will be induced. One example set of biomass equations created for urban tree biomass estimates is presented in Nowak and others (2002).

Conclusions and Recommendations

The CRM produced biomass estimates that feature nationally consistent specific gravities, and biomass consistent with volumes. However, because CRM

is based on volume equations which still differ by FIA unit, biomass estimates for the same species and diameter can differ by unit. If we had a consistent national volume approach, biomass based on volumes would be nationally consistent also. Additional items were identified as work that was needed to be done before biomass equations based on CRM were completed. A key assumption of this approach is that merchantable bole in Jenkins and others (2003) is equivalent in definition to the bole in the volume equations. The validity of this assumption should be further investigated. Biomass estimates in terms of tons per acre based on this approach were almost always less than the regionally based estimates and the Jenkins and others (2003) estimates. This is a curious result that may be worth investigating further.

There were additional research questions identified as a result of this preliminary analysis. Did the perceived over-prediction of biomass in the Jenkins and others (2003) biomass equations for larger trees, and lack of a deduction for damaged and standing dead trees, over-estimate total biomass? Are the small tree adjustments in CRM under-predicting biomass? Are there problems with the specific gravities used in the CRM? Would the biomass estimates be significantly different if specific gravity was based on samples from the field rather than using average compiled specific gravities?

The Jenkins and others (2003) biomass equations were developed and adopted for producing tree biomass carbon estimates because of regional differences in approaches by FIA units and database limitations. Adopting CRM immediately will hinder use of current U.S. Forest Service carbon estimates and tools based on the Jenkins and others (2003) equations, because consistency across time is critical. Because CRM is fundamentally based on volume, when volume estimates change, then biomass and carbon estimates based on the CRM will change. Since volume updates are planned in the near future in some regions, adopting CRM now means carbon estimates will change as the volume estimates are updated.

Adopting new approaches that are an improvement to existing protocols is inevitable and underway. A planned, coordinated, supported and funded national effort across FIA units and with other interested scientific experts to develop tree level volume, biomass, and carbon equations would increase the credibility and usefulness of the resulting biomass estimates, providing "high-quality, consistent, and reliable data." Ideally, for the long term, a several-year effort involving a team of scientists that allows for data mining of existing studies and data collection for validation data, sets selection criteria, works through inconsistencies, and garners support of our users will be well worth the investment. Equations for calculating tree biomass for carbon in urban forests, agroforestry systems, and perhaps subtropical and tropical forests of U.S. territories and biomass for bioenergy plantations could also be considered for inclusion in such a study.

Acknowledgments and Attributions

Pat Miles contributed the pre-publication database of specific gravity; Brad Smith contributed his idea of using a biomass expansion factor approach, and agreed to contribute the specific gravity database as well; Jim Smith developed and wrote the example and checked calculations; Mark Hansen provided the analysis for the component ratio method and figures and estimates, and substantial rewrite edits; Linda Heath conducted the literature review, developed the outline, wrote the first draft, dealt with reviewers, and takes full responsibility for the recommendations.

Literature Cited

Bechtold, William A.; Patterson, Paul L., eds. 2005. The enhanced forest inventory and analysis program - national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Brown, Sandra; Gillespie, Andrew J.R.; Lugo, Ariel E. 1989. Biomass estimation methods for tropical forests with applications to forest inventory data. Forest Science. 35(4): 881-902.

Cochran, P.H.; Jennings, J.W.; Youngberg, C.T. 1984. Biomass estimators for thinned second-growth ponderosa pine trees. Res. Note PNW-RN-415. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 6 p.

Flewelling, J.W.; Raynes, L.M., 1993. Variable-shape stem-profile predictions for western hemlock. Part I. Predictions from DBH and total height. Canadian Journal of Forest Research. 23: 520–536.

Gholz H. L.; Grier C. C.; Campbell A.G.; Brown A.T. 1979. Equations for estimating biomass and leaf area of plants in the Pacific Northwest. Research Paper 41. Corvallis, OR: Forest Research Laboratory, Oregon State University. 37 p.

Hahn, J. T.; Hansen, M. H. 1991. Cubic and board foot volume models for the Central States. Northern Journal of Applied Forestry. 8:47-57.

Hansen, Mark. 2002. Volume and biomass estimation in FIA: national consistency vs. regional accuracy. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; Moser, J.W., eds. Proceedings of the third 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: 109-120.

Hoover, Coeli; Rebain, Stephanie. 2008. The Kane Experimental Forest carbon inventory: Carbon reporting with FVS. In: Havis, Robert N.; Crookston, Nicholas L., comps. 2008. Third forest vegetation simulator conference; 2007 February 13-15; Fort Collins, CO. Proc. RMRS-P-54. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: 17-22.

Heinz Center. 2002. State of the nation's ecosystems report. Cambridge, UK: Cambridge University Press. 288 p.

Heinz Center. 2005. Web only update. http://www.heinzctr.org/ecosystems/2002report/index.html (11 Dec 2008).

Heinz Center. 2008. The state of the nation's ecosystems 2008: measuring the land, waters, and living resources of the United States. Washington, DC: Island Press. 368 p.

Houghton J.T., Meira Filho L.G., Lim B., Treanton K., Mamaty I., Bonduki Y., Griggs D.J.; Callander B.A., eds. 1997. Revised 1996 intergovernmental panel on climate change guidelines for national greenhouse inventories. Paris: IPCC/OECD/IEA. (see page 5.18)

Jenkins, J.C.; Chojnacky, D.C.; Heath, L.S.; Birdsey, R.A. 2003. National scale biomass estimators for United States tree species. Forest Science. 49(1): 12-35.

Jenkins, J.; Chojnacky, D.; Heath, L.; Birdsey, R. 2004. Comprehensive database of diameter-based biomass regressions for North American tree species. Gen. Tech. Rep. NE-319. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station.

Lambert, M.-C.; Ung, C.-H.; Raulier, F. 2005. Canadian national tree aboveground biomass equations. Canadian Journal of Forest Research. 35: 1996-2018.

Muukkonen, P. 2007. Generalized allometric volume and biomass equations for some tree species in Europe. European Journal of Forest Research. 126: 157-166.

NCASI, Inc. 2008. COLE: Carbon OnLine Estimator, Version 2.0. Available: http://ncasi.uml.edu/COLE [11 December 2008].

Nowak, David J.; Crane, Daniel E.; Stevens, Jack C.; Ibarra, Myriam. 2002. Brooklyn's urban forest. Gen. Tech. Rep. NE-290. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 107 p.

Office of Management and Budget. 2004. Final information quality bulletin for peer review..Available: http://www.whitehouse.gov/omb/inforeg/peer2004/peer_bulletin.pdf [11 Dec 2008].

Pearson, Timothy R.H.; Brown, Sandra L.; Birdsey, Richard A. 2007. Measurement guidelines for the sequestration of forest carbon. Gen. Tech. Rep. NRS-18. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 42 p.

Penman J., Gytarsky M., Hiraishi T., Krug, T., Kruger D., Pipatti R., Buendia L., Miwa K., Ngara T., Tanabe K., Wagner F., eds. 2003. Good practice guidance for land use, land-use change and forestry. Hayama, Japan: Intergovernmental Panel on Climate Change (IPCC), IPCC/IGES.

Pillsbury, N.H.; Kirkley, M.L.1984. Equations for total, wood, and saw-log volume for thirteen California hardwoods. Res. Note PNW-RN-414. Portland, OR: U.S. Dept. of Agriculture, Forest Service, Pacific Northwest Research Station.52 p.

Potter, Christopher; Gross, Peggy; Klooster, Steven; Fladeland, Matthew; Genovese, Vanessa. 2008. Storage of carbon in U.S. forests predicted from satellite data, ecosystem modeling, and inventory summaries. Climatic Change. 90(3): 269-282.

Raile, G.K. 1982. Estimating stump volume. Res. Pap. NC-224. St Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 4 p.

Schroeder, P.; Brown, S.; Mo, J.; Birdsey, R.; Cieszewski, C. 1997. Biomass estimation for temperate broadleaf forests of the United States using inventory data. Forest Science. 43(3): 424-434.

Somogyi, Z.; Cienciala, E.; Mäkipää, R.; Muukkonen, P.; Lohtonen, A.; Weiss, P. 2007. Indirect methods of large-scale forest biomass estimation. European Journal of Forest Research. 126: 197-207.

Smith, J.E.; Heath, L.S.; Jenkins, J.C. 2003. Forest volume-to-biomass models and estimates of mass for live and standing dead trees of U.S. forests. Gen. Tech. Rep. NE-298. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 57 p.

Smith, James E.; Heath, Linda S., Skog, K.E.; Birdsey, Richard A. 2006. Methods for calculating forest ecosystem and harvested carbon with standard estimates for forest

types of the United States. Gen. Tech. Rep. NE-343. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 216 p.

Smith, J.E.; Heath, L.S.; Nichols, M.C. 2007. U.S. forest carbon calculation tool: forestland carbon stocks and net annual stock change. Gen. Tech. Rep. NRS-13. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. [CD-ROM]. Available: http://nrs.fs.fed.us/pubs/2394 [11 December 2008].

Smith, W. Brad. 1991. Assessing removals for North Central forest inventories. Res. Pap. NC-299. St Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 48 p.

Snell, J.A. Kendall; Little, S.N. 1983. Predicting crown weight and bole volume of five Western hardwoods. Gen. Tech. Rep. PNW-GTR-151. Portland, OR: U.S. Dept. of Agriculture, Forest Service, Pacific Northwest Research Station. 42 p.

Ung, C-H; Bernier, P.; Guo, X-J. 2008. Canadian national biomass equations: new parameter estimates that include British Columbia data. Canadian Journal of Forest Research. 38: 1123-1132.

U.S. Department of Agriculture. 2008. U.S. agriculture and forestry greenhouse Gas inventory: 1990-2005. Tech. Bull. No. 1921. 161 p.

U.S. Department of Agriculture, Forest Service. 2008. FIA data and tools. http://fia.fs.fed.us/tools-data/default.asp. [19 Nov 2008].

U.S. Department of Agriculture, Forest Service. 2007. USDA Forest Service strategic plan: FY 2007-2012. FS-880. Washington, DC. 32 p. Available: http://www.fs.fed.us/publications/strategic/fs-sp-fy07-12.pdf [11 Dec 2008].

U.S. Department of Agriculture, Forest Service. 2008. The forest inventory and analysis database: database description and users manual version 3.0 for phase 2, revision 1. Washington,DC: U.S. Department of Agriculture, Forest Service, Washington Office. Available: http://www.fia.fs.fed.us/library/database-documentation/FIADB_user_manual_v3-0.pdf.

U.S. Department of Agriculture, Forest Service. 2008. Tools for carbon inventory, management, and reporting. http://www.nrs.fs.fed.us/carbon/tools/ [11 December].

U.S. Department of Energy. 2006. Methods for calculating forest ecosystem and harvested carbon with standard estimates for forest types of the United States. In: Section 1, Part I; forestry appendix to chapter 1, emission inventory guidelines, technical guidelines for voluntary reporting of greenhouse gas program: 2-227:Available: http://www.pi.energy.gov/enhancingGHGregistry/documents/PartIForestryAppendix.pdf (11 December 2008).

U.S. Environmental Protection Agency. 2008. Inventory of U.S. greenhouse gas emissions and sinks: 1990–2006. EPA 430-R-08-005. Washington, DC: U.S. Environmental Protection Agency, Office of Atmospheric Programs. Available: http://epa.gov/climatechange/emissions/usinventoryreport.html [9 May 2008].

Wirth, C.; Schumacher, J.; Schulze, E-D. 2004. Generic biomass functions for Norway spruce in Central Europe—a meta-analysis approach toward prediction and uncertainty estimation. Tree Physiology. 24: 121-139.

Estimating Diesel Fuel Consumption and Carbon Dioxide Emissions from Forest Road Construction

Dan Loeffler[1], Greg Jones[2], Nikolaus Vonessen[3], Sean Healey[4], Woodam Chung[5]

Abstract: *Forest access road construction is a necessary component of many on-the-ground forest vegetation treatment projects. However, the fuel energy requirements and associated carbon dioxide emissions from forest road construction are unknown. We present a method for estimating diesel fuel consumed and related carbon dioxide emissions from constructing forest roads using published results from a study designed to measure road construction costs together with machine productivity and fuel consumption rates. Our resulting estimate of diesel fuel required per mile of road constructed on slopes up to 50% using a cut-fill construction method is 590 gallons, with 13,400 pounds of carbon dioxide emitted per mile of road built. Using a full bench road construction method on slopes greater than 50% where volume of material handled and moved is very sensitive to hill slope and soil type, we estimated between 3,265 and 8,000 gallons of diesel fuel are required per mile of road emitting between 74,400 to 182,700 pounds of carbon dioxide.*

Keywords: forest roads, carbon, carbon accounting, forest management, road construction, forest products, diesel emissions

Introduction

In 2007 the Chief of the US Forest Service outlined three ways in which forests, including national forest system lands, can be used to address climate change. The first is to manage forests in ways that make them more resistant to fires, insects, and disease resulting in more resilient forest stands. Second, the

[1] College of Forestry and Conservation, The University of Montana, drloeffler@fs.fed.us
[2] USDA Forest Service Rocky Mountain Research Station, jgjones@fs.fed.us
[3] Department of Mathematical Sciences, The University of Montana, Nikolaus.Vonessen@umontana.edu
[4] USDA Forest Service Rocky Mountain Research Station, seanhealey@fs.fed.us
[5] College of Forestry and Conservation, The University of Montana, Woodam.Chung@umontana.edu

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Forest Service should reduce its own carbon footprint, which includes generating more heat from woody biomass, a renewable source of energy that offsets the use of non-renewable fossil fuels. And third is to use the nation's forests to reduce the buildup of greenhouse gases, with support for carbon markets that ultimately convert forests into a carbon sink (Kimball 2007). All three of these proposals share a common theme, as stated by the Chief: "…protecting the existing carbon sink through forest conservation and increasing carbon sequestration through reforesting degraded land, improving forest health, and supporting sustainable forest management" (Kimball 2007). Accomplishing almost any aspect of this agenda will require some form of on-the-ground wood fiber removal.

While carbon storage in forest products is viewed as a means to defer disturbance-related emissions (Skog and Nicholson, 2000), the forest operations enabling this deferral almost always involve the release of fossil carbon, including harvesting and hauling products, and constructing the forest roads, either temporary or permanent, over which raw products are initially hauled. Healey and others (this volume) used historical harvest records and some assumptions about product carbon dynamics to calculate the magnitude and timing of carbon sequestration related to harvesting in Ravalli County, Montana. Healey and others also digitized a county-wide visual assessment of new roads apparent in sequential Landsat satellite imagery. These new-road maps, used together with spatially co-registered slope data, will in the future provide an application for the forest road construction emission factors discussed here. While forest operations release fossil carbon, little attention has been devoted to measuring carbon emissions associated with the various aspects of forest operations and forest products procurement.

The literature discussing carbon accounting methods and guidelines for harvested wood product flows between carbon pools is well established (Birdsey 2006; IPCC 2003, 2006). The Consortium for Research on Renewable Industrial Materials modeled fossil fuel consumption for stump-to-truck harvesting of wood products in the US Northwest and Southeast (CORRIM 2006) and Markewitz (2006) provides a detailed methodology for tracking fossil fuel consumption during silvicultural activities. However, the literature directly discussing the fossil fuel requirements and related emissions to construct the forest roads over which forest operations equipment travel is extremely sparse. A small portion of the forest road construction literature is financially-based, but contains little or no information about the fuel consumption underlying road construction costs (Balcom 1988; Layton and others 1992; Erickson and others 1992; USFS 2007).

Because forest access road construction is a necessary and critical component of many forest vegetation treatment projects, the range of fuel energy requirements and associated emissions from road construction are needed for an accurate carbon accounting of forest vegetation treatments. Here we present a methodology for estimating diesel fuel consumption and corresponding carbon

dioxide emissions associated with building forest roads, and discuss its benefits and limitations.

Methods and Results

There are two common ways to construct a forest road in mountainous terrain – the cut-fill method and the full bench method. The cut-fill method is used on gentle to moderate hill slopes ranging from approximately 0% – 50% and the full bench method is employed for steeper slopes. Using the cut-fill method, the builder would *cut* into hillside approximately half of the total road width, and then use the material removed by that cut as *fill* to construct the remaining half of the road on the downhill slope (Figure 1). When the full bench method is employed, the builder cuts into the hillside the entire width of the road, essentially creating a *bench* in the hill serving as the base of the road (Figure 2). Historically bulldozers have been the primary equipment used to construct forest roads. However, according to Forest Service Northern Region engineers most forest road building contractors switched to using hydraulic excavators in the mid to late 1980's (pers. comm. Rich Raines 3 October 2008; pers. comm. Marcia Hughey 15 October 2008). Excavators were found to be much more versatile and efficient for building forest roads, able to incorporate all aspects of road building into one single pass (Balcom 1988).

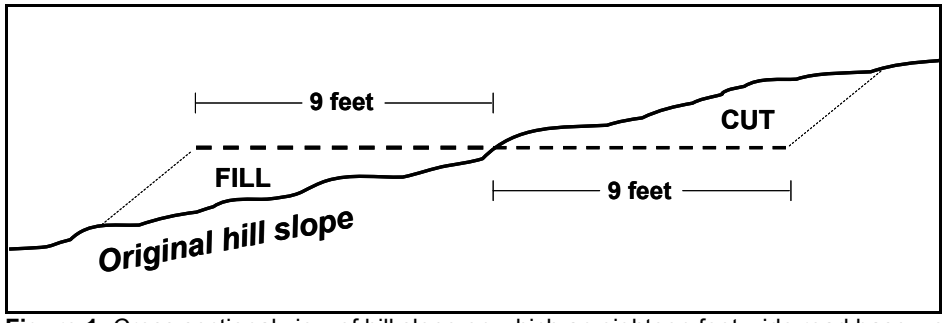


Figure 1: Cross sectional view of hill slope on which an eighteen feet wide road base would be constructed with the cut-fill method.

In this paper we have derived estimates of fuel consumption and resulting carbon dioxide emissions from building forest roads using information from Balcom (1988) and the Caterpillar Performance Handbook (CAT 1989, 2007). We estimate fuel consumption and emissions for each of the following major road building activities for both the cut-fill and full bench methods: 1) pioneering, 2) clearing and grubbing, and 3) sub-grade excavation. However, the following methodology is not limited to using the results from our selected sources. Rather, we have conceptualized a framework for estimating fuel consumption and emissions that is not limited to our selected sources.

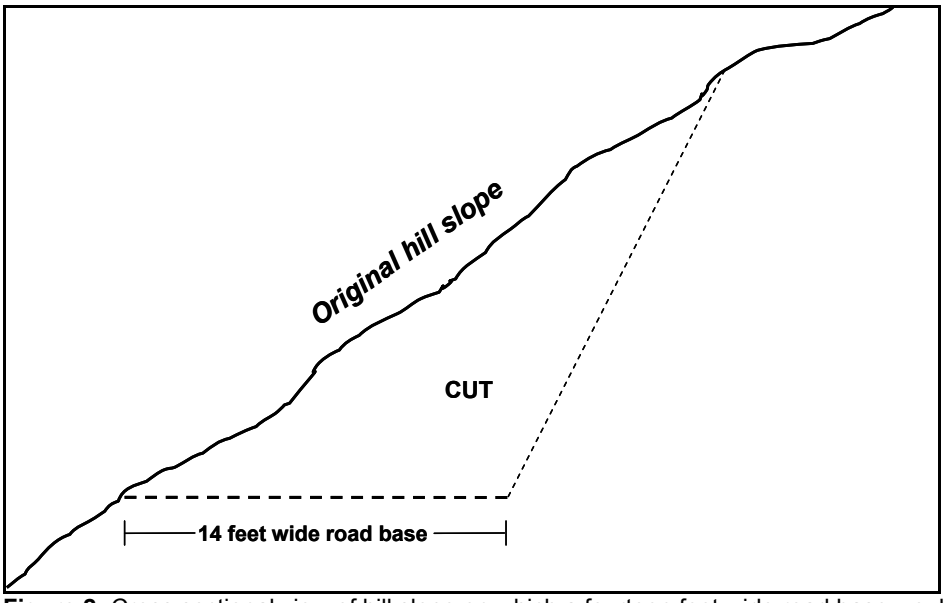


Figure 2: Cross sectional view of hill slope on which a fourteen feet wide road base would be constructed with the full bench method.

Cut-Fill Road Construction

Balcom (1988) conducted a time-motion study of forest road construction costs in Oregon using both crawler tractors and hydraulic excavators. This time-motion study provides results of feet per hour for building roads using the cut-fill method on hill slopes up to 50%. Although several types of machines were analyzed by Balcom, we isolated the results from the Caterpillar 235 hydraulic excavator and acquired that machine's fuel consumption rate of 8 gallons per hour from the 1989 Caterpillar Performance Handbook (CAT 1989) assuming 72% utilization, the midpoint in the utilization range listed for forestry operations. Average construction rates in linear feet per hour reported by Balcom were used to estimate pioneering, clearing and grubbing, and sub-grade excavation. Table 1 displays the average production per hour and fuel consumption per linear foot of road constructed for each of the three major road building activities listed above. Our estimates show that using a hydraulic excavator to construct forest roads with the cut-fill method on gentle slopes consumes approximately 0.11140 gallon of diesel per linear foot of road constructed, or approximately 590 gallons per mile.

We then combined the diesel fuel consumption estimate with emissions data for internal combustion diesel engines reported by the US Environmental Protection Agency (EPA 1995) and diesel energy content reported by the US Energy Information Administration (EIA 2008), which resulted in a carbon dioxide emission factor of 22.796 pounds per gallon of diesel. Combining this with molecular weights for carbon (12) and oxygen (16) yields the carbon dioxide and carbon equivalent results displayed in Table 2.

Table 1: Diesel fuel consumption per linear foot of forest road construction using the cut-fill method on slopes less than 50% for eighteen feet wide roads.

Road construction activity	Production (feet/hour)[a]	Diesel consumption (gallon/foot)	Diesel consumption (gallons/mile)
Pioneering	582	0.01375	73
Clearing and grubbing	129.5	0.06178	326
Sub-grade excavation with sidecasting	223	0.03587	189
Total of all activities		0.11140	588

[a] From Balcom (1988)

Approximately 2.5 pounds of carbon dioxide are emitted from diesel fuel burned per linear foot of forest road constructed on slopes less than 50%, or 13,400 pounds per mile. The carbon equivalent using the carbon-to-carbon dioxide ratio of 12/44 equals roughly 0.7 pound of carbon per linear foot of road construction, or 3,650 pounds of carbon per mile.

Table 2: Carbon dioxide emissions per linear foot of forest road construction using the cut-fill method on slopes less than 50% for eighteen feet wide roads.

Road construction activity	Carbon dioxide emissions (pounds/foot)	Carbon dioxide emissions (pounds/mile)	Carbon equivalent (pound/foot)	Carbon equivalent (pounds/mile)
Pioneering	0.31345	1,655	0.08548	451
Clearing and grubbing	1.40834	7,436	0.38409	2,028
Sub-grade excavation	0.81769	4,317	0.22301	1,177
Total of all activities	2.53947	13,408	0.69258	3,657

Full Bench Road Construction

Contrary to gentle slopes for which the cut-fill method would be appropriate, the amount of material that needs to be handled and moved to construct a full bench road in steep terrain is very sensitive to percent hill slope and soil type. Because of this, measuring the amount of cubic material handled and moved is critically important. To estimate the amount of cubic material for any given hill slope and cut slope, we multiply the cross sectional area that would be cut into the hill by the linear distance of road constructed (Douglas 1999). The cross sectional area of the hill cut out to build the road (bold area in Figure 3) is calculated with the following equation (see Appendix A):

$$Area = \frac{1}{200} \cdot w^2 \cdot \frac{c \cdot s}{c - s} \qquad \text{[Equation 1]}$$

In this equation, $Area$ is the cross sectional area in square feet, w denotes the width of the road base in feet, and s and c denote the percent hill slope and percent cut slope, respectively, where $0 < s < c$. Note that for a slope of, for example, 35%, $s = 35$; more notably $s \neq .35$ (the same condition also holds for c).

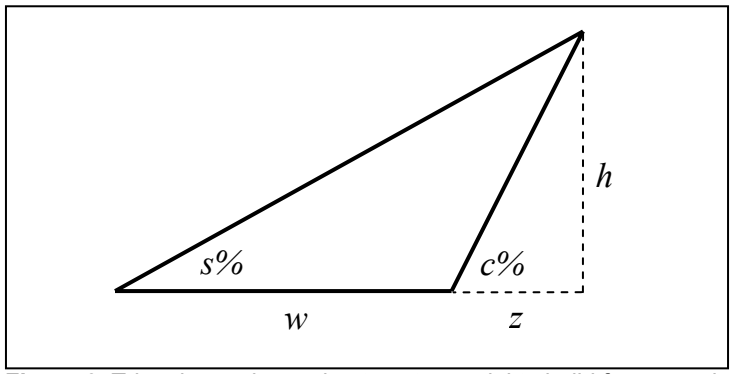


Figure 3: Triangle used to estimate cut material to build forest roads.

Next, to calculate total cubic feet volume of material handled and moved, we multiply the results from Equation 1 by the linear road distance measured in feet (d) and a material swell factor (sf)[6], which accounts for the percent increase in material volume due to air voids introduced into the material when disturbed:

$$TotalVolume = d * sf * Area \qquad \text{[Equation 2]}$$

Appendix B presents an approach for calculating total volume of material to be handled and moved for constructing roads having varying percent hill slopes.

To estimate fuel consumption and emissions for handling and moving the material calculated with the above equations, we used average production rates of cubic material moved per hour with a hydraulic excavator and two dump trucks from Balcom (1988). We assumed the same fuel consumption and utilization rate as above for the excavator. To estimate fuel consumption for endhauling the cut material $(TotalVolume)$, we used the average dump truck production rates reported by Balcom and assumed the use of two Caterpillar D25D articulated dump trucks requiring 4.7 gallons of diesel per hour each (CAT 1989). Additionally, according to a US Forest Service transportation planner, spreading the endhauled material at a waste site is also a necessary component of full bench road construction (pers. comm. Fred Bower, 13 January 2009). For this we assumed a Caterpillar D7 track type dozer is used 4 hours daily (pers. comm. Bob Greil, road construction contractor, 29 January 2009) requiring 8 gallons of diesel per hour (CAT 1989). We further assumed a material swell factor of 1.3.

The resulting fuel consumption and emissions estimates per cubic foot of handled material from road construction on slopes greater than 50% are displayed

[6] If material swell is 30% then the swell factor is $sf = 1+.30 = 1.3$.

in Table 3. The estimates are based upon the same diesel fuel emissions and energy content as with the cut-fill method. The estimates show approximately 0.007 gallon of diesel fuel is required per cubic foot handled and moved to construct a forest road using the full bench method. Carbon dioxide emissions are about 0.17 pounds per cubic foot and the carbon equivalent is roughly 0.05 pound per cubic foot.

Table 3: Diesel fuel consumption and carbon dioxide emissions per cubic foot of handled material with full bench method on slopes greater than 50%.

Road construction activity	Production (cubic feet/hour)	Diesel consumption (gallon/cubic foot)	Carbon dioxide emissions (pound/cubic foot)	Carbon equivalent (pound/cubic foot)
Pioneering, Clearing and grubbing, Sub-grade excavating (excavator)	2926.8[a]	0.00273	0.06223	0.01697
Endhauling (2 dump trucks)	2948.4[a]	0.00319	0.07272	0.01983
Waste site spreading (dozer)	5896.8[b]	0.00136	0.03100	0.00846
Total of all activities		0.00728	0.16595	0.04526

[a]From Balcom (1988)
[b]Assumes dozer operation is half the time as the other equipment (pers. comm. Bob Greil, road construction contractor 29 January 2009)

Tables 4 and 5 display diesel fuel consumption and carbon dioxide emissions by incremental hill slopes. The values are derived from total cubic feet of material to handle and move from Equations 1 and 2 and the estimate of diesel consumed per cubic foot from Table 3. We also assumed a fourteen foot wide road and 200% cut slope (pers. comm. Bob Greil, road construction contractor, 29 January 2009). Our estimates of total diesel fuel consumption for building forest roads on hill slopes greater than 50% range from approximately .62 – 1.5 gallons per linear road foot, and roughly 3,260 – 8,000 gallons per mile. Carbon dioxide emissions range from approximately 74,400 pounds per mile at 50% hill slope to 182,700 pounds per mile at an extreme hill slope of 90%. The carbon equivalent ranges from 20,300 – 49,800 pounds per mile.

Table 4: Diesel fuel consumption estimates by percent hill slope greater than or equal to 50% assuming a cut slope of 200% and fourteen feet wide roads.

Road construction method	Hill slope (percent)	Material to move per linear road foot		Diesel fuel consumption per linear road foot (gallons)				Gallons per mile
		(cubic yards)	(cubic feet)	Excavator	Dump trucks	Dozer	Total	
Full-bench	50	3.15	84.93	0.23187	0.27094	0.11551	0.61831	3,265
	55	3.58	96.65	0.26385	0.30831	0.13144	0.70360	3,715
	60	4.04	109.20	0.29812	0.34835	0.14851	0.79498	4,197
	65	4.54	122.68	0.33492	0.39135	0.16685	0.89312	4,716
	70	5.08	137.20	0.37456	0.43767	0.18659	0.99882	5,274
	75	5.66	152.88	0.41736	0.48769	0.20792	1.11297	5,876
	80	6.29	169.87	0.46374	0.54187	0.23102	1.23663	6,529
	85	6.98	188.33	0.51414	0.60077	0.25613	1.37105	7,239
	90	7.72	208.47	0.56913	0.66503	0.28352	1.51768	8,013

Table 5: Diesel fuel consumption, carbon dioxide emissions and carbon equivalent estimates by percent hill slope greater than or equal to 50% assuming a cut slope of 200% and fourteen feet wide roads.

Road construction method	Hill slope (percent)	Total diesel fuel consumption (gallons/foot)	Total diesel fuel consumption (gallons/mile)	Total carbon dioxide emissions (pounds/foot)	Total carbon dioxide emissions (pounds/mile)	Total carbon equivalent (pounds/foot)	Total carbon equivalent (pounds/mile)
Full-bench	50	0.61831	3,265	14.10	74,422	3.84	20,297
	55	0.70360	3,715	16.04	84,687	4.37	23,097
	60	0.79498	4,197	18.12	95,686	4.94	26,096
	65	0.89312	4,716	20.36	107,499	5.55	29,318
	70	0.99882	5,274	22.77	120,220	6.21	32,787
	75	1.11297	5,876	25.37	133,960	6.92	36,535
	80	1.23663	6,529	28.19	148,844	7.69	40,594
	85	1.37105	7,239	31.25	165,023	8.52	45,006
	90	1.51768	8,013	34.60	182,673	9.44	49,820

Discussion

It is common practice for road engineers to use published machine productivity equations to estimate costs or other related information. Here we have combined mathematical estimates with published production information to estimate fuel consumption and carbon dioxide emissions from building forest access roads. However, the results presented above rely on published studies designed to estimate forest road construction costs. While such studies are based upon field collected data, and general machine production rates were presented, machine fuel consumption during road construction was not investigated. Here we have estimated fuel consumption and emissions from forest road construction by combining computed average production estimates with estimates of machine-specific fuel consumption rates.

We recognize that our estimates have limitations. First, as Erickson (1992) described, forest road construction costs are difficult to estimate due to site-specific variations; therefore it is logical that fuel consumption would also be difficult to estimate due to similar variations. Second, applying the limited information provided by Balcom should be with caution, as conditions such as soil type can significantly impact the necessary cut slope and overall production. Third, there is no way to account for operator experience and production, or control for job-specific variations, such as culvert installment, turnout construction, seeding and stabilization, rolling dip construction, etc. We also cannot reasonably account for machine positioning and re-positioning, idling or other down time exclusive of the basic utilization rate used in our estimation process. Here we assumed that the amount of cut material equals the fill needed to build the road base with the cut-fill method, requiring no relocation of fill material. However, actual hill slope angles are not linear as displayed in Figure 3 and thus can require moving either more or less material for any given hill slope percent.

Additionally, in this analysis we have not included estimates of fuel consumption or emissions for road reconstruction, grading and maintenance, prism obliteration, employee commute to and from the job site, equipment mobilization via a lowboy tractor trailer, or delivery of supplies. We have limited our estimates to the basic elements of road construction, and suggest much more effort be devoted to this and other aspects of forest management as they relate to carbon accounting.

Literature Cited

Balcom, J. 1988. Construction Costs for Forest Roads. Forest Research Laboratory, Oregon State University, Corvallis. Research Bulletin 64 21 p.

Birdsey, Richard A. 2006. Carbon accounting rules and guidelines for the United States Forest Sector Journal of Environmental Quality 35:1518-1524.

CORRIM. 2006. The Consortium for Research on Renewable Industrial Materials. Life-Cycle Impacts of Forest Resource Activities in the Pacific Northwest and Southeast United States. Available at: http://www.corrim.org. Accessed: 11 December 2008.

Douglas, R.A. 1999. Delivery: The transportation of raw natural resource products from roadside to mill. University of New Brunswick. 202p. ISBN 1-55131-009-0

Erickson, Michael D.; Hassler, Curt C.; LeDoux, Chris B. 1992. Economic comparisons of haul road construction versus forwarding versus longer skid distances. In: 1992 international winter meeting, American Society of Agricultural Engineers. Paper No. 92-7513. Nashville, TN: St. Joseph, MI.

CAT. 2007. Caterpillar Performance Handbook Edition 37. Peoria: CAT, 2007.

CAT. 1989. Caterpillar Performance Handbook Edition 20. Peoria: CAT, 1989.

EIA. 2008. US Energy Information Administration. Available at: http://www.eia.doe.gov/basics/conversion_basics.html. Accessed 31 December 2008.

EPA. 1995. US Environmental Protection Agency. Compilation of Air Pollutants Emission Factors, Volume I: Stationary Point and Area Sources. Report AP-42, Office of Air and Radiation, US Environmental Protection Agency, Research Triangle Park, NC, 1995.

Healey SP, Morgan TA, Songster J, Brandt J. 2008. Determining landscape-level carbon emissions from historically harvested forest products. In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2008. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

IPCC. 2006. Intergovernmental Panel on Climate Change. 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Available at: http://www.ipcc.ch. Accessed 11 December 2008.

IPCC. 2003. Intergovernmental Panel on Climate Change. Good Practice Guidance for Land Use, Land Use Change and Forestry. Available at: http://www.ipcc.ch. Accessed 11 December 2008.

Kimball, Gail. 2007. "Climate Change, Kids, and Forests: What's the Connection?" Speech to Society of Environmental Journalists, Annual Conference, Stanford, CA. 7 September 2007. Available at: http://www.fs.fed.us/news/2007/speeches.shtml. Accessed 11 December 2008.

Layton, Deborah, A.; LeDoux, Chris 0.; Hassler, Curt C. 1992. Cost estimators for construction of forest roads in the central Appalachians Res. Pap. NE 665. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 4 p.

Markewitz Daniel. 2006. The carbon costs of silvicultural activities: Impacts on carbon sequestration in forests. Forest Ecology and Management 2006; 236(23):153-161.

Skog, K.E. and Nicholson, GA. 2000. Carbon sequestration in wood and paper products. In: The impact of climate change on America's forests: A technical document supporting the 2000 USDA Forest Service RPA Assessment. Joyce, LD and R Birdsey, tech. eds. Gen. Tech. Rept. RMRS-GTR-59. Fort Collins, CO: USDA Forest Service. Pp 79-88, Chapter 5.

USFS. 2007. US Department of Agriculture, Forest Service. Cost Estimating Guide for Road Construction. USDA Forest Service Northern Region Engineering. Available at: http://www.fs.fed.us/r1/projects/costguide_index.shtml. Accessed 12 December 2008.

Appendix A

To verify the area in Equation 1, we first note that the hill slope s in percent equals 100 multiplied by rise over run: $s = 100\frac{h}{w+z}$ (see Figure 3). Similarly, $c = 100\frac{h}{z}$. The area of the triangle is $\frac{1}{2}wh = \frac{1}{2}w^2\frac{h}{w}$. Also,

$$\frac{h}{w} = \frac{h^2}{h[(w+z)-z]} = \frac{h^2}{h(w+z)-hz} = \frac{\frac{h^2}{z(w+z)}}{\frac{h(w+z)-hz}{z(w+z)}} = \frac{\frac{h}{z}\cdot\frac{h}{w+z}}{\frac{h}{z}-\frac{h}{w+z}}.$$

Using that $\frac{h}{z} = \frac{c}{100}$ and $\frac{h}{w+z} = \frac{s}{100}$, we obtain $\frac{h}{w} = \frac{1}{100}\cdot\frac{c\cdot s}{c-s}$. Consequently, the area of the triangle is $\frac{1}{2}wh = \frac{1}{2}w^2\frac{h}{w} = \frac{1}{200}w^2\frac{c\cdot s}{c-s}$.

Appendix B

When building a forest road, the hill slope is usually not constant over any significant distance. To use the data from Tables 1, 2, 4 and 5 to estimate fuel consumption and emissions in this more general setting, proceed as follows. Estimate the total length of forest road to be built in d miles (or feet). Then determine the total length of the road pieces that will be built on gentle to moderate hill slopes < 50% and denote this length by d_{mod}. Next, determine the total lengths of the road segments that will be built on hill slopes of approximately 55%, 60%, 65%, etc. Denote these lengths by $d_{55\%}$, $d_{60\%}$, $d_{65\%}$, etc. Therefore

$$d = d_{\text{mod}} + d_{55\%} + d_{60\%} + d_{65\%} + \cdots + d_{90\%}.$$

Next denote the estimate for fuel consumption or carbon emissions from forest road construction on gentle to moderate slopes from Table 1 or 2 (per foot or mile) by E_{mod}. Similarly, for a given hill slope of s% on which the full bench construction method would be used, denote the corresponding estimate for fuel consumption or emissions from Table 4 or 5 by $E_{s\%}$. The total estimate *TE* for fuel consumption or carbon dioxide emissions for building the entire forest road is then given by

$$TE = d_{\text{mod}} * E_{\text{mod}} + d_{55\%} * E_{55\%} + d_{60\%} * E_{60\%} + d_{65\%} * E_{65\%} + \cdots + d_{90\%} * E_{90\%}.$$

Determining Landscape-Level Carbon Emissions from Historically Harvested Forest Products

Sean P. Healey[1], Todd A.Morgan[2], Jon Songster[2], Jason Brandt[2]

Abstract: *Resources have been developed in the literature to enable landowners to estimate the carbon sequestration timeline of forest products derived from their land. These tools were used here to estimate sequestration and emissions related to harvests carried out in Ravalli County from 1945 to 2007. This county-level accounting of product carbon release can later be combined with county-level estimates of growth and emissions from fire and other sources to provide landscape-level insights into relationships between forest carbon exchange and processes like harvest and fire.*

Keywords: Forest Products, Carbon Sequestration, Carbon Accounting

Introduction

It is broadly recognized that durable forest products may represent significant storage of forest carbon. Although the sum of stored product carbon and re-growing stand carbon may take many years to reach pre-harvest stand storage levels (Harmon et al., 1990), harvest products nevertheless delay the release of carbon from disturbed stands. Accurate assessment of the relationship between forest dynamics and the carbon cycle must account for the storage of forest carbon in products that are either in use or in landfills (Schlamadinger and Marland, 1996). Efforts are underway in the Rocky Mountain Research Station to monitor several interrelated components of landscape-scale forest carbon dynamics by combining remote sensing with carbon-tracking tools developed to help individual landowners register sequestration related to management. These tools, detailed by Smith et al. (2006), include heuristics for estimating the timeline of decay for forest products in various regions within the United States. Since decay of old forest products can contribute to current emissions, it is important to understand historical harvest trends of areas being monitored. The goal of this paper is to outline a strategy for combining product disposition heuristics such as those published by Smith et al. (2006) with county-level historical timber harvest information to develop estimates of recent emissions stemming from historical harvest products. The resulting estimated rates of product carbon release will be combined in future work with estimated effects of ecosystem processes such as growth and disturbance to provide a fine-scale picture of forest carbon exchange related to landscape dynamics.

Much regional and national carbon cycle monitoring relies upon what is called the "stock change" approach – tracking inventory-based estimates of ecosystem carbon stores

1. RMRS Inventory and Monitoring Program, Ogden, UT – seanhealey@fs.fed.us
2. University of Montana Bureau of Business and Economic Research, Missoula, MT

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

over time – plus accounting for carbon sequestered in products (e.g. Skog and Nicholson, 2000; Heath et al., 2003; Woodbury et al., 2007). Stock change approaches consider the competing factors that affect ecosystem carbon storage only in aggregate: stable stock levels may actually imply stasis, or they may imply offsetting trends in factors such as fire, growth, and harvest. An alternative monitoring approach that does address the differential effects of growth and harvest was described by Smith et al. (2006) to help landowners understand and potentially register gains in carbon sequestration due to their management decisions. Like Birdsey (1996), Smith et al. (2006) provided "look-up" tables that give regional averages for ecosystem carbon accumulation and product carbon sequestration. Work is underway at the US Forest Service's Rocky Mountain Research Station to produce similar look-up tables for other processes such as fire.

While stand-level look-up tables address dynamics such as the role of disturbance that are ignored in stock-change approaches, stand-level approaches cannot work at broader scales unless important details are known about every stand in the forest. Choosing the appropriate cell in any given look-up table, for example, requires that the region, species group, and growing stock volume or age be known. Additionally, one must know how and when disturbance has reset carbon densities among different stand pools in each stand. Remote sensing can provide estimates of these stand parameters across large areas. Satellite-derived species group maps are available across the country (Ruefenacht et al., 2008), volume and age estimates can be mapped when inventory and satellite data are combined (Kimes et al., 1996; Healey et al., 2006), and many disturbances can be mapped with a high level of precision (Cohen and Goward, 2004).

Like other approaches for tracking forest carbon dynamics, efforts to use remote sensing in conjunction with stand-level look-up tables depend upon accurate assessment of the disposition of product carbon. Historical harvest monitoring data are available at the county level in timber-producing areas throughout the United States. One source of this harvest information is Timber Product Output (TPO) monitoring, which is carried out nationally on behalf of the U.S. Forest Service's Forest Inventory and Analysis program (FIA). These data contribute to national timber harvest analyses presented in RPA reports (Forest and Rangeland Renewable Resources Planning Act, 1974 -- see Smith et al., 2004). Using historical harvest data along with product disposition rates published by Smith et al. (2006) and some assumptions about how these rates have changed over time, one may estimate recent rates of product carbon emissions at the county level. This paper details how such estimates may be made, using Ravalli County in western Montana as an illustration.

Methods

Smith et al. (2006) provided heuristics for the timing and movement of product carbon as it passes from the "in use" pool into landfills and finally into the atmosphere. Table 1 reprints these heuristics for the northern Rocky Mountain region. The methods described here allow product dynamics implicit in these tables to be combined with historical records of annual harvest volume to track how much carbon is released each year from

Table 1: Product Disposition Rates from Smith et al. (2006) with relevant derived annual changes.

Years after production	In use	Landfill	Energy	Emitted without energy	Year-on-year emission	Year-on-year disposal
	---------- Fraction of original harvest -------				----- Annual Change -----	
0	0.704	0	0.209	0.087	0.087	0.087
1	0.664	0.019	0.223	0.094	0.007	0.026
2	0.628	0.036	0.235	0.101	0.007	0.024
3	0.595	0.051	0.247	0.107	0.006	0.021
4	0.567	0.065	0.256	0.112	0.005	0.019
5	0.541	0.077	0.265	0.118	0.006	0.018
6	0.517	0.088	0.273	0.122	0.004	0.015
7	0.495	0.098	0.28	0.127	0.005	0.015
8	0.474	0.107	0.287	0.131	0.004	0.013
9	0.455	0.116	0.294	0.135	0.004	0.013
10	0.438	0.124	0.3	0.139	0.004	0.012
11	0.425	0.1296	0.304	0.142	0.003	0.0086
12	0.412	0.1352	0.308	0.145	0.003	0.0086
13	0.399	0.1408	0.312	0.148	0.003	0.0086
14	0.386	0.1464	0.316	0.151	0.003	0.0086
15	0.373	0.152	0.32	0.154	0.003	0.0086
16	0.3644	0.1558	0.3226	0.1562	0.0022	0.006
17	0.3558	0.1596	0.3252	0.1584	0.0022	0.006
18	0.3472	0.1634	0.3278	0.1606	0.0022	0.006
19	0.3386	0.1672	0.3304	0.1628	0.0022	0.006
20	0.33	0.171	0.333	0.165	0.0022	0.006
21	0.3234	0.1738	0.335	0.167	0.002	0.0048
22	0.3168	0.1766	0.337	0.169	0.002	0.0048
23	0.3102	0.1794	0.339	0.171	0.002	0.0048
24	0.3036	0.1822	0.341	0.173	0.002	0.0048
25	0.297	0.185	0.343	0.175	0.002	0.0048
26	0.2918	0.187	0.3444	0.1768	0.0018	0.0038
27	0.2866	0.189	0.3458	0.1786	0.0018	0.0038
28	0.2814	0.191	0.3472	0.1804	0.0018	0.0038
29	0.2762	0.193	0.3486	0.1822	0.0018	0.0038
30	0.271	0.195	0.35	0.184	0.0018	0.0038
35	0.248	0.204	0.356	0.192	0.0016	0.0034
40	0.229	0.211	0.36	0.2	0.0016	0.003
45	0.213	0.217	0.364	0.207	0.0014	0.0026
50	0.198	0.222	0.367	0.213	0.0012	0.0022
55	0.185	0.227	0.369	0.219	0.0012	0.0022
60	0.174	0.231	0.371	0.225	0.006	0.01
65	0.163	0.235	0.372	0.23	0.005	0.009
70	0.154	0.238	0.373	0.235	0.005	0.008
75	0.146	0.241	0.373	0.24	0.005	0.008
80	0.138	0.244	0.373	0.244	0.004	0.007
85	0.131	0.247	0.373	0.249	0.005	0.008
90	0.124	0.25	0.373	0.253	0.004	0.007
95	0.118	0.253	0.373	0.256	0.003	0.006
100	0.112	0.255	0.373	0.26	0.004	0.006

products produced in each previous year. This accounting of product carbon release depends upon good records of historical harvests and upon systematic tracking of the carbon in each year's product cohort as it progresses through the timeline of disposition described in Table 1. Both of these accounting components are discussed below.

Underlying Timber Product Output Data

Timber Product Output (TPO) studies are conducted nationally for FIA to estimate industrial and non-industrial uses of roundwood across the United States (e.g. White et al., 1980; Keegan et al., 2001). All primary wood-using mills in each state are periodically canvassed regarding the amounts, types, and origin of the wood products they use. Logging utilization studies that are also carried out for FIA provide an important link between this industrial data and inventoried volume.

TPO data showing annual removals in board feet (bf) from Ravalli County in four points in time (1988, 1993, 1998, and 2004) were the starting point for a county-level harvest timeline. County-level harvest numbers for other years between 1961 and 2007 were derived from a combination of sources: Montana Department of Natural Resources and Conservation (DNRC) cut-by-county reports for private lands, DNRC harvests for state lands, and USFS Cut & Sold Reports for the Bitterroot National Forest. Earlier removals (1945-1960) were estimated using available state-level harvest information. Since the distribution of harvest among ownership groups was known in both state and county datasets, it was possible to:

1. Use the 1961-1969 county-level data to establish the relative importance of each ownership group in the county-level cut (federal Bitterroot National Forest 95.4%; private forest 2.6%; other, mostly state, forest 2%)
2. Come up with an ownership-weighted estimate of 1945-1960 Ravalli County harvest as a function of the state-level harvest for those years.

Annual harvest numbers were converted to cubic feet (cf) from board feet (bf) TPO data (total bf / total cf) from different dates (see Spoelma et al., 2008). The conversion for 2004-2007 (3.69 bf/cf) was taken from the 2004 survey, while the 1998-2003 conversion (3.71) was taken from the 1998 survey, 1993-1997 used the 1993 conversion (4.79), 1988-1992 used the 1988 conversion (4.93), and dates prior to 1988 used an arbitrarily selected conversion of 5.0 bf/cf. It is likely that declining log sizes have led to a declining ratio of board-foot volume to cubic-foot volume. This relationship must be accounted for in the inference of cubic harvest volume from county-level board foot harvest records. Calculated cubic volumes were converted to units of carbon (first to pounds, then to metric tonnes, or megagrams) using the "Northern Rocky Mountains" factor published in Table 5.1 of Skog and Nicholson (2000). The resulting product carbon mass for each year (1945-2007, see Figure 1) was the input for subsequent inquiry into product disposition.

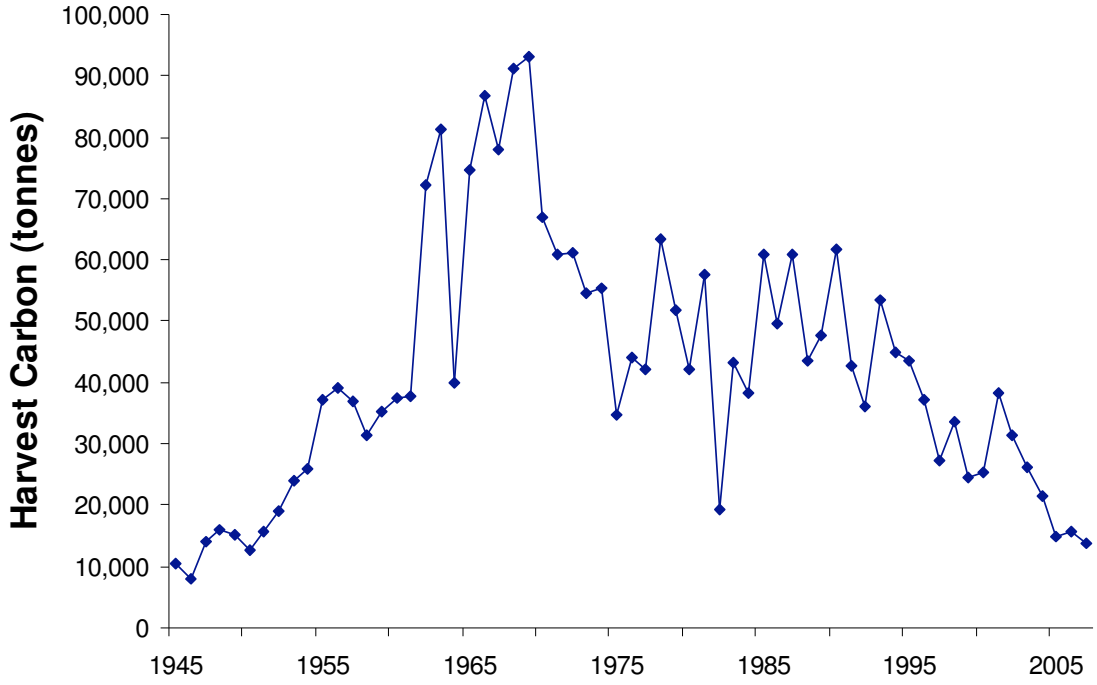


Figure 1: Harvest products from Ravalli County, MT, 1945-2007. Mass of carbon comes from administrative harvest volume records, transformed from board feet to cubic feet using TPO survey data and further translated to carbon using conversion factors in Skog and Nicholson (2000).

Estimating Emissions from Decomposition of Harvest Products

In general, assumptions regarding forest product longevity were taken from Table 6 in Smith et al. (2006), which shows changing fractions of the carbon in industrial roundwood among 4 different pools: 1) in use, 2) in landfills, 3) emitted with energy re-capture, and 4) emitted without energy capture. This table (partially illustrated here in Table 1) was based upon: national and regional assessments of wood harvest and end uses (partially summarized by Birdsey, 1996; Smith et al., 2004); basic assumptions about recycling rates and efficiencies (Skog and Nicholson, 1998), and studies of carbon decay limits and decay rates (Freed and Minz 2003, de Silva Alves et al. 2000). The disposition of the roundwood harvest volume from Ravalli County for each year from 1945 to 2007 was tracked over time using this table with the goal of monitoring all product carbon emissions for each year from 1985 to 2007. This study period corresponds with the Landsat TM era, and will complement satellite-derived information about growth and fire in future work.

While the dynamics of product life cycles have changed in the past and will continue to change in the future, generalized carbon flow heuristics may be applied over time with attention to major shifts in use or storage (Miner, 2006). In this study, the only implemented deviation from the dynamics discussed by Smith et al. (2006) was separate treatment of forest products discarded prior to 1980. Legislation in the 1970s mandated that dumps, which until then had been the primary means of waste disposal, be phased out by 1986 (Skog and Nicholson, 2000). Since landfills contain only a limited amount

of oxygen, and therefore facilitate much less aerobic decomposition than dumps, the shift from dump- to landfill-disposal was consequential with regard to post-disposal emission rates. For the sake of simplicity, the move to landfill disposal was represented in this study as a discrete event in 1980. While the tables in Smith et al. (2006) show relatively slow emission rates consistent with the anaerobic conditions in most landfills, Skog and Nicholson (2000) estimated that 65% of the material sent to dumps was burned and/or emitted relatively quickly, with the remaining material decomposing over a 96-year cycle.

The amount of carbon entering the product pools in each year from 1945 to 2007 was tracked independently, and emissions from each of these "origin" years were calculated and summed for each year of the study period (1985-2007). For material harvested and processed after the shift in 1980 to the landfill dynamics addressed in the Smith et al. tables, calculation of carbon emitted each year was simply derived from the by-year change in the "emitted without energy" column (see Table 1) from one "year after production" to the next. Linear interpolation was used in this table to fill in values not supplied by Smith et al. Consider as an example the determination of how much carbon was emitted in 1989 from material harvested in 1982. If the roundwood harvested in 1982 (Year 0) was 10000 tonnes, and Table 1 indicates a fractional increase from 12.2% to 12.7% in carbon emitted without energy capture from Year 6 to Year 7, the calculated emission for 1989 from 1982 products will be 50 tonnes (0.5% of 10000).

For products produced and disposed of prior to the 1980-shift to landfills (i.e. 1945-1979), the dump dynamics outlined by Skog and Nicholson (2000) were implemented; 65% of disposed material was supposed emitted within the first year, with the remainder emitted uniformly over a 96-year period. Table 1 was used to determine the rate at which product carbon reached the dump; the disposal rate was assumed to be equal to the fraction of product carbon falling out of the "in use" category minus the fraction "emitted with energy re-capture." This year-by-year fraction of disposed carbon (*D* below) was used to track the arrival of part of each year's product cohort at the dump in each successive year. Further, emissions of dump contents were calculated for each year *A* as follows:

$$\sum_o TDC \qquad \text{(Equation 1)}$$

where *O* is each product origin year between 1945 and Year *A*, *T* is the mass in metric tonnes of carbon in the roundwood originally processed in year *O*, *D* is the fraction of *T* expected to reach the dump in Year *A* based upon Table 1. *C* is a constant:

$$C = RF \qquad \text{(Equation 2)}$$

where *R* is percentage of *TD* remaining following the initial burning at the dump (0.35 here), and *F* is the fraction of the assumed life cycle represented by Year *A* (1/96 here). Since the amount of carbon in dumps has been capped since 1980, and since that carbon is assumed to be emitted evenly on a 96-year cycle, emissions from dumped 1945-1979

products will be a constant until the first cohort of dumped products is envisioned to be completely emitted (i.e. 2041).

Many pre-1980 products were in use following 1980, of course, and their disposal must be accounted for via the landfill dynamics represented in Table 1. The percentage of products in use in 1980 from each harvest year since 1945 was determined using the "in use" column of Table 1. These products were assumed to be emitted using a rate of 0.17% per year. This rate, while somewhat arbitrary, is the 100-year average rate of emission without energy re-capture from Table 1 if the most volatile first year is excluded. First-year emissions were not considered in this emission rate because fact none of the emissions in this category (produced prior to 1980, disposed and emitted following 1980) could have occurred in the first year following harvest.

In summary, a fraction of each year's harvest products was estimated to be emitted in each subsequent year using product life cycle heuristics found in Table 1 and in the work of Skog and Nicholson (2000). Emissions from dumps and landfills from each product cohort were summed for each year to produce an estimate of overall emission of carbon resulting from forest products dating back to 1945. These emissions were plotted over time for the study period 1985-2007.

Results

As discussed above, county-level harvest volumes were tracked through the product cycle to disposal and eventual emission using three different scenarios: 1) landfill dynamics (see Table 1) for all harvests after 1979; 2) dump disposal (after Skog and Nicholson, 2000) for products discarded prior to 1980; and 3) landfill dynamics for products produced prior to 1980 but discarded thereafter. Table 2 shows part of the accounting (1961-1979) of the rate at which each year's forest products reached the dump; quantities in tonnes of carbon have been reduced by 65% to account for initial rates of burning (Skog and Nicholson, 2000). Similar accounting was done for estimated forest product levels from 1945 to 1960 (not shown). The total mass of carbon in dumps resulting from 1945-1979 forest products is estimated to be 148,288 tonnes, resulting in an annual emission of 1545 tonnes. Since the quantity of carbon in dumps has been capped since 1980 and emissions are distributed evenly over a 96-year period, calculated dump emissions will stay at this level until 2041. Emissions from older forest products still in use in 1980 (and thus assumed to have gone through landfills) was estimated to be 1101 tonnes per year.

Emissions from products estimated to have been discarded after 1980 were determined through the "emissions without energy re-capture" column in Table 1. Figure 2 shows the composite emissions from dumps and landfills. Total emissions declined throughout the study period from above 12000 tonnes per year to less than 9000 tonnes per year. The decline in emission in Figure 2 is not as pronounced as the decline in harvest volumes seen in Figure 1 because the amount of carbon accumulated in landfills since 1980 has

Table 2: Rate of Deposition into Current Dump Storage.

		Year Sent to the Dump																		
Year Produced	**Metric Tonnes C**	**1961**	**1962**	**1963**	**1964**	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**	**1971**	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**	**1978**	**1979**
		---------- Tonnes of carbon not initially burned and now decomposing on a 96-year cycle ----------																		
1961	37741	1149	343	317	277	251	238	198	198	172	172	159	114	114	114	114	114	79	79	79
1962	72195		2198	657	606	531	480	455	379	379	328	328	303	217	217	217	217	217	152	152
1963	81156			2471	739	682	596	540	511	426	426	369	369	341	244	244	244	244	244	170
1964	39841				1213	363	335	293	265	251	209	209	181	181	167	120	120	120	120	120
1965	74749					2276	680	628	549	497	471	392	392	340	340	314	225	225	225	225
1966	86896						2646	791	730	639	578	547	456	456	395	395	365	262	262	262
1967	77920							2373	709	655	573	518	491	409	409	355	355	327	235	235
1968	91235								2778	830	766	671	607	575	479	479	415	415	383	275
1969	93162									2837	848	783	685	620	587	489	489	424	424	391
1970	67067										2042	610	563	493	446	423	352	352	305	305
1971	60752											1850	553	510	447	404	383	319	319	276
1972	61128												1861	556	513	449	407	385	321	321
1973	54447													1658	495	457	400	362	343	286
1974	55255														1683	503	464	406	367	348
1975	34621															1054	315	291	254	230
1976	44144																1344	402	371	324
1977	42090																	1282	383	354
1978	63412																		1931	577
1979	51685																			1574

been growing. Thus, emissions from products produced and discarded at the end of the study period are augmented by emissions from products originating earlier in the period. In general, annual variation in overall product carbon emission is most sensitive to recent harvesting trends (i.e. black bars in Figure 1) because: disposal rates are highest soon after harvest (Table 1), and because this accounting system treats emissions from pre-1980 products as a constant.

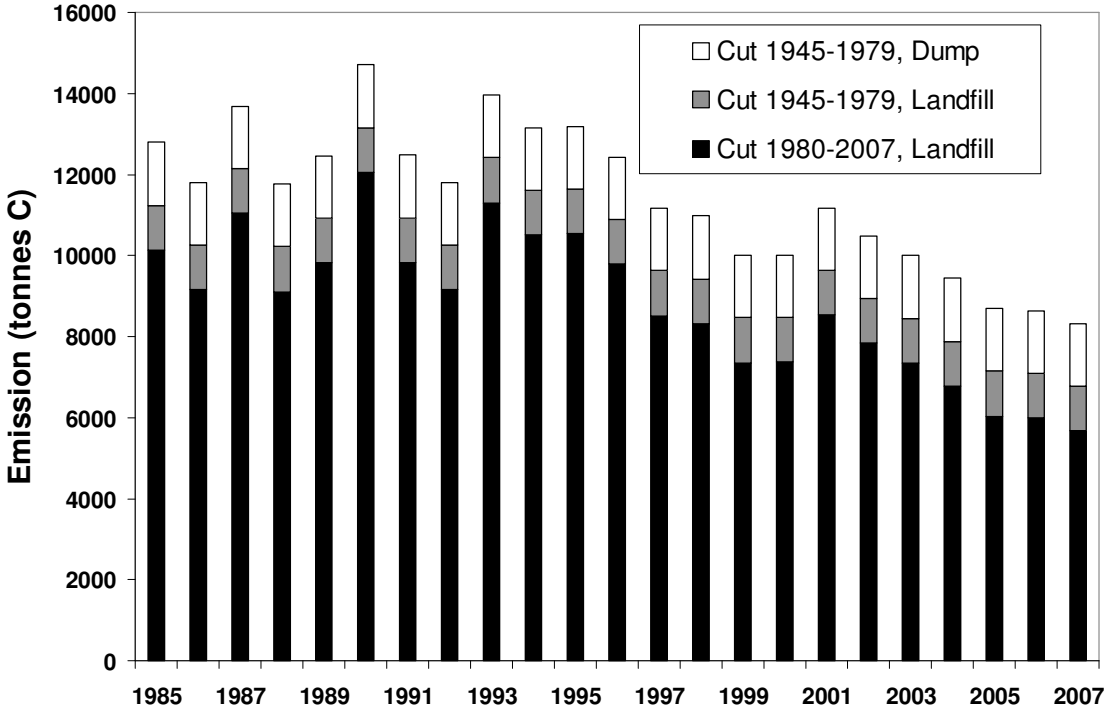


Figure 2: Annual product carbon emissions in Ravalli County, 1985-2007. Three product pathways were considered: produced after 1980 and sent to a landfill (black), produced 1945-1979 and sent to a landfill (gray), and produced 1945-1979 and sent to a dump (white).

Discussion

The legacy of carbon associated with historical forest products must be considered when assessing the role management plays in forest carbon cycles. Because the forest product cycle of processing, use and disposal can defer release of carbon into the atmosphere, harvest can be seen as a relatively benign form of disturbance. Deferral of product emissions is not permanent, however, and accurate comparison of the effects of harvest vs. fire and other factors, as is the goal of currently ongoing research, depends upon adequate estimates of current emissions from products manufactured over preceding decades.

Remote sensing using historical Landsat imagery, as calibrated with inventory data, can provide temporally and spatially consistent harvest information (Healey et al., 2007). Indeed, in countries where there is no systematic monitoring of forest industry output, remotely-sensed estimates of volume removal could form the basis of historical product assessments. In the United States, however, the most credible county-level estimates of harvest trends come from administrative records related to taxes and TPO industry surveys. Mill-based survey assessments are presumed to have little error, and are available across much of the country for the last 50 years (Smith et al., 2004). The long continuous county-level time series used here may not be available in some parts of the country where the timber industry is less important than in western Montana, but in these

cases, TPO surveys (typically produced once every 5 or so years) might be extrapolated to fill in the time series.

While the use of TPO data obviates modeling error that would be present if harvest numbers were provided by remote sensing, two potential sources of error should be emphasized in the historical product carbon accounting approach described here. One source of uncertainty may be potential mismatch between the regional spatial scale of the reference data used in Smith et al.'s product dynamic tables (see Table 1), and the relatively small portion of the Rocky Mountain region studied here. Specifically, the species composition of Ravalli County's forests and the mix of its product types may deviate from regional averages. Secondly, while the disposition rates in Table 1 may accurately portray current product life cycles, application of those rates to historical harvests may mask important changes in the rate that product carbon is emitted. The system discussed here accounts for a shift in destination from dumps to landfills, for example, but it does not address potentially important changes in recycling rates (Skog and Nicholson, 2000).

The emissions described in this paper do not distinguish between carbon dioxide and methane. Since the global warming potential of methane (over a 100-year period) is estimated to be 21 times that of carbon dioxide (IPCC, 1996), the effective of Ravalli County's product carbon emissions would actually be higher if methane contributions were translated into carbon dioxide equivalents. However, rate of capture of landfill methane has increased recently and is currently about 50% (Skog, 2008). So, the effect of methane on emissions in terms of carbon dioxide equivalents would be to raise the estimates given here, but in a way that continues to decline as more methane is captured.

Nevertheless, the processes used to create the historical emission estimates in Figure 2 are straightforward and could easily be modified to accommodate life cycle heuristics that are better adapted either spatially or temporally to the unit of analysis. Resulting county-level product emission estimates will be compatible in spatial and temporal scale with estimates of carbon flux related to other processes like growth and fire. Synthesis of different carbon cycle components and analysis of their behavior in relation to forest management should yield fundamental ecosystem-specific information about the relationship between management and landscape-level carbon movement.

References

Birdsey, R.A. 1996. Carbon storage for major forest types and regions in the coterminous United States. In: Sampson, N.; Hair, D., eds. Forests and global change. Volume 2: forest management opportunities for mitigating carbon emissions. Washington, DC: American Forests: 1-25, Appendixes 2-4.

de Silva Alves, J.W.; Boeckx, P.; Brown, K. [and others]. 2000. Chapter 5 – Waste. In: Penman, J.; Kruger, D.; Galbally, I. [and others], eds. Good practice guidance and uncertainty management in national greenhouse gas inventories. Hayama, Kanagawa,

Japan: Institute for Global Environmental Strategies for the Intergovernmental Panel on Climate Change: 5.1-5.32.

Freed, R.; Minz, C. 2003. Memorandum: Revised Input Data for WOODCARB. Memorandum dated August 29, 2003. ICF Consulting, Washington, DC.

Harmon, M.E.; Ferrell, W.K.; Franklin. J.F. 1990. Effects on carbon storage of conversion of old-growth forests to young forests. Science 247, 699701

Heath, L.S.; Smith, J.E.; Birdsey, R.A. 2003. Carbon trends in U. S. forest lands: a context for the role of soils in forest carbon sequestration. In: Kimble, J.M.; Heath, L.S.; Birdsey, R.A.; Lal, R., eds. The potential of US forest soils to sequester carbon and mitigate the greenhouse effect. New York: CRC Press: 35-45.

IPCC 1996. Climate Change 1995: The Science of Climate Change. Intergovernmental Panel on Climate Change; J.T. Houghton, L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg, and K. Maskell, eds.; Cambridge University Press. Cambridge, U.K.

Kimes, D.S.; Holden N.; Nickerson, J.E.; Mckee, W.A. 1996. Extracting forest age in a Pacific Northwest Forest from thematic mapper and topographic data. Remote Sensing of Environment, 56(2): 133-140.

Miner R. 2006. The 100-year method for forecasting carbon sequestration in forest products in use. Mitigation and Adaptation Strategies for Global Change. 1573-1596. DOI: 10.1007/s11027-006-4496-3.

Ruefenacht B.; Finco, M.V.; Nelson, M.D.; Czplewski, R.; Helmer, E.H.; Blackard, J.A.; Holden, G.R.; Lister, A.J.; Salajanu, D.; Weyermann, D.; Winterberger, K. 2008. Conterminous U.S. and Alaska Forest Type Mapping Using Forest Inventory and Analysis Data. Photogrammetric Engineering and Remote Sensing, 74(11): 1379-1388.

Schlamadinger, B.; Marland, G. 1996. The role of forest and bioenergy strategies in the global carbon cycle. Biomass and Bioenergy 10(5/6): 275-300.

Skog, K.E.; Nicholson, G.A. 1998. Carbon cycling through wood products: the role of wood and paper products in carbon sequestration. Forest Products Journal. 48(7/8): 75-83.

Skog, K.E.; Nicholson, G.A. 2000. Carbon sequestration in wood and paper products. In: The impact of climate change on America's forests: A technical document supporting the 2000 USDA Forest Service RPA Assessment. Joyce, LD and R Birdsey, tech. eds. Gen. Tech. Rept. RMRS-GTR-59. Fort Collins, CO: USDA Forest Service. Pp 79-88, Chapter 5. www.fpl.fs.fed.us/documents/pdf2000/skog00b.pdf (accessed 12/28/08).

Skog, K.E. 2008. Sequestration of carbon in harvested wood products for the United States. Forest Products Journal. 58(6): 56-72.

Spoelma, T.P.; Morgan, T.A.; Dillon, T.; Chase, A.L.; Keegan, C.E. III; DeBlander, L.T. 2008. Montana's forest products industry and timber harvest, 2004. Resour. Bull. RMRS-RB-8. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 36 p.

Smith J.E.; Heath, L.S.; Skog, K.E.; Birdsey, R.A. 2006. Method for Calculating Forest Ecosystem and Harvested Carbon with Standard Estimates for Forest Types of the United States. Forest Service, Northeastern Research Station, General Technical Report NE-343. 216 p.

Smith, W.B.; Miles, P.D.; Vissage, J.S.; Pugh, S.A. 2004. Forest resources of the United States, 2002. Gen. Tech. Rep. NC-241. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 137 p.

Woodbury, P.B.; Smith, J.E.; Heath, L.S. 2007. Carbon sequestration in the U.S. forest sector from 1990 to 2010. Forest Ecology and Management 241:14-27.

Wulder, M.A.; Skakun, R.S.; Kurz, W.A.; White, J.C. 2004. Estimating time since forest harvest using segmented Landsat ETM+ imagery. Remote Sensing of Environment, 93(1-2): 179-187.

Phase 3 Indicators: Past, Present, and Future



Lichens, Ozone, and Forest Health – Exploring Cross-indicator Analyses with FIA Data

Susan Will-Wolf[1]
Sarah Jovan[2]

Abstract: *Does air pollution risk represented by a lichen bioindicator of air pollution, an ozone bioindicator, or a combination of both, correlate with forest health as reflected by condition of tree crowns and other variables? We conducted pilot analyses to answer this question using Forest Inventory and Analysis (FIA) data from the Sierra Nevada region of California and the New England region; they have very different environments. We started by addressing procedural and statistical issues. Several steps were required to assemble data sets for addressing this question. Currently, incomplete data availability limits the ability of an analyst not affiliated with the FIA program to conduct similar investigations. Statistical issues include data characteristics, data screening for suitability, and how to combine different variables in the same analysis. We calculated combined plot-level air pollution risk indexes from lichens and ozone data, and combined plot-level 'forest stress' indexes from two or more tree health indicators. Consultations with other FIA indicator advisors and with FIA analysts having experience with the tree variables we used were extremely important. For the Sierra Nevada region no significant correlations were found between pollution risk and 'forest stress' indexes. The next step there is to decide whether it seems worthwhile to pursue other related avenues investigating the same question. For the New England states, some correlations were significant but so weak that their biological relevance is questionable. For one of three ecoregion subgroups examined, correlations were significant and strong enough to warrant follow-up studies. In general we concluded that crown dieback and vigor variables were more useful than other tree variables, and summarizing to plot-level based on tree basal area was more useful than using tree counts. Development of more sophisticated plot-level indexes of 'forest stress' should be part of all follow-up investigations.*

Keywords: Air pollution, bioindicator, California, forest health, forest stress, lichens, ozone, Phase 3, risk index, tree crowns, tree mortality, New England

[1]FIA lichen indicator co-advisor, East. Department of Botany, University of Wisconsin-Madison
[2]FIA lichen indicator co-advisor, West. USDA FS PNW Station, Portland, Oregon and Oregon State University

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

We decided to explore whether estimated air pollution risk in forests (represented by a lichen bioindicator of air pollution, an ozone bioindicator, or a combination of both) correlates with forest health as reflected by condition of tree crowns and other variables. The U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis Program (FIA) collects both tree mensuration data and data addressing forest ecosystem health on a subset of its full national grid (phase 3 or P3 plots). Lichens are included in the program as useful forest health indicators for air quality impact, biodiversity, and forest ecosystem integrity. An ozone bioindicator is included to help assess the level of this known risk to some vascular plants. A number of variables included in the program can be used to assess tree health on FIA plots, including growth and mortality, crown condition, and tree damage variables (FIA 2008f).

We conducted pilot analyses to answer our question using FIA data from the Greater Sierra Nevada region of California and the New England region (including New York state) – parts of the USA with very different environments. Since this is a very preliminary investigation, our project included evaluation of the procedural and statistical issues involved in asking such a question. What steps are required to assemble data sets for addressing this question? How might these steps be seen by a data user from outside the Forest Service trying to do similar analyses based on FIA data available online? What statistical issues arise related to using the different indicators in the same analyses? What issues arise when we attempt to generate a combined air pollution response index from lichens and ozone indexes, or a combined plot-level 'forest stress' index from two or more tree variables? When and where are consultations critical - with FIA Indicator Advisors and FIA Analysts having experience with these variables?

Our starting assumption was that air pollution tracked by lichen air quality indexes and the ozone bioindicator index might contribute to low-level stress that would render trees more susceptible to reduced health or damage from other agents. It is possible this might happen well below levels of air pollution that would directly damage most trees.

Selected variables

We selected several FIA variables for our pilot project. These variables are primary or derived variables related to either pollution risk or tree health. Most are defined in online FIA documents. Other published documents linked to the FIA program explain how to use and interpret the variables.

Pollution risk variables

We used two different kinds of pollution risk variables. Lichens are sensitive to several kinds of Nitrogen- and Sulfur-based air pollutants, and lichen air quality

indexes track response of lichens to all of these pollutants together. The ozone interpolated biotic index (IBI) is calculated based on response of sensitive vascular plants; it assesses potential risk from only this single pollutant. Both lichen air quality indexes and the ozone IBI apply to an entire FIA plot.

Lichen air quality indexes: Lichen air quality indexes are based on lichen species composition at sites, are derived from models specific to a particular geographic region (Will-Wolf and Neitlich in review), and apply only within that region. They are not currently defined in online FIA database documentation; users refer to published descriptions of each model and to Will-Wolf (in review) for guides to usage. Lichens are sensitive to two major classes of air pollution: 1) mostly neutral pH/alkaline N pollution from agricultural sources (figure 1A), and 2) acidic N and S pollution that comes primarily from urban and industrial sources (figure 1B). In addition, N from both sources can have a fertilizing effect on the ecosystem. Responses of lichens to the two classes of pollution differ, and the most important kind of pollution may differ between regions (Will-Wolf et al. in review). Thus lichen air quality indexes from different regions may reflect response to different kinds of air pollution. In both regions included in this study, a higher index number indicates poorer air quality.

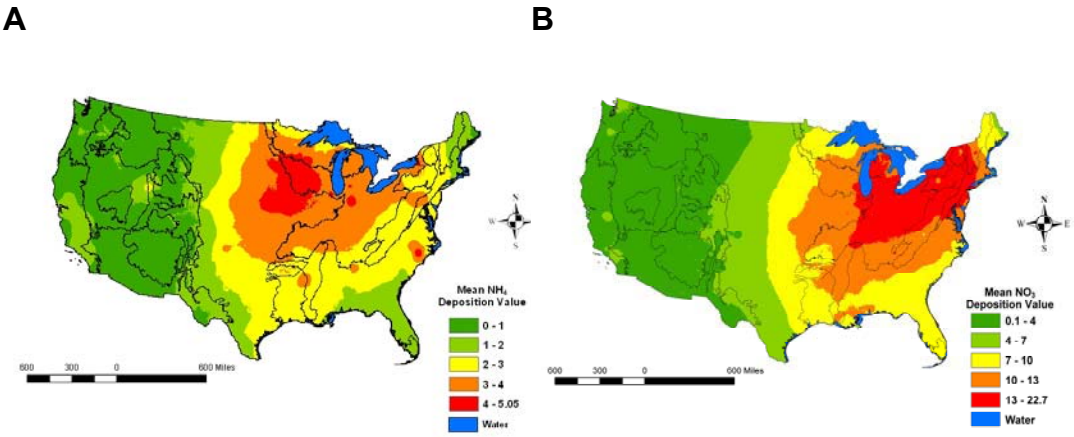


Figure 1: Pollution wet deposition modeled from 1998-2004 National Atmospheric Deposition Program data (Coulston and others 2004). Black lines mark boundaries of ecoregion groups (some noted in figures 3 and 4). **A.** NH_4^+ represents effects of agriculture; **B.** NO_3^- represents urban/ industrial effects. Note amounts for both are much higher in the East. Reproduced from Will-Wolf et al. (in review).

Greater Sierra Nevada (GSN) region of California – The published Lichen Air Quality index (LichenAirGSN for this project) for this region (Jovan & McCune 2006) tracks mostly response to neutral/alkaline agricultural nitrogen pollution (Fenn et al 2003, Jovan 2008). The amount of pollution in this region is underestimated in figure 1 because those maps are based only on wet deposition and dry deposition is an important contributor there (Fenn et al. 2008).

New England (NE) region – This region has two provisional Lichen Air Quality indexes (Will-Wolf unpublished) that track response primarily to acidic urban/industrial pollution. LichenAirNE-A is relative abundance of eight pollution-tolerant lichen indicator species, and LichenAirNE-B is the former index adjusted for latitude. The latter takes into account north-south gradients in climate as well as regional background air pollution (figure 1), leaving this index more responsive to relative local pollution level regardless of regional context.

Ozone interpolated biosite index (IBI): This index is based on level of damage to sensitive species at special field plots. It is calculated by spatial interpolation for FIA plots (FIA 2008a). It is diagnostic for one pollutant, and is applicable and comparable across the country (Coulston et al. 2004, Smith et al. 2007); ozone risk is quite similar for the two pilot project regions (figure 2). A higher number indicates greater potential risk for ozone injury.

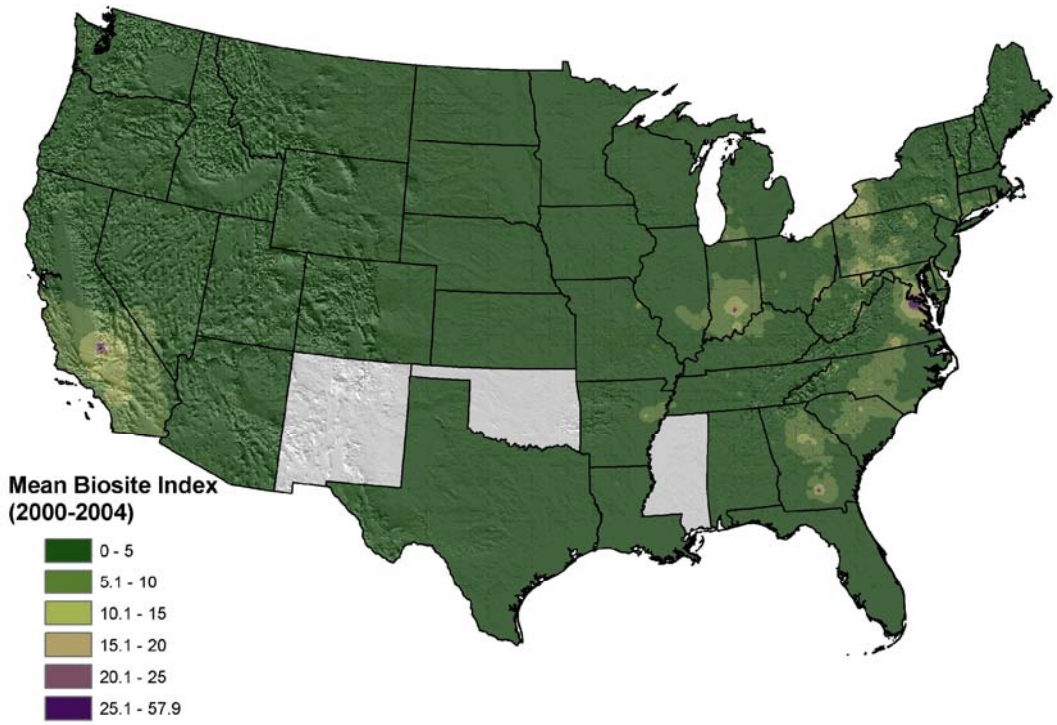


Figure 2: This map of Ozone Interpolated Biosite Index (IBI) illustrates estimated risk of ozone injury (map by John Coulston).

Tree health variables

Crown condition, growth, mortality, and damage are all classes of FIA tree variables (FIA 2008b) that are good candidates for evaluating tree health. Relatively few plots for which we have lichens data also have damage data, so we did not consider damage variables further. We needed to aggregate values

representing tree health across all trees at the plot into 'forest stress' indexes for comparison with pollution risk indexes. For this pilot study, we conducted no calibration based on species of tree. In most cases stressed trees were expressed as proportion of either total number of trees or total basal area at the plot. Total basal area was also used as a surrogate for stand age.

Crown variables: From crown condition documents (Bechtold and Randolph 2007, Schomaker et al. 2007) and consultation with current crowns indicator advisors William Bechtold and KaDonna Randolph, we identified several crown variables suitable to represent tree stress. Our 'stressed tree' cutoff levels are based on both advice from Bechtold and Randolph and on previous work by others; our choices of cutoff levels are our responsibility alone.

Crown dieback (trees ≥5in/12.7cm diameter) – This variable may be aggregated across species to plot-level. It is considered a strong 'tree stress' indicator. Both average dieback and proportion of trees with dieback ≥10% (indicating stress) are calculated for 'forest stress' indexes.

Crown vigor (trees <5in/12.7cm diameter) – This variable may be aggregated across species to plot-level. This is considered a strong 'tree stress' indicator. Proportion of small trees (saplings) in vigor class 3 (indicating stress) is calculated for a 'forest stress' index.

Crown density (trees ≥5in/12.7cm diameter) – Density for healthy trees varies by species, so this variable should be related to a species standard before aggregating to plot-level. Steinman et al. (2000) used crown density ≤35% as a conservative indicator of a stressed tree, so we used proportion of trees stressed based on this cutoff value as a weak 'forest stress' index.

Foliage transparency (trees ≥5in/12.7cm diameter) – Density for healthy trees varies by species, so this variable should be related to species standard before aggregating to plot-level. Steinman et al. (2000) used foliage transparency ≥35% as a conservative indicator of a stressed tree, so we used proportion of trees stressed based on this cutoff value as a weak 'forest stress' index.

Growth: Growth for healthy trees varies by species, so this variable should be related to a species standard before aggregating to plot-level. We found no examples in literature for reasonably conservative indicators of a stressed tree based on growth alone, so we did not pursue this variable further.

Mortality: – This variable may be aggregated across species to plot-level. We calculated the proportion of trees that died up to 5 years before an inventory year as an estimate of tree stress, to be consistent with mortality variable definitions (FIA 2008b). Expected mortality might vary with stand age (Bechtold, pers. comm.), so the proportion might need to be adjusted for stand age. Also, mortality became an optional variable to report starting in 2003, and some states did not report this after 2002. For both these reasons, we consider this a weak 'forest stress' index.

Data for analysis

Data were assembled for the two pilot study areas, then were screened for a variety of issues. Plot variables were assembled from screened data. Data sets were assembled in Excel (2003); statistical analyses used SPSS (2007).

Study areas

The Greater Sierra Nevada region of California (figure 3) is mostly one ecoregion province. Lichens data are available from 1998-2001 and 2003.

The New England states (figure 4) include Connecticut, Maine, Massachusetts, New Hampshire, New York, Rhode Island, and Vermont. This region is much larger than the California region; we investigated patterns both for the whole region and for three ecoregion group subdivisions from another project (Will-Wolf et al. in review). Lichens data are available from 1994 and 1998-2004.

Assembling data sets

FIA data are in the public domain (FIA 2008c) and are made available in a timely manner, but difficulties reconciling some data matching issues mean not all measured data used for this project are currently available to the public online. No 2000-2002 data are currently available online, pending resolution of data matching issues. Neither lichen nor ozone indexes (derived variables) are archived in the central database and posted online, pending development of FIA database structures for derived variables. They must be obtained by both FIA and public data users from lichen indicator advisors (currently the coauthors of this document) or the FIA region. It is assumed that both database and data matching issues that limit data availability online are temporary; all data will eventually be publicly available in some form.

We started with a list of all plots in each region for which lichens data are available. Plots were screened based on the following criteria. Plots with no lichens found (cannot calculate a lichens air quality index for them) or with flags for poor fit to the regional model were excluded. Eastern plots with live tree basal area (BA) below 5 m^2/ha (21.8 ft^2/ac) were excluded; their lichen air quality indexes might be biased due to limited substrate (no BA cutoff needed for western plots; Will-Wolf in review). Some plots were excluded because tree data were not easily available. Plots with fewer than 3 trees rated for crowns were excluded as giving inadequate representation of forest health. An ozone IBI can be calculated by interpolation for all our plots; no screening was needed based on ozone data.

Lichen air quality index values for the greater Sierra Nevada region of California were calculated by Sarah Jovan. Two provisional lichen air quality indexes for the New England region were calculated by Susan Will-Wolf. The ozone IBI was calculated via interpolation for all project plots by John Coulston

using exact plot locations. Tree data were extracted from FIA database files and FHM legacy database files mostly by Randy Morin, and also by Susan Will-Wolf.

Available lichen plots and plots excluded based on lichen data criteria are summarized in table 1; plots excluded based on tree data and remaining plots available for this project are summarized in table 2. Plots in four ecoregion subgroups of the New England region are listed in table 2.

Table 1: Lichen plots available and excluded based on lichen data screening.

	Lichen plots: years	**No lichens; plots excluded**	**Plots surveyed in 2 years; 1 year excluded**	**Plots excluded - tree BA too low[a]**
California Sierra Nevada region	146 plots: 1998-2001, 2003	3	9[b]	
New England region	668 plots: 1994, 1998-2004	4	71[c]	53

[a]NE plots only; lichen air quality index might be biased.

[b]1998 plots retained

[c]plots from several later years retained

Table 2: Lichen plots available and excluded based on tree data screening.

	Plots with lichens data[a]	**Plots excluded: tree data unavailable**	**Plots excluded: too few trees rated for crowns**	**Plots used for this project**
California Sierra Nevada region	134	2 plots: 1998-2001 25 plots: 2003	8	99
New England region	540	4 plots: 1994-1999 178 plots: 2000-2002	6	352
Adirondacks subregion				146
Laurentian subregion				125
Eastern Deciduous subregion				80
Appalachian subregion				1

[a]Plots remaining from table 1.

Assembling plot variables

We used single variables calculated from the program databases as described above. We also used composite variables calculated from combining different single variables. For composite variables we used different approaches depending on the variables used. Our calculation procedures took into account a number of statistical issues; we outline these first.

Statistical issues: To avoid bias in our cross-indicator analyses, we modified differently-scaled variables to be used in the same statistical procedure. We tried two 'low tech' options to equalize scales for variables. One choice was to add and/or multiply values of some variables by constants to match ranges of the other variables to be used in the same analysis. A second choice was to rank-transform all variables to equalize ranges. This second choice at the same time corrects for any problems with data distributions.

Pollution - lichen and ozone plot indexes: These are the single or composite variables we used for analysis. We used the same approach for all three LichenAir… indexes in both regions. Since high values of both lichen and ozone indexes suggest higher pollution risk, we added them together for a composite index. We tried three kinds of composite pollution index, one with original indexes weighted equally, and others with one or the other weighted double. One set of indexes was calculated using modified variable range – these are listed below. A second parallel set of variables was calculated using rank-transformed data.

- LichenAir… [region suffix]: Range modified to match Ozone IBI in region.
- OzoneIBI: No region suffix needed; IBI is nationally consistent.
- LichenAir…+OzoneIBI: Composite index from simple addition.
- 2xLichenAir…+OzoneIBI: Composite index, LichenAir… weighted double.
- LichenAir…+2xOzoneIBI: Composite index, OzoneIBI weighted double.

Forest stress - crowns and mortality plot variables: Most of the forest stress indexes are composite variables; only average crown dieback seems a useful single index. All variables are calculated so that a higher value suggests greater stress on the forest at the plot. We found that % recent mortality (based on counts or BA of trees) in either region was not correlated with total plot BA, our best estimator for stand age.

- AvgDbk is Average Crown dieback class (for all trees rated) – a single variable

All composite variables are % of trees ('Ct' for count added after % in variable name) or % of basal area ('BA' added after % in variable name) on a plot with various combinations of variables indicating tree stress. Since numerator and denominator were in the same units and numbers were converted to unitless proportions, there was no need to rescale values for composite variables.

- %...DbkVgr are proportion (live trees only) showing stress based on crown dieback or crown vigor class cutoffs. These are strong combined forest stress indexes.
- %...DbkVgrM are proportion (all trees) showing stress based on crown dieback or crown vigor class cutoffs or recent mortality. These are slightly weaker combined forest stress indexes, since mortality was not available for all years.
- %...Stress are proportion (all trees) showing stress based on any of the crown variable cutoffs or recent mortality. No extra weighting was given based on multiple indications of stress to a tree. These are weak combined forest stress indexes, since mortality was not available for all years and crown density and foliage transparency should be compared with species standards.

Greater Sierra Nevada Region Analyses

Sierras results

No correlation of any of the pollution risk indexes with any 'forest stress' index was significant. The strongest was a Spearman rank correlation of *rho* = 0.157 (p=0.121, $rho^2 \approx 0.03$) for [2 x rank of LichenAirGSN + rank of OzoneIBI] with [%BA-DbkVgr (crown dieback ≥10% or crown vigor code 3)].

Despite the lack of significant results, we observed some patterns worth reporting qualitatively to support possible future work. Correlations were stronger for composite pollution risk variables based on ranked data than for composite variables with LichenAirGSN rescaled to match OzoneIBI. Correlations were stronger for LichenAirGSN than for OzoneIBI, as were those for combined indexes weighting LichenAirGSN more heavily. Combined pollution risk indexes gave slightly stronger correlations than either single index. Correlations with 'forest stress' indexes based on crown dieback or vigor (the strongest of our tree stress variables) were stronger than others. Correlations with 'forest stress' indexes based on BA were stronger than those based on tree counts. Composite pollution risk variables from ranked data gave stronger correlations than those from rescaled data, and nonparametric correlations were mostly stronger than parametric correlations. These findings suggest probably most variables failed to meet assumptions for parametric statistics (not tested).

Sierras discussion

The Greater Sierra Nevada region appears to have relatively low S and N air pollution as compared with the New England region (figure 1). However, dry deposition has been shown to be a strong source of S and N pollution more important than wet deposition in West Coast states (Fenn et al 2003, 2008). Moisture may actually mitigate pollution effects on lichens there (Jovan and

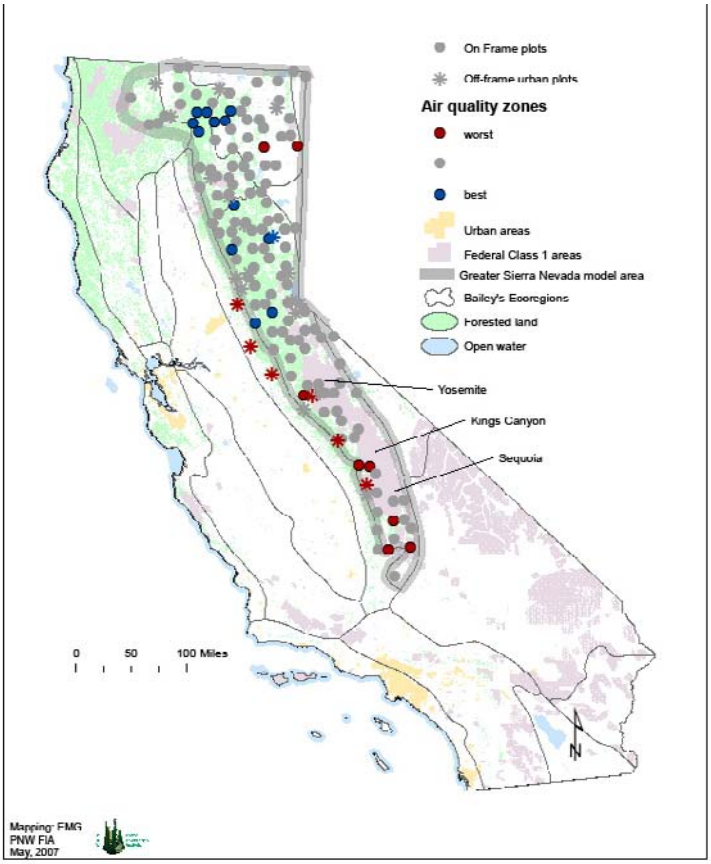


Figure 3: Greater Sierra Nevada Region of California. Data used in this project are from standard (on-frame) FIA plots. Figure modified from Jovan (2008)

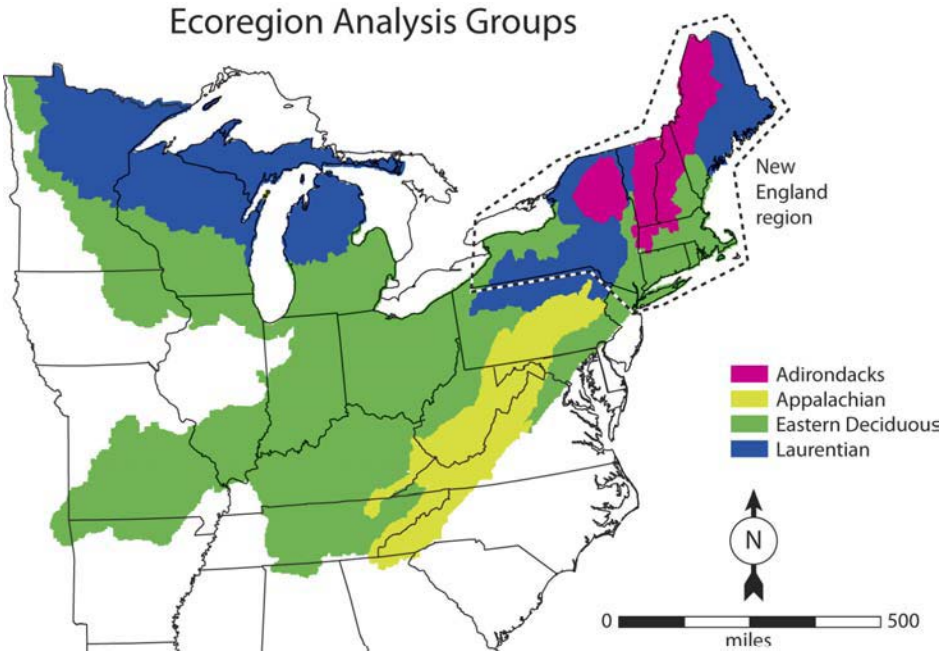


Figure 4: New England region with ecoregion subgroups. (modified from Will-Wolf et al. in review).

Carlsberg 2006). So the apparently lower pollution loads shown in figure 1 for the West Coast as compared with the Northeast might be somewhat misleading. LichenAirGSN is primarily an index of response to agricultural nitrogen pollution (Jovan and McCune 2006).

The Sierras and New England regions appear to have similar ozone risk (figure 2). Ozone is one of several variables correlated with effects on lichens in an adjacent region of California (Jovan and McCune 2005).

Clearly our simple approaches are not adequate to establish there is potential for pollution risk variables to be related to forest stress in this region. Since most of the N pollution is related to agriculture, it may be having a mostly fertilizing effect on forests that is not reflected in our selected measures of tree stress. It might be useful to test more sophisticated estimation of tree stress on a plot. For instance, it might for this small region be useful and not too difficult to calculate tree stress and growth indexes by comparing with species-based standards and aggregate those to plot level for correlation with pollution risk. The presence of so many federal class 1 areas in this region (figure 3) is a reason further research should be conducted on correlation between pollution risk and forest stress indexes, even with little support from this pilot study.

New England Region Analyses

A brief summary of New England region analyses is followed by summaries of results for three ecoregion subgroups (figure 4).

New England region results

New England entire region: A few correlations of 'forest stress' indexes with LichenAirNE-B or composite indexes that include it were statistically significant or approached significance (table 3) for the New England region as a whole. The variation in variables accounted for by the correlations was no more than 1-2% (estimated by r^2); these results have little biological importance. More variation was accounted for by the best nonsignificant correlation for the Greater Sierra Nevada region. Statistical significance for the weak NE region results stems from the large sample size. Correlations for pollution risk indexes and 'forest stress' indexes not included in table 3 were much weaker than those shown.

Despite few significant results and the lack of any biologically important results, we observed some patterns worth reporting qualitatively to support possible future work. LichenAirNE-B in general gave stronger correlations with 'forest stress' indexes than did LichenAirNE-A, suggesting that it is important to control for the effect of regional gradients in climate and air pollution. Correlations were also stronger for LichenAirNE-B than for OzoneIBI or any combined indexes. The one combined index that approached significance included LichenAirNE-B weighted double (table 3). These patterns suggest forest

stress is potentially more strongly related to urban/industrial pollution than to ozone risk. Correlations with 'forest stress' indexes based on crown dieback or vigor were stronger than others. Parametric correlations were stronger than non-parametric correlations. Combined pollution risk indexes based on rescaled data gave stronger correlations than those based on ranked data. Both patterns suggest assumptions for parametric statistics were probably met for New England region data (not tested).

Table 3: New England region: Pearson correlations between selected pollution risk and 'forest stress' indexes. Correlation coefficients are followed by probability. We include r^2 in parentheses to indicate percent of variation involved in the correlation. Dashes indicate lack of significant results.

'Forest stress' indexes	**Pollution risk indexes**	
N = 352	**LichenAirNE-B**	**2 x LichenAirNE-B +OzonelBl**
AvgDBK	0.126 p=0.018 (r^2=0.02)	0.102 p=0.055 (r^2=0.01)
%BA-DbkVgr	0.107 p=0.044 (r^2=0.01)	-
%BA-DbkVgrM	0.118 p=0.027 (r^2=0.01)	-

Table 4: Eastern Deciduous ecoregion group: Pearson correlations between selected pollution risk and 'forest stress' indexes. Correlation coefficients are followed by probability. A p>0.01 is considered significant; a 0.01<p<0.1 is considered approaching to weakly significant. We include r^2 in parentheses to indicate percent of variation involved in the correlation.

'Forest stress' indexes	**Pollution risk indexes**	
N = 80	**LichenAirNE-B**	**2 x LichenAirNE-B +OzonelBl**
AvgDBK	0.240 p=0.032 (r^2=0.06)	0.205 p=0.068 (r^2=0.04)
%BA-DbkVgr	0.341 p=0.002 (r^2=0.12)	0.284 p=0.011 (r^2=0.08)
%BA-DbkVgrM	0.328 p=0.003 (r^2=0.11)	0.273 p=0.014 (r^2=0.07)
%Ct-Stress	0.280 p=0.017 (r^2=0.08)	0.233 p=0.038 (r^2=0.05)
%BA-Stress	0.270 p=0.016 (r^2=0.07)	0.228 p=0.042 (r^2=0.05)

New England ecoregion groups: Based on results for the entire region, we correlated only LichenAirNE-B, Ozone IBI, and combined indexes with 'forest stress' indexes within three ecoregion groups. After excluding plots, only one plot remained in the Appalachian ecoregion. For the Adirondacks and Laurentian ecoregions, all correlations were weaker than for the entire region, and no additional or contrary patterns were observed. So we consider them no further.

For the Eastern Deciduous ecoregion there were several significant correlations. The two strongest correlations accounted for 11-12% of variation between the two indexes; these are strong enough to perhaps approach biological relevance (table 4). The strongest correlation still shows a fairly weak relationship (figure 5), but it does suggest pollution may indeed be a relevant contributor to stress on forests in this region.

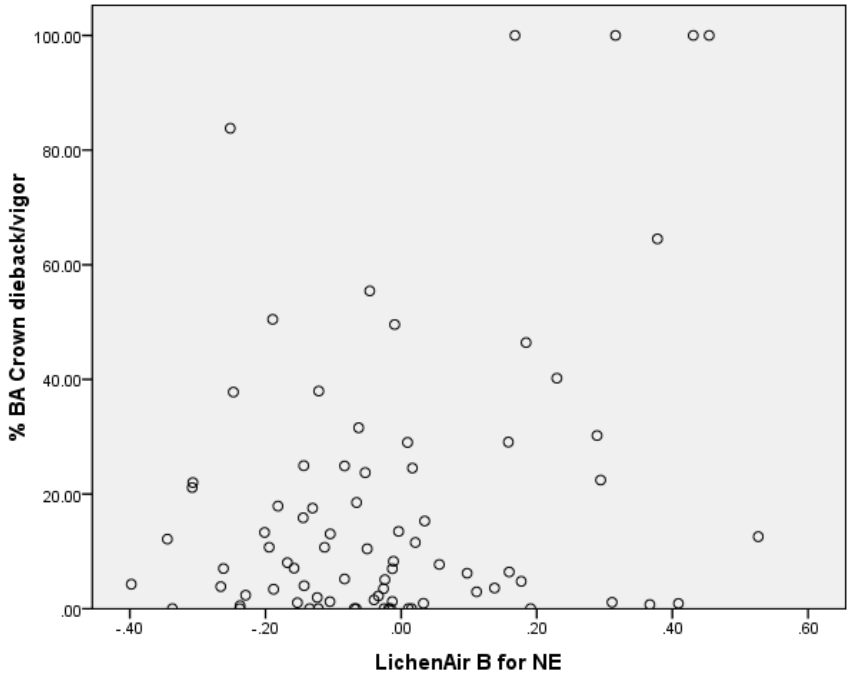


Figure 5: Scatterplot for the 80 Eastern Deciduous ecoregion group plots of LichenAirNE-B vs. %BA-DbkVgr (% of live tree BA with crown dieback ≥10% or crown vigor code 3). The Pearson correlation for this strongest of the within-group relationships (table 4) is 0.341, p=0.002, r^2=0.12.

In general, LichenAirNE-B and combined pollution risk indexes with LichenAirNE-B weighted double had stronger correlations with 'forest stress' indexes than did OzoneIBI and other combined pollution risk indexes. Correlations with LichenAirNE-B were always stronger than for combined pollution indexes. Correlations based on rescaled data were stronger than those based on ranked data. 'Forest stress' indexes based on BA mostly gave stronger correlations than those based on tree counts. The combined index %BA-DbkVgr (from crown dieback and crown vigor, our two best 'tree stress' variables) gave the strongest correlation with pollution risk of all the 'forest stress' indexes for the Eastern Deciduous ecoregion subgroup.

New England region – discussion

For the New England region as a whole and for two of the ecoregion subgroups, the answer to our question is similar to that for the Sierras. Clearly our simple approaches are not adequate to establish there is potential for pollution risk variables to be related to forest stress in most of this region.

In contrast, for the Eastern Deciduous ecoregion subgroup there does indeed appear to be at least some relationship of pollution risk indexes to our simple 'forest stress' indexes. The Eastern Deciduous subregion is subject to heavier pollution loads than the other two subregions, while the Laurentian subregion probably has the widest range of pollution loads (compare figures 1 and 4). Our results suggest absolute pollution load might be the more important factor promoting the finding of a significant correlation between pollution risk indexes and 'forest stress' indexes.

More sophisticated estimation of tree stress on the plot would certainly be useful to calculate. It would for the entire NE region be much more difficult to calculate tree stress and growth indexes by comparing with species-based standards because there are many more tree species recorded in the region. Another possibility might be for the %...Stress indexes to weight contribution of a single tree based on number of stress indications for that tree.

Conclusions and follow-up

This very preliminary pilot study suggests simple approaches to linking pollution risk indexes to forest stress may not work in most regions. However they do show some promise in one subregion with relatively high levels of pollution. Our study was somewhat limited by data availability, and it benefited greatly from consultation with other FIA experts. It does seem it would be worthwhile to pursue development of more sensitive forest stress indicators in follow-up studies.

Data availability and consultation

Limited data availability did affect our project. Temporary unavailability of the ozone and lichen indexes online was a minor inconvenience easily overcome with consultation. Ozone data were made available rapidly, and all lichen data were available to the authors from previous projects. Substantially more tree and lichens data will be available for analysis after existing data mismatches are resolved, especially for 2000-2002 data. Data collected before 2000, when FHM plot data collection was transferred to the FIA program, were one-half to two thirds of the data available and suitable for our pilot study. Any further

documentation of changes and online crosswalks between data sets for the years 1994 to present will enhance the value of all current data for analysis.

Assembling the data sets and data screening were by far the most time-consuming steps for this project. Online tools for selecting and sorting FIA data according to user-defined criteria would substantially aid both FIA and outside analysts to pursue a variety of exploratory investigations. Database scripts for calculating plot-level summaries such as BA, number of trees, and forest composition summaries would aid investigation of plot-level forest health questions.

Consultation with FIA indicator experts about variables other than lichens indexes was critical to the success of this project. Even after all data availability issues are resolved, researchers both inside and outside the FIA program would benefit from substantial consultation with FIA personnel to ask similar questions using FIA data. Online documents explaining use of the ozone indicator were completely adequate to support investigating our question. Consultations with ozone indicator advisor John Coulston (FIA-Southern Region; ~3 hrs; consulted, provided data and maps) and past advisor Gretchen Smith (University of Massachusetts-Amhearst; ~1 hr; consulted) were essential to guide us to ask the right questions and to obtain data. Consultation with crown indicator advisors William Bechtold (FIA-Southern Region; ~1.5 hrs) and KaDonna Randolph (FIA-Southern Region; ~1 hr) was very important to supplement information available in online documents about appropriate use of crown variables. Bechtold also advised on use of growth and mortality variables; little information was available online about appropriate use of these variables for estimating plot-level forest health. Randall Morin (FIA-Northern Region; ~2-3 hr; consulted, provided map and 2000-2002 tree data) consulted on historical use of crown variables and mortality to assess forest stress, and consulted on efficient management of tree data. Elizabeth LaPoint (FIA-Northern Region; ~1 hr) provided crosswalk tables for linking lichens data with tree data (1994-2002 data especially).

Evaluation of project for follow-up

We recommend follow-up investigation soon of linkage between pollution risk and forest stress for the Eastern Deciduous ecoregion, since we found evidence for such linkage in that subgroup of the New England region. There was no evidence from our simple pilot study for such a linkage in the other regions and subregions we investigated.

Results of our pilot project suggest that crown dieback and crown vigor seem to be the best variables to use for calculating plot-level ‘forest stress’ indexes related to pollution risk. Calculating these indexes based on % BA of trees appears to be better than calculating based on % of tree counts. More sophisticated estimation of plot-level forest stress seems essential to address our question in most regions. Relating crown density and foliage transparency to standards for each species seems a logical, but very time-consuming step to take. In some areas it might be worthwhile to try using different cutoff standards for

conifers vs hardwoods. Another possibility is to weight contribution of single trees to plot-level indexes based on the number of stress indications recorded for each tree.

Use of damage variables should be investigated where possible. Calculation of % BA of trees with any kind of damage avoids problems with identifying cause of damage or number of damages. With this approach FHM legacy data and all other available damage data would be appropriate to combine and use in analyses for this purpose.

We found that in the Greater Sierra Nevada region combination pollution indexes seemed better. However, in the New England region a single pollution index seemed to work better. This suggests a general approach will need to be adapted for different regions; neither one kind of pollution risk index, nor one kind of 'forest stress' index is likely to work best for all regions of the USA.

Acknowledgements

As mentioned above, FIA personnel or affiliates William Bechtold, John Coulston, Elizabeth LaPoint, Randall Morin, KaDonna Randolph, and Gretchen Smith each gave help and/or advice critical to the completion of this project. However all conclusions and errors are the responsibility of the authors. Use of Excel and SPSS software does not constitute endorsement by the U.S. Department of Agriculture, Forest Service. This project was supported by funds from two sources: from cooperative agreement SRS 06-CA-11220145-079 between the U.S. Department of Agriculture, Forest Service, Southern Research Station and University of Wisconsin-Madison; also from Joint Venture Agreement 08-JV-11261979-349 between the U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station and Oregon State University.

References

Bechtold W.; Randolph, K. 2007. FIA Crown Analysis Guide. V. 1. 3/6/2007. Pdf file [online] http://srsfia2.fs.fed.us/crowns/index.shtml and http://www.srs.fs.usda.gov/pubs/27730 [November 2008].

Campbell, S.J.; Wanek, R.; Coulston, J.W. 2007. Ozone injury in west coast forests: 6 years of monitoring. Gen. Tech. Rep. PNW-GTR-722. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 53 p

Coulston, J.W.; Riitters, K.H.; Smith, G.C. 2004. A Preliminary Assessment of Montreal Process Indicators of Air Pollution for the United States. Environmental Monitoring and Assessment. 95: 57-74.

Excel. 2003. Microsoft® Office Excel 2003 v. 11.8231.8221 SP3. Part of Microsoft Office Professional Edition 2003. Copyright © 1985-2003 Microsoft Corporation.

Fenn, M.E.; Baron, J.S.; Allen, E.B.; Rueth, H.M.; Nydick, K.R.; Geiser, L.; Bowman, W.D.; Sickman, J.O.; Meixner, T.; Johnson, D.W.; Neitlich, P. 2003. Ecological effects of nitrogen deposition in the western United States. BioScience. 53(4): 404-420.

Fenn, M.E.; Jovan, S.; Yuan, F.; Geiser, L.; Meixner, T.; Gimeno, B.S. 2008. Empirical and Simulated Critical Loads for Nitrogen Deposition in California Mixed Conifer Forests. Environmental Pollution. 155(3): 492-511

Forest Inventory and Analysis Program [FIA]. 2008a. The forest inventory and analysis database: database description and users guide version 3.0 (Phase 3 indicators). U.S. Department of Agriculture, Forest Service, Washington Office. Ozone p 6-49, lichens p 50-80. [online] http://www.fia.fs.fed.us/library/database-documentation/FIADB_user_manual_v3-0.pdf [November 2008].

Forest Inventory and Analysis Program[FIA]. 2008b. The forest inventory and analysis database: database description and users manual version 3.0 for Phase 2, revision 1. U.S. Department of Agriculture, Forest Service, Washington Office. Tree tables p 72-134. [online] http://www.fia.fs.fed.us/library/database-documentation/FIADB_user_manual_v3-0.pdf [November 2008].

Forest Inventory and Analysis National Program [FIA]. 2008c. The forest inventory and analysis database: FIA DataMart. (download data and documentation). [online] http://fiatools.fs.fed.us/fiadb-downloads/fiadb3.html and http://fia.fs.fed.us/ through the 'FIA Data and Tools' link. Online data are updated at intervals. [November 2008].

Forest Inventory and Analysis National Program [FIA]. 2008d. Forest Inventory and Analysis National Core Field Guide. FIA Field Methods for Phase 3 Measurements, version 3.0. [online] http://fia.fs.fed.us/library/field-guides-methods-proc/ and hard copy on file with: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, Rosslyn Plaza, 1620 North Kent Street, Arlington, VA 22209. Field guides are updated at intervals. [November 2008].

Forest Inventory and Analysis National Program [FIA]. 2008e. The forest inventory and analysis program: Other data. (1991-1999 data and documentation) [online]. http://www.fia.fs.fed.us/tools-data/other_data/default.asp and http://fia.fs.fed.us/ through the 'Other Data' link. Online data are updated at intervals. [November 2008].

Forest Inventory and Analysis National Program [FIA]. 2008f. The forest inventory and analysis program: Program features. Forest health indicators. [online]. http://fia.fs.fed.us/program-features/indicators/ [November 2008].

Jovan, S. 2008. Lichen bioindication of biodiversity, air quality, and climate: baseline results from monitoring in Washington, Oregon, and California. Gen. Tech. Rep. PNW-GTR-737. Portland, Oregon: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 124 p. [online] http://www.fs.fed.us/pnw/pubs/pnw_gtr737.pdf [November 2008].

Jovan, S; Carlsberg, T. 2006. Nitrogen content of Letharia vulpina tissue from forests of the Sierra Nevada, California: Geographic patterns and relationships to ammonia estimates and climate. Environmental Monitoring & Assessment 129: 243-251.

Jovan, S.; McCune, B. 2005. Air-quality bioindication in the greater Central Valley of California, with epiphytic macrolichen communities. Ecological Applications. 15: 1712-1726.

Jovan, S.; McCune, B**. 2006.** Using epiphytic macrolichen communities for biomonitoring ammonia in forests of the Greater Sierra Nevada, California. Water, Air, and Soil Pollution. 170: 69–93.

Smith, G.C.; Smith, W.D.; Coulston, J.W. 2007. Ozone bioindicator sampling and estimation. Gen. Tech. Rep. GTR-NRS-20. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 34 p.

Steinman, J. 2000. Tracking the health of trees over time on forest health monitoring plots. In: Hansen, M. and Burk, T., eds. Integrated tools for natural resources inventories in the 21st century; 1998 August 16-20; Boise, ID. Gen. Tech. Rep. NC-212. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 743 p.

Schomaker, M.E.; Zarnoch, S.J.; Bechtold, W.A.; Latelle, D.J.; Burkman, W.G.; Cox, S.M. 2007. Crown-condition classification: a guide to data collection and analysis. Gen. Tech. Rep. GTR-SRS-102. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 78 p. [online] http://www.srs.fs.usda.gov/pubs/27730. [November 2008].

SPSS. 2007. SPSS 16.0 for Windows, standard version, release 16.0.1. SPSS, Inc.

Will-Wolf, S. in review. Analyzing lichen indicator data in the Forest Inventory and Analysis Program. Gen. Tech. Rep. Portland, Oregon: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Will-Wolf, S.; Ambrose, M.J.; Morin, R.S. in review. Relationship of a lichen species diversity indicator to environmental factors across Continental United States. In: Conkling, B.L.; Ambrose, M.J., eds. Forest Health Monitoring 2007 national technical report. Gen. Tech. Rep. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. Ch. 3.

The Power of FIA Phase 3 Crown-Indicator Variables to Detect Change

William A. Bechtold[1]
KaDonna C. Randolph[2]
Stanley J. Zarnoch[3]

ABSTRACT: *The goal of Phase 3 Detection Monitoring as implemented by the Forest Inventory and Analysis Program is to identify forest ecosystems where conditions might be deteriorating in subtle ways over large areas. At the relatively sparse sampling intensity of the Phase 3 plot network, a rough measure of success for the forest health indicators developed for this purpose is the ability to detect meaningful change when indicator data are pooled across two or three States. The statistical power of a test is the probability of detecting a difference of a certain magnitude when it indeed exists, and is defined as 1 minus the type II error (i.e., 1-beta). This study applies statistical power analysis to Phase 3 Crown-Indicator variables (crown density, foliage transparency, and crown dieback) to determine how many plots are necessary to detect various degrees of change at various levels of statistical power.*

KEYWORDS: Statistical power analysis, forest health indicators, tree crown condition, tree crown health, tree health indicators, tree crown measurement

Introduction

The conceptual approach to forest health monitoring in the U.S. includes a component to detect long-term regional changes (Detection Monitoring), a component for assessing the practical importance and impact of observed changes (Evaluation Monitoring), and a component for conducting process-level research of cases that remain unresolved (Intensive Site Monitoring). The goal of Detection Monitoring is to identify forest ecosystems where conditions might be deteriorating in subtle ways over large areas (Riitters and Tkacz 2004). The Forest Inventory and Analysis (FIA) Phase 3 sampling frame (Bechtold and Patterson 2005) contributes to this goal by systematically sampling a set of forest health indicators over space and time. Core tables used in FIA reports (e.g., Turner and

[1] Research Forester, USDA Forest Service, Southern Research Station, Asheville NC.
[2] Mathematical Statistician, USDA Forest Service, Southern Research Station, Knoxville TN.
[3] Mathematical Statistician, USDA Forest Service, Southern Research Station, Asheville NC.

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

others 2008) document the status and trends associated with these indicators, and analysts look for signals that suggest potential problems.

Most of the indicators currently implemented on FIA Phase 3 plots were developed by the Forest Health Monitoring (FHM) Program in collaboration with the U.S. Environmental Protection Agency (Alexander and Palmer 1999). One of the original goals for indicator development was to provide statistically unbiased estimates of status, trends, and relationships with quantifiable confidence limits over regional and national scales (Hunsaker and Carpenter 1990). The national strategic plan of the FHM program[4] states that Detection Monitoring is designed to identify forest health changes at multiple spatial scales to adequately describe disturbance events that vary in magnitude. Riitters and Tkacz (2004) say that "Detection Monitoring accepts a high rate of false positives (i.e., a high Type I error rate) as the price of not overlooking change (i.e., a low Type II error rate)."

Statistical power analysis (Castelloe 2000) is one practical tool that can be used to evaluate statistical rigor, identify the spatial scale at which an indicator is functional, and address Type I and Type II error rates. This technique clarifies uncertainties about the ability of an indicator to detect a significant effect by examining interactions among the factors contributing to statistical power. Of particular interest to FIA, power analysis can identify areas or situations where sampling intensity is insufficient. Where grid intensification is an option, it can help to establish the appropriate level of intensification. When migrating Phase 3 indicators to the Phase 2 grid is under consideration, it can quantify the benefits to be gained.

Although the application of power analysis to forest health indicators has been under-utilized, it has not been ignored completely. Smith and others (1996) used it to evaluate the benefits of adding an annual overlap to panel rotation schedules when the FHM Program was collecting indicator data. Conkling and others (2002) conducted a power analysis to evaluate the use of FHM mensuration and soils data to detect changes in forest carbon budgets.

The objective of this paper is to evaluate the statistical power of FIA's Crown-Condition Indicator variables (Schomaker and others 2007) and relate the results to the Phase 3 sampling grid.

Methods

Power Analysis

Environmental monitoring often involves hypothesis testing, which is a statistical procedure designed to test the premise that a population displays some

[4] Available at http://fhm.fs.fed.us/annc/strategic_plan03.pdf [Date accessed: July 28, 2008].

effect (H_1), against the null hypothesis that there is no effect (H_0). In Detection Monitoring, a typical alternative hypothesis is that there is some difference in indicator values between groups of observations drawn from a population at different points in time. Or the groups may be from the same time period but represent different regions, different species, or stands that have experienced different disturbances.

There are two types of error associated with hypothesis testing. Type I error is a false positive, where the null hypothesis is rejected and the test incorrectly concludes there is some significant effect. The probability of a Type I error is designated as alpha (α). The value of α is selected by the analyst, who decides what the risk of Type I error should be. Alpha commonly is set at 0.05, which restricts the chance of a false positive to 5 percent or less. Type II error is a false negative, where the test misses a true problem by failing to reject a false null hypothesis. The probability of a Type II error is designated as beta (β). The statistical power of a test is the probability of detecting a difference of a certain magnitude when it indeed exists, and is defined as ($1-\beta$).

Most statistical tests are designed to minimize α, while β is often overlooked. Although the consequences of a false positive may be costly in terms of initiating unnecessary Evaluation Monitoring studies, the failure to recognize a significant forest health problem could be disastrous. In Detection Monitoring, it is important to know if the statistical power of an indicator is adequate to support acceptable α and β probability levels.

We used the SAS POWER procedure (SAS Institute 2004) to examine the relationships among Type I and II error, sample size, data variability, and effect size. We chose the SAS TWOSAMPLEMEANS option to determine the number of plots necessary to detect a difference in mean crown indicator values between two independent samples. Independent samples would come into play when checking for differences among regions during the same time period, or when checking for differences among independent panels[5] at different times. We selected the SAS PAIREDMEANS option to determine the number of plots necessary to detect a difference in mean values between paired observations. The paired approach would be of interest when checking for differences involving survivor trees from remeasured plots.

A variety of input specifications, outputs, and options are available for both the independent and paired procedures. The output of interest for this analysis is the number of plots. The inputs are the α probability level, the power level ($1-\beta$), the effect size (mean difference between groups), the underlying data distribution (normal vs. log normal), the standard deviation of the observations; and, for the paired analysis, the correlation coefficient. Any of these input values can be manipulated to determine the subsequent effect on the number of plots required to attain the input specifications. The desired α probability value, power level, and

[5] Details about the FIA panelized inventory system are available in Bechtold and Patterson (2005).

effect size are simply specified by the analyst. The underlying data distributions, standard deviations, and correlation coefficients can be estimated from data already on hand, or from a pilot study.

Crown-Condition Data

This analysis focuses on the three main indicators of tree crown health recorded on FIA Phase 3 plots—crown density, foliage transparency, and crown dieback. Crown density is the amount of crown biomass (crown stem, branches, twigs, shoots, buds, foliage, and seeds) that blocks light penetration through the crown. Foliage transparency is the amount of skylight visible through small holes in the live portion of the crown where foliage occurs. Crown dieback is recent mortality of branches with fine twigs that begins at the terminal portion of a branch and proceeds inward toward the trunk. These three indicators are analyzed as continuous variables, although they are actually recorded in 5-percent classes, ranging from 0 to 99 percent. More details about the indicators are available in Schomaker and others (2007); additional details about related FIA sampling protocols are available in the Phase 3 field guide (U.S. Department of Agriculture Forest Service 2007).

The data distributions, standard deviations, and correlation coefficients used in this analysis were obtained from FHM plots measured between 1992 and 1999. FHM data were used because more remeasured plots were available from which to obtain correlation coefficients for paired plots. Although the FHM plot network was integrated into the FIA sampling grid in 2000, some regions still have limited numbers of remeasured panels. For the purpose of power analysis, only approximations of the data attributes are needed, and the measurement protocols for the crown indicators measured by FIA remain unchanged from those used by FHM.

By 1999, the FHM Program was measuring plots in 24 eastern and 8 western States. For the East, the latest measurements of each plot sampled between 1994 and 1999 were used to calculate standard deviations from independent observations. To obtain correlation coefficients from paired observations, the latest measurement of each plot sampled during this period was paired with its previous observation, which was usually 4 years earlier. However, due to differences in the way States and panels were implemented, the remeasurement period was shorter than 4 years in some States. For the West, the latest measurements of each plot sampled between the years 1992 and 1999 were used for the independent observations. Again, the latest observation of each plot measured during this period was paired with its previous observation, which ranged from 1 to 4 years depending on the State. This plot selection method was designed to maximize the numbers of observations available to estimate standard deviations and correlation coefficients of crown indicators by species group; the result was not constrained to represent the actual distribution of species groups in

the populations of eastern and western forests. Correlations were slightly higher between plot pairs measured less than 4 years apart, but judged insufficient to preclude the use of the additional data for the purpose of power analysis.

Power analyses are most commonly applied to experimental designs, but extension of the technique to survey designs is appropriate when the intention is to use survey data for hypothesis testing. Some adaptation was necessary to accommodate FIA's complex design. The ratio-of-mean estimators used by FIA produce variance estimates that accommodate the unbalanced clustering of trees on plots, as well as the use of partial plots in the estimation process (Bechtold and Patterson 2005). However, calculations of standard deviations and correlation coefficients from this design are not as straight-forward. We approximated these statistics by first calculating the mean crown condition for each plot, and then computing the standard deviations and correlation coefficients from the plot-level means. We compared the variance estimates from both methods and obtained similar results, so the simplified approach should have a negligible affect on prospective hypothesis tests and the power analyses associated with them.

Results

Crown Density

Crown density in the East averages 47 percent with a standard deviation of 7.5 (table 1). By species group, the means range from 42 to 53 percent, with most standard deviations clustered around 10 or 11. Crown density statistics are similar in the West, except the standard deviations are slightly higher (table 2). Correlation coefficients between paired plots, where mean crown density was computed from survivor trees, are also similar between the two regions—equaling 0.42 in the East and 0.48 in the West. Histograms of data distributions and skewness statistics indicated that crown density was approximately normally distributed.

Table 3 shows the numbers of plots required to detect differences in crown density for various combinations of power-test specifications that conform to the ranges of statistics presented in tables 1 and 2. The effect size (mean difference) was set to detect changes in crown density values from 10 to 15, with standard deviations also fluctuating between 10 and 15. The α specification alternates from 0.01 to 0.05, and the power level $(1 - \beta)$ from 0.8 to 0.9. Given these parameters, the number of plots necessary to detect a change in crown density from independent observations ranges from 18 to 138, depending on the stringency of the specifications. When paired observations are analyzed, the number of plot pairs needed to detect the specified effects ranges from 8 to 54 when the correlation coefficient is 0.25, and from 6 to 37 when the correlation

coefficient is set to 0.5. The SAS code used to produce table 3 is provided in Appendix A.

Foliage Transparency

Foliage transparency in the East averages 18 percent with a standard deviation of 6 (table 1). By species group, the means range from 13 to 21 percent, with most standard deviations between 5 and 10. The correlation between paired observations was less pronounced than the results for crown density, with a correlation coefficient of 0.19 for all species combined. The correlation coefficients for many species groups were not significant at the 0.05 level. Foliage transparency statistics for the West (table 2) were very similar to the East.

Histograms of the foliage transparency data revealed some skewness caused by outliers in the right tail, and it was not clear whether this variable was normally distributed. Skewness tests indicated that the distribution was not normal, but with large numbers of observations, statistical tests often find deviations from normality that are statistically significant but practically unimportant. Choosing the alternative that yields conservative results is usually the most prudent course of action in borderline cases. We chose the normal classification because more observations were required to attain the specified power than the non-normal alternative. Randolph (2006) also noted that the distribution of this variable was uncertain, but justified a normal classification on the basis that statistical tests are robust against skewness and outliers.

The same α and power $(1-\beta)$ specifications used for crown density were also used for transparency, but the standard deviations were reduced to 5 and 10 (table 3). The effect size was also reduced to detect changes that fluctuate from 5 to 10 because mean transparency was much smaller than mean crown density (approximately half). Ideally, effect size should be tied to biological thresholds, but none are currently available for many forest health indicators (including the crown variables), so expert judgment must be substituted. Given these specifications, the number of plots necessary to detect a change in foliage transparency from independent observations ranges from 12 to 242. For paired observations, the number of plot pairs needed to detect the specified effects ranges from 6 to 93, when the correlation coefficient is 0.25, and from 5 to 63 when the correlation coefficient is 0.5.

Crown Dieback

Crown dieback differs from density and transparency in that most trees have little or no dieback, so the data have a severely skewed log normal distribution. Histograms and skewness statistics corroborate this. The SAS POWER procedure

requires different input variables when the underlying data distribution is log normal (see Appendix A). For the specification of effect size, the ratio of the two group means is substituted for the mean difference, and the coefficient of variation (CV) is substituted for the standard deviation.

Dieback in the East averages 4.0 percent with a CV of 1.3 (table 1). By species group, mean dieback varies from 1.6 to 5.9 percent, with CVs ranging from 1.0 to 3.5. Crown dieback means and coefficients of variation are similar in the West (table 2). At 3.3 percent, mean crown dieback is slightly lower in the West because there is a lower proportion of hardwood species there, and dieback tends to be more prevalent in hardwoods.

The same α and power $(1-\beta)$ specifications used previously are used here again, but a relatively large mean ratio is required to detect a meaningful effect when the means are close to 0. The mean ratios were thus set to 2 and 2.5 (specifying changes of 200 and 250 percent), while the CVs varied from 1.5 to 2. Given this input, the number of plots necessary to detect a change in crown dieback from independent observations spans from 48 to 204 (table 3). When paired observations are analyzed, the number of plot pairs needed to detect the specified effects range from 16 to 61 when the correlation coefficient is 0.25, and from 11 to 36 when the correlation coefficient is set to 0.5.

Relating Results to the FIA Plot Network

Once the number of plots needed to detect a significant difference is known, relating this information to the FIA plot network is simplified by the systematic nature of the grid, where the base sampling intensity is 1 plot per 6,000 acres for Phase 2 and 1 plot per 96,000 acres for Phase 3.

The following formula can use this information to determine how much area must be impacted by a forest health problem before the FIA plot network is able to detect it:

$$IA = n(E)\frac{P_t}{P_n} \qquad (1)$$

where

$IA =$ the minimum size of an impact area detectable by the FIA plot grid in the area of interest,

$n =$ the number of plots or plot pairs (i.e., grid points) needed to detect an impact (from the power analysis). For independent observations, n is the number of plots; for paired observations n is the number of plot pairs.

$E =$ the plot expansion factor in the area of interest (e.g., 6,000 for Phase 2 or 96,000 for Phase 3),

P_t = the total number of panels per measurement cycle in the area of interest, and
P_n = the number of available panels in the area of interest.

When the detectable impact area is known, basic FIA area statistics can be used to evaluate whether the grid intensity is sufficient to recognize the problem. For example, suppose that we want to know if the sampling grid is adequate to detect a change in crown dieback from remeasured plots. Because mean crown dieback is low, usually less than 5 percent, the analyst decides that a doubling of mean crown dieback should not be overlooked. This could mean that a lot more trees have a little dieback, or some subset of trees developed a lot of dieback. Scrutinizing the data from tables 1 and 2, we see that the coefficient of variation for dieback averages about 1.8, and the correlation coefficient between paired observations is usually around 0.3. Plugging these numbers into a power analysis where α is fixed at .05 and the power $(1-\beta)$ is set to 0.9, we determine that it will take at least 36 plot pairs to flag the specified change as statistically significant. This number can then be used for *n* in equation 1. The crown indicator is measured on the Phase 3 plot network, so $E = 96{,}000$. FIA has 5 panels in most Eastern States, so $P_t = 5$, and we want the results to apply to a full set of remeasured panels, so $P_n = 5$. Entering these values into equation 1 shows there is sufficient power to detect an impact affecting 3.5 million acres of forest (table 4).

The first two columns in table 4 list the current distributions of total land and forest land by region and State from the 2007 RPA statistics[6]. If we divide the detectable impact area for crown dieback (3.5 million acres) by the forest area in column 2, we can see what percentage of the plot grid in a given State or region must be impacted in order to detect a statistically significant difference. Wherever this number is more than 100 percent, the plot grid in the area of interest is not sufficient to detect the specified impact. The sampling intensity for crown dieback is sufficient in most States, but there are several States in the Northeast and in the Great Plains where the grid is obviously insufficient. The sampling intensity of Rhode Island would have to be increased by a factor of 10 to be able to detect the specified change.

Table 4 includes similar analyses for crown density and foliage transparency, where the data in tables 1 and 2 were scrutinized closely to select one set of power-analysis input values that represent a plausible scenario for each crown variable. The input specifications for plausible scenarios for these two variables revealed that 21 plot pairs were needed to detect an unusual change in density, and 31 were needed for transparency. These two variables required fewer plots than crown dieback, but still the grid intensity was not adequate in a few States with relatively little forest.

[6] Available at http://fia.fs.fed.us/program-features/rpa/default.asp [Date accessed: July 28, 2008].

Discussion

Adequacy of the Phase 3 Plot Network

When the crown indicators were being developed during the 1990s it was speculated that the sampling intensity was probably adequate for regions or groups of States, but not sufficient to analyze individual States. Results from tables 3 and 4 indicate that about 100 forested plots would be adequate to identify a problem in a majority of prospective change scenarios when independent observations are used. About 50 plots are adequate when using plots with paired observations from two points in time. On the Phase 3 plot network, 50 plots translate into 4.8 million acres. There are 15 States that have less than 4.8 million acres of forest, so each of these States would have to be combined with other States to yield a minimally adequate sample. Considering that crown indicators are often analyzed by species group, and that any given species occurs only on a subset of the total forest plots, then the original conjecture is confirmed that the sampling intensity is not adequate for analysis of many individual States. Power analysis verifies that the FIA plot grid yields an adequate sample at the regional level, with the possible exception of the Great Plains, where 60 percent of the forest would have to be affected. In other regions, less than 4 percent of the forests would need to be impacted in order to detect a problem (table 4).

Equation 1 can be incorporated directly into an Excel spread sheet set up like table 4 to quickly check the adequacy of the sampling grid for a wide variety of scenarios. As presented, table 4 shows how much intensification, if any, is needed in each State to support analyses at the State level. This information can be used to guide State-level intensifications. Moving some of the forest health indicators to the Phase 2 sampling grid has lately been of interest. Changing the expansion factor (E) in equation 1 from 96,000 to 6,000 yields a measure of the additional power to be gained from this modification.

The detectable impact areas in table 4 are based on a full complement of panels. At current rates of implementation it takes 5 to 10 years (depending on the State and region) to obtain a complete measurement cycle, and twice as long to obtain paired observations. So there is a period of vulnerability where sample size may be inadequate during the 10 to 20 years needed to ramp up to a full complement of remeasured panels. Equation 1 can be used to investigate the issue by reducing the number of panels available for analysis.

The grid adequacy examples discussed so far are all related to the overall grid, where plots are contiguous over States and regions. The power analysis can be refined to more specific, non-contiguous plots by applying equation 1 to almost any FIA area table, such as area by forest type. This approach informs the analyst if there is enough power to analyze a specific forest type (or any other area-related attribute such as elevation or disturbance history).

The detectable impact areas discussed thus far are based on forest area. The detectable impact area can be converted to total land area by dividing the detectable forest area by the percentage of forest in the area of interest. For example, 3.5 million acres of *forest* must be impacted by the crown dieback example in table 4. This means that a problem which is not confined to forests (e.g., climate change) must be impacting a larger area than just the forest. In the Northeast, where the forests constitute 67 percent of the total land area, about 5.2 million acres of total land area must be affected to recognize the problem. This information could be used in simulation models, where the FIA grid might be checked for adequacy to recognize problems dispersed across the total landscape.

Independence

FIA plots can contribute observations to more than one group, so it is important to consider whether the groups being compared consist of independent, paired, or mixed observations. Independence is guaranteed only when the groups are drawn from mutually exclusive sets of plots, which is usually the case when comparing different regions or different panels. Purely paired analyses are achieved only when each plot is required to contribute an observation to both groups, which is commonly the case when the groups are based on survivor trees from plots measured at two points in time. In other situations, the groups will be a mixture of independent and paired observations. For example, a plot can experience multiple disturbances. For an analysis designed to compare two disturbances, a single plot has the potential to contribute an observation to one or both disturbance groups, resulting in a mixture.

The SAS TWOSAMPLEMEANS option is most appropriate for independent observations, but setting the correlation coefficient to 0 in the PAIREDMEANS option yields approximately the same result. So PAIREDMEANS can be modified to accommodate mixtures of independent and paired observations by adjusting the correlation coefficient. For example, if 40 percent of the observations in a prospective analysis are expected to be paired, the correlation coefficient could be reduced to 40 percent of the value that would be used if all plots were paired. Keep in mind when interpreting results that PAIREDMEANS presents output in terms of plot pairs (*npairs*). When mixtures of independent and paired plots are analyzed with PAIREDMEANS, the total number of grid points (i.e., independent plots plus plot pairs) necessary to support the analysis can be calculated by adjusting *npairs* for the proportion of paired plots:

$$n = 2(npairs)(1 - prop) + (npairs)(prop) \qquad (2)$$

where

npairs = the number of paired plots output from PAIREDMEANS, and

prop = the expected proportion of paired plots in a prospective analysis.

All of the results presented here assume that numbers of observations are balanced between groups. Statistical power decreases as designs become unbalanced. This is not an issue with paired observations, but FIA's survey design can become unbalanced when analyses are based on independent or mixed observations. For independent observations, the TWOSAMPLEMEANS procedure has a weighting feature to accommodate unbalanced designs. This feature is not available with PAIREDMEANS, so the effects of an unbalanced design are not readily obtainable when mixtures of independent and paired observations are processed with PAIREDMEANS. Sensitivity to imbalance can be approximated with TWOSAMPLEMEANS, and estimated power reductions could then be extrapolated to results from PAIREDMEANS. For example, the number of plots required for independent observations in table 3 increases by an average of 11 percent when the sample size in each group is unbalanced by a ratio of 2:1. The number of plots resulting from a similar analysis of mixed observations with PAIREDMEANS would then be increased by this percent.

Strategies to Increase Statistical Power

Increasing sample size is one obvious way to gain statistical power. Grid intensification is the preferred approach if economically feasible. Sample size can also be increased by combining areas, but this must be done carefully. Adding adjacent States will increase power if those areas have the same problem as the original location. If not, the analysis could be compromised because adding unaffected areas will dilute any effect that may have been present at the original location. Sometimes adding more territory is not an option. For example, Hawaii is so isolated that it would not be practical to combine Hawaii with any other State. Given this, it is noteworthy that the base Phase 3 sampling intensity in Hawaii would not support an analysis of crown indicators (table 4).

Power can also be gained by tailoring analyses to more specific subsets of the data through stratification. Forest health impacts that have "clumped" distributions display a higher variance than impacts that are evenly dispersed. Statistical power is adversely affected by increased variance. Stratification can help by reducing variation within strata. If a clumped distribution is suspected, post-stratify the data to isolate the clumps if possible.

Power might be gained by taking advantage of the "double sampling for stratification" technique that FIA uses for standard inventory estimates (Bechtold and Patterson 2005). This technique reduces the variance of many FIA inventory estimates, and variance reduction always increases statistical power. However, the Phase 3 indicators have not yet been incorporated into FIA's processing engine, so the degree to which this procedure might reduce the variance of crown indicators is uncertain.

Species groups are another important aspect of stratification. Grouping species together has the potential to increase power by increasing the number of plots available for analysis, but a forest health problem that impacts only a few species in a group may go undetected. If a particular species is known to be sensitive to a prospective threat, that species should be isolated. Care must always be taken to avoid different species mixes between the two groups of observations being tested, because effect size could be an artifact of the difference in species mix.

The results presented here are based on two-tailed tests, where the alternate hypothesis (H_1) is that the mean of Group 1 differs from the mean of Group 2, so the analyst can determine whether an indicator is improving or deteriorating. Additional power can be attained by using one-tailed tests, where H_1 is modified to specify that the mean of Group 2 has either increased or decreased. On average, when one-tailed tests are used, the number of plots is reduced to 83 percent of the required observations listed in table 3. Note that the alternative hypotheses (H_1) should be specified prior to any testing. It is not valid to look at the data and then decide if the variable of interest should be increasing or decreasing.

Conclusions

Comprehensive power analyses of FIA's Phase 3 indicators are possible with only three pieces of information derived from actual data—the indicator's frequency distribution (normal vs. log normal), its standard deviation, and a correlation coefficient for paired observations. The analyst can then solve for detectable effect size, α level, statistical power $(1-\beta)$, or numbers of observations required. Supplied with the number of plots required to detect a significant effect, the adequacy of the FIA grid to support prospective analyses can be readily evaluated.

For most plausible scenarios involving the crown indicator, about 100 plots (or 50 paired plots) would be adequate. Paired plots resulting from repeated observations over time reduces the number of grid points to less than half the observations that would be required from independent observations. These results are consistent with previously unverified assumptions that the crown indicators as implemented on the Phase 3 grid could reliably support analyses at the regional level, but probably not for many individual States.

We recommend that power analyses be conducted for all indicators where statistical hypothesis testing is an option. The information needed to accomplish this is readily available from existing datasets for indicators that have already been implemented. As new indicators are proposed, power analysis should become a formal part of the indicator development process.

Literature Cited

Alexander, S.A.; Palmer, C.J. 1999. Forest Health Monitoring in the United States: first four years. Environmental Monitoring and Assessment 55: 267-277.

Bechtold, W.A.; Patterson, P.L. eds. 2005. The enhanced Forest Inventory and Analysis Program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 85 p.

Castelloe, J.M. 2000. Sample Size Computations and Power Analysis with the SAS System Paper 247. Cary, NC: SAS Institute. Available at: http://www.ats.ucla.edu/stat/sas/library/powersamplesize.pdf [Date accessed: July 28, 2008].

Conkling, B.L.; Hoover, C.M.; Smith, W.D.; Palmer, C.J. 2002. Using forest health monitoring data to integrate above and below ground carbon information. Environmental Pollution: Vol. 116, Supplement 1, March 2002, Pages S221-S232.

Hunsaker, C.T.; Carpenter, D.E. (Eds.), 1990. Environmental Monitoring and Assessment Program. Ecological Indicators. EPA 600/3-90/060. US Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.

Randolph, K.C. 2006. Descriptive statistics of tree crown condition in the Southern United States and impacts on data analysis and interpretation. Gen. Tech. Rep. SRS-94. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 17 p.

Riitters, K.; Tkacz. B. 2004. Forest health monitoring. pp. 669-683 in Wiersma, B., ed. Environmental Monitoring. CRC Press, Boca Raton, FL. 792 p.

SAS Institute Inc. 2004. SAS OnlineDoc® 9.1.3. Cary, NC: SAS Institute Inc.

Schomaker, M.E.; Zarnoch, S.J.; Bechtold, W.A.; Latelle, D.J.; Burkman, W.G.; Cox, S. M. 2007. Crown-condition classification: a guide to data collection and analysis. Gen. Tech. Rep. SRS-102. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 78 p.

Smith, W.D.; Gumpertz, M.L.; Catts, G.C. 1996. An analysis of the precision of change estimation of four alternative sampling designs for forest health monitoring. Forest Health Monitoring Tech. Rep. Series. Research Triangle Park, NC: Forest Health Monitoring National Office. 24 p.

Turner, J.A.; Oswalt, C.M.; Chamberlain, J.L [and others]. 2008. Kentucky's Forests, 2004. Resour. Bull. SRS-129. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 101 p.

U.S. Department of Agriculture Forest Service. 2007. Forest Inventory and Analysis National Core Field Guide (Phase 3), version 4.0. Washington, DC: U.S. Department of Agriculture Forest Service, Forest Inventory and Analysis, Washington, D.C.

Available at: http://socrates.lv-hrc.nevada.edu/fia/dab/databandindex.html#4.%20%20Current%20National%20Core%20Field [Date accessed: July 28, 2008].

Appendix A. SAS Power Analysis Program

```
/*
This SAS program performed the Power Analysis resulting in Table 3.

This program produces numbers of observations (ntotal or npairs) as the
dependent variable. The analysis can be revised to make any of the associated
variables (power, alpha, stddev, cv, meandiff, meanratio, or corr) into the
dependent variable by specifying the number of observations available and setting
the dependent variable of interest to null.
*/

title1 'Crown Density - independent observations';
     proc power; twosamplemeans
     outputorder  = syntax
      test        = diff
      dist        = normal
     power        = .80 .90
     alpha        = .01 .05
     stddev       = 10 15
     meandiff     = 10 15
     ntotal       = .
     ;
title1 'Foliage Transparency - independent observations';
      proc power; twosamplemeans
     outputorder = syntax
      test        = diff
      dist        = normal
     power        = .80 .90
     alpha        = .01 .05
     stddev       = 5 10
     meandiff     = 5 10
     ntotal       = .
     ;
title1 'Crown Dieback - independent observations';
     proc power; twosamplemeans
     outputorder = syntax
      test        = ratio
     dist         = lognormal
     power        = .80 .90
     alpha        = .01 .05
     cv           = 1.5 2
     meanratio    = 2 2.5
     ntotal       = .
     ;
```

```
title1 'Crown Density - paired observations';
    proc power; pairedmeans
    outputorder = syntax
    test        = diff
    dist        = normal
    corr        = .25 .5
    power       = .80 .90
    alpha       = .01 .05
    stddev      = 10 15
    meandiff    = 10 15
    npairs      = .
    ;

title1 'Foliage Transparency - paired observations';
    proc power; pairedmeans
    outputorder = syntax
    test        = diff
    dist        = normal
    corr        = .25 .5
    power       = .80 .90
    alpha       = .01 .05
    stddev      = 5 10
    meandiff    = 5 10
    npairs      = .
    ;
title1 'Crown Dieback - paired observations';
    proc power; pairedmeans
    outputorder = syntax
    test        = ratio
    dist        = lognormal
    corr        = .25 .5
    power       = .80 .90
    alpha       = .01 .05
    cv          = 1.5 2
    meanratio   = 2 2.5
    npairs      = .
    ; run;
```

Table 1. Means (m), standard deviations (s), coefficients of variation (cv), and correlation coefficients (r) of crown data collected by the Forest Health Monitoring Program in 24 eastern States[a], by species group, 1994-1999.

Species Group	Independent[b] observations		Paired[c,d] observations		Crown density			Foliage transparency			Crown dieback		
	n trees	n plots	n trees	n plots	m	s	r[e]	m	s	r[e]	m	cv	r[e]
Softwoods:													
South. yellow pine	6,648	441	2,729	219	42.1	8.6	0.37 *	18.8	7.1	0.41 *	2.1	1.9	0.46 *
East. white pine	1,157	180	604	99	48.6	10.6	0.51 *	20.4	5.8	0.29 *	2.7	1.5	0.72 *
Red pine	521	37	347	27	50.3	10.5	0.31	18.4	4.0	0.24	1.6	1.5	0.08
Spruce/fir	2,459	258	1,720	193	51.6	10.3	0.61 *	16.7	4.6	0.21 *	4.4	1.5	0.18 *
Hemlock	776	130	408	69	53.1	10.8	0.64 *	17.7	5.7	0.35 *	3.0	1.3	0.67 *
Other Softwoods	1,867	241	1,462	142	48.5	13.7	0.36 *	19.4	7.1	0.27 *	5.0	1.6	0.38 *
Hardwoods:													
Oak	6,088	855	2,532	410	47.0	8.4	0.32 *	17.0	6.2	0.15 *	4.2	1.3	0.24 *
Elm	576	228	229	99	44.1	11.1	0.30 *	20.2	7.8	0.30 *	5.9	1.9	0.35 *
Hickory	1,196	391	401	137	51.5	10.4	0.48 *	14.8	6.3	-0.06	2.5	2.3	0.14
Birch	1,724	416	1,043	263	50.6	9.9	0.40 *	18.0	5.4	0.15 *	4.5	1.1	0.28 *
Maple	6,481	916	3,428	502	48.4	9.1	0.46 *	16.8	5.5	0.15 *	3.9	1.5	0.55 *
Beech	730	201	320	95	51.1	11.6	0.58 *	15.3	4.5	0.17	3.7	1.5	0.30 *
Sweetgum	1,032	262	440	126	48.2	10.9	0.41 *	13.4	5.7	0.24 *	3.0	2.5	-0.01
Tupelo-blackgum	690	237	341	102	47.1	11.8	0.62 *	16.1	9.5	0.03	3.5	2.9	0.81 *
Ash	1,303	340	680	168	46.7	11.6	0.50 *	19.0	8.4	0.67 *	5.0	2.1	0.82 *
Quaking Aspen	1,113	192	801	136	47.2	10.3	0.45 *	21.4	5.7	0.07	4.9	1.0	0.51 *
Basswood	388	98	277	63	48.3	10.7	0.44 *	18.4	6.1	-0.10	4.1	2.2	0.24
Yellow Poplar	1,022	282	498	141	51.7	10.2	0.16	15.3	7.7	0.08	2.1	3.5	0.20 *
Walnut	114	65	44	27	44.4	12.9	0.51 *	19.3	11.6	0.57 *	5.3	3.1	0.00
Other Hardwoods	2,900	770	1,310	367	45.5	10.6	0.36 *	18.9	8.1	0.12 *	4.5	1.9	0.18 *
All softwoods	13,428	957	7,270	549	46.9	10.8	0.58 *	18.6	6.4	0.37 *	3.0	1.6	0.46 *
All hardwoods	25,357	1,515	12,344	813	47.7	7.8	0.43 *	17.4	6.3	0.35 *	4.5	1.5	0.47 *
All species	38,785	1,625	19,614	890	47.2	7.5	0.42 *	17.7	5.6	0.19 *	4.0	1.3	0.36 *

[a] States included: AL, CT, DE, GA, IL, IN, ME, MD, MA, MI, MN, MO, NH, NJ, NY, NC, PA, RI, SC, TN, VT, VA, WV, and WI.

[b] Independent observations from the latest set of 4 panels sampled between 1994 and 1999 in each State were used to calculate means (m), standard deviations (s), and coefficients of variation (cv).

[c] Paired observations from the latest set of remeasured panels sampled between 1994 and 1999 in each State. Plot means from survivor trees were paired with their previous values to calculate correlation coefficients (r). The number of remeasured panels available varies by State.

[d] The numbers of observations represent pairs. The total numbers of observations are twice the listed numbers of pairs.

[e] Asterisks indicate significant correlation coefficients, where p-values are $\leq .05$.

Table 2. Means, standard deviations (s), coefficients of variation (cv), and correlation coefficients (r) of crown data collected by the Forest Health Monitoring Program in 8 western States[a], by species group, 1992-1999.

Species Group	Independent[b] observations		Paired[c,d] observations		Crown density			Foliage transparency			Crown dieback		
	n trees	n plots	n trees	n plots	m	s	r[e]	m	s	r[e]	m	cv	r[e]
Softwood timberland:													
Douglas fir	3,932	380	2,003	219	50.8	11.4	0.43 *	14.5	4.0	0.10	1.8	1.8	0.23 *
Ponderosa pine	1,467	192	985	134	46.4	11.4	0.42 *	17.3	6.1	0.24 *	2.1	2.9	0.08
True fir	3,084	315	1,915	204	51.6	12.4	0.43 *	13.5	4.6	0.11	2.6	1.7	0.48 *
Western hemlock	962	85	455	45	51.5	10.7	0.38 *	13.9	3.4	0.12	1.7	1.8	0.52 *
Sugar pine	102	27	76	22	47.8	9.6	0.25	18.1	6.9	0.03	1.1	2.0	-0.21
Western white pine	82	30	46	17	48.8	12.2	0.69 *	18.4	5.3	0.58 *	3.8	1.4	0.65 *
Spruce	1,255	138	870	90	50.7	10.5	0.42 *	12.3	4.2	-0.02	2.1	1.3	0.29 *
Western larch	173	38	77	22	50.1	14.3	0.36	20.0	5.9	-0.06	2.2	1.5	0.40
Incense cedar	209	41	179	33	47.3	13.4	0.47 *	15.8	4.7	0.17	1.5	2.5	0.05
Lodgepole pine	2,818	166	2,011	108	43.6	10.0	0.51 *	16.3	4.1	0.17	3.0	1.3	0.51 *
Western red cedar	431	54	237	29	48.6	12.8	0.49 *	18.1	5.8	0.26	1.7	1.9	0.37 *
Other softwoods	702	126	396	82	43.5	12.9	0.63 *	16.3	7.2	0.46 *	3.0	1.9	0.20
Softwood woodland	3,321	270	961	100	51.8	13.7	0.64 *	12.3	4.2	0.40 *	4.3	1.4	0.14
Hardwood timberland:													
Quaking aspen	1,246	83	852	51	36.0	9.0	0.06	21.3	6.2	-0.05	5.2	1.6	0.31 *
Cottonwood/poplar	69	11	64	9	47.9	11.1	0.56	17.6	6.9	0.54	2.4	1.0	0.04
Red alder	363	42	203	18	48.5	14.4	0.19	23.3	13.1	-0.10	5.2	2.8	0.32
Timberland oak	1,068	106	874	91	38.0	10.3	0.52 *	18.8	8.5	-0.07	5.8	1.7	0.41 *
Other hardwoods	921	100	628	63	43.3	12.1	0.18	19.1	8.1	-0.05	4.4	2.1	-0.04
Hardwood woodland	3,321	86	125	29	43.2	14.5	0.58 *	18.2	7.4	-0.10	7.8	1.5	0.66 *
All softwoods	18,538	927	10,211	526	49.5	11.0	0.49 *	14.4	4.8	0.28 *	2.8	1.4	0.20 *
All hardwoods	4,310	361	2,746	219	41.0	11.8	0.30 *	19.5	7.6	0.01	5.6	1.6	0.46 *
All species	22,848	1,000	12,957	581	47.2	11.1	0.48 *	15.0	5.1	0.18 *	3.3	1.3	0.35 *

[a] States included: CA, CO, ID, NV, OR, UT, WA, and WY.

[b] Independent observations from the latest set of 4 panels sampled between 1992 and 1999 in each State were used to calculate means (m), standard deviations (s), and coefficients of variation (cv).

[c] Paired observations from the latest set of remeasured panels sampled between 1992 and 1999 in each State. Plot means from survivor trees were paired with their previous values to calculate correlation coefficients(r). The number of remeasured panels available varies by State.

[d] The numbers of observations represent pairs. The total numbers of observations are twice the listed numbers of pairs.

[e] Asterisks indicate significant correlation coefficients, where p-values are $\leq$ .05.

Table 3. Numbers of independent and paired observations required to detect a statistically significant signal, given the listed combinations of input specifications for 3 crown variables (crown density, foliage transparency, and crown dieback).

Variable	Power analysis input specifications					Plots required		
						Independent	Paired[c,d]	
	Data Distribution	Power level (1-*B*)	Alpha level (*a*)	Data variability[a]	Effect size[b]	n plots	n plots (r=.25)	n plots (r=.50)
Crown density	Normal	0.8	0.01	10	10	52	21	16
		0.8	0.01	10	15	26	12	9
		0.8	0.01	15	10	110	43	30
		0.8	0.01	15	15	52	21	16
		0.8	0.05	10	10	34	14	10
		0.8	0.05	10	15	18	8	6
		0.8	0.05	15	10	74	29	20
		0.8	0.05	15	15	34	14	10
		0.9	0.01	10	10	64	26	19
		0.9	0.01	10	15	30	14	11
		0.9	0.01	15	10	138	54	37
		0.9	0.01	15	15	64	26	19
		0.9	0.05	10	10	46	18	13
		0.9	0.05	10	15	22	10	7
		0.9	0.05	15	10	98	38	26
		0.9	0.05	15	15	46	18	13
Foliage transparency	Normal	0.8	0.01	5	5	52	21	16
		0.8	0.01	5	10	16	8	7
		0.8	0.01	10	5	192	74	51
		0.8	0.01	10	10	52	21	16
		0.8	0.05	5	5	34	14	10
		0.8	0.05	5	10	12	6	5
		0.8	0.05	10	5	128	50	34
		0.8	0.05	10	10	34	14	10
		0.9	0.01	5	5	64	26	19
		0.9	0.01	5	10	20	10	8
		0.9	0.01	10	5	242	93	63
		0.9	0.01	10	10	64	26	19
		0.9	0.05	5	5	46	18	13
		0.9	0.05	5	10	14	7	5
		0.9	0.05	10	5	172	65	44
		0.9	0.05	10	10	46	18	13
Crown dieback	Log normal	0.8	0.01	1.5	2	118	39	25
		0.8	0.01	1.5	2.5	70	24	16
		0.8	0.01	2.0	2	160	48	29
		0.8	0.01	2.0	2.5	94	29	18
		0.8	0.05	1.5	2	80	26	16
		0.8	0.05	1.5	2.5	48	16	11
		0.8	0.05	2.0	2	108	32	19
		0.8	0.05	2.0	2.5	64	20	12
		0.9	0.01	1.5	2	150	49	30
		0.9	0.01	1.5	2.5	88	30	19
		0.9	0.01	2.0	2	204	61	36
		0.9	0.01	2.0	2.5	118	36	22
		0.9	0.05	1.5	2	106	35	21
		0.9	0.05	1.5	2.5	62	21	13
		0.9	0.05	2.0	2	144	43	25
		0.9	0.05	2.0	2.5	84	25	15

[a] For variables with a normal distribution the measure of data variability is the standard deviation. For variables with a log normal distribution the measure of data variability is the coefficient of variation.

[b] For variables with a normal distribution the effect size is specified as the difference between two means. For variables with a log normal distribution the effect size is specified as the ratio between two means.

[c] Paired observations require the additional specification of a correlation coefficient (r). These results show the numbers of observations required when the correlation coefficient is set to 0.25 and 0.50.

[d] The numbers of plots listed represent pairs. Each pair represents 1 plot (i.e., grid point) with two observations.

Table 4. Total land area, forest land area, and detectable impact area for three crown-indicator scenarios involving remeasured (paired) plots, by region and State.

Region / State	RPA area[a]		Detectable impact area[b]					
			Crown density[c] (21 paired plots)		Foliage transparency[d] (31 paired plots)		Crown dieback[e] (36 paired plots)	
	Total land	Forest land	Forest Area[f]	Percent Forest[g]	Forest Area[f]	Percent Forest[g]	Forest Area[f]	Percent Forest[g]
	-------- M acres --------		M acres	Percent	M acres	Percent	M acres	Percent
Northeast:								
Connecticut	3,101	1,794	2,016	112	2,976	166	3,456	193
Delaware	1,251	383	2,016	527	2,976	777	3,456	903
Maine	19,752	17,673	2,016	11	2,976	17	3,456	20
Maryland	6,256	2,566	2,016	79	2,976	116	3,456	135
Massachusetts	5,018	3,171	2,016	64	2,976	94	3,456	109
New Hampshire	5,740	4,850	2,016	42	2,976	61	3,456	71
New Jersey	4,748	2,132	2,016	95	2,976	140	3,456	162
New York	30,217	18,669	2,016	11	2,976	16	3,456	19
Pennsylvania	28,683	16,577	2,016	12	2,976	18	3,456	21
Rhode Island	669	356	2,016	567	2,976	836	3,456	971
Vermont	5,920	4,618	2,016	44	2,976	64	3,456	75
West Virginia	15,415	12,007	2,016	17	2,976	25	3,456	29
Total	126,767	84,796	2,016	2	2,976	4	3,456	4
North Central:								
Illinois	35,608	4,525	2,016	45	2,976	66	3,456	76
Indiana	22,980	4,656	2,016	43	2,976	64	3,456	74
Iowa	35,842	2,879	2,016	70	2,976	103	3,456	120
Michigan	36,275	19,545	2,016	10	2,976	15	3,456	18
Minnesota	51,024	16,391	2,016	12	2,976	18	3,456	21
Missouri	44,093	15,078	2,016	13	2,976	20	3,456	23
Ohio	26,207	7,894	2,016	26	2,976	38	3,456	44
Wisconsin	34,791	16,275	2,016	12	2,976	18	3,456	21
Total	286,819	87,243	2,016	2	2,976	3	3,456	4
South:								
Southeast:								
Florida	35,026	16,147	2,016	12	2,976	18	3,456	21
Georgia	37,114	24,784	2,016	8	2,976	12	3,456	14
North Carolina	31,128	18,447	2,016	11	2,976	16	3,456	19
South Carolina	19,207	12,746	2,016	16	2,976	23	3,456	27
Virginia	25,626	15,766	2,016	13	2,976	19	3,456	22
Total	148,102	87,889	2,016	2	2,976	3	3,456	4
South Central:								
Alabama	32,435	22,693	2,016	9	2,976	13	3,456	15
Arkansas	33,324	18,830	2,016	11	2,976	16	3,456	18
Kentucky	25,426	11,970	2,016	17	2,976	25	3,456	29
Louisiana	27,880	14,222	2,016	14	2,976	21	3,456	24
Mississippi	30,026	19,622	2,016	10	2,976	15	3,456	18
Oklahoma	43,954	7,665	2,016	26	2,976	39	3,456	45
Tennessee	26,390	14,480	2,016	14	2,976	21	3,456	24
Texas	167,693	17,273	2,016	12	2,976	17	3,456	20
Total	387,127	126,756	2,016	2	2,976	2	3,456	3
Great Plains:								
Kansas	52,488	2,106	2,016	96	2,976	141	3,456	164
Nebraska	49,206	1,245	2,016	162	2,976	239	3,456	278
North Dakota	44,337	724	2,016	278	2,976	411	3,456	477
South Dakota	48,434	1,682	2,016	120	2,976	177	3,456	205
Total	194,465	5,757	2,016	35	2,976	52	3,456	60

Continued…

Table 4 continued. Total land area, forest land area, and detectable impact area for three crown-indicator scenarios involving remeasured (paired) plots, by region and State.

Region / State	RPA area[a]		Detectable impact area[b]					
			Crown density[c] (21 paired plots)		Foliage transparency[d] (31 paired plots)		Crown dieback[e] (36 paired plots)	
	Total land	Forest land	Forest Area[f]	Percent Forest[g]	Forest Area[f]	Percent Forest[g]	Forest Area[f]	Percent Forest[g]
	-------- M acres --------		M acres	Percent	M acres	Percent	M acres	Percent
Intermountain:								
Arizona	72,764	18,671	2,016	11	2,976	16	3,456	19
Colorado	66,390	22,612	2,016	9	2,976	13	3,456	15
Idaho	52,909	21,430	2,016	9	2,976	14	3,456	16
Montana	93,306	25,014	2,016	8	2,976	12	3,456	14
Nevada	70,446	11,089	2,016	18	2,976	27	3,456	31
New Mexico	77,674	16,682	2,016	12	2,976	18	3,456	21
Utah	52,497	17,962	2,016	11	2,976	17	3,456	19
Wyoming	62,062	11,445	2,016	18	2,976	26	3,456	30
Total	548,047	144,905	2,016	1	2,976	2	3,456	2
Pacific Northwest:								
Oregon	61,181	30,169	2,016	7	2,976	10	3,456	11
Washington	42,609	22,279	2,016	9	2,976	13	3,456	16
California	99,599	32,817	2,016	6	2,976	9	3,456	11
Hawaii	4,111	1,748	2,016	115	2,976	170	3,456	198
Alaska	365,042	126,869	2,016	2	2,976	2	3,456	3
Total	572,542	213,883	2,016	1	2,976	1	3,456	2

[a] Source: 2007 RPA statistics (http://fia.fs.fed.us/program-features/rpa/default.asp).

[b] The forest area that must be impacted in order to detect a forest health problem, given a specified number of plots.

[c] 21 paired plots are necessary to detect a significant change given: a normal data distribution; power level (1-B)=0.9; a=0.05; standard deviation=12; correlation coefficient=0.4; mean difference=10.

[d] 31 paired plots are necessary to detect a significant change given: a normal data distribution; power level (1-B)=0.9; a=0.05; standard deviation=6.5; correlation coefficient=0.2; mean difference=5.

[e] 36 paired plots are necessary to detect a significant change given: a log normal data distribution; power level (1-B)=0.9; a=0.05; coefficient of variation=1.8; correlation coefficient=0.3; ratio of mean difference=2.

[f] Impact area=n(E)*(Pn/Pt) where n = the specified number of plots; E=plot expansion factor=96,000; Pn=available panels=5; Pt=total panels=5.

[g] Percent forest impact area = detectable forest impact area / forest area * 100. This is the percentage of forest area in the listed State or region that must be impacted by a disturbance in order to detect a significant change. Values greater than 100 indicate that the sampling intensity is insufficient.

Phase 3 Indicators: Data in Action



From Detection Monitoring to Evaluation Monitoring—A Case Study Involving Crown Dieback in Northern White-Cedar

KaDonna Randolph[1]
William A. Bechtold[2]
Randall S. Morin[3]
Stanley J. Zarnoch[4]

***Abstract**: The Forest Inventory and Analysis (FIA) Phase 3 plot network is a crucial part of the U.S. Forest Health Monitoring program's detection monitoring system, where select indicators are monitored for signals that may indicate deteriorating forest health. When a negative signal is identified, evaluation monitoring provides a mechanism whereby a potential problem can be further investigated. Elevated crown dieback was observed among northern white-cedar (Thuja occidentalis L.) trees in Maine and Michigan on FIA Phase 3 plots measured between 2000 and 2004. We present results of this potential problem through the detection and evaluation monitoring process. We discuss the advantages and disadvantages of using FIA data to frame the problem, and share lessons learned from all phases of the project—including problem identification, project implementation, and results presentation.*

Keywords: Crown condition, FHM, FIA, forest health, Maine, Michigan, *Thuja occidentalis.*

Introduction

The U.S. Forest Service, Forest Health Monitoring (FHM) program institutes a tiered strategy to monitor the status, changes, and trends in forest health across the United States (Riitters and Tkacz 2004). The first tier, known as detection monitoring, evaluates the status and change in the condition of forested ecosystems through analysis of data collected from nationally standardized aerial and ground surveys. The second tier, known as evaluation monitoring (EM), seeks to determine the extent, severity, and cause of any deteriorating forest conditions observed through detection monitoring. A third tier, intensive site monitoring,

[1] United States Forest Service; Southern Research Station; Forest Inventory and Analysis; 4700 Old Kingston Pike; Knoxville, TN 37919 USA; krandolph@fs.fed.us

[2] United States Forest Service; Southern Research Station; 200 Weaver Blvd; Asheville, NC 28804 USA; wabechtold@fs.fed.us

[3] United States Forest Service; Northern Research Station; Forest Inventory and Analysis; 11 Campus Drive, Suite 200; Newtown Square, PA 19073 USA; rsmorin@fs.fed.us

[4] United States Forest Service; Southern Research Station; 200 Weaver Blvd; Asheville, NC 28804 USA; szarnoch@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

investigates cause-effect relationships by linking detection monitoring to ecosystem process studies (U.S. Department of Agriculture 2003). Detection monitoring activities were begun in 1990 by FHM. In 2000, the network of FHM detection monitoring plots was integrated into the U.S. Forest Service, Forest Inventory and Analysis (FIA) program. Now known as Phase 3 plots, the detection monitoring plots are a subset of FIA's total network of Phase 2 ground plots (McRoberts 2005). In addition to these plots, the U.S. Forest Service, Forest Health Protection (FHP) unit also participates in detection monitoring by conducting annual aerial and ground-based surveys for forest insects, disease, and invasive plants.

Our objectives for this study were to describe the EM process, describe the role of FIA in EM, and review our experience with an EM project that originated from an unexpected trend observed in FIA Phase 3 data.

FHM Evaluation Monitoring Process

The FHM program funds EM projects under two broad categories: base and fire plan. The purpose of base EM projects is to investigate any issues or concerns identified during detection monitoring. The purpose of fire plan EM projects is to investigate and explain the extent, severity, and/or cause of a fire-related phenomenon observed during detection monitoring. EM proposals are first screened by regional FHM managers and then forwarded to a national selection committee, which includes the FHM Program Manager and representatives from the Forest Service's National Forest System, State and Private Forestry, and Research and Development programs. Selection criteria include:

- linkage to detection monitoring,
- significance in terms of geographic scale,
- biological impact and/or political importance, and
- feasibility of successful project completion within 1 to 3 years.

Listings and descriptions of EM projects funded since 2004 are posted at http://fhm.fs.fed.us/em/funded/proposals_base.shtm.

Problem Identification and Presentation of Results

National reports on forest health conditions are produced annually by the FHM program (e.g., Ambrose and Conkling 2007). State-level reports that include forest health indicators are produced by FIA on a 5-year basis (e.g. Turner and others 2008). Annual Forest Health Highlights (e.g., Washington State Department of Natural Resources 2007) are produced for each State by FHP staff and State partners. All of these reports are potential sources for identifying unusual or deteriorating forest health conditions that require further study.

Failure to identify a problem that truly exists has the potential for disaster. Thus, the EM process allows a wide range of potential forest health problems to be examined so that serious problems are not overlooked. Successful EM projects include not only those that confirm a significant forest health problem but those that nullify a suspected forest health problem as well. Both results improve knowledge of the resource in question and, in the case of confirmation, provide managers an opportunity to mitigate the threat.

Once an EM project is funded, the FHM management team requires annual progress reports. Results based on these reports are also presented in poster format at annual FHM working group meetings. Beginning with the 2008 report, summaries of recently completed EM projects will be included in the FHM national reports. This provides a structured forum for EM results and guarantees a published outlet to EM investigators. The format for these summaries is similar to an extended abstract, but it also allows one or two tables or figures per summary. This abbreviated format is not expected to interfere with other presentation and publication outlets, which are encouraged as well.

Role of FIA

Using FIA data as the basis for EM projects has several benefits. First, researchers have access to a large, long-term, broad-scale dataset. FIA has been conducting inventories of the Nation's forest land for over 70 years. Up through the late 1990s, statewide inventories focused primarily on timber-based variables and were completed approximately once every 6 to 8 years in the South and 11 to 18 years in the rest of the country (Gillespie 1999). Since then FIA has switched to a panelized annual inventory system (Bechtold and Patterson 2005) and worked to reduce inventory cycles to 5 years in the East and 10 years in the West. Further, the integration of the FHM detection monitoring plots in 2000 broadened FIA's survey to include more nontimber variables. These FIA data, including the FHM detection monitoring data, are available to the public through FIA's Spatial Data Service (SDS) centers or online databases at http://www.fia.fs.fed.us/tools-data/default.asp.

All FHM data from 1990 to 1999 and FIA data from 2000 and beyond are available for analysis. One obstacle to integrating these data, however, is that even though FHM plots were merged into the FIA Phase 3 network, changes in protocols make it difficult to follow individual trees across the 2000 integration date. For many plots, algorithms must be used to match trees. Population inferences based on matched FHM-FIA detection monitoring data may be tenuous because there is a question about the extent to which the subsample of matched trees accurately represents the population due to the inability to match all trees.

A second benefit of using the FIA data is repeated measurements of permanent plots. These repeated measurements allow for longitudinal studies at the plot or

individual tree level. Though such studies may be complicated by missing or temporally irregular observations, most current statistical software packages can overcome this difficulty with careful model specifications. Almost all of the detection monitoring plots were measured more than once between 1990 and 1999, and since 2000 most of the FIA Phase 3 plots in the Eastern United States also have been remeasured. Within a few years, States in the Western United States will begin remeasurement as well.

A third benefit to using FIA data for detection monitoring and EM is that each plot is georeferenced with latitude and longitude coordinates. This allows researchers to look for changes and trends across space as well as over time. There are, however, strict regulations regarding the release of plot locations. Exact locations are protected as confidential information under the Food Security Act of 1986, Public Law 99-198 [H.R. 2100], so coordinates that accompany the publicly distributed plot data have been perturbed (McRoberts and others 2005). Researchers requiring exact plot locations should contact SDS for possible accommodation before submitting EM project proposals.

Experiences with EM Project NE-07-01

Discovering Dieback among Northern White-Cedar

As part of the 2006 national FHM report, spatial patterns of crown conditions by species group were evaluated to identify potential forest health problems within the coterminous United States (Randolph in press). Crown density, foliage transparency, and crown dieback averages were calculated by plot for individual species groups if the plot contained five or more trees (diameter $\geq$5.0 inches) in a given species group. All available data from FIA Phase 3 plots measured between 2000 and 2004 were included in the analyses. Spatial clusters of plots with high crown dieback, high foliage transparency, or low crown density averages relative to the other plots were identified as areas with potential forest health problems. Clusters of plots with northern white-cedar (*Thuja occidentalis* L.) crown dieback[1] averaging 10 percent or more were discovered in Maine and northern Michigan. Plot averages for northern white-cedar dieback met this threshold for 33 percent and 19 percent of the plots with at least five northern white-cedar trees in Maine and Michigan, respectively. Such elevated levels of dieback were of concern because unlike hardwood trees, conifers often do not exhibit crown dieback unless the tree is under serious stress (Millers and others 1992).

In general, northern white-cedar is a species resistant to serious injury from insects and disease, so reasons for the elevated levels of dieback were unclear. Local foresters, entomologists, and pathologists were questioned about potential

[1] Crown dieback is defined as "recent mortality of branches with fine twigs, which begins at the terminal portion of a branch and proceeds toward the trunk" (Schomaker and others 2007).

causes. Most had not noticed a deterioration of northern white-cedar, though its general poor form and harsh growing conditions were often cited. Drought was mentioned as a suspected cause in Maine, where one of the worst droughts ever recorded there occurred between 1999 and 2002 (Lombard 2004). Dry conditions also occurred between 1998 and 2002 in the Upper Peninsula and northern Lower Peninsula of Michigan (Steinman 2004). Thus, drought was considered to be a potential explanation. Other suggested causes included unfavorable winter weather (Johnston 1990), other severe weather events, and silvicultural practices.

Stand-level condition data (e.g. disturbances, stand age, presence of water) were examined for correlations that might provide an explanation for the elevated levels of crown dieback. Residual stress from harvesting and wind disturbances explained the elevated dieback levels on two plots in Maine. However, harvesting and detrimental weather events did not seem broadly applicable reasons for the elevated levels of crown dieback.

Continuing the investigation, we examined plot averages by broad species group to determine if the elevated levels of crown dieback were limited to the northern white-cedars. Crown dieback averages were calculated at plot-level for groupings of hardwoods and non-northern white-cedar softwoods (hereafter, "other softwoods"). We observed that high dieback averages for northern white-cedars were not necessarily accompanied by elevated averages among the hardwoods and other softwoods on the same plots. We therefore decided that additional ground work and more indepth data analyses warranted an EM project proposal.

We submitted a base EM project proposal that recommended additional field work to verify the original field crew assessments and to search for additional evidence of disturbances that may have been below the thresholds at which they are recorded by FIA field crews. We also proposed further analysis of the 2000-2004 FIA data and of FHM data collected during the 1990s. The proposal was accepted and funded for 2 years (Randolph 2008).

Field Verification and Data Analysis

During the summer of 2007, we selected plots to revisit which had average northern white-cedar crown dieback > 10 percent (based on the 2000-2004 FIA surveys). Thereby, we identified 15 plots (6 in Michigan and 9 in Maine). To compare stand and tree conditions, we also revisited five additional plots with an average cedar crown dieback of 5 percent or less. Special approval was granted by FIA administrators to visit these plots. Working closely with the FIA crews in Maine and Michigan, visits were scheduled to coincide with their regularly planned visits as much as possible. This was done for three reasons. The experience and local knowledge of the field crews quickened our ability to find the plots. In addition, coincident visits limited the number of times the landowners were

contacted, and reduced the potential for plot degradation. Given the time allotted for field work we were able to visit 11 plots with the FIA field crews and 7 others independently, for a total 18 plot visits (figure 1).

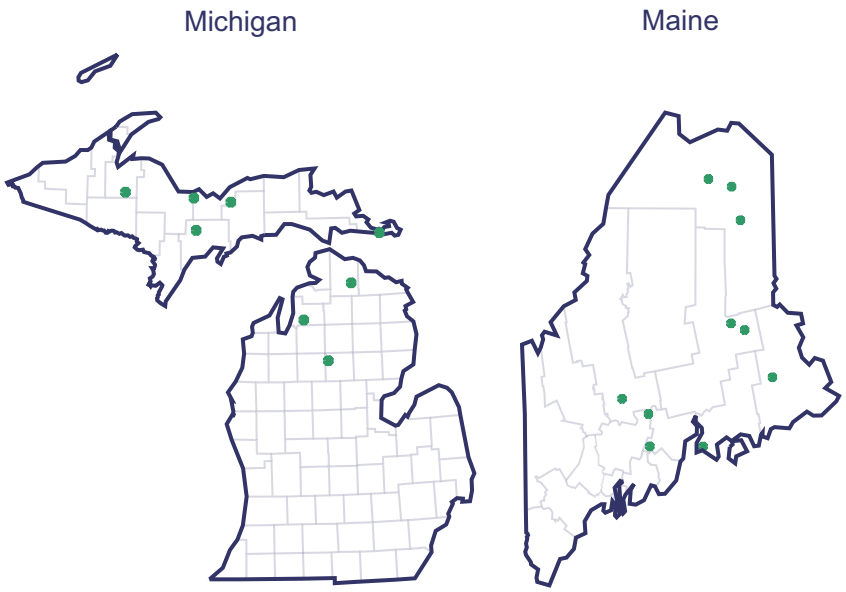


Figure 1: Approximate locations of the plots visited in Michigan and Maine during the summer of 2007 for EM project NE-07-01.

These plot visits enabled us to verify the elevated levels of northern white-cedar crown dieback and to observe local growing conditions. Disturbances on two of the selected plots that previously had been identified by FIA field crews were validated also. In addition, on two other plots we found localized wind damage <1-acre in size and a combination of wind and flooding damage on still another. Besides the plot-level disturbances, many northern white-cedars were leaning, had exposed roots, or exhibited strips of dead cambium around the bole. These conditions gave insight into the distressed conditions of many trees. Readers are directed to Randolph and others (2008) for a detailed summary of the plot visits.

The plot visits did not identify any single cause for the high levels of crown dieback but did help identify variables to include in subsequent data analyses such as stand age, disturbance type, temperature, and precipitation. We began the data analysis by obtaining all detection monitoring plots measured in Maine and Michigan between 1990 and 2005 through the national SDS center. FHM data collection began in 1990 in Maine and in 1994 in Michigan; and it ended in both States in 1999. These timeframes are hereafter referred to as the "FHM time period." A total of 259 plots (123 in Maine and 136 in Michigan) were obtained for this period of FHM data collection. Northern white-cedar was present on 37 percent of the plots in Maine and 18 percent of the plots in Michigan. All of the

plots in Maine and 122 plots in Michigan were measured at least twice during the FHM time period. Some plots were measured as many as nine times in Maine and four times in Michigan. FIA data collection began in both States in 2000. Data from a total of 417 plots (201 in Maine and 216 in Michigan) were obtained for the period of FIA data collection, 2000-2005 ("FIA time period"). Northern white-cedar was present on 34 percent of the plots in Maine and 22 percent of the plots in Michigan. Repeated measurements were made on 37 plots in Maine and 42 plots in Michigan. Of those remeasured plots, northern white-cedar was present on 11 in Maine and 10 in Michigan. Using a tree-matching algorithm, we confidently were able to match 495 northern white-cedar trees from 54 plots established in the FHM time period with their corresponding assessments from the FIA time period.

We then used the available data to determine (1) if the level of crown dieback in northern white-cedar was significantly higher than it was in other species, and (2) if there had been a change over time in the average level of northern white-cedar crown dieback. An analysis of variance model, with plot-level average crown dieback as the response variable and taxa (hardwoods, northern white-cedar, and other softwoods), measurement year, and the taxa * measurement year interaction as the explanatory factors, was applied to answer question (1). This was done separately by State for each time period (FHM and FIA). Data from the two time periods were not combined due to complications from being unable to follow some FHM plots across the 2000 integration date. To accommodate the repeated measurements, we analyzed the data using the SAS® procedure MIXED (Littell and others 1996) with a REPEATED statement. Correlations between repeated measurements are typically larger for observations with shorter measurement intervals. To account for this, an autoregressive order 1 covariance structure was used to model the covariance structure within subjects for the FIA time period. Because of the unequally spaced time intervals between measurements, the spatial power law covariance structure was used for the FHM time period. Pairwise, least-squares mean difference comparisons between the taxa means were made with a Bonferroni adjusted comparisonwise alpha level of 0.0167 (overall familywise alpha = 0.05).

Preliminary analyses revealed that a number of plots were influencing the results. Weighting the plot observations by the number of trees sampled alleviated much of the undue influence. However, one plot in Michigan continued to have considerable influence. This plot (which we visited during our field work) was located in an area that experienced extreme flooding and wind damage (figure 2). We concluded that this plot was atypical of the other plots visited in Michigan and decided to drop it from the dataset. This case of one plot exerting great influence on the analyses highlights the need for researchers to be watchful for such possibilities.



Figure 2: Conditions of the atypical plot in Michigan with flooding and wind damage.

For question (2), the subset of individual northern white-cedar trees with matched observations spanning the FHM-FIA time periods were modeled as a linear function of measurement year, by State. Again, because of the unequally spaced time intervals between measurements, we used the spatial power law covariance structure to model the covariance structure within subjects.

Conclusions about Northern White-Cedar Crown Dieback

We concluded for question (1) that when significant differences were present, northern white-cedar crown dieback levels were higher than the crown dieback levels of other softwood species and lower than the dieback levels of hardwood species, although such differences sometimes were dependent upon measurement year. In Maine, taxa, measurement year, and the taxa * measurement year interaction were all significant at the alpha = 0.05 level during the FHM time period, but only taxa was significant during the FIA time period (table 1). During the FHM time period, average dieback for northern white-cedar was less than average dieback for the hardwoods and greater than average dieback for the other softwoods, but in most years was not significantly different from either group (figure 3). During the FIA time period, average northern white-cedar dieback was significantly higher than the average dieback of the other softwoods (figure 4). In Michigan, taxa was the only significant factor at the alpha = 0.05 level during the FHM time period, whereas all three factors were significant during the FIA time period (table 1). During the FHM time period, average northern white-cedar dieback

Table 1: Results (p-values) for the analysis of variance testing the effect of taxa and measurement year on average crown dieback in Maine and Michigan between 1990 and 2005

	Maine		Michigan	
Factor	**1990-1999[a]**	**2000-2005[b]**	**1994-1999[a]**	**2000-2005[b]**
Taxa	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Year	< 0.0001	0.0939	0.0754	0.0005
Taxa * year	0.0118	0.4335	0.1584	0.0015

[a] FHM time period.
[b] FIA time period.

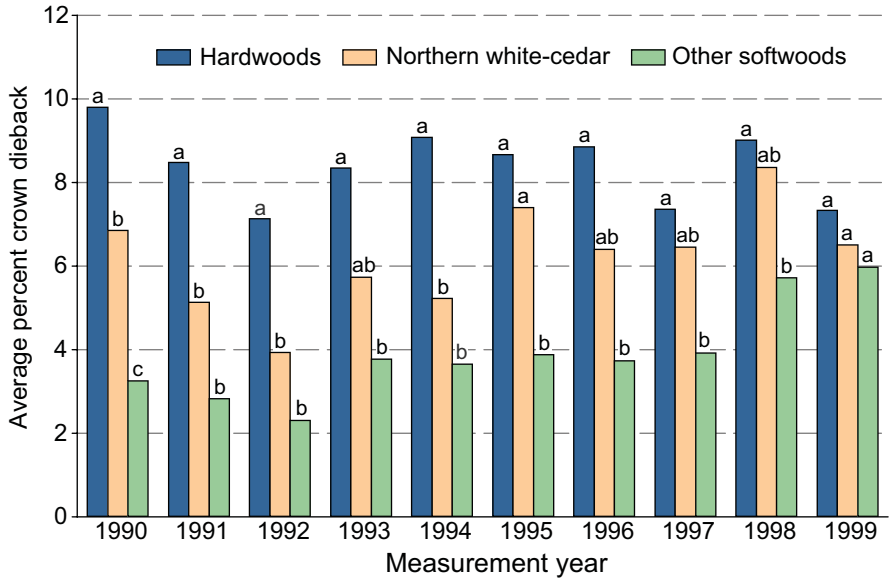


Figure 3: Least-squares mean estimates of average percent crown dieback by taxa and year in Maine. For each year, bars with different letters are significantly different at the Bonferroni comparison-wise alpha value of 0.0167.

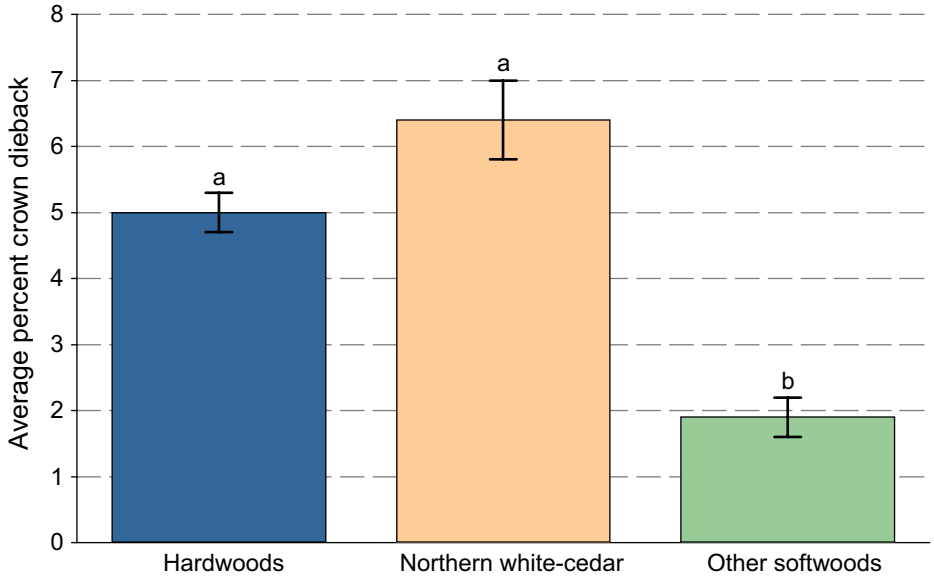


Figure 4: Least-squares mean estimates of average percent crown dieback in Maine during the FIA time period (2000–2005), by taxa, with standard error bars. Taxa means with different letters are significantly different at the Bonferroni comparison-wise alpha value of 0.0167.

was significantly higher than the average dieback of the other softwoods (figure 5). During the FIA time period, significant differences among the taxa occurred in 2002 only, when average dieback for northern white-cedar was significantly higher than the averages of the hardwoods and other softwoods (figure 6).

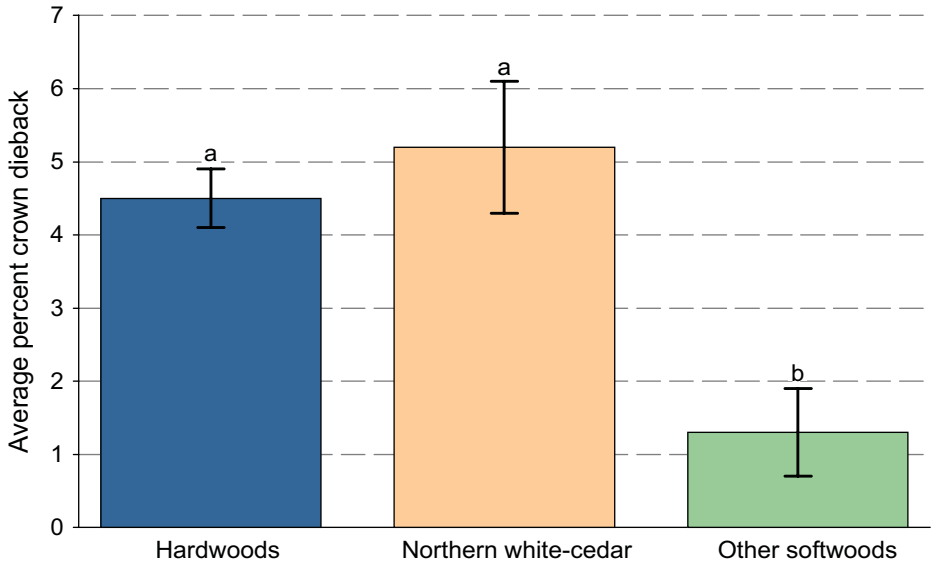


Figure 5: Least square mean estimates of average percent crown dieback in Michigan during the FHM time period (1994–1999), by taxa, with standard error bars. Taxa means with different letters are significantly different at the Bonferroni comparison-wise alpha value of 0.0167.

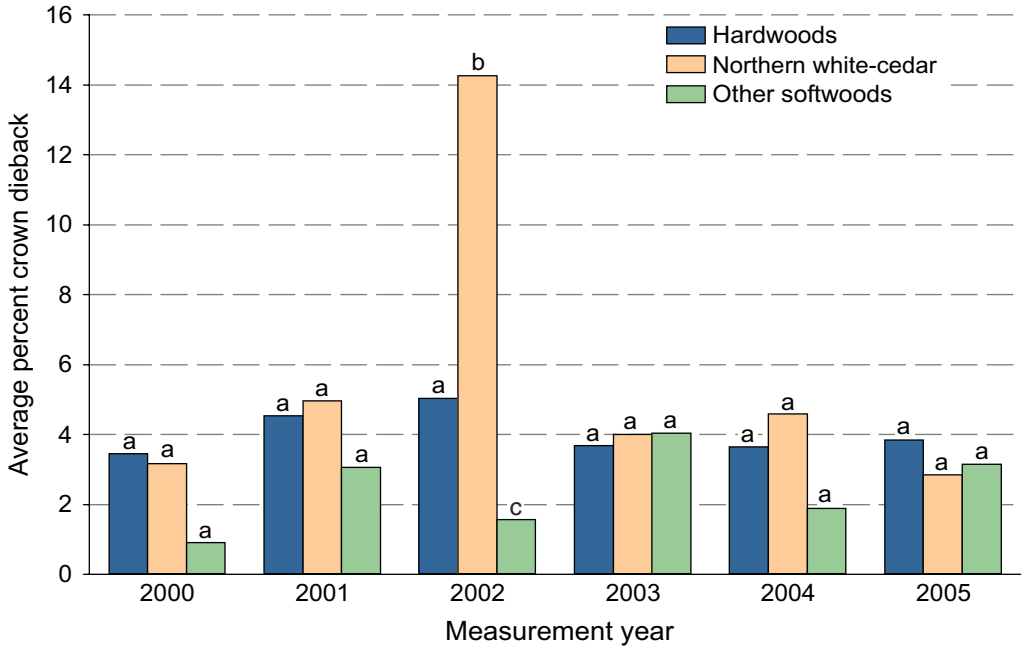


Figure 6: Least-squares mean estimates of average percent crown dieback by taxa and year in Michigan. For each year, bars with different letters are significantly different at the Bonferroni comparison-wise alpha value of 0.0167.

For question (2), we were not able to conclude from the available data that there was a significant increase in northern white-cedar crown dieback in either State during the timeframe we examined. Measurement year was nonsignificant in the model predicting crown dieback in both Maine (p-value = 0.5962) and Michigan (p-value = 0.1949). The answers to additional questions, e.g., about specific stand and weather conditions associated with high levels of northern white-cedar, will require additional research.

Conclusion

Though we failed to reject the null hypothesis that northern white-cedar dieback was unchanged, we did verify that the species tends to have different average levels of dieback than other taxa. This finding supports the argument that average crown conditions tend to be species-specific, and that species differences should be taken into account when analyzing crown-condition data (Zarnoch and others 2004).

There are many advantages of using FIA data as the basis for EM projects. The broad scale, long-term data provide a wealth of information that can be analyzed to understand the past, present, and future ecological conditions of our forests. Because FIA has strict definitions and data collection protocols, researchers engaging in detection and EM activities are encouraged to collaborate or consult with FIA personnel throughout their projects' duration.

Acknowledgments

Funding for this EM project (NE-EM-07-01) was provided in part by the U.S. Forest Service, Forest Health Monitoring Program. Appreciation is extended to the National FIA Spatial Data Service Center, everyone who assisted in the field work, and Jim Steinman and Manfred Mielke who provided helpful comments on an earlier draft of this manuscript.

References

Ambrose, M.J.; Conkling, B.L., eds. 2007. Forest health monitoring: 2005 national technical report. Gen. Tech. Rep. SRS–104. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 76 p.

Bechtold, W.A.; Patterson, P.L., eds. 2005. The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS–80. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 85 p.

Gillespie, A.J.R. 1999. Rationale for a national annual inventory program. Journal of Forestry. 97(12): 16-20.

Johnston, W.F. 1990. Northern white-cedar. In: Burns, R.M.; Honkala, B.H., tech. coords. Silvics of North America: 1. Conifers; 2. Hardwoods. Agric. Handb. 654. Washington, DC: U.S. Department of Agriculture Forest Service: 580-589.

Littell, R.C.; Milliken, G.A.; Stroup, W.W.; Wolfinger, R.D. 1996. SAS® system for mixed models. Cary, NC: SAS Institute, Inc. 633 p.

Lombard, P.J. 2004. Drought conditions in Maine, 1999-2002: a historical perspective. Water-Resources Investigations Report 03-4310. Augusta, ME: U.S. Department of the Interior, Geological Survey. 36 p.

McRoberts, R.E. 2005. The enhanced Forest Inventory and Analysis program. In: The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures; Bechtold, William A.; Patterson, Paul L., eds. Gen. Tech. Rep. SRS–80. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station: 1-10.

McRoberts, R.E.; Holden, G.R.; Nelson, M.D. [and others]. 2005. Estimating and circumventing the effects of perturbing and swapping inventory plot locations. Journal of Forestry. 103(6): 275-279.

Millers, I.; Anderson, R.; Burkman, W.; Hoffard, W. 1992. Crown condition rating guide. Newtown Square, PA: U.S. Department of Agriculture Forest Service, Northeastern Area State and Private Forestry; Atlanta: U.S. Department of Agriculture Forest Service, Southern Region. 37 p.

Randolph, K. 2008. Evaluating elevated levels of crown dieback among northern white-cedar (*Thuja occidentalis*) trees in Maine and Michigan. [Online]. U.S. Forest Service Forest Health Monitoring Program evaluation monitoring project NE-EM-07-01. http://fhm.fs.fed.us/em/funded/08/ne_em_07_01.pdf. [July 24, 2008].

Randolph, K.; Ostrofsky, B.; Steinman, J. [and others]. 2008. Field observations of northern white-cedar (*Thuja occidentalis*) crown dieback in Maine and Michigan (NE-EM-07-01) [Online]. http://fhm.fs.fed.us/posters/posters08/northern_white_cedar.pdf. [August 22, 2008].

Randolph, K.C. [In press]. Crown condition. In: Ambrose, M.J.; Conkling, B.L (eds.). (in press). Forest health monitoring 2006 national technical report. Gen Tech. Rep. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station.

Riitters, K.; Tkacz, B. 2004. The U.S. forest health monitoring program. In: Environmental Monitoring. Boca Raton, FL: CRC Press: 669-683.

Schomaker, M.E.; Zarnoch, S.J.; Bechtold, W.A. [and others]. 2007. Crown-condition classification: a guide to data collection and analysis. Gen. Tech. Rep. SRS–102. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 78 p.

Steinman, J. 2004. Forest health monitoring in the Northeastern United States: disturbances and conditions during 1993-2002. NA–TP–01–04. Newtown Square, PA: U.S. Department of Agriculture Forest Service, Northeastern Area, State and Private Forestry. 46 p.

Turner, J.A.; Oswalt, C.M.; Chamberlain, J.L. [and others]. 2008. Kentucky's forests, 2004. Resour. Bull. SRS–129. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 101 p.

U.S. Department of Agriculture Forest Service. 2003. Program description. Forest health monitoring factsheet series. [Online]. http://www.fhm.fs.fed.us/fact/03/prog_desc.pdf. [July 24, 2008].

Washington State Department of Natural Resources. 2007. Forest health highlights in Washington—2007. [Online]. http://fhm.fs.fed.us/fhh/fhh_07/wa_fhh_07.pdf. [August 19, 2008].

Zarnoch, S.J.; Bechtold, W.A.; Stolte, K.W. 2004. Using crown condition variables as indicators of forest health. Canadian Journal of Forest Research. 25: 462-469.

Can Live Tree Size-Density Relationships Provide a Mechanism for Predicting Down and Dead Tree Resources?

Christopher W. Woodall[1] and James A. Westfall[2]

***Abstract**: Live tree size-density relationships in forests have long provided a framework for understanding stand dynamics. There has been little examination of the relationship between the size-density attributes of live and standing/down dead trees (e.g., number and mean tree size per unit area, such information could help in large-scale efforts to estimate dead wood resources. The goal of this study was to examine the relationship between standing live, standing dead, and downed dead trees in the context of size-density attributes using a national inventory of forests. Our results indicated that from the lowest to the highest live tree relative stand density, the mean biomass/ha of live trees increased by more than 2,000 percent while the mean biomass/ha of standing dead and downed dead trees increased 295 and 75 percent, respectively. Correlations between downed dead wood and stand/site attributes reached their highest level ($r > 0.60$) when a stand's relative density exceeded 80 percent. We propose a model for highly stocked stands whereby downed and dead wood biomass may be predicted based on live/dead tree size-density attributes, stand age, and climatic factors. We also provide an alternative model for moderate/low stocked stands whereby potential maximum live biomass may serve as a limit to dead wood resources with stochastic events (e.g., wind/mortality disturbances) as high-impact variables. Overall, the size-density attributes of live/dead trees may help guide the estimation of downed and dead wood attributes in forests.*

Keywords: Downed dead wood, stand density index, size-density, self-thinning, coarse woody debris

Predicting Dead Wood Resources at Large Scales

Forest detritus may be defined as dead organic material in forest ecosystems. For this study, forest detritus will be limited to standing and downed dead woody materials (DDW). Estimates of forest detritus attributes are critical to numerous scientific fields such as carbon accounting (Smith and others 2004, Woodall and others 2008), wildlife habitat assessment (e.g., Bull and others 1997, Harmon and others 1986, Maser and others 1979), and fuel loading estimation (Woodall and

[1]Research Forester, U.S. Department of Agriculture, Forest Service, Northern Research Station, St. Paul, MN 55108; fax: 651-649-5140, phone: 651-649-5141, e-mail: cwoodall@fs.fed.us
[2]Research Forester, U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Monleon 2008). Detritus provides a diversity (stages of decay, size classes, and species) of habitat for fauna ranging from large mammals to invertebrates (Bull and others 1997, Harmon and others 1986, Maser and others 1979). Plants use the microclimate of moisture, shade, and nutrients provided by DDW to establish and regenerate (Harmon and others 1986). Due to the possibility of dwindling detritus habitat for native species and increasing fuel loadings across the United States, comprehensive large-scale inventories of DDW have been established for habitat assessments/wildlife conservation efforts and fire hazard mitigation efforts (e.g., Marshall and others 2000, Rollins and others 2004, Tietje and others 2002, Woodall and Monleon 2008). Worldwide, there has been increased effort in recent years to inventory detrital resources to address greenhouse gas offset accounting and biodiversity concerns (Kukeuv and others 1997, Woldendorp and others 2002, Woodall and others 2008). In 2001, the U.S. began implementing a nationwide inventory of DDW on a subset of inventory plots where standing live/dead trees are measured. An impetus exists to predict DDW for all national inventory plots based on standing live and dead tree attributes (Woodall and others 2008).

To date, efforts to model DDW attributes have been focused at large scales using remotely sensed information and gradient models (e.g., Rollins and others 2004) and at small scales by trying to relate DDW to stand/site attributes (e.g., McCarthy and Bailey 1994, Pyle and Brown 1999, Rubino and McCarthy 2003). Spetich and Guldin (1999) found that DDW accumulation corresponded with increasing site productivity—a function of increased biomass corresponding with increased DDW volume over time. In contrast, Norden and others (2004) found no correlation between DDW volume and basal area in temperate broadleaved forests. Despite the development of models to estimate relationships between forest detritus and stand/site attributes, a sizeable knowledge gap remains in understanding fundamental relationships between forest detritus and basic stand attributes. How does DDW vary by levels of standing live tree density? Can the size/density attributes of both live and dead trees help predict of DDW resources?

Size-Density Relationships in Forest Stands

The size-density of live trees has formed a basis for interpreting/predicting forest stand dynamics for decades (for example see Drew and Flewelling 1979, Gingrich 1967, Krajicek and others 1961, Reineke 1933, Woodall and others 2005). The concept of self-thinning forms the theoretical basis for developing indices of live-tree size-density attributes. Self-thinning is based on the premise that as mean plant size per unit area increases, the number of individuals per unit area decreases (Enquist and others 1998). An inherent component of the self-thinning process is density-induced tree mortality. The forest detritus of standing dead and downed dead wood (DDW) must originate from the mortality/branch shedding of live trees. How closely related are the size-density attributes of live trees in any given forest stand to the attributes of forest detritus?

The goal of this study was to examine the trends in stand-level DDW attributes in relation to the size-density relationships of standing live/dead trees

and selected site factors (e.g., climate and stand age) in forests of the United States. The study had three specific objectives: 1) to estimate mean biomass (tonnes/ha) of standing live, standing dead, and DDW by classes of relative density; 2) to test for correlations between standing live, standing dead, DDW biomass, 30-year mean annual maximum temperature, 30-year mean annual minimum temperature, 30-year mean annual precipitation, and stand age by classes of relative density; and 3) to develop conceptual models for estimating DDW by stand/site factors for stands with high and moderate/low relative density.

Data and Methods

The FIA program is responsible for inventorying the forests of the U.S., including both standing trees and DDW on permanent sample plots established across the country (Bechtold and Patterson 2005). Sample plots are established at an intensity of approximately 1 plot per 2,400 ha. If the plot lies in a forested area, field crews visit the site and measure tree and site variables ranging from tree sizes to forest types. FIA standing live/dead tree inventory plots consist of four 7.32-m fixed-radius subplots for a total plot area of approximately 0.07 ha. All standing trees greater than 12.25 cm in diameter at breast height (d.b.h.) are inventoried on the plot, while trees less than 12.25 cm dbh and greater than 2.54 cm d.b.h. are measured on a 2.07-m fixed radius microplot on each subplot. DDW sampling methods on FIA plots are detailed by Woodall and Monleon (2008). DDW with a transect diameter greater than 7.60 cm are sampled on each of three 7.32-m horizontal distance transects radiating from each FIA subplot center at 30, 150, and 270 degrees; DDW pieces of this size are termed coarse woody debris (CWD). Data collected for every CWD piece include transect diameter, length, small-end diameter, large-end diameter, decay class, and species. Fine woody debris (FWD) are DDW pieces with a transect diameter less than 7.60 cm and are sampled on the 150-degree transect on each subplot. Fine woody debris with transect diameters less than 2.54 cm were tallied separately on a 1.83-m slope-distance transect (4.27 m to 6.09 m on the 150-degree transect). Fine woody debris with transect diameters of 2.55 to 7.59 cm were tallied on a 3.05-m slope-distance transect (4.27 m to 7.32 m on the 150-degree transect).

The Forest Inventory and Analysis program of the U.S. Forest Service inventoried standing and down tree attributes across most of the United States between 2003 and 2006 on a total of 4,221 permanent inventory plots. For every inventory plot, the biomass/ha of standing live and dead trees was determined using procedures detailed by Bechtold and Patterson (2005). Plot-level estimates of DDW were calculated using procedures detailed by Woodall and Monleon (2008, section 3.1). To account for data collection errors across the Nation, extreme outliers were removed using 25 times the interquartile range for all classes of FWD (IQR=6.88 tonnes/ha) and CWD (IQR=9.95 tonnes/ha). The relative density of live trees on every plot was determined using the Stand Density Index (SDI).

SDI was first proposed by Reineke (1933) as a stand density assessment tool based on size-density relationships observed in fully stocked pure or nearly pure

stands. A metric version of SDI is defined as the equivalent trees per hectare at a quadratic mean diameter of 25 cm and is formulated as:

$$SDI = tph\ (DBHq/25)^{1.6} \quad [1]$$

where *SDI* is stand density index, *tph* is number of trees per hectare, and *DBHq* is quadratic mean diameter (cm) at breast height (1.3 m) (Long 1985). SDI has been widely used in even-aged stands because it is independent of species composition. The SDI of even-aged monocultures is typically compared to an empirically observed, species-specific maximum SDI for determining the stand's relative density. Maximum SDI (SDImax) may be defined as the maximum density (tph) that can exist for a given mean tree size (25 cm) in a self-thinning population (Long 1985). To determine relative density (RD), the SDI of any particular stand is compared to the SDImax characteristic of the stand's species composition. Woodall and others (2005) proposed a methodology that estimates the SDImax for any stand based on the mean specific gravity of all trees in a stand to estimate its unique SDImax. By using the summation method (Shaw 2000) to determine the current density of a stand and the Woodall and others (2005) model to predict a SDImax, the RD of all study plots was determined (current SDI/SDImax).

Finally, three climatic variables were selected for correlation with stand-level variables in this study: 30-year mean annual precipitation (PRECIP), 30-year mean annual maximum temperature (TMAX), and 30-year mean annual minimum temperature (TMIN). Data for PRECIP, TMAX, and TMIN were obtained from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) dataset (4-km grid cell size; PRISM Group 2004). Each of these three variables is represented by a 30-year climate normal. As such, annual precipitation is the mean annual total precipitation from 1971 to 2000. TMAX and TMIN are the mean daily temperature extremes for that period.

Correlations and Means by Classes of Stand Relative Density

The RD of forest stands increased with the mean biomass/ha of standing live trees (Table 1). From an RD of less than 0.10 to more than 0.90, the mean live tree biomass (tonnes/ha) increased by nearly 2,080 percent, while standing dead tree and DDW increased by approximately 295 percent and 75 percent, respectively (Table 1). It appears that the greatest rates of increase in biomass for both standing and DDW were from the moderate to high RD levels (i.e., from a RD of 0.7 to 0.9). In contrast, the biomass of standing live trees had its greatest rate of increase when RD was below 0.5. Using the same RD classes, correlations were conducted between DDW and a selection of stand/site attributes (Table 2). Generally, as a stand's RD increased so did stand/site correlations with DDW. For RDs between 0.00 and 0.10, no correlation coefficient exceeded 0.25. In contrast, the majority of correlation coefficients exceeded 0.40 when RDs

Table 1: Mean biomass (tonnes/ha) and associated standard errors (tonnes/ha) for standing live/dead and downed dead woody materials in forests of the United States by classes of relative density, 2003-2006

Relative Density	Downed, dead woody material	Std. Error	Standing dead	Std. Error	Standing live	Std. Error
0.0-0.1	13.62	0.90	4.29	0.70	10.19	0.40
0.1-0.2	12.81	0.76	3.71	0.34	30.08	0.67
0.2-0.3	14.08	0.74	4.97	0.40	53.22	1.18
0.3-0.4	16.28	0.81	7.15	0.50	78.50	1.48
0.4-0.5	16.82	0.70	8.59	0.57	111.12	2.17
0.5-0.6	16.98	0.88	8.82	0.61	132.02	2.43
0.6-0.7	17.86	1.28	12.74	1.96	163.54	4.69
0.7-0.8	13.63	0.81	7.48	0.71	168.33	5.87
0.8-0.9	21.88	3.42	16.47	2.64	212.85	11.70
0.9-1.0	23.77	2.25	16.94	3.50	222.22	21.27

Table 2: Pearson's correlation coefficients between estimates of downed dead woody material biomass (tonnes/ha) and other stand/site attributes in forests of the United States by classes of relative density, 2003-2006 (Italicized coefficients have p-values> 0.05)

Relative Density	Standing live biomass	Standing dead biomass	30-year mean max. temp	30-year mean min. temp.	30-yr mean annual precipitation	Stand age
0.0-0.1	0.19	0.21	-0.11	*-0.07*	0.14	*-0.03*
0.1-0.2	0.29	0.20	-0.19	-0.12	0.15	*0.02*
0.2-0.3	0.20	0.21	-0.24	-0.18	0.12	*0.02*
0.3-0.4	0.21	0.32	-0.21	-0.14	0.13	*0.01*
0.4-0.5	0.24	0.25	-0.23	-0.21	*0.07*	0.13
0.5-0.6	0.25	0.27	-0.25	-0.21	0.17	0.27
0.6-0.7	0.39	0.36	-0.23	-0.22	0.04	0.21
0.7-0.8	0.25	0.36	-0.23	-0.19	*0.15*	0.24
0.8-0.9	0.68	0.65	-0.26	*-0.20*	*0.17*	0.56
0.9-1.0	0.61	0.56	-0.14	*0.01*	0.42	0.55

exceeded 0.80. For example, stand age only had a correlation coefficient with DDW of -0.03 (p-value> 0.05) when relative density was below 0.10. When a plot's RD was between 0.80 and 0.90, the same correlation had a coefficient of 0.56 (p-value < 0.001). Based on these results, we hypothesize that only stand-level live tree biomass increases consistently with increases in a stand's stocking (i.e., RD). Stand-level biomass for both standing and DDW stand-level only increased at very high levels of stocking, which was further confirmed by DDW correlation results. When constructing models to estimate DDW resources based on stand/site attributes, we propose two models based on a stand's RD: 1) low/moderate and 2) highly stocked.

Conceptual Models of Downed Dead Wood Accretion

A conceptualization of DDW accretion may be developed using a live tree size-density diagram as a framework (Fig. 1). Lightly or moderately stocked stands (in terms of live tree size or biomass) have unpredictable DDW due to management and/or stochastic disturbance events (location D and C in Fig. 1). For stands located in this live tree size-density zone, perhaps the maximum DDW

biomass can be predicted based on an estimate of where the maximum size-density self-thinning line is located. The estimate of maximum live tree stand biomass can be reduced by stochastic disturbance and management events to reflect a stand's unique DDW. Although this approach may be "reverse-engineering" of DDW estimates, it provides a conceptual framework that DDW cannot exceed the maximum live tree size/biomass on a site that has been impacted by stochastic disturbance events. When stands are past the zone of imminent mortality and experiencing self-thinning, DDW resources may be fairly predictable using stand and site attributes (e.g., live tree size/biomass and annual precipitation/temperature) (location A and B in Fig. 1). DDW prediction may be relatively straightforward where stands have not been disturbed by stochastic events and management effects.

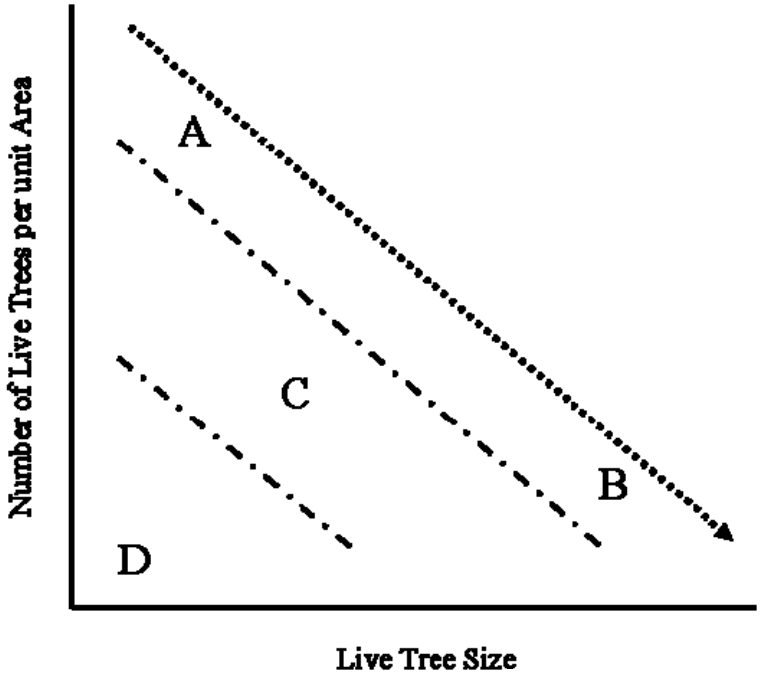


Figure 1: Live tree size density diagram with notable locations in the development of downed dead wood resources: A) high relative density, small live tree size, predictable dead wood resources, B) high relative density, large live tree size, predictable dead wood resources, C) moderate relative density, medium-size live trees, unpredictable dead wood resources, and D) very low relative density, small-size live trees, and highly unpredictable dead wood resources.

Conclusions

Live tree size density attributes of forest stands may provide a framework for understanding and estimating DDW resources in forests across the United States. For stands that are highly stocked in terms of the maximum size-density relationship, DDW resources may be predicted with a reasonable level of confidence due to relatively strong correlations with stand/site attributes. For stands with low/moderate stocking of live trees, an alternative model is proposed

whereby the maximum potential DDW biomass is predicted with deductions for highly improbable but high impact disturbance events. We suggest continued research in the area of stochastic event impacts on DDW resources.

Literature Cited

Bechtold, W.A.; Patterson, P.L., eds. 2005. Forest Inventory and Analysis National Sample Design and Estimation Procedures. Gen. Tech. Report. SRS-GTR-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Bull, E.L.; Parks, C.G.; Torgersen, T.R. 1997. Trees and logs important to wildlife in the Interior Columbia River Basin. Gen. Tech. Rep. PNW-GTR-391. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 55 p.

Drew, T.J.; Flewelling, J.W. 1979. Stand density management: an alternative approach and its application to Douglas-fir plantations. Forest Science. 25: 518-532.

Enquist, B.J.; Brown, J.H.; West, G.B. 1998. Allometric scaling of plant energetics and population density. Nature. 395: 163-165.

Gingrich, S.F. 1967. Measuring and evaluating stocking and stand density in upland hardwood forests in central states. Forest Science. 13: 39-53.

Harmon, M.E.; Franklin, J.F.; Swanson, F.J.; Sollins, P.; et al. 1986. Ecology of coarse woody debris in temperate ecosystems. Advances in Ecological Research. 15: 133-302.

Krajicek, J.E.; Brinkman, K.A.; Gingrich, S.F. 1961. Crown competition—a measure of density. Forest Science. 7: 35-42.

Kukeuv, Y.A.; Krankina, O.N.; Harmon, M.E. 1997. The forest inventory system in Russia. A wealth of data for western researchers. Journal of Forestry. 95: 15-20.

Long, J.N. 1985. A practical approach to density management. Forest Chronicle. 61: 23-27.

Marshall, P.L.; Davis, G.; LeMay, V.M. 2000. Using line intersect sampling for coarse woody debris. Tech. Rep. TR-003. Vancouver, BC: British Columbia Ministry of Forests, Vancouver Forest Region. 34 p.

Maser, C.; Anderson, R.G.; Cromack Jr., K.; 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. Hand. 553. Washington, DC: U.S. Department of Agriculture: 78-95.

McCarthy, B.C.; Bailey, R.R. 1994. Distribution and abundance of coarse woody debris in a managed forest landscape of the central Appalachians. Canadian Journal of Forest Research. 24: 1317-1329.

Norden, B.; Ryberg, M.; Gotmark, F.; et al. 2004. Relative importance of coarse and fine woody debris for the diversity of wood-inhabiting fungi in temperate broadleaf forests. Biological Conservation. 117: 1-10.

PRISM Group. 2004. Parameter-elevation Regressions on Independent Slopes Model [online]. Available online at: http://prism.oregonstate.edu/ [December 1, 2008].

Pyle, C.; Brown, W.M. 1999. Heterogeneity of wood decay classes within hardwood logs. Forest Ecology and Management. 114: 253-259.

Reineke, L. H. 1933. Perfecting a stand-density index for even-aged stands. Journal of Agricultural Research. 46: 627-638.

Rollins, M.G.; Keane, R.E.; Parsons, R.A. 2004. Mapping fuels and fire regimes using remote sensing, ecosystem simulation, and gradient modeling. Ecological Applications. 14: 75-95.

Rubino, D.L.; McCarthy, B.C. 2003. Evaluation of coarse woody debris and forest vegetation across topographic gradients in a southern Ohio forest. Forest Ecology and Management. 183: 221-238.

Shaw, J.D. 2000. Application of Stand Density Index to irregularly structured stands. Western Journal of Applied Forestry. 15: 40-42.

Smith, J.E.; Heath, L.S.; Woodbury, P.B. 2004. How to estimate forest carbon for large areas from inventory data. Journal of Forestry.102: 25-31.

Spetich, M.A.; Guldin, J.M. 1999. The spatial distribution of dead trees across Arkansas timberlands. Southern Silvicultural Research Conference. 10: 53-57.

Tietje, W.D.; Waddell, K.L.; Vreeland, J.K.; et al. 2002. Coarse woody debris in oak woodlands of California. Western Journal of Applied Forestry. 17: 139-146.

Woldendorp, G.; Keenan, R.J.; Ryan, M.F. 2002. Coarse woody debris in Australian forest ecosystems. Canberra, ACT, Australia: Department of Agriculture, Fish and Forestry. 75 p

Woodall, C.W.; Miles, P.D.; Vissage, J.S. 2005. Determining maximum stand density index in mixed species stands for strategic-scale stocking assessments. Forest Ecology and Management. 216: 367-377.

Woodall, C.W.; Monleon, V.J. 2008. Sampling protocols, estimation procedures, and analytical guidelines for down woody materials indicator of the Forest Inventory and Analysis Program. Gen. Tech . Rep. NRS-22. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 68 p.

Woodall, C.W., Heath, L.S., Smith, J.E. 2008. National inventories of dead and downed forest carbon stocks in the United States: opportunities and challenges. Forest Ecology and Management. 256: 221-228.

Vegetation Inventory Data: How Much is Enough?

Bethany Schulz[1], Sonja Oswalt[2], and W. Keith Moser[3]

***Abstract**: Recognition of the value of forest vegetation data has increased in recent years, especially when it is collected using consistent methods over many forest types. Because the cost of collecting large datasets is substantial, managers must balance the cost of collection with the utility of the conclusions that may be drawn from the data analyses. There is no single standard for collecting vegetation data; sampling protocols should be developed to address clearly defined analysis objectives. We compare the utility of the established Phase 3 Vegetation Diversity and Structure Indicator data with the proposed vegetation data to be collected with the Forest Inventory and Analysis Program's Phase 2 Vegetation Profile.*

Keywords: biomass, carbon pools, wildlife habitat potential, fuel characterization, diversity, species distribution, plant community classification, spatial scale

Introduction

Recognition of the value of forest understory vegetation inventory data has increased in recent years. When collected in a consistent fashion over large regions, vegetation inventory data allow for quantitative assessments of existing conditions across broad areas. Repeated visits to permanent plots permit change and trend analyses. The focus of data collection may be relatively simple – such as total foliage cover by height layers or the abundance of general growth habits – or more detailed vegetation composition, such as which species are present or dominate the area sampled. The cost of data collection can be substantial and demands a planning process with clearly defined objectives and a balance of cost and utility. This is especially important when the inventory is designed to monitor trends over time.

The Forest Inventory and Analysis (FIA) program has traditionally conducted timber inventories of the nation's forests. The enhanced program, organized in three phases, is well-suited for collecting vegetation data at different scales and intensities. Phase 1 (P1) uses remotely sensed data to stratify the landscape by coarse physiognomic filters; at its most basic level, P1 might stratify by forest vs.

[1] Bethany Schulz is a research ecologist, Pacific Northwest Research Station, Forestry Sciences Laboratory, 3301 C Street, Suite 200, Anchorage AK 99503;
[2] Sonja Oswalt is a forest resource analyst, Southern Research Station, Forest Inventory and Analysis, 4700 Old Kingston Pike, Knoxville, TN 37919;
[3] W. Keith Moser is a research forester, Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

non-forest lands. Phase 2 (P2) data include the plot-based observations of traditional tree variables, with plots established approximately every 2430 hectares (6000 ac) on forested lands. With this spatial density of plots, population estimates can be derived for some county-sized areas (Bechtold and Patterson 2005), but estimates are more commonly derived for larger land management units (e.g., National Forests, National Parks) or state-wide reports. Moving beyond its commodity-driven origin, many researchers have recognized the value of FIA data to analyze tree species composition and structure that influences wildlife habitat, range, recreation, hydrology and more (Rudis 1991, 2003). Phase 3 (P3) data are collected on a subset of P2 plots (1 out of 16 plots; approximately 1 plot every 38 880 hectares [96,000 ac]) and include additional measurements for monitoring forest health conditions (Bechtold and Patterson 2005). These data are used to establish valuable baseline conditions and detect more detailed changes not assessed in P2 and to provide indications of potential impacts to ecosystem functions that may be worthy of additional investigation. The P3 grid spatial intensity affords population estimates at regional and national levels.

Each FIA regional program has some history of collecting non-tree vegetation data in conjunction with timber inventories. In fact, some of these programs have long histories of collecting data on understory plants (O'Brien 2003, O'Brien et al. 2003). Although methods have been similar among programs, they have varied enough that is difficult to compare or combine data across regional boundaries.

Forest Inventory and Analysis' P3 Vegetation Indicator (P3VEG) provides a method to collect data on all vascular plants growing on forested plots (U.S. Department of Agriculture, Forest Service 2005). Estimations of vascular plant species richness and the distribution and abundance of those species, including the relative abundance of introduced species, may be calculated using the P3 vegetation data. Pattern recognition, such as indicator species analysis and presence/absence of introduced species, may also be performed using P3 species composition data. In addition, composition data can be used to compare differences in species mix with differences in the physical attributes across plots and to develop plant community classifications (Schulz et al. 2008). Data collection requires a dedicated crew member with specialized botanical skills.

Because of the growing recognition of the value of forest understory vegetation inventory data, a team was recently assigned to develop a core-optional method that could be used on P2 plots by any FIA unit to yield comparable data across regions. In order to minimize demands on time, training and staffing and gain efficiencies of resources, data collected with the P2 Vegetation Profile (P2VEG) will be limited to structure, recorded as cover by growth habit by layer, with an option to collect additional information on the "most abundant" species. The objective of this effort is to produce estimates of biomass and structural characteristics that will allow for evaluation of carbon pools, wildfire fuel hazard, wildlife habitat suitability, forage availability, and grazing potential in a consistent way across regions. The option to collect information on the most abundant species would afford refinements in the above estimates as well as allow for plots to be classified beyond forest type using pre-defined community classifications.

It makes sense in efficiency to take advantage of FIA's infrastructure to collect these valued vegetation data in addition to the traditional tree inventory: logistics for training and moving field crews across large regions, and the data management to collect, edit, process, manage and store data are in place. However, the costs of an extra crew member on plot to collect additional variables can be significant and should not be done without careful consideration.

Clearly defined objectives are required when allocating resources to vegetation data collection. Although there is little reason to collect more data than required, it is also inefficient to leave the plot without enough data to address the issues at hand or looming in the future. When planning vegetation inventories, it is important to consider the uses for vegetation inventory data, the level of detail, thoroughness, and spatial scale required to produce desired estimates, and the benefits of data at each level of detail.

We discuss a variety of objectives for collecting vegetation data and compare how the established P3VEG and proposed P2VEG measurements can support these objectives. Differences between the two methods are summarized in table 1.

Table 1: Summary of differences in scale, measured structure, and species data collected with P3VEG and the proposed P2VEG methods.

Method	Scale	Structure	Species Identify	Species Abundance	Arrangement
P3VEG	1 plot / 38 880 ha	Total foliar cover by layer	All vascular plants	Total cover	Cover by layer
P2VEG	1 plot / 2430 ha	Cover of growth habit by layer	4 most abundant per growth habit with cover of at least 3%	Total cover	Tallest layer

Objectives for Collecting Vegetation Inventory Data

There are many potential uses for vegetation inventory data. The basic inventory objective – how much of what is where? – can include:

- Biomass and carbon pools
- Fuel characteristics
- Wildlife habitat
- Diversity
- Species distributions
- Plant community types (species composition and structure)

Both P3VEG and the proposed P2VEG methods, where implemented, will yield cover and height distribution measures that can be used to estimate biomass and describe structure, which are key elements for the first three objectives listed above. Species abundance data can help refine these estimations and assessments, even when species data are limited to only the most abundant species present. Data from thorough species inventories are used to assess both the species

richness and frequency components of diversity, to define both species distributions and range, and to develop plant community classifications.

Biomass and carbon pools

The ability to estimate biomass and describe vegetation structure is central to estimating standing carbon pools, characterizing fuel conditions, and assessing wildlife habitat potential. Estimations of carbon pools in forests is essential to understanding carbon cycles, and changes in stored carbon pools, which are critical as society develops policies to mitigate emissions of CO_2 and other greenhouse gasses. Without complete inventories on forest carbon pools and their dynamics, it will be difficult to develop effective policy and monitoring systems to manage these pools and sustain the services of forest ecosystems to carbon sequestration (Ingerson and Loya 2008). Researchers are continually refining tools to estimate carbon stocks in U.S. forests (Smith et al. 2007), although uncertainties associated with carbon pools in non-tree forest components remain admittedly high (Smith and Heath 2008). Above-ground live tree biomass is estimated at less than one half of total carbon found in forests (Ingerson and Loya 2008). Soils and forest floors are hold a large proportion of carbon in most forest, followed by standing dead and downed wood. Understory vegetation is generally assumed to contribute a small fraction to the overall carbon stock in most forest ecosystems (Birdsey 1992). However, better calibrated estimates will be possible with the availability of direct measurements of abundance and height of understories from many forest types and to clarify the dynamics.

Fuel characterization

Data on vegetation structure and composition are essential to characterize fuel. Understory vegetation influences fire behavior through the quantity of burnable biomass, the vertical structure and arrangement of vegetation, and the species present (Riccardi et al 2007). Some growth habits and species are particularly combustible while other species are very hard to ignite and can act as fire breaks (U.S. Department of Agriculture, Forest Service [Online]). Most fuel analysis tools require data that describe how much vegetation is present and how the existing vegetation is arranged (ladder fuels). For further detailed analysis, information on the species present is necessary.

Many of the same models are currently used to quantify biomass for both carbon pools and fuel characterization. Current tools vary in required data inputs (e.g., the Carbon Calculator Tool [CCT] [Smith et al.] Fire and Fuels Extension [FFE] [Reinhardt and Crookston 2003] of the Forest Vegetation Simulator [Dixon 2002], the Fire Effects Monitoring and Inventory Protocol [FIREMON] [Lutes et al 2006], and the Fuel Characteristic Classification System [Ottmar et al 2007]). The most accurate estimates of biomass can be made when abundance, height, and bulk density of the species are known, but these data are rarely available.

Thus, most models include basic approximations that are based on tree cover data to estimate biomass of understory vegetation. Some use direct measurements of total cover by life form to account for biomass, but use approximated values for bulk density of understory components. Measured attributes of percent canopy cover for herbaceous, grassy, and shrubby growth habits could aid in further calibration of biomass estimates for both carbon modeling and fuel characterization.

Wildlife habitat

Forest management plans often consider impacts and maintenance of particular wildlife habitat elements. Important features for assessing wildlife habitat include overall vegetation structure; cover by growth forms, canopy complexity, presence of dead standing or downed trees, and plant species composition data (Thomas and Verner 1986). Habitat may be assessed at a number of scales – broad regional, landscape, or fine (plot level) scales. At broad scales, course filters such as dominant vegetation type, successional stages, and canopy closure are used to assess habitat conditions for particular species. At finer scales, habitat assessments are highly dependent on specific features required by a species or guild of species at spatial and temporal scales (Noon et al. 2003). Some species have particular habitat requirements and others are more generalists. It is impossible to design a single vegetation inventory that is suitable for every species. However, additional information generated with the core-optional FIA P2VEG methods could be extremely useful for the development of more specific models to predict habitat features at spatial scales that are useful for land use policy makers.

Diversity

Established baseline levels of diversity are critical for assessing changes over time in response to natural succession, disturbance events, or global climate change. Diversity can be evaluated at a variety of scales from ecological regions to genetic materials. The thoroughness of data collection affects the scale at which diversity can be assessed. The FIA program's P1 and core P2 samples can be used to describe regional-scale tree diversity of land cover, forest type, and structure, but provide very limited information about the vegetation under the forest canopy. The proposed FIA P2 vegetation plot measurements can assess diversity in terms of structure (canopy complexity) and growth habit distribution.

Species richness, the number of species present over a standard area, is a fundamental and easily understood measure of diversity. Overall species richness and species richness of each growth habit can only be addressed with a complete inventory of species on standard-sized sampling areas. A complete inventory also allows for the estimation of baseline species richness, comparing native to introduced species richness, and the examination of diversity patterns in measured

stands and across large regions. Although standard P2 data and the proposed P2VEG data can provide information on regional forest type diversity, growth habit, and structural diversity, only the P3VEG can address overall species richness, species richness by-growth habits, and the diversity of native and introduced species.

Species distribution

Determining the distribution of individual species or groups of species is a common use of vegetation inventory data. Investigators may be concerned with invasive species, indicator species, or species with particular characteristics (e.g., important as wildlife forage or cover, species that burn readily or act as suppressants to fire, subsistence use). Species composition data from multiple plots yield frequency and distribution data, affording assessments of where an individual species occurs, how abundant it is, and if it co-occurs regularly with other species. The types of possible analyses of species distribution again depend on the thoroughness of the inventory of species.

For some purposes, data collection is limited to a finite list of species; for example, the top 20 most unwanted invasive species. Researchers can reach some conclusions about how widely those particular species are distributed, but can not determine how those species interact with other species present such as the impacts of introduced species on native flora. For other purposes, species data are limited to the most abundant or dominant species present. This method is informative about the dominant species present and can be useful for assessing wildlife habitat quality and fuel characterizations. However, researchers will not know where a species is absent or present with an abundance below a designated threshold (e.g. a species must be present with a cover of at least 3% to be recorded). This information gap is limiting when assessing species distributions.

When the inventory includes all species present, data can be used to examine any species of interest, including species co-occurrence with other species. If many plots over large areas are included in the sample, the distribution of any species found in the sample can be estimated. Studying the patterns of species distributions and co-occurrences is extremely useful for predicting where species occur in places not sampled.

Finally, FIA Vegetation inventory data based on a grid sampling design is not usually very informative about rare or endangered species distribution – these species are most often found in rare or unique habitats – which systematic sample grids will usually miss.

For analyzing species distributions, the P3 VEG all-species inventory is more valuable than the P2 VEG. Although there is some value in list-based or most abundant data, the conclusions that can be drawn from the limited species data are restricted. Analyses of broad scale species distribution, relative cover of introduced species, indicator species, or patterns of species co-occurrence are most informative when all species are recorded and assessed.

Plant Community Types

Classification of existing vegetation uses species data and data derived from site physical characteristics to group ecologically-like items together. Mapped classes of vegetation describe the landscape and provide vital information concerning the ecological systems a land manager must consider before taking some management action. The ability to describe the plant community where a forest health issue has been observed is an important communication tool. Although FIA plot locations are confidential and scientists cannot reveal exactly where an invasive plant species was located, they can describe the forest plant community, thus providing a detailed search tool for locating areas where problems could occur.

There is no single standard technique used to develop vegetation classifications, but there are efforts to standardize the data required (Tart 2005). The newly revised Federal Geographic Data Committee's National Vegetation Classification System (NVCS) defines standards for classification plots (with enough data that help define vegetation types) and occurrence plots, (plots with fewer data but sufficient to document the occurrence of a previously defined vegetation type) (FGDC 2007).

There is also a hierarchy of classification levels. At the upper level, physiognomic and ecological factors are used to define broad combinations of dominant general growth forms adapted to basic physical conditions. At the mid-level, physiognomic and floristic characteristics define the groups, similar to the forest types used by FIA. The lower level units are defined within the above two groups, with the alliance level defined by a characteristic range of species composition, habitat conditions, and diagnostic species usually found in the uppermost or dominant stratum of vegetation, reflecting regional to sub-regional climate, hydrology, moisture/nutrient conditions, and disturbance regimes. The association is a finer level of detail defined by the characteristic range of species composition, diagnostic species occurrences, and habitat conditions reflecting the local topo-edaphic climax, geological substrate, and hydrologic conditions.

NVCS specifies the detail of data required to develop classifications. The complete species inventory data collected using the P3VEG protocol is sufficient to develop classifications. The limited number of species that can be collected with the proposed P2VEG method restricts its use to occurrence plots – that is, to describe the plant community based on classifications developed from plots with full species inventories. The addition of some species information can increase our understanding of the distribution of classified plant communities – allowing for the mapping of vegetation types with more detail than just forest cover type.

Summary

The plot-based system of the FIA program provides an excellent platform for collecting understory vegetation data. Between the proposed core-optional P2VEG and the P3VEG measurements, FIA has the potential to address many current and emerging issues that tree data alone cannot (table 2). To answer the question "how much vegetation inventory data is enough?" investigators must consider the objectives, spatial scales, and level precision and accuracy required.

Table 2 – Utility of FIA vegetation measurements to address objectives[1]

Objective (How much of what is where?)	P2 Vegetation Profile (1 plot / 2430 ha)		P3 Vegetation Indicator (1 plot / 38 880 ha)	
	Cover by growth habit	Most abundant species	Total foliar cover*	All species
Biomass[2] (carbon/fuel)	Good	Good	Good	More than enough
Structure (fuel/wildlife)	Good	Good	Good	More than enough
Diversity				
Structure	Good	Okay	Good	Good
Growth habit	Good	Okay	Derived	Good
Species richness	---	Not possible	---	Required
Species distribution	---	Presence of most abundant only	---	Best: presence/absence with abundance
Plant community classification	---	Can be used with pre-existing keys to alliance level	---	Can be used to build to association level.

[1]Assumes at least 30 plots per forest type to reduce variance: small land management units may require intensification of sample grid to increase confidence of estimates.
[2]Assumes bulk density of growth habit types or species are known or can be approximated: Biomass (kg/m^2) = Height (m) x Cover (%/100) x Bulk Density (kg/m^3)

When evaluating the utility of each set of measures to address specific objectives, the difference in spatial scale at which they are collected is just as important as the detail of the data collected.

The spatially dispersed P3 vegetation indicator data is more detailed, providing data for the development of vegetation classifications, assessing species distributions and impacts of introduced species on native plant communities, as well as a better detection measure of change while assessing regional and national trends. This information is used to make regional or national assessments for FIA reporting as well as by others concerned with national or regional assessments of the state of forest ecosystems (e.g., Resource Planning Act, Heinz Center's State of the Nation's Ecosystems, Wilderness Society Science and Policy briefs) and is not designed for establishing baseline data at local levels.

Although the proposed core-optional P2 vegetation profile methods will provide general structure and less complete species information, it will be collected on a higher spatial density than the P3 vegetation indicator and will be potentially useful at more localized scales. Land managers can use the data in forest planning, monitoring forest plan standards and effectiveness, monitoring and management of wildlife habitat, monitoring and management of fuels. However, sample size may need to be increased above the intensity of the P2 grid in order to decrease variance and increase confidence in the precision of the estimates (O'Brien et al 2003).

Beyond making population estimations based on direct measurements as discussed here, implementing both P2VEG and P3VEG data collection would provide a wealth of data to help refine current tools and models for estimating carbon pools, describing fuel characteristics, and identifying potential wildlife habitat. Forest plant community classes can be identified and described, improving manager's ability to map areas of concern and improve their ability to monitor the effectiveness of management plans.

Literature Cited

Bechtold, W.A.; Patterson P.L., eds. 2005. The enhanced Forest Inventory and Analysis Program–national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Birdsey, R.A. 1992. Carbon storage and accumulation in the United States forest ecosystems. General Technical Report GTR-WO-59. Washington D.C. U.S. Department of Agriculture, Forest Service, Washington Office. 51 p. **http://www.nrs.fs.fed.us/pubs/5215 (15 December 2008).**

Dixon, G.E. 2002. Essential FVS: A User's Guide to the Forest Vegetation Simulator. Internal Rep. Fort Collins, CO: U. S. Department of Agriculture, Forest Service, Forest Management Service Center. Available online at: http://www.fs.fed.us/fmsc/fvs/documents/gtrs_essentialfvs.php (11 December 2008).

Federal Geographic Data Committee [FGDC]. 2007. National vegetation classification standard, version 2–working draft. FDGC-STD-005. Reston, VA: Vegetation Subcommittee, FGDC Secretariat, U.S. Geological Survey. 124 p. http://www.fgdc.gov/4/projects/FGDC-standards-projects/vegetation/ (23 November, 2008).

Ingerson, A.; Loya, W. 2008. Measuring Forest Carbon: Strengths and weaknesses of available tools. Science and Policy Brief. Washington, D.C. The Wilderness Society. 19 p.

Lutes, D.C.; Keane, R.E.; Caratti, J.F.; Key, C.H.; Benson, N.C.; Sutherland, S.; Gangi, L.J. 2006. FIREMON: Fire effects monitoring and inventory system. General Technical Report RMRS-GTR-164-CD. [CD-ROM] Fort Collins, CO. US Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Noon, B.R.; Murphy, D.D.; Beissinger, M.L.; Shaffer, M.L.; Dellasala, D. 2003. Conservation planning for U.S. National Forests: condicting comprehensive biodiversity assessments. BioScience 53: 1217-1220.

O'Brien, R. A. 2003. New Mexico's Forest Resources, 2000 Resour. Bull. RMRS-RB-3. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 117 p.

O'Brien, R. A.; Johnson, C. M.; Wilson, A. M.; Elsbernd, V. C. 2003. Indicators of rangeland health and functionality in the Intermountain West. Gen. Tech. Rep. RMRS-GTR-104. Ogden, UT: U.S.Department of Agriculture, Forest Service, Rocky Mountain Research Station. 13 p.

Ottmar, R.D.; Sandberg, D.V.; Riccardi, C.L.; Prichard, S.J. 2007. An overview of the Fuel Characteristic Classification System — Quantifying, classifying, and creating fuelbeds for resource planning. Can. J. For. Res. 37: 2383–2393.

Reinhardt, E.; Crookston, N. L. (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.

Riccardi, C.L.; Ottmar, R.D.; Sandberg, D.V; Andreu, A. Elman, E.; Kooper, K.; Long, J. 2007. The fuelbed: a key element of the Fuel Characteristic Classification System. Can. J. For. Res. 37: 2394-2412.

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. 2003. 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.

Tart, D.; Williams, C.; DiBendedetto, J.; Crow E.; Girard, M.; Gordon, H.; Sleavin, K.; Manning, M.; Haglund, J.; Short, B.; Wheeler, D. 2005. Section 2: Existing vegetation classification protocol. In: Brohman, R.; Bryant, L., eds. Existing vegetation classification and mapping guide. Gen. Tech. Rep. WO-67. Washington, DC: U.S. Department of Agriculture, Forest Service, Ecosystem Management Coordination Staff: 2-1–2-34

Thomas, J.W.; and Verner, J. 1986. Forests. In: Cooperrider, A.Y.; Boyd, R.J.; Stuart, H.R., eds. Inventory and monitoring of wildlife habitat. Denver CO: U.S. Department of the Interior, Bureau of Land Management, Service Center. 73-91.

Smith, J.E.; Heath, L.S. 2008. Carbon stocks and stock changes in U.S. Forests. In: U.S. agriculture and forestry greenhouse gas inventory: 1990-2005. U.S. Department of Agriculture. Technical Bulletin No. 1921. Washington D.C.: Office of the Chief Economist: 65-80, C1-C7.

Smith, J.E.;Heath, L.S.; Nichols, M.C. 2007. U.S. Forest carbon calculation tool: Forest-land carbon stocks and net annual stock change. Gen. Tech. Rep. NRS-GTR-13. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 28 p.

Schulz, B.K.; Bechtold, W.A.; Zarnoch, S.J. 2008. Sampling and estimation procedures for the vegetation diversity and structure indicator. Gen. Tech. Rep. PNW-GTR-781. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 53 p.

U.S. Department of Agriculture, Forest Service [USDA FS]. 2005. Forest inventory and analysis national core field guide. Volume 2: Field data collection procedures for phase 3 plots. Version 4.0. U.S. Washington DC: Internal report. On file with: U.S. Department of Agriculture, Forest Service, Forestry Sciences Lab, 3041 Cornwallis

Rd, Research Triangle Park, NC 27709. http://fia.fs.fed.us/library/field-guides-methods-proc/. (3 December 2008).

U.S. Department of Agriculture, Forest Service [Online]. Fire Effects Information System, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (Producer). Available: http://www.fs.fed.us/database/feis. (3 December 2008).

Combining Panel Data



Tree-level imputation techniques to estimate current plot-level attributes in the Pacific Northwest using paneled inventory data

Bianca N.I. Eskelson[1], Hailemariam Temesgen[2], Tara M. Barrett[3]

Abstract: *The Forest Inventory and Analysis program (FIA) of the US Forest Service conducts a nationwide annual inventory. One panel (20% or 10% of all plots in the eastern and western United States, respectively) is measured each year. The precision of the estimates for any given year from one panel is low, and the moving average (MA), which is considered to be the default estimator, can result in biased estimates of current conditions. An alternative to the MA is sought, and studies comparing different alternatives to the MA approach for estimating current forest attributes in the Pacific Northwest are lacking. Paneled data from national forests in Oregon and Washington were used to explore nearest neighbor (NN) imputation methods to project all panels to a common point in time. When using the most recent ground measurements of the panels measured in prior years as ancillary data, tree-level NN imputation outperformed the MA estimator in estimating basal area/ha, stems/ha, volume/ha, and biomass/ha in terms of bias and root mean square error (RMSE) and plot-level NN imputation in terms of RMSE. When basal area/ha, stems/ha, volume/ha, and biomass/ha were summarized by three species groups, tree-level NN imputation outperformed plot-level NN imputation in terms of both bias and RMSE. Tree-level NN imputation outperformed the MA in terms of bias and RMSE for estimating basal area/ha, stems/ha, volume/ha, and biomass/ha for species group 'pine' and provided comparable results in terms of bias and RMSE for species groups 'Douglas-fir' and 'other.'*

Keywords: moving average, nearest neighbor imputation, panel, plot-level, tree-level

Introduction

Information on current forest condition is essential to assess and characterize resources and to support management and policy decisions. The 1998 Farm Bill mandates the US Forest Service to conduct annual inventories to provide annual updates of each state's forest. Only 10% or 20% of all plots in the western and eastern United States, respectively, are measured annually. Because only a portion

[1] Oregon State University; Department of Forest Engineering, Resources and Management; 204 Peavy Hall; Corvallis, OR 97331-5706 USA; e-mail: bianca.eskelson@oregonstate.edu.
[2] Oregon State University; Department of Forest Engineering, Resources and Management; 204 Peavy Hall; Corvallis, OR 97331-5706 USA.
[3] United States Forest Service; Pacific Northwest Research Station; Forest Sciences Lab; 3301 "C" Street, Suite 200, Anchorage, AK 99503-3954 USA.

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

of the full sample is measured annually, the precision of the estimates for any given year is low. To achieve higher precision, the Forest Inventory and Analysis program (FIA) uses a moving average (MA) as default estimator which combines the data of multiple panels. In the presence of trend, biased estimates will result, if the MA is applied to the end of the period to estimate current conditions. Other approaches to combine data from all panels include: 1) updating unmeasured panel data to the current year with growth models (Lessard et al. 2001); 2) using time series models (Johnson et al. 2003); 3) mixed estimation (Van Deusen 1996); or 4) filling in missing panel data using tree- and plot-level imputation techniques (Gartner and Reams 2001, 2002, McRoberts 2001). Since spatial, temporal, and forest characteristics differ within and among regions it is unclear if any single technique will provide satisfactory results for all regions (Patterson and Reams 2005). It may be necessary to evaluate different methods for a variety of issues and regions. Studies comparing different alternatives to the MA approach for estimating current forest attributes in the Pacific Northwest (PNW) are lacking.

Nearest Neighbor (NN) imputation methods are donor-based, which means that the imputed value was either observed for another unit or was calculated as the average of values from more than one unit. NN imputation can be performed on different levels. Eskelson et al. (2009) have shown that plot-level imputation, that is plot-level attributes (e.g., basal area/ha) are imputed, can provide more accurate results than the MA approach. They found the randomForest (RF) imputation method (Crookston and Finley 2008), which is an extension of classification and regression tree (CART) methods (Breiman 2001), to outperform other NN imputation methods. Imputation can also be performed at the tree-level, that is tree-level attributes (e.g., diameter at breast height (DBH in cm)) are imputed, and the results of the tree-level imputation are then summarized for each plot (e.g., imputed DBH is used to calculate basal area/ha).

Depending on the intended use, tree- and plot-level imputation techniques differ in their predictive abilities and suitability (Gartner and Reams 2002). Plot-level and tree-level NN imputation techniques might have a similar relationship to each other as whole stand growth models, which might not apply in heterogeneous conditions (Curtis and Hyink 1985), have with single-tree growth models, which can provide more detailed information about stand dynamics and structure (Burkhart 1992). Tree-level nearest neighbor (NN) imputation techniques have been successfully used to estimate tree volumes and heights (Korhonen and Kangas 1997), single-tree biomass (Fehrmann et al. 2008) as well as 5-year diameter growth and bark thickness (Sironen et al. 2001, 2003, 2008).

The objectives of this study are to: 1) use paneled data from the PNW to estimate current forest attributes (see Table 1) using tree-level imputation methods and compare their performance against the MA and the estimates based only on the data from the current panel; 2) examine the performance of tree-level imputation methods to estimate current forest attributes by species groups; and 3) compare tree-level and plot-level imputation results.

Methods

Data

The data used in this study consist of 618 plots from six national forests that were collected as part of the Pacific Northwest Region's Current Vegetation Survey (CVS) of the US Forest Service. The plots were installed between 1993 and 1997 and remeasured in 2000. The particular national forests sampled were the Colville (28), Mt. Hood (111), Ochoco (82), Rogue River (70), Wallowa-Whitman (199), and Winema (128).

Five plots are installed in each basic CVS sampling unit, which is one hectare (ha) in size. Each plot consists of three permanent circular, nested subplots of different sizes in which trees are measured depending upon their DBH. For a detailed description of the CVS inventory see Max et al. (1996). Tree height (HT in m) is only subsampled and missing HTs were filled using height models developed in Barrett (2006) for live trees with DBH of 12.7 cm or larger. Volume and biomass equations from the US Forest Service were used to calculate gross cubic-meter volume and total gross oven dry weight biomass (USDA 2000). For each plot, basal area in m^2 per ha (BA), stems per ha (SPH), volume in m^3 per ha (VOL), and biomass in tons per ha (BIOT) were calculated and summarized (Table 1). BA, SPH, VOL, and BIOT were also calculated for each of the following three species groups: 1) 'Douglas-fir'; 2) 'pine' including all occurring pine species; and 3) 'other' including other conifers and hardwoods. Basal area in larger trees (BAL in m^2) was calculated for each tree. Ingrowth for each plot was determined by calculating BA, SPH, VOL, and BIOT for all trees that were present in 2000 but not present at the first measurement occasion. BA and SPH were also calculated for small trees with DBH larger than 2.54 cm and smaller than 12.7 cm.

Table 1: Summary of plot-level variables in 2000.

Variable	Minimum	Mean	Maximum	Std
Basal area (m^2/ha)	0.24	24.32	105.35	19.00
SPH (stems/ha)	1	305	1517	221
Volume (m^3/ha)	0.66	224.82	1444.74	221.04
Biomass (tons/ha)	0.58	134.09	800.64	132.64
Canopy cover (%)	0	54	97	29
Slope (%)	0	23	83	17
Elevation (m)	274	1389	2377	321
Annual precipitation (ln cm) (scaled * 100)	577	683	817	48
Annual mean temperature (ºC) (scaled * 100)	60	579	1067	166

The data set comprises 30,709 trees in 33 species. The most common species in decreasing order are Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco), ponderosa pine (*Pinus ponderosa* C. Lawson), grand fir (*Abies grandis* (Douglas ex D. Don) Lindl.), lodgepole pine (*Pinus contorta* Douglas ex Louden), white fir

(*Abies concolor* (Gord. & Glend.) Lindl. ex Hildebr.), and western hemlock (*Tsuga heterophylla* (Raf.) Sarg.) (see Eskelson et al. 2009 for details).

In NN imputation methods, ancillary variables are those variables that are measured on all units. Thematic Mapper (TM) images from 2000 were extracted from the national land-cover database 2001 [NLCD 2001 (Homer et al. 2004)] and used as ancillary data. The raw imagery bands 1 to 5 and band 7 (TM1, TM2, TM3, TM4, TM5, TM7) as well as the Tasseled Cap (TC) transformations (TC1 – TC6) (Kauth and Thomas 1976) were used. The normalized difference vegetation index (NDVI) and three commonly used band ratios (band 4 to band 3 (R43), band 5 to band 4 (R54), and band 5 to band 7 (R57)) were calculated. Percent canopy cover was extracted from the NLCD 2001 (Homer et al. 2004).

The following climate and topography variables for plot locations were additionally used as ancillary data: Annual precipitation and mean annual temperature (Table 1) [Data source: DAYMET Daily Surface Weather Data and Climatological Summaries (Thornton et al. 1997, Thornton and Running 1999)], elevation (EL in m) and transformations (EL^2, ln(EL)) [Data source: CVS inventory], and slope (%) and aspect (degrees) and transformations (cosine(aspect), sine(aspect), cosine(aspect)*slope, and sine(aspect)*slope) [Data source: 30 m digital elevation model using Arc Workstation GRID surface functions and commands (Environmental Systems Research Institute 1991)]. These climate, topography, and satellite variables have been successfully used as ancillary data for NN imputation methods in previous studies (e.g., Eskelson et al. 2009, Ohmann and Gregory 2002).

Imputation techniques

Panel data is a special case of inventory data with measurements taken at different times. All plots were remeasured in 2000. In order to mimic a panel system with the available data 25% of the plots (154) were randomly assigned to *P4* and the remaining 75% of the plots (464) were assigned to *P1*, *P2*, and *P3* based on their year of installation. This resulted in *P1*, *P2*, and *P3* having different sizes for each iteration (Table 2).

Table 2: Number of plots measured by year of installation and corresponding panel assignment. All plots listed were remeasured in 2000.

Year of Installation	# of Plots	Assigned Panel
1993	7	1
1994	229	1
1995	223	2
1996	158	3
1997	1	3

The variables of interest (Y) in this study were BA, SPH, VOL, and BIOT. Their observed mean value in the year 2000 was calculated as:

$$\overline{Y}_{OBS} = \sum_{i=1}^{n} Y_i / n \qquad [1]$$

where Y_i is the observed Y value of the i^{th} plot in 2000 and $n = 618$. The observed mean value was used as best available estimate of the true mean.

For each plot in *P4*, BA, SPH, VOL, and BIOT were calculated using the tree data from *P4*. The mean values of Y for the year 2000 (SAMPLE25 estimator) were calculated as:

$$\overline{Y}_{SAMPLE25} = \sum_{i:Y_i \in P4} Y_i / n_4 \quad [2]$$

where Y_i is the observed Y value of the i^{th} plot, and n_4 is the number of plots in *P4*.

The MA estimator, the FIA default method is:

$$\overline{Y}_{MA(4)} = 0.25 * \overline{Y}_{t-3,i} + 0.25 * \overline{Y}_{t-2,i} + 0.25 * \overline{Y}_{t-1,i} + 0.25 * \overline{Y}_{t,i} \quad [3]$$

where $\overline{Y}_{t-3,i}$, $\overline{Y}_{t-2,i}$, $\overline{Y}_{t-1,i}$, and $\overline{Y}_{t,i}$ are the mean values of the variables of interest of *P1*, *P2*, *P3*, and *P4*, respectively. The MA takes into account that the panels include different numbers of plots. Instead of equal weighting of the panels a weighted version of [3] is proposed:

$$\overline{Y}_{WMA(4)} = w_{t-3} * \overline{Y}_{t-3,i} + w_{t-2} * \overline{Y}_{t-2,i} + w_{t-1} * \overline{Y}_{t-1,i} + w_t * \overline{Y}_{t,i} \quad [4]$$

where w_{t-3}, w_{t-2}, w_{t-1}, and w_t are the weights of *P1*, *P2*, *P3*, and *P4*, respectively. Larger weights were chosen for *P3* and *P4* ($w_{t-1} = w_t = 0.3$) than for *P1* and *P2* ($w_{t-3} = w_{t-2} = 0.2$). MA(4) and WMA(4) will be referred to as MA and WMA, respectively.

Instead of using the previous measurements to fill in the Y values for *P1*, *P2*, and *P3*, as is done with MA and WMA, the current Y values of *P1*, *P2*, and *P3* were imputed using tree-level RF and plot-level RF imputation. Target data are units that have ancillary variables measured only (e.g., trees or plots in *P1*, *P2*, and *P3*). Reference data are units where both variables of interest and ancillary variables were measured (e.g., trees or plots in *P4*). RF imputation was employed using the yaImpute R package (Crookston and Finley 2008). Details on RF imputation can, for example, be found in Hudak et al. (2008).

For tree-level RF, the target trees were assumed to be non-sampled trees lacking inventory data in 2000. DBH, HT, and mortality for each target tree were imputed using DBH, HT, and BAL at the previous measurement (DBHocc1, HTocc1, and BALocc1) as ancillary data. The reference trees constituted the pool of potential trees with inventory and ancillary data (*P4*), which could be selected to impute the DBH, HT, and mortality for the target trees. Ingrowth of BA, SPH, VOL, and BIOT was imputed at the plot-level using BA and SPH of small trees at

the previous measurement as well as the available climate, topography, and satellite data as ancillary data. BA, SPH, VOL, and BIOT were calculated for each plot based on the imputed tree data and the imputed ingrowth.

For plot-level RF the previous measurements of the four variables of interest (BAocc1, SPHocc1, VOLocc1, BIOTocc1) were used as ancillary data since this was found to provide better imputation results than using climate, topography, and satellite data in a previous study. For more details see Eskelson et al. (2009).

For both tree-level and plot-level RF the overall mean of the variables of interest for the year 2000 was estimated as:

$$\overline{Y}_{IMP} = \left[\sum_{i:Y_i \in P1} Y_{imp,i} + \sum_{i:Y_i \in P2} Y_{imp,i} + \sum_{i:Y_i \in P3} Y_{imp,i} + \sum_{i:Y_i \in P4} Y_{t,i} \right] / n \qquad [5]$$

where $Y_{imp,i}$ is the imputed Y value for the i^{th} plot and 'IMP' refers to either tree-level RF or plot-level RF.

SAMPLE25, MA, WMA, and the tree-level and plot-level RF imputation methods were compared based on the overall means of the variables of interest in 2000 (see Equations 2 – 5). The five estimation methods were also compared based on their performance of estimating the four variables of interest by species groups 'Douglas-fir', 'pine', and 'other.'

The basis of evaluation was accuracy, as expressed by the root mean square error (RMSE), and bias, calculated as the mean difference between the estimates (Equations 2 – 5) and the observed mean values (Equation 1) from $m = 200$ iterations of randomly splitting the data. Two hundred iterations were considered sufficient because other studies have found RMSE and bias to stabilize at around 200 iterations (e.g., Arner et al. 2004). Both RMSE and bias were expressed as percent of the observed mean for each variable of interest:

$$Bias\ \% = \frac{\sum_{i=1}^{n} \frac{\left(est_i - obs_i\right)}{m}}{\frac{\sum_{i=1}^{n} obs_i}{m}} * 100 \qquad [6]$$

$$RMSE\ \% = \frac{\sqrt{\sum_{i=1}^{n} \frac{\left(est_i - obs_i\right)}{m}}}{\frac{\sum_{i=1}^{n} obs_i}{m}} * 100 \qquad [7]$$

Results

The SAMPLE25 estimator provided virtually unbiased estimates for all four variables of interest. Its RMSE values ranged from 4.89% for SPH to 6.58% for BIOT. The MA estimates had a negative bias with values from -1.93% for VOL to -2.58% for SPH. The MA estimator provided very precise estimates with the bias contributing most to the RMSE. The WMA estimator reduced both bias and RMSE for BA and SPH. For WMA the bias of VOL and BIOT estimates was positive and the RMSE was larger than those for MA (Table 3).

Plot-level RF imputation resulted in small negative bias and smaller RMSE values than those of the MA estimator. In terms of RMSE, plot-level RF imputation only outperformed the WMA for VOL and BIOT (Table 3).

Table 3: Imputation results.

	BA		SPH		VOL		BIOT	
Method	**% bias**	**% RMSE**	**% bias**	**% RMSE**	**% bias**	**% RMSE**	**% bias**	**% RMSE**
SAMPLE25	0.05	5.29	-0.20	4.89	0.20	6.53	0.26	6.58
MA	-2.53	2.60	-2.58	2.63	-1.93	2.08	-1.97	2.12
WMA	0.59	1.03	-1.54	1.72	2.52	2.74	2.62	2.83
plot-level RF	-0.44	1.50	-0.73	2.52	-0.26	1.78	-0.22	1.66
tree-level RF	0.44	1.09	-0.60	1.31	0.43	1.36	0.42	1.35

Tree-level RF imputation produced a small positive bias in BA, VOL, and BIOT but a small negative bias in SPH. Its RMSE values were smaller than those of the MA and the plot-level RF imputation. Tree-level RF imputation outperformed the WMA estimates in terms of bias and RMSE for SPH, VOL, and BIOT. The variance contributed most to the RMSE for both tree- and plot-level imputation (Table 3).

By species group the SAMPLE25 estimator provided virtually unbiased results (0.62% or less). RMSE values ranged from 8.52% for ‘pine’ BA to 11.17% for ‘other’ BIOT (Table 4).

The MA estimator resulted in a larger negative bias for the four variables of interest for species group ‘pine’ which contributed most to the RMSE values of more than 9%. For the species group ‘Douglas-fir’ and ‘other,’ MA resulted in small bias with absolute values ranging from 0.30% to 1.17% and RMSE values ranging from 1.00% to 1.68% (Table 4).

WMA estimates were biased for all three species groups with the bias being largest for ‘pine.’ The bias contributed most to the RMSE values, which exceeded the RMSE values of the MA estimates and the RMSE values for ‘pine’ for the SAMPLE25 estimates (Table 4).

Table 4: Tree- and plot-level imputation results by species group.

	BA 'Douglas-fir'		BA 'pine'		BA 'other'		SPH 'Douglas-fir'		SPH 'pine'		SPH 'other'	
Method	**% bias**	**% RMSE**	**% bias**	**% RMSE**	**% bias**	**% RMSE**	**% bias**	**% RMSE**	**% bias**	**% RMSE**	**% bias**	**% RMSE**
SAMPLE25	0.41	9.94	0.29	8.52	-0.33	10.10	-0.39	10.21	0.49	10.04	-0.56	9.99
MA	0.40	1.14	-9.85	9.89	-0.70	1.18	-0.30	1.00	-9.28	9.35	0.61	1.15
WMA	6.63	6.79	-13.49	13.57	3.86	4.12	2.21	2.64	-14.51	14.61	4.95	5.14
plot-level RF	2.34	9.65	2.19	9.75	-4.01	8.25	2.05	11.01	2.01	10.36	-4.13	9.01
tree-level RF	-0.54	1.63	0.98	1.68	-0.39	1.21	-0.57	2.25	-1.54	2.80	-0.93	1.74

	VOL 'Douglas-fir'		VOL 'pine'		VOL 'other'		BIOT 'Douglas-fir'		BIOT 'pine'		BIOT 'other'	
Method	**% bias**	**% RMSE**	**% bias**	**% RMSE**	**% bias**	**% RMSE**	**% bias**	**% RMSE**	**% bias**	**% RMSE**	**% bias**	**% RMSE**
SAMPLE25	0.62	10.99	0.23	8.84	-0.15	11.05	0.56	10.68	0.23	8.62	0.02	11.17
MA	-0.77	1.50	-9.27	9.31	0.48	1.19	-1.17	1.68	-9.36	9.40	0.48	1.21
WMA	6.88	7.09	-11.89	11.98	5.59	5.82	6.06	6.28	-12.15	12.23	5.95	6.17
plot-level RF	2.00	9.94	1.94	10.36	-3.17	8.57	1.81	9.63	2.18	10.24	-3.04	8.39
tree-level RF	-0.84	2.05	3.74	4.02	-1.25	1.97	-3.37	3.76	8.36	8.54	-0.91	1.81

Plot-level RF imputations resulted in a smaller bias than with WMA for all species groups. However, RMSE values for RF exceeded those of WMA for all but 'pine' (Table 4).

Tree-level RF imputation outperformed SAMPLE25, WMA, and plot-level RF imputation in terms of RMSE. Compared to MA, tree-level RF imputation provided smaller RMSE values for 'pine' and slightly larger RMSE values for 'Douglas-fir' and 'other' (Table 4).

Discussion

The performance of the MA estimator in terms of the variance-bias trade-off was as expected. As in most other studies (e.g., Arner et al. 2004, Johnson et al. 2003, Van Deusen 2002), the large bias was found to be more than compensated for by the high precision. Hence, MA provided better estimates in terms of accuracy than SAMPLE25. MA is a temporal 'midpoint' estimator yielding biased estimates at the end of a time-series in the presence of trend (Roesch and Reams 1999). Giving more weight to the more recently measured panels resulted in the WMA estimator, which improved the estimates for BA and SPH in terms of bias and hence, also in terms of RMSE.

MA by species groups outperformed WMA in terms of both bias and RMSE. The larger weights applied for *P3* and *P4* for WMA increased the negative bias for species group 'pine' and resulted in large positive bias for all variables of interest for species groups 'Douglas-fir' and 'other.' This indicates that weights applied to the WMA which improve the MA for estimating BA, SPH, VOL, and BIOT do not necessarily improve the MA when the variables of interest are summarized by species group. Choosing appropriate weights for the WMA requires the knowledge of the trend inherent in the data. If the trend inherent in BA, SPH, VOL, and BIOT differs from the trend of the variables of interest summarized by species group, different weights need to be chosen for the WMA. Objective ways for choosing appropriate weights are still lacking. Panels that do not change much should receive larger weights than panels that exhibit a lot of change. Knowledge about change could possibly be acquired from remotely sensed data, growth models, or other information on, for example, fire or insect outbreaks.

Tree-level RF imputation outperformed MA in terms of bias and RMSE for estimating BA, SPH, VOL, and BIOT. This is due to the lag bias inherent in the MA estimator. Tree-level imputation attempts to update the tree data, which results in a smaller bias than that observed for MA. Compared to the WMA, which tries to adjust the lag bias of the MA estimator, the improvement of tree-level RF imputation is less pronounced. If the lag bias of the MA could be adjusted, MA might outperform RF tree-level imputation in terms of both bias and RMSE since the MA estimates are more precise than those of the tree-level RF imputation.

When the variables of interest were summarized by species groups, the MA slightly outperformed tree-level RF imputation in terms of RMSE for species groups ‘Douglas-fir’ and ‘other’ because the MA resulted in low bias for those variables. For species group ‘pine’ MA resulted in large bias and therefore tree-level RF imputation provided much better results for the variables of interest for species group ‘pine’ in terms of bias and RMSE.

Tree-level RF imputation outperformed plot-level imputation for estimating BA, SPH, VOL, and BIOT as well as for estimating the variables of interest summarized by species groups. The results of this study suggest that tree-level RF imputation should be preferred over plot-level RF imputation for estimating total BA, SPH, VOL, and BIOT or for estimating BA, SPH, VOL, and BIOT by species group. The same considerations for choosing single-tree growth models over whole-stand growth models probably apply for choosing tree-level NN imputation over plot-level NN imputation and depend mainly on the demands of the user.

In this study, tree-level variables were imputed using reference trees irrespective of whether the tree species of reference and target trees matched. Imputing only within tree species or species group might improve the results for tree species such as Douglas-fir, ponderosa pine, grand fir, lodgepole pine, white fir, and western hemlock which occur frequently in the data set. However, results for rare tree species would definitely degrade with decreasing number of possible reference trees. Overall results could possibly be improved by imputing tree-variables for frequent tree species within tree species but using the complete reference data set for rare tree species.

Conclusions

This study has shown that tree-level RF imputation has the potential to provide better results in terms of bias and accuracy for estimating plot-level attributes such as BA, SPH, VOL, and BIOT than can be achieved with the SAMPLE25, MA, and WMA estimators, or plot-level RF imputation.

Giving more weight to most recently measured panels by using a WMA improved the estimates for BA, SPH, VOL, and BIOT compared to the MA estimates. When the variables of interest were summarized by species group, MA outperformed WMA in terms of bias and accuracy. More research is warranted for finding objective methods for choosing appropriate weights.

Tree-level RF imputation outperformed MA and WMA in terms of bias and accuracy when BA, SPH, and VOL were estimated. When the variables of interest were summarized by species group, MA provided slightly better estimates in terms of accuracy than tree-level RF imputation.

Tree-level imputation outperformed plot-level imputation. This might be due to the fact that tree-level NN imputation requires more information and is based on a more detailed representation of the stand than plot-level imputation.

The results of the tree-level NN imputation methods tested in this study provide a good argument to further develop the application of tree-level NN imputation techniques for estimating current forest attributes from paneled inventory data.

Acknowledgements

We thank Janet Ohmann and Matt Gregory for sharing their ancillary data from the NLCD and DAYMET with us and for their insights on the data. We appreciate the help that Jim Alegria, Carol Apple, Bob Brown, and Melinda Moeur provided in obtaining national forest data, and thank Kurt Campbell for assistance with volume and biomass equations. We also thank Frank Roesch, Steen Magnussen, and Emilie Grossmann for their helpful review comments.

Literature Cited

Arner, S.L., J.A. Westfall, and C.T. Scott. 2004. Comparison of annual inventory designs using forest inventory and analysis data. Forest Science 50(2): 88-203.

Barrett, T.M. 2006. Optimizing efficiency of height modeling for extensive forest inventories. Canadian Journal of Forest Research 36:2259-2269.

Breiman, L. 2001. Random forests. Machine Learning 45(1): 5-32.

Burkhart, H.E. 1992. Tree and stand models in forest inventory. P. 164-170 in *Proceedings of Ilvessalo Symposium on National Forest Inventories*. IUFRO S4.02. Finnish Forest Research Institute Research Papers 444, Helsinki, Finland.

Crookston, N.L. and A.O. Finley. 2008. yaImpute: An R package for kNN Imputation. Journal of Statistical Software 23(10):1-16.

Curtis, R.O. and D.M. Hyink. 1985. Data for growth and yield models. P. 1-5 in *Proceedings – Growth and yield and other mensurational tricks: A regional technical conference*, Van Hooser, D.D. and N. Van Pelt (eds.). USDA Forest Service General Technical Report INT-GTR-193, Ogden, UT.

Eskelson, B.N.I., H. Temesgen, and T.M. Barrett. 2009. Estimating current forest attributes from paneled inventory data using plot-level imputation: a study from the Pacific Northwest. Forest Science 55(1):64-71.

Environmental Systems Research Institute [ESRI]. 1991. Cell-based modeling with GRID. ESRI Inc., US, Redlands, California.

Fehrmann, L., A. Lehtonen, C. Kleinn, and E. Tomppo. 2008. Comparison of linear and mixed-effect regression models and a k-nearest neighbour approach for estimation of single-tree biomass. Canadian Journal of Forest Research 38:1-9.

Gartner, D. and, G.A. Reams. 2001. A comparison of several techniques for estimating the average volume per acre for multipanel data with missing panels. P. 76-81 in *Proceedings of the second annual Forest Inventory and Analysis Symposium*, Reams, G.A., R.E. McRoberts, and P.C. Van Deusen (eds.). USDA Forest Service General Technical Report SRS-GTR-47, Asheville, NC.

Gartner, D., and G.A. Reams 2002. A comparison of several techniques for imputing tree level data. P. 103-108 in *Proceedings of the Third Annual Forest Inventory and*

Analysis Symposium, McRoberts, R.E., G.A. Reams, P.C. Van Deusen, and J.W. Moser (eds.). USDA Forest Service General Technical Report NC-GTR-230, St. Paul, MN.

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan. 2004. Development of a 2001 National Landcover Database for the United States. Photogrammetric Engineering and Remote Sensing 70(7):829-840.

Hudak, A.T., N.L. Crookston, J.S. Evans, D.E. Hall, and M.J. Falkowski. 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sensing of Environment 112(5): 2232-2245.

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

Kauth, R.J., and G.S. Thomas. 1976. The tasseled cap—a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. PP. 4B41-4B51 in *Proceedings of the symposium on machine processing of remotely sensed data*, Purdue University. West Lafayette, IN.

Korhonen, K.K., and A. Kangas. 1997. Application of nearest-neighbour regression for generalizing sample tree information. Scandinavian Journal of Forest Research 12:97-101.

Lessard, V.C., R.E. McRoberts, and M.R. Holdaway. 2001. Diameter growth models using Minnesota forest inventory and analysis data. Forest Science 47(3):301-310.

Max, T.A., H.T. Schreuder, J.W. Hazard, J. Teply, and J. Alegria. 1996. The Region 6 vegetation inventory and monitoring system. USDA Forest Service General Technical Report PNW-RP-493. 22 p.

McRoberts, R.E. 2001. Imputation on model-based updating techniques for annual forest inventories. Forest Science 47(3):322-330.

Ohmann, J.L., and M.J. Gregory. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, U.S.A. Canadian Journal of Forest Research 32:725-741.

Patterson, P.L., and G.A. Reams. 2005. Combining panels for forest inventory and analysis estimation. P. 69-74 in *The enhanced forest inventory and analysis program—National sampling and estimation procedures*, Bechtold, W.A., and P.L. Patterson (eds.). USDA Forest Service General Technical Report SRS-GTR-80. Asheville, NC.

Roesch, F.A., and G.A. Reams. 1999. Analytical alternatives for an annual inventory system. Journal of Forestry 97(12):33-37.

Sironen, S., Kangas, A., Maltamo, M., and J. Kangas. 2001. Estimating individual tree growth with k-nearest neighbour and k-most similar neighbour methods. Silva Fennica 35(4):453-467.

Sironen, S., A. Kangas, M. Maltamo, and J. Kangas. 2003. Estimating individual tree growth with nonparametric methods. Canadian Journal Forest Research 33:444-449.

Sironen, S., A. Kangas, M. Maltamo, and J. Kangas. 2008. Localization of growth estimates using non-parametric imputation methods. Forest Ecology and Management 256:674-684.

Thornton, P.E. and S.W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. Agricultural and Forest Meteorology 93:211-228.

Thornton, P.E., S.W. Running, and M.A. White. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. Journal of Hydrology 190:214-251.

USDA 2000. Western Oregon 1995-1997, August 15th, 2000. A CD available online from USDA/ FS PNW Research at http://www.fs.fed.us/pnw/fia/; last accessed Sep.10, 2008.

Van Deusen, P.C. 1996. Incorporating predictions into an annual forest inventory. Canadian Journal of Forest Research 26(9):1709-1713.

Van Deusen, P.C. 2002. Comparison of some annual forest inventory estimators. Canadian Journal of Forest Research 32:1992-1995.

Opportunities to Improve Monitoring of Temporal Trends with FIA Panel Data

Raymond L. Czaplewski[1] and Michael T. Thompson[2]

***Abstract**: The Forest Inventory and Analysis (FIA) Program of the Forest Service, Department of Agriculture, is an annual monitoring system for the entire United States. Each year, an independent "panel" of FIA field plots is measured. To improve accuracy, FIA uses the "Moving Average" or "Temporally Indifferent" method to combine estimates from multiple panels that were measured during recent years. However, time-series estimators better serve monitoring objectives than temporally indifferent methods.*

This paper reviews the Kalman filter, which is a linear, minimum variance, sequential, model-based, time-series estimator based on the simple composite estimator. The Kalman filter combines predictions from a population dynamics model with the observed time-series of annual FIA panel estimates. This combination of design-based and model-based methods reduces serious risks from model bias, yet preserves the gain in precision from the model. Alternative models in the Kalman filter represent alternative hypotheses that may be ranked based on their relative agreement with design-based panel estimates. For example, does a model that includes the expected consequences of climate change on average rates of tree growth, regenerations and mortality better fit the annual FIA design-based panel estimates than a model that assumes no climate change?

The Kalman filter is presented in a tutorial style that relies more on graphical examples than mathematical equations. Hopefully, this genre builds awareness and confidence in this somewhat unfamiliar statistical estimator. The Kalman filter and Moving Average estimators are compared with hypothetical simulations of changing populations. A final set of examples is based on annual FIA panel estimates for the State of Colorado from 2002 to 2007, where epidemic levels of mountain pine beetle infestation are causing catastrophic tree mortality in lodgepole pine forests.

Three analysis questions are addressed. Is there an observable trend in population parameters over time? Does the trend make sense? Is the trend significant relative to the uncertainty in the population estimates?

Keywords: Forest monitoring, Kalman filter, composite estimator, *Pinus contorta*, *Dendroctonus ponderosae*, population dynamics model, AIC, information-theoretics.

[1] United States Forest Service; Rocky Mountain Research Station; Inventory, Monitoring, and Analysis Program; 240 W. Prospect; Fort Collins, CO 80526 rczaplewski@fs.fed.us; http://www.fs.fed.us/rm/ogden/

[2] United States Forest Service; Rocky Mountain Research Station; Inventory, Monitoring, and Analysis Program; 507 25th Street; Ogden, UT 84401; mtthompson@fs.fed.us; http://www.fs.fed.us/rm/ogden/

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. Forest Inventory and Analysis (FIA) Symposium 2008; October 21-23, 2008; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

The U.S. Forest Service Research and Development branch, through its Forest Inventory and Analysis (FIA) program, has long been committed to the delivery of current, consistent, and credible information about the status, condition, and trends of America's forests across all land ownerships and conditions. FIA has made impressive progress during the past 10 years in annual updates to these forest inventories. However, the recent FIA strategic plan for 2007-2011 (U.S. Forest Service, 2007a) stresses the importance of monitoring changes over time in forest conditions, which goes beyond an annually updated forest inventory. Monitoring changes over time is critically important to substantive strategic analyses of the nation's forests and the long-term relevance of the FIA program.

FIA has used "periodic surveys" since its inception in the 1930's. Bechtold and Patterson (2005:82) define a periodic survey as is a "*strategy whereby a set of inventory panels is measured simultaneously over a short time frame, often 1 to 3 years in the case of FIA, and there is a time lag, often many years, before the panels are remeasured.*" During recent years, FIA has re-engineered itself by moving from periodic surveys to annual surveys, in which a small sub-sample of field plots is re-measured every year in every county. Relative to periodic surveys, annual surveys are expected to improve the ability to detect and interpret changes in forest conditions (Gadbury and others 2004). McRoberts (2005) highlights the need for statistical estimators that combine population estimates or sample units measured in multiple years, and this might utilize model-based updating techniques.

Current Statistical Methods for Annual FIA Inventories

The annual FIA design is organized around a system of 5 "panels" (*i.e.*, independent sub-samples) in the eastern USA and 10 panels in the western USA. A panel is a systematic sub-sample of all permanent FIA primary sampling units (Phase-2, or "P2", field plots) for which field measurements are conducted on 2 or more occasions. Each FIA panel is composed of a spatially balanced, systematic, interpenetrating sub-sample of all FIA field plots (Reams and others, 2005). In simple terms, FIA treats each panel of field plots as an independent, equal-probability sample of the entire population. All FIA field plots in the first panel are measured during the course of 1 or more years[3]. After essentially all plots in the first panel are measured, field measurements of the second panel commence. Afterwards, the 3rd panel is measured, and so forth for the remaining panels. This sequence of field plot measurements is repeated over time until all FIA panels are measured. This typically takes 5 to 10 years, depending on the region of the country and available funding. After the initial 5 to 10 years, this same sequence

[3] The annual FIA design originally envisioned measurement of a single panel within a single calendar year. However, within the limits of available funds, it often requires more than 12 months to measure a single panel. This is termed "panel creep."

of annual panel measurements is repeated in subsequent time periods. The result is a cyclic time-series of cross-sectional and longitudinal sample survey observations of the population from independent panels of FIA field plots. FIA does not use split or overlapping panels (Kish 1987).

FIA does not currently have an officially endorsed estimator that combines multiple panel data to improve statistical precision, and it has not yet been determined if any single estimator will be fully satisfactory for all regions and forest conditions (Patterson and Reams 2005). FIA is currently investigating the Moving Average, the Temporally Indifferent Method and model-based estimators The latter includes mixed estimators (*e.g.*, Van Deusen 2002), Kalman filters (*e.g.*, Brockwell and Davis 1996), and various time series models (*e.g.*, Johnson and others 2003). The Moving Average and the Temporally Indifferent methods are relatively straightforward, and they closely resemble the estimators used by FIA during the past half-century for periodic inventories (Scott and others 2005, Patterson and Reams 2005).

Patterson and Reams (2005) discuss a simple approach to monitoring change. The net difference between 2 sequential, but different, panels is one simple estimate of change. Since each sequential panel is an independent sample by design, the variance of the difference is the sum of the variances from each of the 2 panels. Over time, annual estimates of the net differences produce a series of annual estimates for change. However, estimates for the components of net change require re-measurement of FIA plots, which occurs after all annual panels are measured for the first time (i.e., after 5 to 10 years). Also, variance of the net difference can be large at the scale of a state or smaller sub-populations, resulting in considerable uncertainty.

The remainder of this section provides more detail on the Moving Average and Temporally Indifferent methods[4]. Later sections look at the Kalman filter as an example of a model-based time-series estimator.

Moving Average Method: Patterson and Reams (2005) describe the moving average method as follows. *"Let P denote the number of panels to be combined for analysis. Let Y_p denote the true quantity for panel P, where p=1,…,P; and let $\hat{Y}_p$ denote the estimate of Y_p obtained using the appropriate technique from Scott and others (2005). Note that each panel is treated as an independent estimate, which permits (1) the weighting of individual panels; and (2) Phase 1 stratification instruments to differ among panels (i.e., different maps may be used to stratify different panels)."* The Moving Average estimator is given by Patterson and Reams (2005) and Roesch and Reams (1999) as:

[4] The Moving Average and Temporally Indifferent methods are implicitly "model-based" if the assumption is that they are estimators of the current conditions at time t=1 in equation 1.

$$\hat{Y}_{MA,P} = \sum_{p=1}^{P} w_p \hat{Y}_p \quad \text{where} \sum_{p=1}^{P} w_p = 1 \qquad [1]$$

The panel measured most recently is denoted p=1, and panels measured at previous P time periods have values p=2,…,P. Patterson and Reams recommend equal weighting of panels, *i.e.*, w_p=1/P for all p. They give the variance estimator for the Moving Average as the corresponding weighted sum of the variances for each panel estimate p:

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

Patterson and Reams remark that the Moving Average dampens annual fluctuations in the estimates, and estimated changes in the population tend to appear smaller than their true values. The Moving Average can cause a "lag bias" when the population is not at a static steady state, and this can obscure trends over time (Roesch and Reams 1999). However, lag bias can be inconsequential unless there is a rapid, widespread catastrophic event (Johnson and others 2003).

Moving Average Residuals: The residual difference r_{MA} between the estimator for the current panel and the moving average for the last P panels might be used to test the hypothesis that the population is at relative steady state (*i.e.*, Y_1=$Y_{MA,P}$) If this hypothesis is rejected, then there is empirical evidence from the FIA sample that the population measured with panel p=1 does not equal the population moving average over the most recent P panels, *i.e.*, the population has changed somehow during the last P panel measurements. The variance estimator for the residual difference equals:

$$\begin{aligned}
\mathrm{Var}\left(\hat{Y}_1 - \hat{Y}_{MA,P}\right) &= \mathrm{Var}\left(\hat{Y}_1\right) + \mathrm{Var}\left(\hat{Y}_{MA,P}\right) + 2\sum_{p=1}^{P} \mathrm{Cov}\left(\hat{Y}_1, \frac{-1}{P}\hat{Y}_p\right) \\
&= \mathrm{Var}\left(\hat{Y}_1\right) + \mathrm{Var}\left(\hat{Y}_{MA,P}\right) + \frac{-2}{P}\mathrm{Var}\left(\hat{Y}_1\right) \quad \begin{cases} \mathrm{Cov}\left(\hat{Y}_1, \hat{Y}_{p|p\neq 1}\right) = 0 \\ \mathrm{Cov}\left(\hat{Y}_1, \hat{Y}_1\right) = \mathrm{Var}\left(\hat{Y}_1\right) \end{cases} \\
&= \left(1 - \frac{2}{P}\right)\mathrm{Var}\left(\hat{Y}_1\right) + \mathrm{Var}\left(\hat{Y}_{MA,P}\right)
\end{aligned} \qquad [3]$$

The difference between the design-based estimator $\hat{Y}_1$ for the single panel at time t=1 and the Moving Average $\hat{Y}_{MA,P}$ over panels p=1,…,P may be standardized relative to the estimated standard deviation of the difference from equation 3:

$$r_{MA} = \frac{\hat{Y}_1 - \hat{Y}_{MA,P}}{\sqrt{\mathrm{Var}\left(\hat{Y}_1 - \hat{Y}_{MA,P}\right)}} \qquad [4]$$

Given the hypothesis $Y_1=Y_{MA,P}$ in equation 4, r_{MA} has an expected value of 0 with an expected standard deviation of 1 (i.e., unit variance). Assuming the residuals have a normal distribution, there is a 32 percent probability that a standardized residual be less than -1 or greater than +1, and a probability of merely 5 percent that it will be less than -1.96 or greater than +1.96.

A simple metric for the relative efficiency of the Moving Average is its variance relative to the variance for the most current panel (p=1):

$$\beta = \frac{\mathrm{Var}\left(\hat{Y}_{MA,P}\right)}{\mathrm{Var}\left(\hat{Y}_1\right)} \qquad [5]$$

β will generally range between 0 and 1 because the estimated variance of the Moving Average is usually, but not necessarily, less than the estimated variance for any one panel being averaged. The standardized residual r_{MA} in equation 4, and the gain in statistical efficiency β in equation 5, subsequently will be used to directly compare the Moving Average with the Kalman filter.

Temporally Indifferent Method: Patterson and Reams (2005) describe the Temporally Indifferent Method, which is very similar to the Moving Average. Sampling units in all *panels* (p=1,…,P) are pooled as though they were all part of a single large periodic inventory. As with previous FIA periodic inventories, Phase 1 post-stratification is applied across sampling units from all P panels. If the number of sampling units is virtually the same in each panel, and a single source of remotely sensed data are used for post-stratification within each panel, then the Temporally Indifferent Method is algebraically equivalent to the (Moving Average).

Summary: The Moving Average and Temporally Indifferent estimators are simple, and simplicity can be good. The Temporally Indifferent estimator can be applied with the same familiar approach previously used for decades with periodic FIA surveys, and familiarity can be good. The Moving Average and Temporally Indifferent estimators produce useful inventory estimates for the current state of the nation's forests, which are updated annually as new panel data are acquired. However, these simple and familiar methods have inherent limitations in the context of estimation and interpretation of changes over time, which is necessary to address the substantive monitoring questions identified in the FIA strategic plan (U.S. Forest Service 2007a).

An Alternative Estimator for Monitoring

In addition to the Moving Average and Temporally Indifferent methods, Patterson and Reams (2005) list model-based time-series methods as alternative estimation methods that can combine FIA panel estimates over time. Van Deusen (1999) reviews sampling with partial replacement methods, although those have concentrated on periodic surveys. One promising approach for panel data is the mixed estimator (Theil 1971), which has been studied by Van Deusen (1999, 2002, 2008) and Roesch (1999, 2007, 2008) for applications to annual forest inventory and monitoring. In addition, Patterson and Reams (2005) cite Czaplewski (1995) and Brockwell and Davis (1996), who discuss the Kalman filter (see Maybeck 1979) as another alternative. As discussed below, the Kalman filter combines the advantages of model-based and design-based estimators to mitigate risk while reliably improving statistical efficiency.

The remainder of this paper concentrates on the Kalman filter, which has been widely used in engineering applications for 50 years to estimate the states of a system over time. However, the Kalman filter is not well-known in forest inventory and monitoring applications[5]. The Kalman filter is conceptually intuitive, which, hopefully, will soon become apparent. With this goal in mind, the following exposition begins with the univariate Kalman filter, which uses as its basic building block a well known and simple statistical method, namely, the composite estimator.

Composite Estimator

The Kalman filter may be viewed as the sequential application of the composite estimator (Gregoire and Walters 1988). Maybeck (1979) is an often cited and well written source on the Kalman filter. He begins his 3-volume seminal treatise on the Kalman filter with a simple example of the composite estimator, which may also be considered an example of the static Kalman filter. The following is a slightly modified reproduction of Maybeck's introduction from his section 1.5 in Volume 1:

"Suppose that you are lost at sea during the night and have no idea at all of your location. So you take a star sighting to establish your position (for the sake of simplicity, consider a 1-dimensional location). At some time t_1 you determine your location to be z_1. However, because of inherent measuring device inaccuracies, human error, and the like, the result of your measurement is somewhat uncertain. Say you decide that the precision is such that the standard deviation (1-sigma value) involved is σ_{z_1} (or equivalently, the variance or second

[5] For exceptions, see Dixon and Howitt 1979; Gregoire and Walters 1988; Walters and others 1991; Visser and Molenaar 1992; Czaplewski 1995; Van Deusen 1987, 1989; Devall and others 1991; Brakel and Visser 1996; Gove and Houston 1996; Williams and others 2005; Hurteau and others 2007; Gao and others 2008.

order statistic, is $\sigma_{z_1}^2$ *). Thus, you can establish the conditional probability of* $x(t_1)$*, your position at time* t_1 *conditioned on the observed value of the measurement being* z_1*, as depicted in* (figure 1)*. This is a plot of* $f_{x(t_1)|z(t_1)}(x|z_1)$ *as a function of the location x: it tells you the probability of being in any 1 location, based upon the measurement you took. Note that* σ_{z_1} *is a direct measure of the uncertainty: the larger* σ_{z_1} *is, the broader the probability peak is, spreading the probability "weight" over a larger range of x values. For a Gaussian density, 68.3% of the probability "weight" is contained within the band σ units to each side of the mean, the shaded portion in* (figure 1).

"Based on this conditional probability density, the best estimate of your position is

$$\hat{x}(t_1) = z_1 \tag{6}$$

and the variance of the error in the estimate is

$$\sigma_x^2(t_1) = \sigma_{z_1}^2 \tag{7}$$

"Note that $\hat{x}$ *is both the mode (peak) and the median (value with 1/2 of the probability weight to each side), as well as the mean (center of mass).*

"Now say a trained navigator friend takes an independent fix, right after you do, at time $t_2 \sim t_1$ *(so that the true position has not changed at all, and obtains a measurement* z_2 *with a variance* $\sigma_{z_2}^2$*. Because he has a higher skill, assume the variance in his measurement to be somewhat smaller than yours.* Figure 2 *presents the conditional density of your position at time* t_2*, based only on the measured value* z_2*. Note the narrower peak due to smaller variance, indicating that you are rather more certain of your position based on his measurement.*

"At this point, you have 2 measurements available for estimating your position. The question is, how do you combine these data? It will be shown subsequently that, based on the assumptions made, the conditional density of your position at time $t_2 \sim t_1$*,* $x(t_2)$*, given* both z_1 and z_2*, is a Gaussian density with mean* μ *and variance* σ^2 *as indicated in* figure 3*, with*

$$\mu = \left(\frac{\sigma_{z_2}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2}\right) z_1 + \left(\frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2}\right) z_2 \tag{8}$$

$$\frac{1}{\sigma_\mu^2} = \frac{1}{\sigma_{z_1}^2} + \frac{1}{\sigma_{z_2}^2} \tag{9}$$

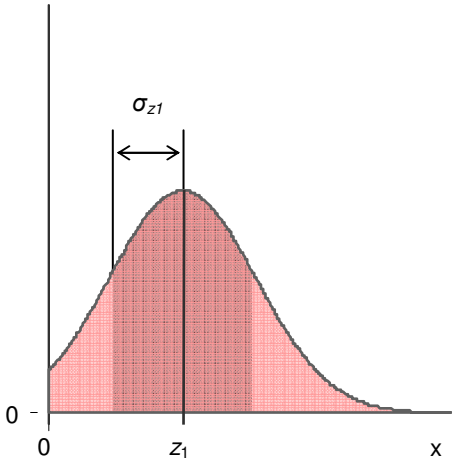


Figure 1: Conditional density of position based on measured value z_1 (facsimile of figure 1.4 in Maybeck 1979). The distribution extends to both ±∞, but it is truncated in this illustration.

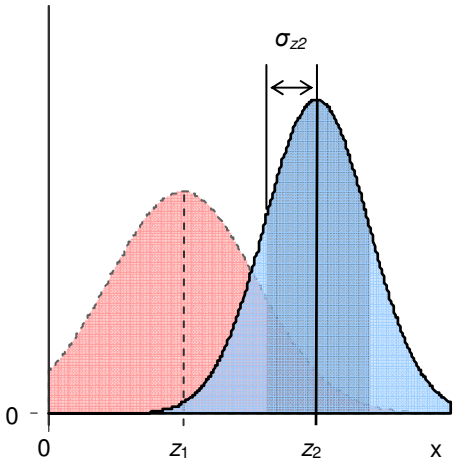


Figure 2: Conditional density of position based on measurement z_2 alone (facsimile of figure 1.5 in Maybeck 1979).

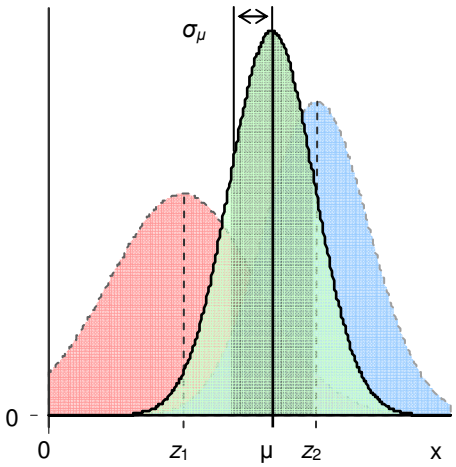


Figure 3: Conditional density of position based on data z_1 and z_2 (facsimile of figure 1.6 in Maybeck 1979).

*"Note that, from (*equation 9*), σ_μ is less than either σ_{z1} or σ_{z2}, which is to say that the uncertainty in your estimate of position has been decreased by combining the 2 pieces of information.*

"Given this density, the best estimate is

$$\hat{x}(t_2) = \mu \tag{10}$$

with an associated error variance σ^2. It is the mode and the mean (or, since it is the mean of a conditional density, it is also termed the conditional mean). Furthermore, it is also the maximum likelihood estimate, the weighted least squares estimate, and the linear estimate whose variance is less than that of any other linear unbiased estimate.[6] *In other words, it is the "best" you can do according to just about any reasonable criterion.*

"After some study, the form of μ given in (equation 8) *makes good sense. If σ_{z1} were equal to σ_{z2}, which is to say you think the measurements are of equal precision, the equation says the optimal estimate of position is simply the average of the 2 measurements, as would be expected* [7]*. On the other hand, if σ_{z1} were larger than σ_{z2}, which is to say that the uncertainty involved in the measurement z_1 is greater than that of z_2, then the equation dictates "weighting" z_2 more heavily than z_1. Finally, the variance of the estimate is less than σ_{z1} even if σ_{z2} is very large: even poor quality data provide some information, and should thus increase the precision of the filter output.*

"The equation for $\hat{x}(t_2)$ can be rewritten as

$$\begin{aligned}\hat{x}(t_2) &= \left(\frac{\sigma_{z_2}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2}\right) z_1 + \left(\frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2}\right) z_2 \\ &= z_1 + [z_2 - z_1]\left(\frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2}\right)\end{aligned} \tag{11}$$

or, in final form that is actually used in Kalman filter implementations [noting that $\hat{x}(t_1) = z_1$]

$$\begin{aligned}\hat{x}(t_2) &= \hat{x}(t_1)(1-\beta) + z_2\beta \\ &= \hat{x}(t_1) + [z_2 - \hat{x}(t_1)]\beta\end{aligned} \tag{12}$$

[6] The Kalman filter may also be derived as an Empirical Bayes estimator (Jazwinski 1970, Meinhold and Singpurwalla 1983, Cressie and Wikle 2002).

[7] This is the same assumption made in the FIA Moving Average (Patterson and Reams 2005), namely, each estimate from the last *P* panels is equally accurate in estimating the current condition of the population (or the condition at the time panel *p*=*P*/2 was sampled).

where[8]

$$\beta = \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2}$$

$$= \frac{\sigma_x^2(t_1)}{\sigma_x^2(t_1) + \sigma_{z_2}^2} \qquad [13]$$

"These equations say that the optimal estimate at time t_2, $\hat{x}(t_2)$, is equal to the best prediction of its value before z_2 is taken, $\hat{x}(t_1)$, plus a correction term of an optimal weighting value times the difference between z_2 and the best prediction of its value before it is actually taken, $\hat{x}(t_1)$. It is worthwhile to understand this "predictor-corrector" structure of the filter. Based on all previous information, a prediction of the value that the desired variables and measurement will have at the next measurement time is made. Then, when the next measurement is taken, the difference between it and its predicted value is used to `correct' the prediction of the desired variables.

"Using the β … *in* (equation 13)*, the variance equation given by* (equation 9) *can be rewritten as*

$$\sigma_x^2(t_2) = \sigma_x^2(t_1) - \sigma_x^2(t_1)\beta \qquad [14]$$

"Note that the values of $\hat{x}(t_2)$ and $\sigma_x^2(t_2)$ embody all of the information in $f_{x(t_2)|z(t_1),z(t_2)}(x|z_1, z_2)$. Stated differently, by propagating these 2 variables, the conditional density of your position at time t_2, given z_1 and z_2, is completely specified.

"Thus we have solved the static estimation problem."

Hopefully, Maybeck's example provides intuitive insight into the simplicity, efficiency, and flexibility of the composite estimator, which is a special case of the Kalman filter.

[8] The symbol "*β*" is used in place of Maybeck's "*K*" to draw the analogy to regression estimators in sample surveys (e.g., Sarndäl and others 1992).

Model-Based Bias with the Composite Estimator

If estimators z_1 and z_2 for population parameter z are both unbiased by design, then the composite estimator, which is a weighted sum of z_1 and z_2, will be design-unbiased. However, the Kalman filter generally assumes that z_1 is based on an estimate from the previous time period, which is "updated" with a prediction model into an estimate for the state of the population at the current time period, namely $z_1=x(t_1)$. This model-based component has the potential to substantially improve precision of the composite estimator, but accuracy can be poor if the model produces biased predictions. In most applications, the model requires assumptions that are difficult to test. Therefore, special attention is required to detect and correct failures in model assumptions, which is the topic of this section. Otherwise, the composite estimator can produce estimates that are apparently very precise, but are, in fact, very inaccurate (*i.e.,* biased).

Successful applications of the Kalman filter require close monitoring of the residual difference between $x(t_1)$ and z_2 to detect likely bias in model predictions (Maybeck 1979). Denote this residual as

$$r(t_2) = \hat{z}_2 - \hat{x}(t_1) \qquad [15]$$

Assuming independence, the expected variance of this residual is:

$$\sigma_r^2(t_2) = \sigma_{z_2}^2 + \sigma_x^2(t_1) \qquad [16]$$

A useful monitoring technique requires standardization of this residual based on its expected variance. If the model is an unbiased estimator, then the standardized residual in equation 17 is expected to have a distribution with mean zero and unit variance:

$$\frac{r(t_2)}{\sqrt{\sigma_r^2(t_2)}} \sim (0,1) \quad \Big| \quad \begin{array}{l} \text{given unbiased} \\ \text{model - based estimator} \\ x(t_1) \text{ and } \sigma_x(t_1) \end{array} \qquad [17]$$

At each time step, assume there is a known model-based prediction of $x(t_1)$ with a known model-based estimate of its variance. Further assume there is a known design-based panel estimate z_2 with a known design-based estimate of its variance. Under these assumptions, the residual may be computed with equation 15, the predicted variance of this residual may be computed with equation 16, and the resulting residual may be standardized as in equation 17.

One process to monitor the reliability of the model-based estimator is with the "size" of this standardized residual. Assuming the standardized residual is

normally (Gaussian) distributed, with mean zero and unit variance (*i.e.*, variance equal to 1 in equation 17), then there is a 68.3 percent probability that the standardized residual will be between -1 and +1, and a 95.4 percent probability that it will be between −2 and +2.

Figure 4A provides an example, in which the standardized residual equals 1, and the standardized residual is assumed normally distributed with mean zero and unit variance. When monitoring residuals under these assumptions, a standardized residual as large as 1 is not too surprising because 31.7 percent (*i.e.*, 1.000-0.683= 0.317) of all residuals are expected to be less than −1 or greater than +1.

As an aside, this example also illustrates that the distribution of random errors in the design- and model-based estimators need not be assumed to be Gaussian normal. Figure 4 assumes a lognormal distribution, in which feasible variables may not have negative values. However, the residual difference between 2 estimates may have a negative value if the model-based estimate exceeds the design-based panel estimate. Therefore, the distribution of standardized residuals may be assumed Gaussian normal, although that is nothing more than another untested assumption.

Figure 4B is another example, in which the model-based estimate is greater than that in figure 4A, but it otherwise shares the same parameter values as those in figure 4A. In this example, the standardized residual equals 2. Under the assumptions in this example, the probability of a standardized residual less than −2 or greater than +2 is only 4.6 percent. This is a relatively low probability, but observation of a residual of this magnitude remains plausible under the assumptions of the residual analysis.

Figures 4C and 4D are more extreme examples. Again, under the assumptions in this analysis of residuals, the probability of the absolute value of the standardized residual exceeding the value of 3 is 0.3 percent in figure 4C, and of exceeding the value 4 in figure 4D is 0.006 percent. Both are evidence that 1 or more assumptions in the residual analysis are suspect. Since the panel estimate z_2 is design-unbiased, the assumptions related to this estimator remain credible, at least in the absence of substantial non-sampling errors. It is more plausible that 1 or more assumptions related to the model-based estimator $x(t_1)$ are incorrect.

The model-based estimate might be biased. However, if the magnitude of the bias is quantitatively predictable, then the model should include a correction term that corrects for the bias. In other words, a model with a bias that could be estimated should not be knowingly used in the Kalman filter. Rather, the bias correction should be incorporated into the base model.

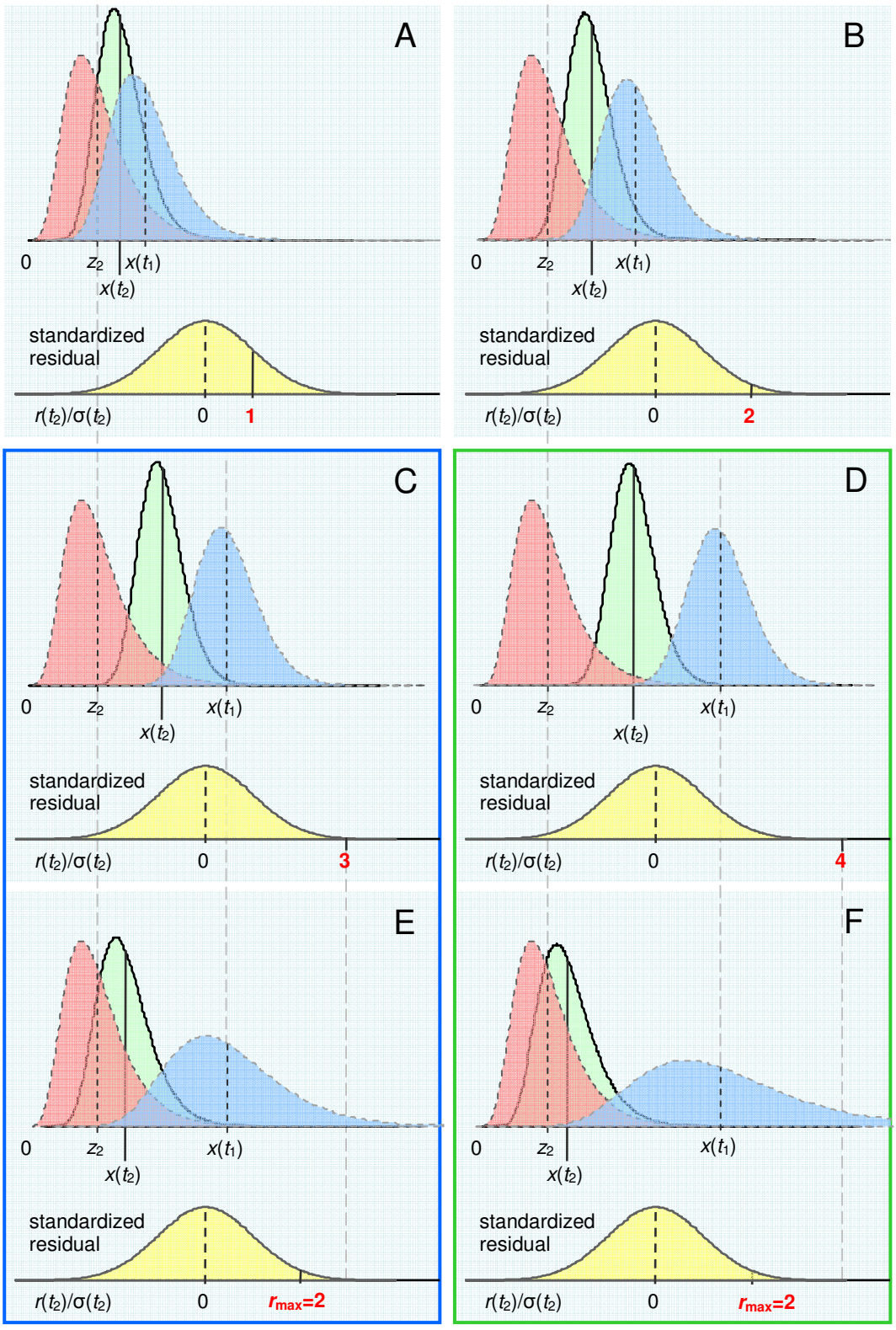


Figure 4: Reliable applications of the composite estimator with a model-based component, which is a special case of the static Kalman filter, requires monitoring residuals and adapting to evidence that the model assumptions are significantly flawed.

The other suspect is the estimated variance of random prediction errors with the model, namely $\sigma_x^2(t_1)$. This variance can be difficult to accurately estimate from *a priori* empirical data when the number of residuals is small or the variance is heteroscedastic over time. One solution would be to assume the model-based variance estimator is biased, and re-estimate its variance so that the realized magnitude of the standardized residual is more plausible given the design-based estimate from the most current FIA panel.

One *ad hoc* adjustment rule could be to increase the estimated variance of the model-based estimator by a scalar factor of c so that the absolute value of its standardized residual becomes a more plausible value, say $r_{\max}$. This assures that the standardized residuals never exceed $\pm r_{\max}$ standard deviation units. For example, if $r_{\max}=2$, then the residual difference between the model-based and design-based estimates would be forced to remain within 2 standard deviation units of 0.

From equations 16 and 17 and this rule, it will be assumed that a less biased estimator for the variance of model prediction errors is $c\sigma_x^2(t_1)$, which is computed as:

$$\frac{|r(t_2)|}{\sqrt{\sigma_{z_2}^2 + c\sigma_x^2(t_1)}} \le |r_{\max}|$$

$$c\sigma_x^2(t_1) = \left(\frac{r(t_2)}{r_{\max}}\right)^2 - \sigma_{z_2}^2 \qquad [18]$$

Figure 4F is an example in which this *ad hoc* solution is applied to the statistics from figure 4D. In figure 4D, the standardized residual has the value of 4 standard deviation units. The variance of the estimator for $x(t_1)$ is re-estimated with equation 18, where the maximum plausible value of the standardized residual is chosen to be $r_{\max}=2$, namely, 2 standard deviation units. This increases the estimated variance of the model prediction error, resulting in a new standardized residual exactly equal to $r_{\max}=2$, which, in turn, increases the value of β in equation 13, which decreases the relative "weight" placed on the model prediction in equation 12. The outcome is a composite estimate that more closely agrees with the design-based estimator of z_2. The new composite estimate in figure 4F has a larger variance estimate than the composite estimate in figure 4D, but the estimate in figure 4F will be closer to the true value if the new assumptions in the model become more accurate. Figure 4E, which has a standardized residual equal to 3 standard deviation units, is a less extreme example relative to figure 4F.

In summary, the model-based composite estimator can be very efficient, but it can also be dangerously biased if the model is inaccurate. Since model parameters are typically fit with historical data, substantial deviations of a population from its

historical conditions will likely cause biased model predictions. Statistical efficiency has an associated risk. There is nothing inherent within the composite estimator, and therefore, within the Kalman filter, that protects against model-based bias[9]. Monitoring residuals and adapting to anomalous outcomes can mitigate the risk of serious bias without foregoing the opportunities for increased efficiencies. Kott (2005) advocates a similar view in his paper on randomization-assisted model-based survey sampling

Kalman Filter Estimator

The Kalman filter can be viewed as the sequential application of the composite estimator for each encounter of new information. In monitoring changes and trends over time with sub-sampled FIA panel data, the first component $z(t)$ is the design-based estimator for the most current FIA panel (Scott and others 2005). The second component $x(t|t-1)$ is a model-based estimator that uses the best FIA estimate from the previous year ($t-1$), which is updated to year t with a model for predicted change between years t and $t-1$. The Kalman filter is simply the sequential application, 1 year at a time (t=1,2,…), of this composite estimator and its underlying model.

Recall from Maybeck's introduction (page 6) that the composite estimator is the simple weighted sum of $x(t|t-1)$ and $z(t)$ (see also Sarndäl and others 1992, section 9.9.1). Using primarily Maybeck's notation in the context of the univariate time-series (equation 12), the composite estimator $x(t|t)$ at time t is the weighted sum of the FIA panel estimator $z(t)$ at time t and the model-based predictor $x(t|t-1)$ of the same population parameter at time t given the composite estimator at time $t-1$:

$$\left.\begin{aligned}\hat{x}(t|t) &= (1-\beta_t)\hat{x}(t|t-1)+\beta_t\hat{z}(t)\\ &= \hat{x}(t|t-1)+\left[\hat{z}(t)-\hat{x}(t|t-1)\right]\beta_t\end{aligned}\right\} \quad \text{where } 0\le\beta_t\le 1 \qquad [19]$$

The "optimal" weight ($1-\beta_t$) placed on the model estimate (equations 12 and 13) at time t is:

$$(1-\beta_t)=\frac{\sigma_z^2(t)}{\sigma_x^2(t|t-1)+\sigma_z^2(t)}$$

$$\beta_t=\frac{\sigma_x^2(t|t-1)}{\sigma_x^2(t|t-1)+\sigma_z^2(t)} \qquad [20]$$

The variance estimator for equation 19, which is the time-series version of equations 9 and 13, is algebraically equivalent to the following expressions:

[9] There is nothing to protect against similar bias with the Moving Average or Temporally Indifferent methods (Roesch, personal communication).

$$\begin{aligned}\sigma_x^2(t|t) &= (1-\beta_t)^2\sigma_x^2(t|t-1) + \beta_t^2\sigma_z^2(t)\\ &= \frac{\sigma_z^2(t)\sigma_x^2(t|t-1)}{\sigma_x^2(t|t-1)+\sigma_z^2(t)}\\ &= \sigma_x^2(t|t-1)-\beta_t\sigma_x^2(t|t-1)\\ &= \beta_t\sigma_z^2(t)\end{aligned} \quad [21]$$

The weight β_t in equation 21 is algebraically equivalent to the scalar reduction in the variance of the design-based panel estimator at time t given the model-based estimate at time t. This is analogous to the "design effect" used to evaluate efficiency of different sampling designs (Maybeck 1979).

The time-series model in the Kalman filter assumes the population total $x(t)$ at time t equals the total $x(t-1)$ at $t-1$, which is multiplied by the scalar rate of change (φ_t) between t and $t-1$. This is the difference equation:

$$x(t) = \varphi_t\ x(t-1) + \varepsilon_w(t) \quad [22]$$

where $\varepsilon_w(t)$ in equation 22 represents the unknown random prediction error from time $t-1$ to t. The model assumes $\varepsilon_w(t)$ is distributed with mean zero and variance $\sigma_w(t)$. The corresponding model-based predictor for use in the Kalman filter is:

$$\hat{x}(t|t-1) = \varphi_t\ \hat{x}(t-1|t-1) \quad [23]$$

$$\sigma_x^2(t|t-1) = \varphi_t^2\ \sigma_x^2(t-1|t-1) + \sigma_w^2(t) \quad [24]$$

In successful applications of the Kalman filter, estimates for the standard deviation of model-based predictions $\sigma_w(t)$ are carefully monitored through the realized residuals and adapted if necessary (see above).

Sequential Recursive Time-Series Estimation with the Kalman Filter

The following example is intended to explain in a more intuitive fashion an application of the Kalman filter for FIA monitoring. Assume a temporal model predicts the population parameter x is exponentially increasing over the time period being analyzed. More specifically, assume, $\varphi_t = 1.5$ in equations 23 and 24. Assume the random prediction error over a single year (equation 22) has a heteroscedastic standard deviation equal to 0.25 times the population parameter x at time t. In prose, the model assumes that the population parameter increases 50

percent each year relative to its value at the previous year, and the random prediction error of this model has a standard deviation of ±25 percent at time t relative to its condition at $t-1$ for all t. With these specific assumptions, the model-based estimator at time t is specified from equations 23 and 24 as:

$$\hat{x}(t|t-1) = \left(\frac{3}{2}\right) \times \hat{x}(t-1|t-1) \quad [25]$$

$$\sigma_x^2(t|t-1) = \left[\left(\frac{3}{2}\right)^2 \times \sigma_x^2(t-1|t-1)\right] + \left[\left(\frac{1}{4}\right) \times x(t-1|t-1)\right]^2 \quad [26]$$

Assume an FIA panel (*i.e.*, sub-sample) is measured at time t=1. The design-based estimate of $x(1|1)$ is produced with standard FIA methods as given by Scott and others (2005). Given this model, figure 5 illustrates how these initial conditions are predicted to change over time in the absence of any further FIA data at times t=2,…,6. The model in equation 25 forecasts that the population parameter will increase exponentially, and the model in equation 26 predicts the uncertainty of this forecast increases nonlinearly over time. The random error distributions are assumed to the skewed lognormal, similar to figure 4, in which negative values are infeasible. Since the Kalman filter is a minimum variance estimator, it does not necessarily depend on the assumption of Gaussian normal or symmetric error distributions.

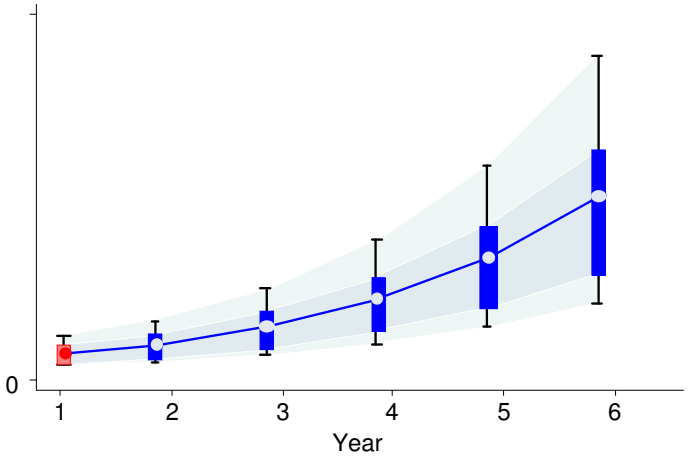


Figure 5: Predictions of the population parameter and the associated uncertainty (standard deviation) from the model in equations 25 and 26 in the absence of new FIA panel estimates at times t=2,…,6.. The initial condition at time t=1, which is portrayed by the red box plot, is the standard FIA design-based estimate from a single panel (Scott and others 2005). The subsequent model-based estimates are illustrated by the blue box plots. The model assumes the random sampling and prediction errors have a lognormal distribution to assure that negative values are infeasible.. The range of the boxes is the 25th and 75th percentiles, and the range of the "whiskers" is the 10th and 90th percentiles.

This model is the component of the Kalman filter that links the time series of observations over time. Using the simple composite estimator, the Kalman filter combines the observation at time *t* with the model prediction at time *t* to produce a more accurate estimate at time *t*.

The Kalman filter produces a composite estimate at a single time *t* based on any direct observations of the population at time *t* (e.g., a design-based estimate from a single FIA panel) plus the independent model forecast of the same population parameter at time *t* (e.g., equations 25 and 26, figure 5). This forecast uses the best estimate of the population parameter at time $t-1$ as initial conditions. The resulting composite estimate embodies the information available in all observations and model predictions up to and including time *t*. The next operation in the Kalman filter sequence uses this best estimate at time *t* to forecast the state of the population at the subsequent time step $t+1$, and the composite estimator combines this forecast with any new independent observations at time $t+1$ into a more accurate estimate at time $t+1$. And so forth. Figure 6 provides an example.

The red box plots represent the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of the assumed lognormal distribution of sampling errors in the estimate of the population attribute. This means that 10 percent of the assumed distribution exceeds the upper whisker on the box plot, and another 10 percent is less than the lower whisker. Thus, there is only about 80 percent of the distribution covered by the range of each box plot.

Simulated Examples of Monitoring Trends over Time

This section provides a variety of examples that compare Kalman filter and Moving Average estimates. The "true" population value at each time *t* is known exactly because they were used to generate the simulated estimates from each panel[10].

The following examples provide a range of departures of the "real world" from the assumed model in the Kalman filter, which reveals examples of the consequences of those departures on the reliability of annual estimates. However, these examples are merely specific realizations. They do not provide valid generalizations about bias or efficiency of the estimators. Such generalizations require deriving their expected values over all possible samples, or at least a very large number of potential samples.

[10] While the true values in these hypothetical populations may be known during these simulations, the true values are not known in FIA sampling and estimation of forest populations.

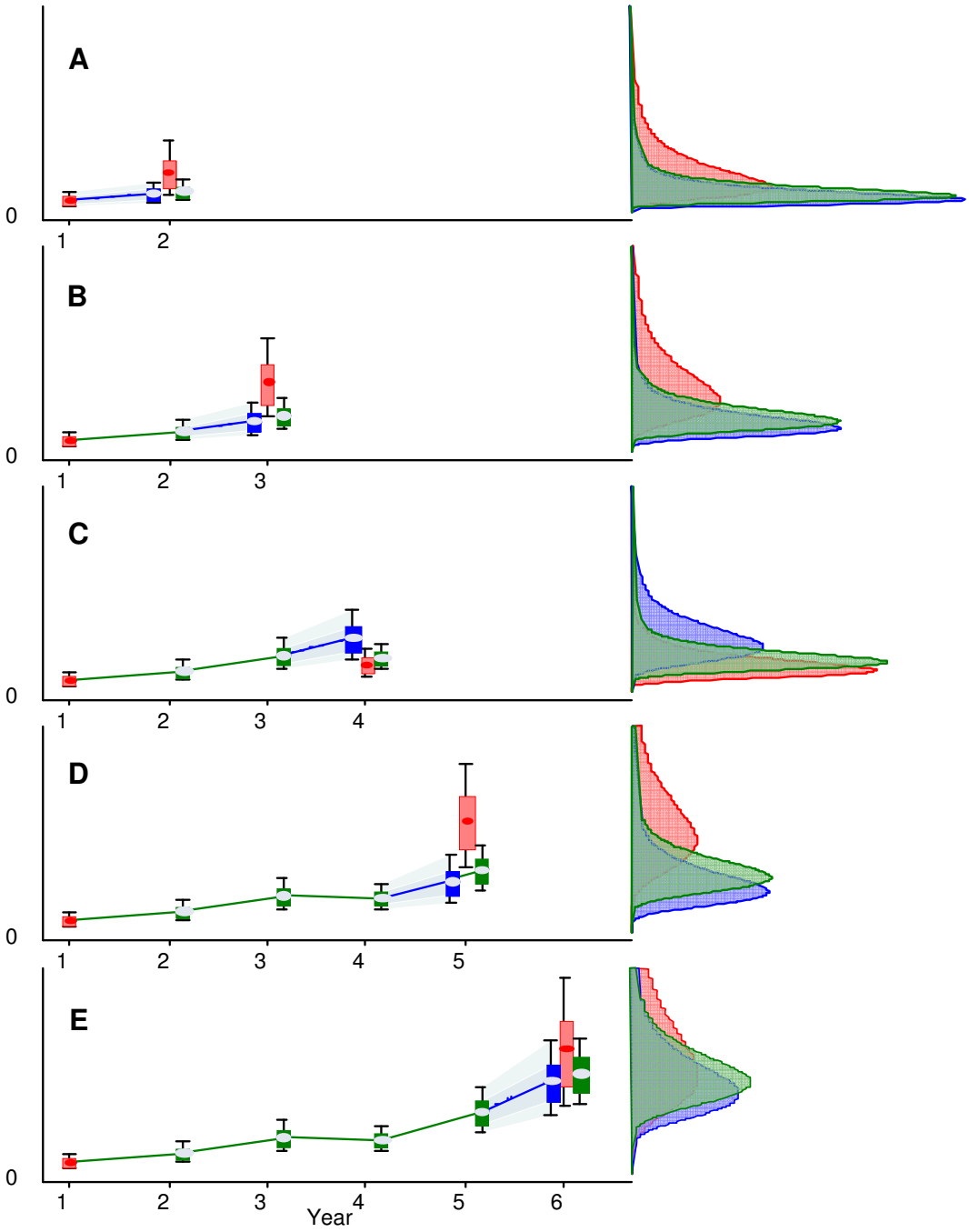


Figure 6: An example of the sequential recursive nature of the Kalman filter. Start with figure 6A. The design-based panel estimate at time *t*=2 is denoted with the red box plot. The model-based estimate (denoted with the blue box plot) is an independent forecast for time *t*=2 based on the best estimate at time *t*=1 (see equations 25 and 26, figure 5). The Kalman filter is the composite estimate (denoted with the green box plot) of these design- and model-based estimates. Figure 6B illustrates the next sequential step at time *t*=3, which uses the same process in the Kalman filter as that at time *t*=2. The design-based estimate at time *t*=1 is no longer needed at time *t*=3 because its contribution to the *t*=3 estimate is completely captured in the Kalman filter estimate at time *t*=2. Figures 6C, 6D and 6E illustrate subsequent steps in this sequential process. Conditional lognormal error distributions on the right-hand side supplement the box plots use the same color convention as in figure 5, where each box plot portrays the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of the modeled distributions.

For sake of simplicity and generality, the model chosen for this section makes few prior assumptions about the simulated dynamics of the populations. Assume a temporal model in the Kalman filter that predicts that the population parameter x is at a constant (static) steady-state over time, i.e., $\varphi_t = 1$ in equations 23 and 24. Assume the random prediction error over a single year (equation 22) has a relatively small but heteroscedastic standard deviation equal to 0.05 times the population parameter x at time t. In prose, the model assumes that the population is nearly at a constant steady-state that has minor random variations each year around a long-term constant value. The model for the standard deviation of this natural random variation is ±5 percent at time t relative to its condition at $t-1$ for all t. Under these assumptions, the model-based estimator at time t, which is a specific case of the general models in equations 23 and 24, is defined as:

$$\hat{x}(t|t-1) = \hat{x}(t-1|t-1) \quad [27]$$

$$\sigma_x^2(t|t-1) = \sigma_x^2(t-1|t-1) + \left[0.05 \times x(t-1|t-1)\right]^2 \quad [28]$$

This particular model in the Kalman filter (equations 27 and 28) approximately matches the model implied by the Moving Average, namely, estimates with the Moving Average are reliable for the current year if there is negligible change in the population over the time period being averaged.

The first example employs a simulated population that is at a true steady-state, in which there is no change over time. A single realization of a simulated sample from this population is given in figure 7A, in which the design-based sampling errors (red box plots) have a coefficient of variation of 15 percent. Figure 7B illustrates a second independent realization, in which the design-based sampling errors have a coefficient of variation of 30 percent. The Kalman filter with the static steady-state model in equations 27 and 28 produces nearly identical estimates $x(t)$ as the Moving Average. Both estimators reduce the standard deviation of their estimates (σ) by approximately 50 percent (i.e., 25 percent decrease in variance σ^2) relative to the corresponding estimate from a single panel. The model in the Kalman filter ultimately reduces the estimation variance by 75 percent ($1-\beta$=1.00–0.25) in figure 7. The time-series of standardized residuals from the 2 estimators are virtually identical, with a distribution that is seemingly consistent with its expected mean of zero, unit variance, and a random temporal pattern. The standardized residuals vary between positive and negative, and they reveal no obvious non-random temporal trends. These results come as no surprise because the model in equations 27 and 28 agrees well with the true trend for this hypothetical population.

In figure 7, as in following figures, the distribution of simulated random errors from the design-based estimator for each FIA panel at time t is denoted by the red box plots and trend lines. Trend lines for the Moving Average estimator are

identified in cyan, and those for the Kalman filter in green. Unlike figure 6, the box plots for error distributions from these 2 estimators are omitted to reduce clutter. Instead, their standard deviations are graphed separately. Likewise, the distributions of model prediction errors over each time step, which are portrayed with the blue box plots in figure 6, are also omitted to reduce clutter. "Saw tooth" patterns in estimated standard deviations for the Kalman filter are a consequence of model prediction errors (equation 24) that are propagated from time $t-1$ to t, immediately followed by the composite estimate (equation 21) at time t.

Relative efficiency shown in figure 7 is defined as the estimated variance of the Kalman filter estimate divided by the computed variance of the design-based estimate from the single FIA panel measured at time t. Relative variance is used as shorthand for the $1-\beta$ term in equations 19 to 21 for the Kalman filter. β represents a value between 0 and 1 that weights the design-based estimate, while the weight placed on the model-based estimate is $1-\beta$. Therefore, as $1-\beta$ nears 1, more weight is placed on the model estimate, which increases statistical efficiency attributable to the model. While there is no analogous interpretation for $1-\beta$ with the Moving Average (equation 5), it is graphed along with the $1-\beta$ for the Kalman filter as a basis for comparison. For both estimators, efficiency increases proportionally as $1-\beta$ approaches 1. The dashed line at 0.9 represents the relative efficiency expected if the sample size in the single annual panel at time t were 10–times larger, as would be the case if a periodic survey were conducted each and every year at time t.

Finally, the bottom panels in figure 7 graph the standardized residuals for the Kalman filter (equations 17 and 21). By definition, standardized residuals are expected to have a zero mean with unit variance and no temporal patterns if the model is accurate[11]. While there is no basis for this same expectation with residuals from the Moving Average (equation 4), they too are graphed along with the Kalman filter residuals for comparison.

[11] An inaccurate model can also produce residuals with the same expected distribution. Therefore, the distribution of residuals, by itself, is not a sufficient basis for a reliable test of model accuracy.

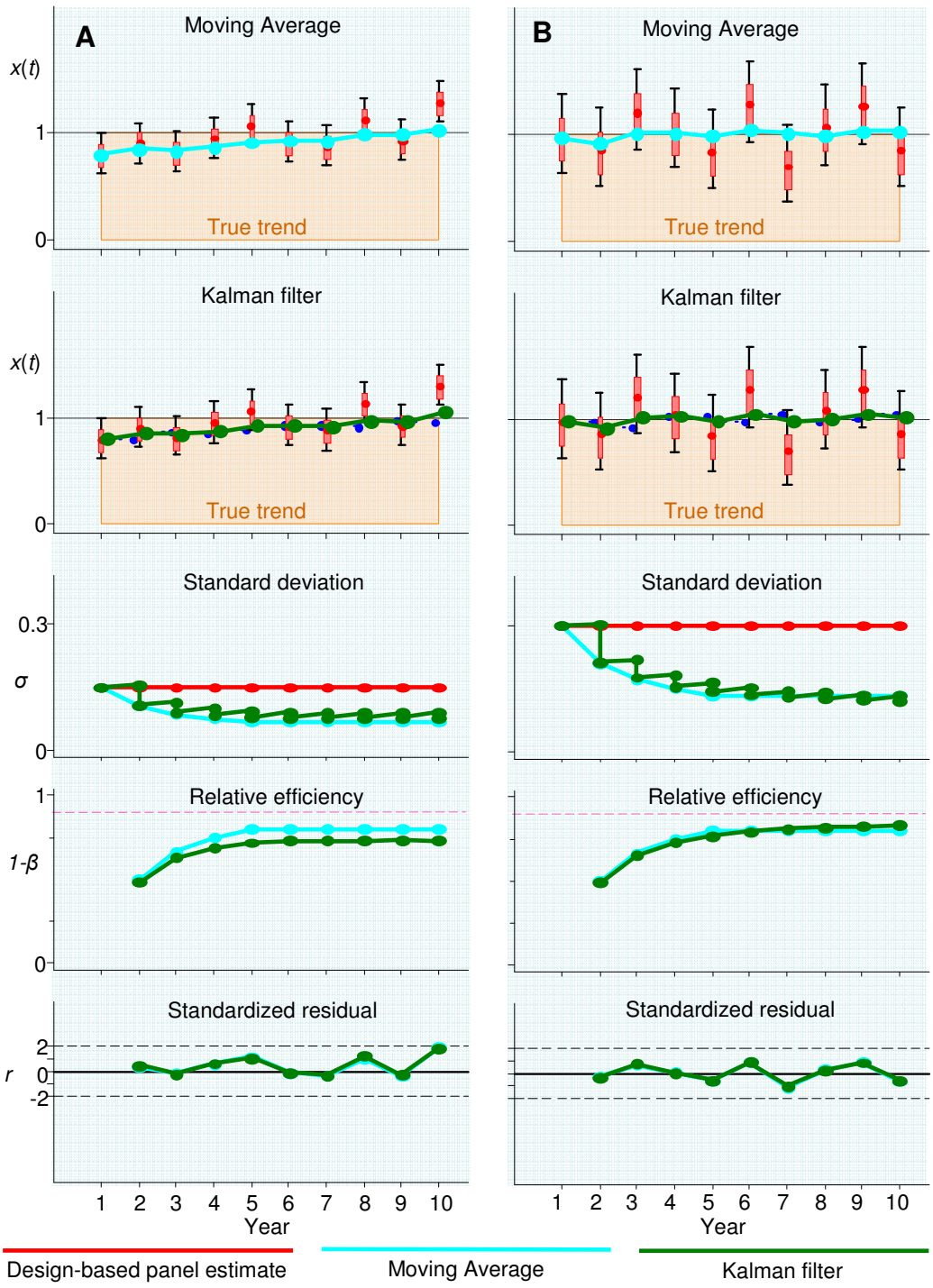


Figure 7: Comparison of Moving Average and Kalman filter estimates for a simulated population that does not change over time. The standard deviation of the estimation error is 15 percent of the population value in figure 7A (left-hand side), and 30 percent in figure 7B (right-hand side). Given these 3 realizations in the time-series of simulated sample estimates, the 2 estimators produce nearly identical results, at least in this case study of a static hypothetical population. The standardized residuals agree well with their expected values of zero mean and unit standard deviation, with no obvious trends over time in positive or negative residuals.

A second hypothetical example is given in figure 8. Unlike figure 7, this population is not static. Instead, there is an abrupt, perhaps catastrophic, 50 percent decline in the population parameter at year t=6. In figure 7A, the 5-year Moving Average exhibits "lag bias" during years $6 \leq t \leq 9$, but it recovers as an unbiased estimator when the 5-year period rolls beyond the abrupt change. However, the estimated standard deviations from the Moving Average remain deceptively low, which produces seemingly very precise estimates during years $6 \leq t \leq 9$, but, in reality, these same estimates are very inaccurate. On the other hand, the Kalman filter produces more accurate estimates during this anomalous time period, even though this particular implementation of the Kalman filter uses the static steady-state model. The Kalman filter estimate departs rapidly from the pre-change *status quo* at year t=6 because the Kalman filter combines the model-based estimate at year t=6 with the relatively precise estimate from the design-unbiased panel estimate at year t=6. This response is reflected by the decrease in "relative efficiency" at year t=6 as less weight is placed on the model-based estimate. However, the standardized residuals for both estimators are suspiciously extreme at year t=6, with eye-popping values of approximately −5 standard deviation units.

Figure 8B shows another estimation realization for the same simulation population. The difference being that the standardized residuals from the Kalman filter are arbitrarily (but consistently) constrained to values within ±2 standard deviations units (r_{max}=2 in equation 18). These Kalman filter estimates are remarkably accurate, especially considering that the static steady-state model is used in this particular implementation of the Kalman filter. This response to major deviations from the static model is associated with lesser weight being placed on the model-based estimates (*i.e.*, reductions in relative efficiency from the model-based estimator) and realistically larger values for the standard deviations. On the other hand, the Moving Average estimates are very inaccurate (*i.e.*, "lag bias") immediately after the abrupt change, with stunningly extreme residuals between −4.9 and −3.2 standard deviation units during years $6 \leq t \leq 8$. Despite these obvious inaccuracies, the estimated standard deviations of the Moving Average estimates remain misleadingly low, which a well recognized known problem that is fully acknowledged by Patterson and Reams (2005).

Even with the constraint on the maximum standardized residual, the time-series of residuals from the Kalman filter in figure 8B exhibit an apparent temporal pattern, with consistently negative values between years $5 \leq t \leq 8$. This provides weak evidence that an alternative to the static steady-state model warrants consideration, even though the Kalman filter produces reasonably accurate estimates with a misspecified model. The model-based Kalman filter can be robust even when the model is inaccurate, especially when independent design-based panel estimates are closely linked to monitoring of residuals to detect deviations from their expected distribution.

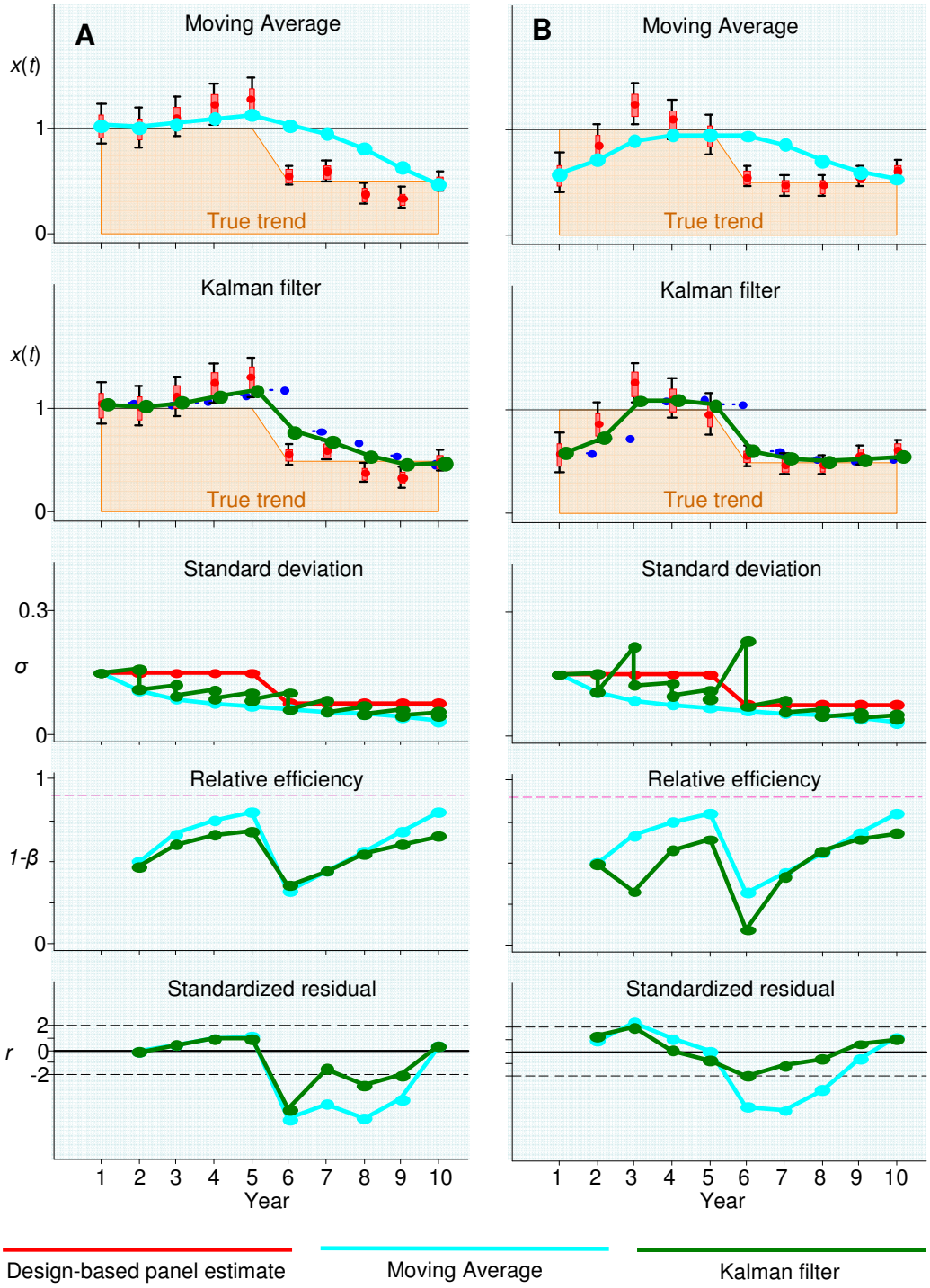


Figure 8: Comparison of Moving Average and Kalman filter estimates simulated population that changes very abruptly at year t=6. The Moving Average estimates in these 2 realizations are obviously inaccurate at year t=6 even though their estimated precisions remain relatively high. In figure 8A, both estimates exhibit improbably negative standardized residuals, although the Kalman filter is slightly better. In figure 8B, the standardized residuals for the Kalman filter are constrained to remain within r_{max}=±2 standard deviation units (equation 18). This improves the fit of the Kalman filter estimates to the true population values in this realization, but suspicious temporal trends in the standardized residuals from the Kalman filter remain.

A more problematic example is given in figure 9. As in figure 8, the population has a strong and variable pattern of change over time, while the model used in this particular implementation of the Kalman filter assumes a static steady-state. In figure 9A, the Moving Average estimates are very inaccurate, while the Kalman filter estimates are even less accurate. The standardized residuals are implausibly positive during years $2 \leq t \leq 6$, and curiously distributed in subsequent years. This is another example of very precise estimates that are very inaccurate. However, the results are much better in figure 9B, where same panel estimates are used, but constraints are imposed on the maximum standardized residuals for the Kalman filter (equation 18). While accuracy is not exceedingly good, the Kalman filter does produce more accurate estimates than the Moving Average, especially for years $8 \leq t \leq 10$. Even so, the standardized residuals from the Kalman filter exhibit strong temporal patterns, with values consistently near +2 standard deviation units during years $2 \leq t \leq 6$, and consistently near −2 standard deviation units during years $7 \leq t \leq 9$. This may be interpreted as evidence that the static steady-state model poorly represents the true temporal trends in the population. More plausible models based on independent information should be investigated.

Finally, figure 10 provides yet another example in which the population is truly changing while the model within the Kalman filter assumes a static steady-state population. The population attribute decreases 5 percent per year in figure 10A, and 10 percent per year in figure 10B. The Moving Average and Kalman filter produce nearly identical estimates. Standardized residuals from the Kalman filter are constrained by equation 18, but they are not excessive, and they reveal no obvious temporal patterns. The realized time-series of estimates and standardized residuals for the exponentially decreasing populations in figure 10 do not appear remarkably different than those for the static population in figure 7. The constraint on the Kalman filter residuals does decrease estimated efficiency relative to the Moving Average. This loss is a cost of mitigating the risk of model-based bias while preserving the potential gains from model-based efficiency with the Kalman filter. Annual panel data and estimators might not offer sufficient accuracy to detect gradual changes in a population, although even slow monotonic trends should eventually become obvious after a long time-series of panel data. However, these results should not be over-interpreted because they are merely a few realizations of sampling and estimation for several hypothetical populations.

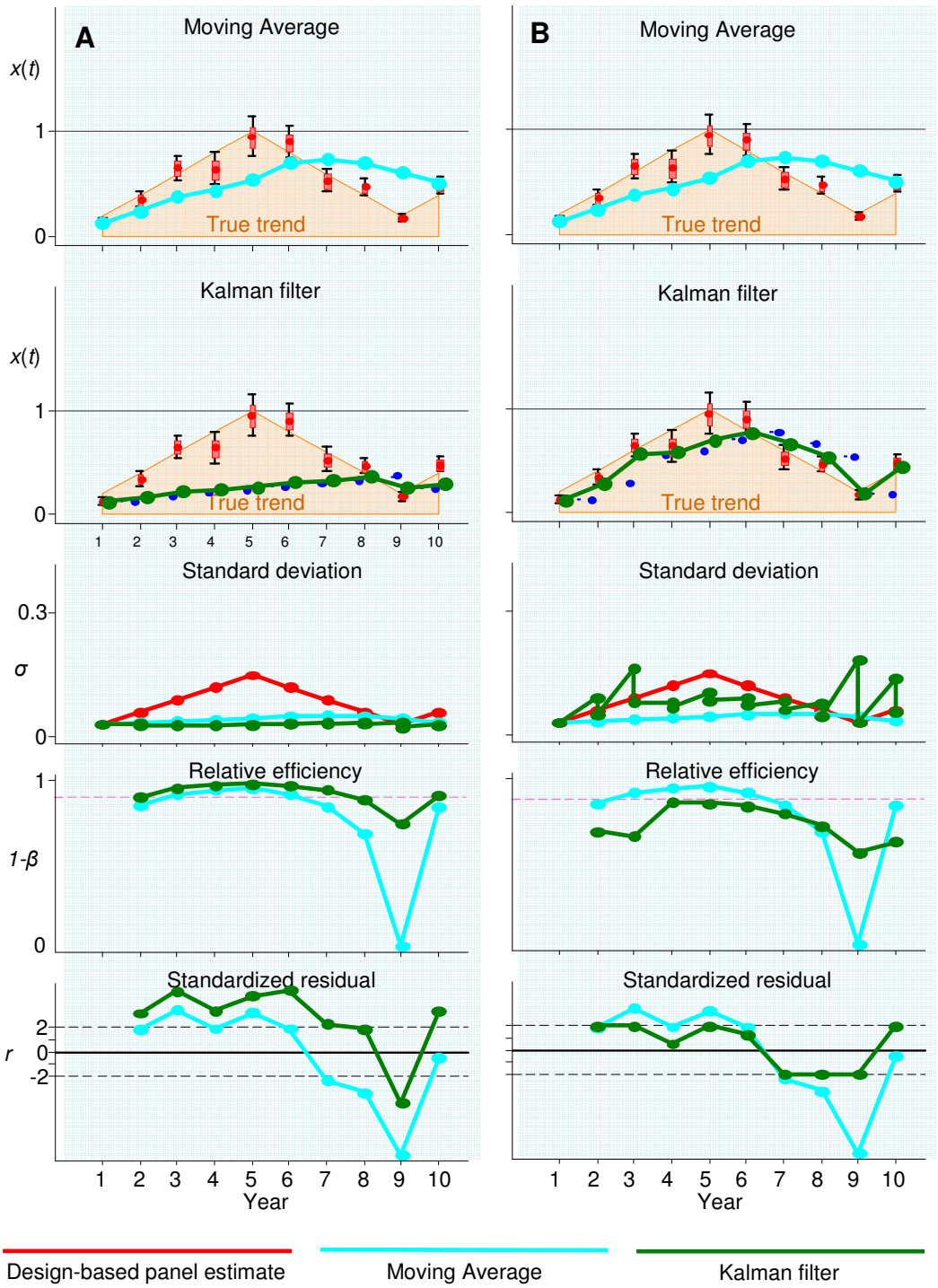


Figure 9: Both the Moving Average and the Kalman filter can appear to be very precise (i.e., low estimated standard deviation of prediction errors). However, their estimates can be very inaccurate (figure 9A). The Moving Average and the static steady-state model within the Kalman filter (equations 27 and 28) do not fit the time-series of FIA panel data very well. Constraining Kalman filter residuals to remain within $r_{max}=\pm 2$ standard deviation units improves the fit in figure 9B, even though the static model remains unchanged. However, a suspicious temporal pattern remains in the residuals (all positive for $2 \le t \le 6$ and all negative for $7 \le t \le 10$). The large residuals might indicate model misspecification rather than rare chance events. An alternative is to use independent information to select a more realistic model without the steady-state assumption.

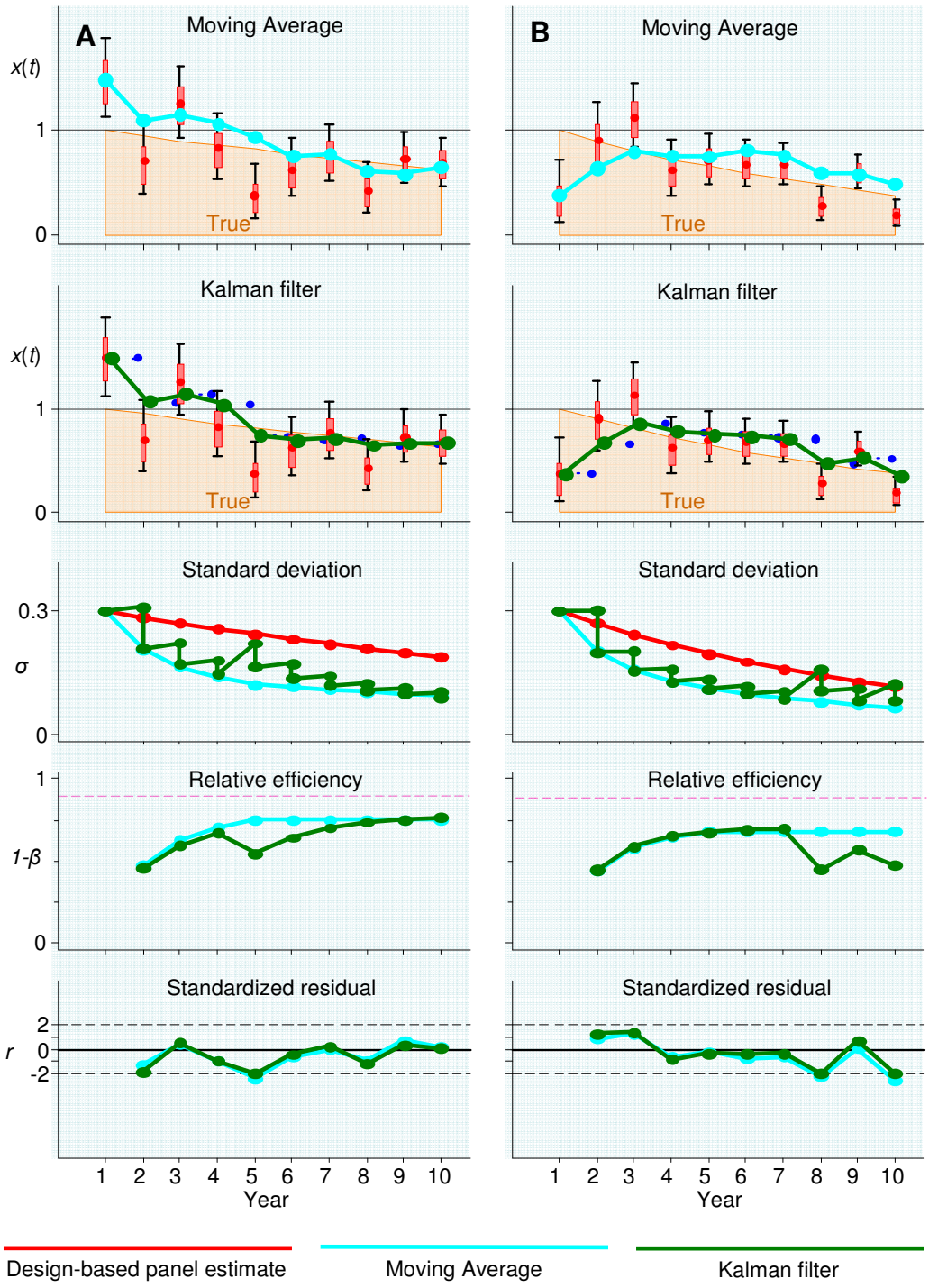


Figure 10: The Moving Average and Kalman filter can produce very similar estimates as in figure 10A, where the population is decreasing 5 percent per year and in figure 10B, where the population is decreasing 10 percent per year. There are no strong clues among the residuals that the static steady-state model (equations 27 and 28) is inaccurate. Annual design-based FIA panel data might not always be insufficient to detect true changes in a population. Constraining the Kalman filter residuals to remain within $r_{max}=\pm 2$ standard deviation units (equation 18) does not notably improve the fit, although it does increase the estimated standard deviation of the predictions and reduce the efficiency offered by the model within the Kalman filter. Mitigation of risks inherent with the model-based Kalman filter (*e.g.*, equation 18) can reduce statistical efficiency.

In "real world" applications, the true population trends are unknown. Therefore, any advantages of the Kalman filter, and the relative costs and benefits from constraints on its residuals, will generally be unknown. However, these examples suggest that the Kalman filter might have desirable characteristics compared to the Moving Average. In ill-behaved situations, it appears that the model-based Kalman filter estimator can be more accurate if residuals are faithfully monitored to reveal model failures. In well-behaved situations, it appears that both estimators can yield very similar results. These preliminary impressions have not been verified with rigorous consideration of the mathematical statistics of the expected values of these 2 estimators, which is the primary tool to make generalizations about efficiency and bias (*e.g.*, Johnson and others 2003).

Monitoring Lodgepole Pine Decline in Colorado

For the last set of examples, we depart from the hypothetical and proceed to an actual time-series of annual FIA estimates. From 2002 to present, a severe epidemic of the mountain pine beetle (*Dendroctonus ponderosae*) has been devastating Colorado's lodgepole pine forests at an alarming rate. Figure 11 illustrates individual tree mortality in the foreground, and the extent of landscape-scale mortality in the background. Figure 12 maps the spread of the epidemic between 2002 and 2007[12].

The following example uses annual FIA panel estimates for Colorado to interpret the magnitude of lodgepole pine mortality caused by the mountain pine beetle outbreak. The annual FIA survey of Colorado uses 10 independent panels, each of which is a $1/10^{th}$ sub-sample of all FIA field plots. The spatial distribution of sample plots within each panel is uniform over the entire state. One panel is measured in the field during each year. Each panel includes approximately 400 forested FIA field plots (table 1).

The annual FIA survey in Colorado began in 2002. There are 6 annual estimates currently available from the initial implementation of the annual FIA survey (see the first two columns in table 1). Each of the 6 design-based sample survey estimates is based on an independent sub-sample of FIA field plots (*i.e.*, a different FIA panel). However, no sample plot occurs in more than 1 panel. Therefore, the data available for monitoring state-wide trends for any single indicator of lodgepole pine condition is limited to 6 observations, namely, 1 statewide design-based panel estimate for each year.

[12] Due to the nature of aerial surveys, the data on the maps in figure 12 will only provide rough estimates of location, intensity and the resulting trend information for agents detectable from the aerial sketchmapping surveys. The data presented on this map should only be used as a partial indicator of mountain pine beetle activity, and should be validated on the ground for actual location and casual agent. Shaded areas show locations where tree mortality or defoliation were apparent from the air. Intensity of damage is variable and not all trees in shaded areas are dead or defoliated.



Figure 11. Since 2002, the epidemic of mountain pine beetle infestations has caused catastrophic levels of lodgepole pine mortality throughout Colorado. Dead trees are shown in the foreground, and landscape-level mortality is apparent on the red slopes in the background. Image courteously of William Ciesla (U.S. Forest Service, retired).

Three indicators related to recent episodic lodgepole pine mortality are considered here: number of live trees, number of mortality trees, and number of damaged trees (table 1). The sampling distributions are assumed to be skewed because negative values are infeasible, and the distributions are assumed to be lognormal.[13] As will be seen shortly, none of the standardized residuals greatly exceeded ±2 standard deviations units; therefore, the r_{max} limit (equation 18) is not applied in this example. Unlike the hypothetical examples above, the true population trends are unknown in the these examples, but independent aerial sketchmapping (*e.g.*, Johnson and Wittwer 2006) demonstrates that the extent of lodgepole pine mortality is extensive (figure 12).

[13] A gamma distribution might also be appropriate, and future analyses should conduct goodness of fit tests to select the most representative distribution.

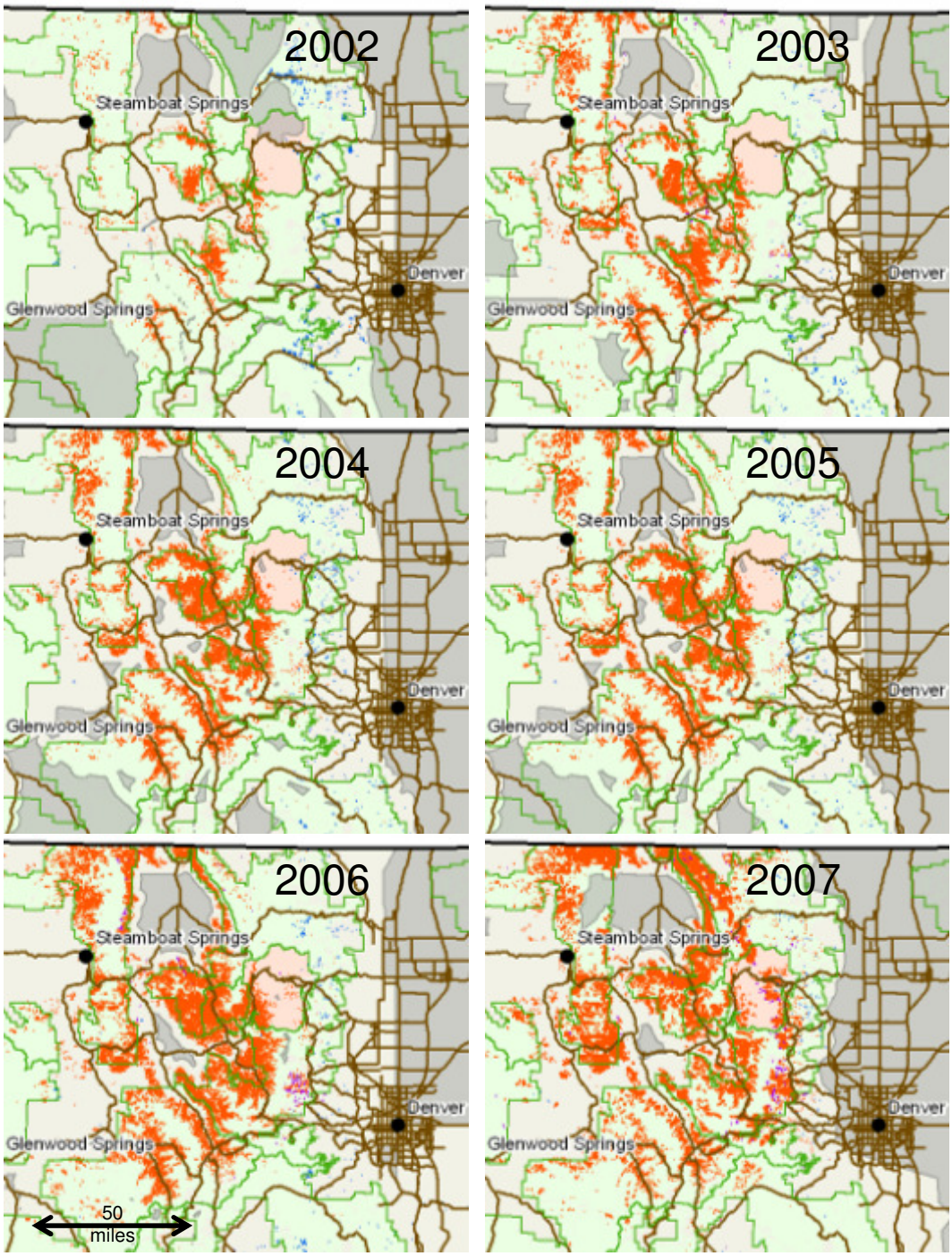


Figure 12. Epidemic spread of lodgepole pine tree mortality between 2002 and 2007 caused by mountain pine beetle in north-central Colorado[12]. The area of infestation is approximately bounded by Denver, Glenwood Springs and the border between the States of Colorado and Wyoming. Areas of heavy mortality are shaded red. These areas were identified through aerial sketchmapping (*e.g.*, Johnson and Wittwer 2006), which is a remote sensing technique that is independent of the FIA annual sample data. National Forests are shaded in green, and areas omitted from each annual survey are shaded in gray. Cartographic products were provided by Jennifer Ross (U.S. Forest Service, Rocky Mountain Region, Forest Health Management Service Center, Lakewood, CO).

Table 1: Lodgepole pine statewide estimates for Colorado from 6 annual FIA panels. These are include all sufficient statistics necessary to apply the Moving Average and Kalman filter estimators in figure 13 to figure 15.

	Design-based estimator		Model-based time-series estimators using annual FIA panels								
	Annual FIA panels[a]		Moving Average[b]				Kalman filter[c]				
Year	Estimated mean	Standard deviation of estimate	Estimated mean	Standard deviation of estimate	Standardized residual[d]	Relative efficiency[e]	Estimated mean	Standard deviation of estimate	Relative efficiency[f]	Standardized residual[g]	Number of forested plots in annual panel
	Average number of live trees per forested acre[h]										
2002	26.30	4.50	26.30	4.50			26.30	4.50			362
2003	27.47	4.25	26.89	3.10	0.13	0.47	26.94	3.15	0.45	0.18	387
2004	36.48	6.17	30.08	2.91	1.14	0.78	29.18	2.99	0.77	1.35	409
2005	28.79	4.91	29.76	2.51	-0.21	0.74	29.06	2.73	0.69	-0.06	407
2006	20.13	3.48	27.83	2.12	-2.25	0.63	25.12	2.31	0.56	-1.92	391
2007	20.49	4.14	26.67	2.09	-1.61	0.74	23.72	2.28	0.70	-0.93	411
	Average annual number of mortality trees per forested acre averaged over 5-year period[i,j]										
2002	0.10	0.06	0.10	0.06			0.10	0.06			362
2003	0.09	0.03	0.10	0.03	-0.06	0.05	0.10	0.03	0.13	-0.56	387
2004	0.16	0.07	0.12	0.03	0.67	0.78	0.15	0.05	0.54	0.06	409
2005	0.13	0.05	0.12	0.03	0.26	0.73	0.16	0.04	0.31	-1.05	407
2006	0.31	0.12	0.16	0.03	1.47	0.93	0.27	0.07	0.67	0.42	391
2007	0.63	0.24	0.26	0.06	1.85	0.94	0.44	0.11	0.80	0.84	411
	Average number of live trees per forested acre damaged by insects[k,i,l]										
2002	0.20	0.10	0.20	0.10			0.20	0.10			362
2003	0.67	0.41	0.44	0.21	0.63	0.74	0.35	0.14	0.88	0.86	387
2004	1.17	0.57	0.68	0.23	0.98	0.83	0.61	0.21	0.86	1.07	409
2005	0.46	0.19	0.63	0.18	-0.71	0.08	0.57	0.17	0.24	-1.16	407
2006	1.88	0.71	0.88	0.20	1.71	0.92	1.04	0.31	0.81	1.31	391
2007	2.11	0.90	1.26	0.27	1.14	0.91	1.69	0.44	0.76	0.54	411

[a] These design-based FIA panel estimates (Scott and others 2005) are sole sufficient statistics used to apply the Moving Average and Kalman filter.

[b] Moving Average defined in Equations 1 and 2.

[c] Univariate Kalman filter defined in equations 19 to 24.

[d] The standardized residual for the Moving Average (r_{MA}) is defined in equation 4.

[e] Relative efficiency (1-β) is defined for the Moving Average in equation 5.

[f] The relative efficiency for the Kalman filter is defined with equation 21.

[g] The standardized residual for the Kalman filter is defined in equation 17.

[h] Kalman filter estimates for total live lodgepole pine tree use the static model with low prediction error (equations 27 and 28, graphed in figure 13).

[i] Kalman filter estimates for lodgepole pine mortality and damage use exponentially increasing model with moderate prediction error (equations 25 and 26).

[j] Graphed in figure 14C.

[k] Graphed in figure 15C.

[l] Tree damage estimates are the basis for examples in figure 5 (2002 only) and figure 6 (2002 to 2007).

Total Numbers of Live Trees

The estimated total numbers of live lodgepole pine trees in Colorado for each year are graphed in figure 13 on page 35. The graph symbolism remains the same as that described on page 20. The initial analysis of this indicator with the Kalman filter used the static steady-state model in equations 27 and 28, which resembles a null hypothesis in which there is zero net change over time.

Both the Moving Average (equations 1 and 2) and the Kalman filter produce very similar time-series estimates (figure 13, table 1). After the initial few years of relatively scant panel data, the variances of the annual estimates average about 33 percent of those for the design-based estimates from each panel, meaning the time-series of Moving Average and Kalman filter estimates have about the same precision as a single sub-sample with 3-times the number of field plots.

The downward trending design-based panel estimates in figure 13 are consistent with the mountain pine beetle epidemic in Colorado. However, the static steady-state model in the Kalman filter fits these data reasonably well. There is nothing obviously askew with the magnitude and temporal trends in the standardized residuals. However, it is very possible that an alternative model would better fit the annual panel estimates, although this possibility was not investigated here.

Tree Mortality

The annual panel estimates for 2006 and 2007 suggest an increasing mortality rate in figure 14 on page 36. However, the spread of the design-based estimation error for any single panel is somewhat broad relative to the apparent trend. Recall from page 20 that the red box plots in figure 14A mark the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of the assumed lognormal distribution with the first 2 moments estimated from the panel data each year. There is a small but plausible chance that the panel estimates in 2006 and 2007 could have been observed even if mortality did not increase beyond the levels in 2002 to 2005. However, the standardized Kalman filter residuals in 2006 and 2007 are suspiciously large (standard deviation units of 1.55 and 2.08 respectively). Under the steady-state model assumed in this case for the Kalman filter, and assuming an expected normal distribution of residuals with zero mean and unit variance, the individual probabilities of each standardized residual are 6 percent and 2 percent respectively. However, had there been an *a priori* hypothesis that the panel estimates would have shown unusually high mortality rates in 2006 and 2007, the single-tailed joint probability of both events would have been $(0.06/2)\times(0.02/2)$, or roughly 0.02 percent. That joint probability is unlikely given the steady-state model and other assumptions.

One alternative hypothesis retains the assumption that the population variable is at a static steady-state (*i.e.*, φ=1 in equation 22), but assumes the annual prediction error is very high instead of very low, which would be characteristic of a dynamic population with annual perturbations that range widely around a constant level. Based on professional judgment, and independent of the observed annual panel estimates, assume a coefficient of variation of 1.00 for the random prediction errors over 1 year prediction interval. Recall that the previous model assumed a coefficient of variation of 0.05. The result is the exact same model for the population variable as equation 27 and figure 14A, but replacing the variance propagation model in equation 28 with:

$$\sigma_x^2(t|t-1) = \sigma_x^2(t-1|t-1) + \left[1.00 \times x(t-1|t-1)\right]^2 \quad [29]$$

The outcome is given in figure 14B. The Kalman filter fits the 2006 and 2007 panel estimates notably better than under the model used for figure 14A. They also fit better than the Moving Average estimates. Furthermore, the time-series of 6 standardized residuals with the modified model does not reveal as strong of an apparently non-random temporal pattern. While the estimated precision from the Kalman filter decreases, the predictions are more plausible given the design-based panel estimates. Unfortunately, the precision with the Kalman filter is not much better than that with the annual FIA panel estimates before the mountain pine beetle epidemic. The model used by this particular Kalman filter contributes little to increased statistical efficiency, at during early stages of the beetle outbreak, and the gain is only marginal thereafter. Most of the information derives from the annual design-based panel estimates, and very little information is gained from model predictions based on past panels.

A third hypothesis is that lodgepole pine mortality is increasing during the beetle epidemic. Based on independent observations, for example forest health reconnaissance with aerial sketchmapping (figure 12) and expert judgment, assume an exponentially increasing model for lodgepole pine mortality between 2002 and 2007, where the mortality rate increases 50 percent (1.5 times) per year, with a coefficient of variation for the random prediction errors over 1-year of 0.25 (*i.e.,* a standard deviation equal to 25 percent the magnitude of the number of mortality trees). This is the same model as that defined in equations 25 and 26 and used to build figure 6, which provides a more detailed example of how the Kalman filter works.

The results for the Kalman filter with the exponentially increasing model from equations 25 and 26 are given in figure 14C and table 1. The fit to the annual design-based panel estimates improves to a modest degree, and surpasses that from the Moving Average. The relative efficiency is intermediate between that from the unrealistic static model (figure 14A) and the uninformative static model (figure 14B). The 6 standardized residuals more resemble their expected

distribution, which has zero mean and unit variance, and their temporal pattern appears more random. However, strong inference regarding the distribution of standardized residuals is rarely possible with only 6 observations.

The model coefficients used in this example are based on expert judgment, and they are not empirically fit to FIA panel data. This process of model parameterization preserves the scientific process in hypotheses formulation and avoids "data mining." With small datasets, there is a risk of misinterpreting a pattern that is caused, in reality, by random processes. For example, sampling error with a small number of annual panel estimates can cause apparent temporal trends in a population that is truly at a static steady-state (e.g., figure 7A and the first 5 years in figure 8). Data mining has a more valid role in the analysis of much larger datasets, where this risk can be less.

This example helps illustrate that no single model or estimator is necessarily "correct." However, there can be a difference when trying to interpret the trends. The Moving Average fits the data (figure 14A), but how does that help answer important analysis questions regarding tree mortality? The static model fits the data when the uncertainty of model predictions is assumed high (figure 14B), but how does that offer any more insights than the Moving Average? The exponentially increasing model also fits the data (figure 14C and table 1), perhaps somewhat better than the alternative hypotheses. However, this final hypothesis suggests that tree mortality measured at time t is about 1.5 times that measured at time t-1, which is an increase of about 50 percent per year. This provides a more meaningful interpretation of the trend data than the alternative models (hypotheses) considered here. If a model is not a meaningful representation of the scientific hypothesis of interest, then "everything is compromised" (Anderson 2008).

Tree Damage

FIA field crews assess damage and insect or pathogen activity that seriously affects live trees with diameter at breast height $\geq$5.0 inches (U.S. Forest Service 2007b). Based on judgment of the field crew, such damage will likely prevent the tree from living to maturity, or surviving 10 more years, if already mature; or the damage will likely reduce the quality of the tree's products (e.g., potentially resulting from lightning strike, excessive lean, tree rot). Whenever feasible, field crews subdivide insect damage into more specific agents, including mountain pine beetles, bark beetles, defoliators, terminal weevils, and *Ips* engraver beetles.

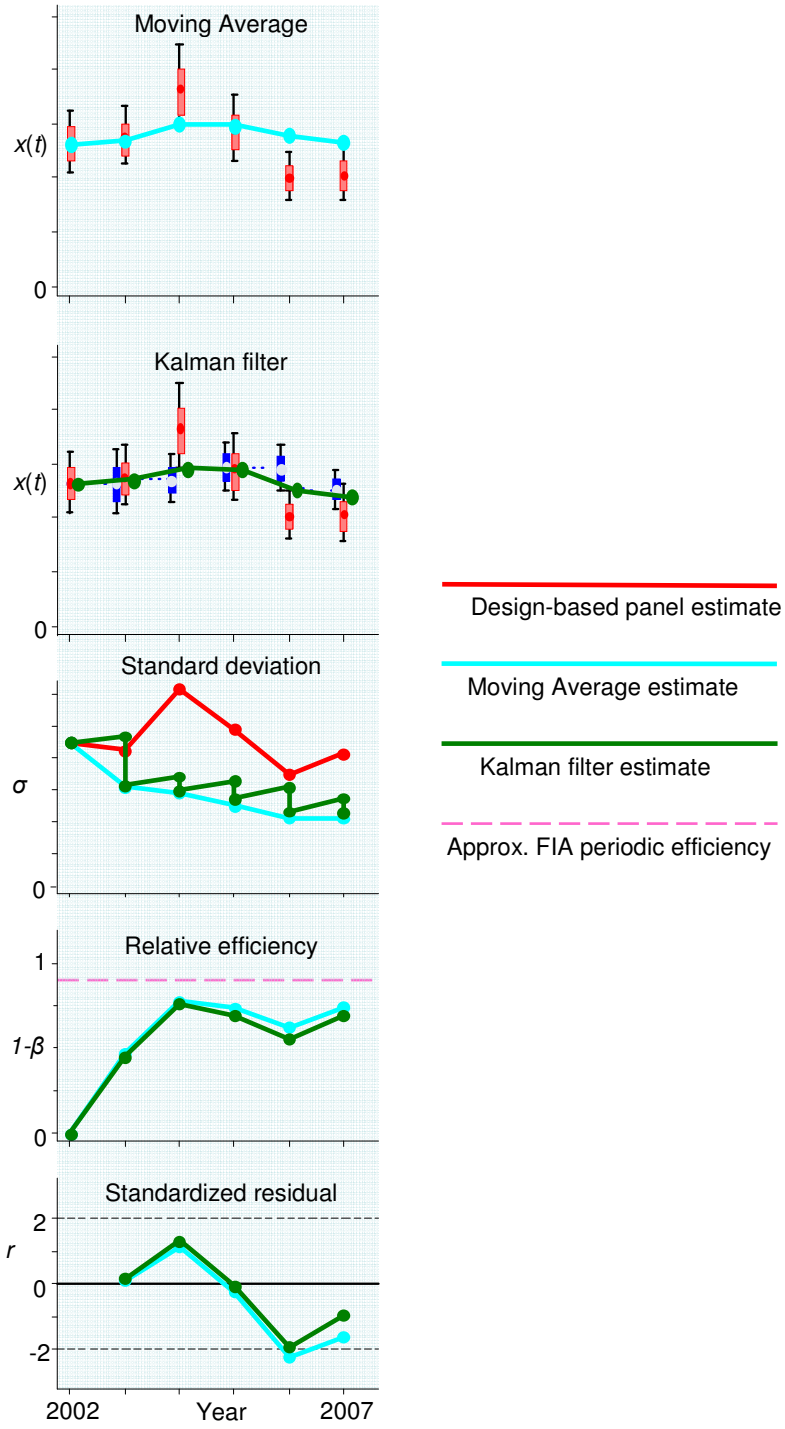


Figure 13: Estimated total number of lodgepole pine trees in Colorado from 2002 to 2007 using independent annual FIA panels. These are actual design-based FIA estimates (Scott and others 2005), and the true population trend is unknown. Both the Moving Average and the Kalman filter yield very similar annual estimates. There is no strong indication from the standardized residuals that the static steady-state model in the Kalman filter (equations 27 and 28) is inaccurate. If there is a true change in the number of lodgepole pine trees in Colorado between 2002 and 2007, then the observed data and the chosen estimators are not powerful enough to detect the change. This figure is discussed on page 32.

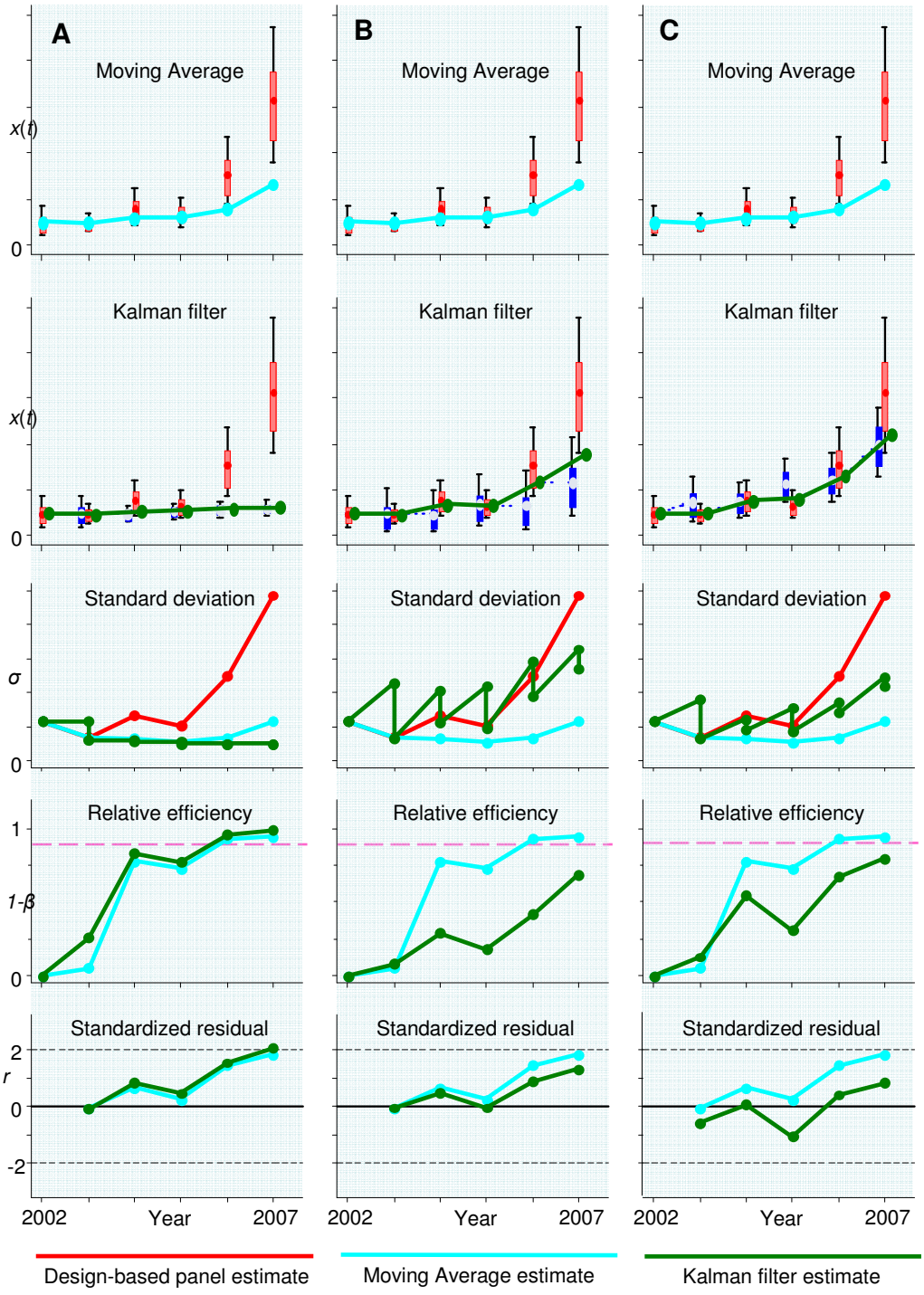


Figure 14: Estimated lodgepole pine tree mortality in Colorado. The Kalman filter with the static steady-state model (equations 27 and 28) does not accurately fit the annual FIA panel estimates in 2006 and 2007 (figure 14A). The fit in figure 14B improves assuming a static model with a much larger random forecasting error (equations 27 and 29). A similar fit in figure 14C is achieved assuming an exponential rate of increase with a moderate degree of forecasting error (equations 25 and 26). The exponentially increasing model has somewhat more plausible residuals. More importantly, it provides the most useful interpretation of the trend in the panel data, which is consistent with the epidemic of pine beetle mortality that is obvious across Colorado. This figure is discussed on page 32.

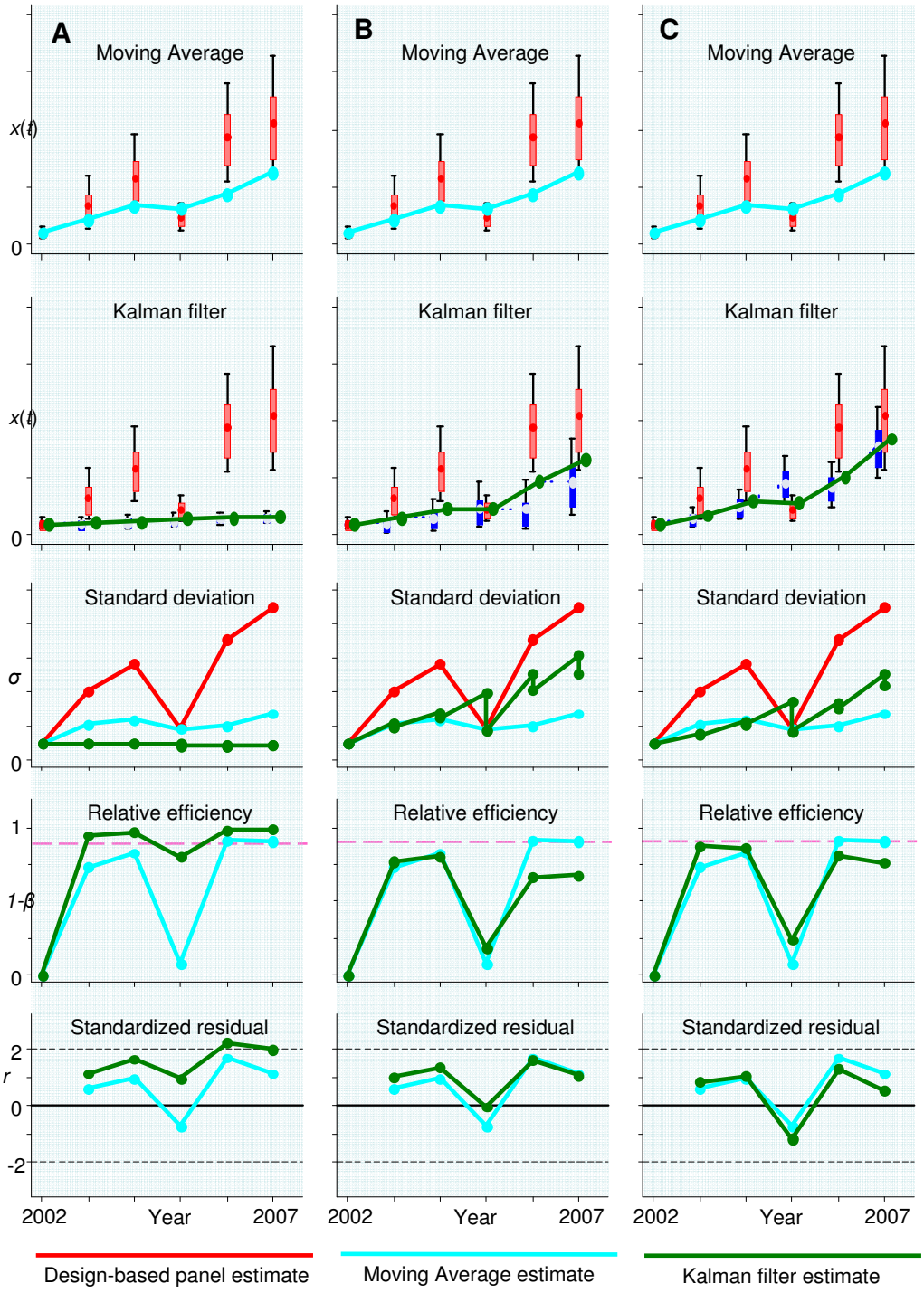


Figure 15: Estimated tree damage in Colorado. Results are very similar to those for tree mortality in figure 14, and the same models are used in figure 15. A multivariate Kalman filter could be formulated that simultaneously predicts total number of trees, tree mortality and tree damage. A multivariate population dynamics model in the Kalman filter might account for the expected biological processes during a pine beetle epidemic. This might improve the accuracy of the time-series estimates and the power to differentiate among alternative models within the Kalman filter. This figure is discussed on page 34.

The estimated time-series of mountain pine beetles damage to live lodgepole pine trees in Colorado is given in figure 15 on the previous page. The results are qualitatively very similar to those for tree mortality. The precise static model (equations 27 and 28) poorly fits the annual FIA panel estimates in figure 15A, while the imprecise static model (equations 27 and 29) fits about as well as the Moving Average (figure 15B), but the model does not markedly help improve statistical efficiency. On the other hand, the exponentially increasing model in equations 25 and 26 fits slightly better (figure 15C and table 1). More importantly, interpretation of the exponentially increasing model in terms of the beetle outbreak is more useful for analyses than the Moving Average and Kalman filter with an imprecise steady-state model.

Discussion

The exposition and examples given above are primarily intended to introduce the reader to the Kalman filter and its relevance to FIA strategic goals and objectives for annual monitoring (U.S. Forest Service 2007a). We hope this helps every reader understand the Kalman filter, at least in an intuitive sense, while providing sufficient, yet simple, mathematical details to serve as an introduction to statisticians. The remaining few sections touch upon more technical issues that warrant future study by FIA analysts and statisticians. The purpose is to suggest possible approaches to improve statistical accuracy, analyses of residuals, quantitative comparisons of alternative hypotheses and statistical estimators, and implementation within FIA information management systems.

Analyzing Trends with Annual FIA Data

Three immediate questions arise when temporal trends are interpreted with annual FIA panel data:

1. From a simple graphical display, is there an observable trend in population parameters over time?
2. If so, does the trend make sense? For example, increased mortality would be expected as the result of a known catastrophic disturbance event, such as bark beetle outbreaks, severe weather, *etc*.
3. Is the trend significant relative to the uncertainty in the population estimates?

For lodgepole pine mortality in Colorado caused by mountain pine beetle, the answer seems to be “yes” for all three questions. The third question has been

addressed in previous studies[14] with analysis of variance and regression. A simple statistical regression with annual panel data verifies there is a significant upward trend in mortality and damage between 2002 and 2007. Analysis of variance identified two combinations of panels that were significantly different from each other. But can the Kalman filter add value beyond the more traditional approaches?

The analyses with the Kalman filter can estimate the temporal trend in tree mortality and damage. Although the Moving Average and regression can accomplish this same task, the results from the Kalman filter may be more interpretable. Furthermore, the Kalman filter can fit the annual panel estimates more accurately. This addresses the first and second questions.

The Kalman filter weights each annual FIA panel estimates with a ratio of variances, which intuitively makes sense and is easy to understand. The Kalman filter can accommodate quantitative models that are based on theory, and alternative models can compare alternative theories and their relative fit to empirical data. This also addresses the second question.

The Kalman filter can compare alternative models with the static steady state model, which resembles a null hypothesis in analysis of variance and regression analyses. This addresses the third question. However, more meaningful hypotheses, in the form of annual transition models (*e.g.*, φ_t in equation 22), may be captured within the Kalman filter. In this context, ranking of alternative hypotheses is discussed in the next section.

Selecting Among Alternative Hypotheses and Estimators

The analyses illustrated in figures 7 to 10 and 13 to 15 employ two different time-series estimators: the Moving Average and the Kalman filter. In addition, the Kalman filter employs intrinsically different temporal models for a population parameter (e.g., static steady state v. exponentially increasing state).

If the Moving Average is used as an unbiased estimator of the current population parameter, then a static or steady-state model is likely implied, in which there is no net change in the population over time. If the Moving Average is assumed to be the minimum variance estimator, then this likely implies temporal homoscadasticity of sampling and prediction errors.

The Kalman filter model used in some of the above analyses shares the same steady-state model with the Moving Average estimator, although the Kalman filter utilizes additional assumptions about the magnitude of time-invariant model prediction errors. Steady-state equations 27 and 28, which are the basis for figure 7 to 13, 14A and 15A, assume a temporally indifferent distribution of random

[14] M.T. Thompson, in review, *Western Journal of Applied Forestry*.

prediction errors with a relatively small standard deviation of annual prediction error (i.e., 5 percent coefficient of variation). Steady-state equations 27 and 29, which are used for figure 14B and figure 15B, assume much more variable annual prediction errors (i.e., 100 percent coefficient of variation). The models used in the Kalman filter analyses for figure 14C and figure 15C, which are defined with equations 25 and 26, hypothesize an exponentially increasing rate of 50 percent per year with modest variability in random annual prediction errors (i.e., 25 percent coefficient of variation).

Which Kalman filter model best fits the annual FIA panel estimates? Is the Kalman filter a more precise estimator than the Moving Average? These questions may be conditionally answered if simplifying assumptions and models are correct. Unfortunately, assumptions and models are virtually always incorrect to varying and unknown degrees. Figure 8A and figure 9A are obvious examples, where it is assumed that prediction error is low because the model is very accurate. Apparently, this assumption is exceedingly overoptimistic because the standardized residuals are suspiciously large[15], with values ranging between -6 and -8 standard deviation units during certain time periods.

The *ad hoc* rule in equation 18 increases the estimate of model prediction error variance based on the analysis of residuals. This rule assures that the residual difference between the model-based prediction and the design-based panel estimate never exceeds 2 standard deviation units, which makes this rule an integral (non-linear) part of the model for prediction errors within the Kalman filter. This rule improves the agreement between the Kalman filter estimates and the true population trends in the hypothetical examples, which is illustrated in figure 8B and figure 9B. But the true trend is never known in actual applications, and this rule was not invoked in the example of annual FIA data for lodgepole pine mortality in Colorado (figures 13 to 15).

Again, how do we select the best hypothesis and estimator among a set of alternatives? Anderson (2008) provides a particularly useful view of comparisons among competing hypothesis and models. This view might help guide analyses of residuals from alternative estimators and models. The following section uses heuristics to briefly explore this train of thought. However, these comments are intended merely to invoke further investigation, and not as a prescription for data analysis.

Entropy and Information-Theoretics: Van Deusen (1999) suggests that alterative hypotheses regarding polynomial temporal trends within the mixed estimator may be compared with the Akaike information criterion (AIC) of Akaike (1974) or the Schwarz information criterion (SIC) of Schwarz (1978)[16]. Burnham and Anderson (2001) review a closely related method from information-theoretics that ranks the fit of alternative models to representative observations,

[15] Deceivingly implausible residuals are possible when the model is accurate, at least in rare cases.

[16] Van Deusen (1999) recommends the Schwarz information criterion in his setting.

where the models represent alternative hypotheses and the fit is measured by observable residuals.

As an embellishment to the likelihood perspective, Burnham and Anderson (2001) describe an AIC method based on residual statistics that are assumed to be normally distributed. This method might be applicable to residuals from the Moving Average (equation 3) and the Kalman filter (equations 15 to 17). If it is applicable, then this method would allow ranking among models (i.e., hypotheses) based on the Moving Average, the steady-state Kalman filter with low prediction error (equations 27 and 28), the steady-state Kalman filter with high prediction error (equations 27 and 29), and the exponentially increasing Kalman filter with moderate prediction error (equations 25 and 26).

The relevant AIC statistic is given by Burnham and Anderson (2001, 2004) as

$$\mathrm{AIC} = n\log\left(\hat{\sigma}_\varepsilon\right) + 2K$$

$$\text{where} \quad \hat{\sigma}_\varepsilon = \sqrt{\frac{\sum \varepsilon_i}{n}}$$

$$\varepsilon_i = \text{residual difference between model and data at point } i$$

$$n = \text{number of residuals}$$

$$K = \text{number of estimated model parameters}$$

[30]

Now consider the afore mentioned heuristics. Let the residuals from the Moving Average estimator be defined as r_{MA} in equation 4, and residuals from the Kalman filter as $r(t_2)$ in equation 15. If each residual were divided by its estimated standard deviation from the design-based panel estimator, then perhaps one might approximate equation 30 with[17]:

$$\hat{\sigma}_\varepsilon = \sqrt{\frac{\sum_{i=2}^{P}\left[r_{MA,i}^2 \Big/ \sigma_{z_i}^2\right]}{P-1}} \quad \text{for the Moving Average}$$

$$\hat{\sigma}_\varepsilon = \sqrt{\frac{\sum_{i=2}^{P}\left[r\left(t_{i|i=1,\cdots,P}\right)^2 \Big/ \sigma_{z_i}^2\right]}{P-1}} \quad \text{for the Kalman filter} \qquad [31]$$

Multivariate vector residuals may be similarly normalized through matrix multiplication by the inverse Cholesky square root of the covariance matrix for estimation errors from the design-based annual panel estimates. Under the

[17] Residuals are not available at the time of initial conditions (t=1).

assumption of negligible model bias and adequate estimation precision, residuals so normalized may be treated as orthogonal with unit variance, and summed together under the assumption that they where all independent and identically distributed. Simultaneous analyses of orthogonal multivariate residuals would provide more observations and stronger evidence when comparing alternative hypotheses and univariate estimators.

Burnham and Anderson (2001) state that the "*AIC* (equation 30) *is not a test in any sense: no single hypothesis (model) is made to be the 'null', there is no arbitrary α level, and there is no arbitrary notion of 'significance'. Instead, there are concepts of evidence and a 'best' inference, given the data and the set of* a priori *models representing the scientific hypotheses of interest. … Akaike's general approach allows the best model in the set to be identified, but also allows the rest of the models to be easily ranked*."

Burnham and Anderson's ranking method only considers hypotheses that are identified explicitly during the analysis. There is no guarantee that some other hypothesis, which was not considered during the analysis, is a better fit to the observed data. For example, the static steady state model in equations 23 and 24 for total number of live lodgepole pine trees provides a reasonable fit the annual FIA panel estimates (figure 13). This model is much like a null hypothesis. However, there is independent evidence from aerial sketchmapping that strongly suggests that the number of live trees should be decreasing (figure 12), not static. An alternative model that predicts a decline in the number of live trees might fit the panel data better that the static model, but the model that predicts a decline was not included in this particular analysis. This is an example of an undesirable omission.

The examples of indictors of lodgepole pine demographics demonstrate the potential inadequacy of tests for a null hypothesis of no change. The null hypothesis may fail to be rejected based on the available data, even when some other hypothesis may have better fit the same data. Therefore, any meaningful analysis should include a set of reasonable *a priori* alternative hypotheses that can be ranked relative to their agreement with independent observations. However, inclusion of a model that mimics a null hypothesis, such as the static steady state model in equations 23 and 24, can provide a relative "baseline" when comparing the differences among AIC statistics for alternative hypotheses. For example, compare 3 models (hypotheses) that predict a percent per year decline in live trees of 25, 50 and 75 percent. The AIC statistics in this hypothetical example are given in table 2. Is the difference in their AIC statistics relatively large or relatively small? Comparisons of these AIC statistics relative to the AIC statistic for the static model (i.e., no annual change) can help answer this question.

In the absence of a model that assumes a static steady state, the AIC statistics in table 2 seem to strongly highlight the model with a 50 percent annual rate of change because the relative AIC for that model is high relative to the other

alternatives (25 and 75 percent per year change). However, the differences among dynamic models (25, 50 and 75 percent per year) are less striking when the static steady state model (0 percent annual change) is used as the baseline[18] The latter comparisons suggest that there is a decline in number of live trees, as would be expected, but there is not strong evidence for estimating the actual rate of change (25, 50 or 75 percent per year).

Table 2 Hypothetical ranking of alternative hypotheses for annual rate of change with the AIC statistic (equations 30 and 31).

Change rate per year	No static model		Static model included[18]	
	AIC	Difference	AIC	Difference
0%			10,000	0
25%	10,200	0	10,200	200
50%	10,250	50	10,250	250
75%	10,225	25	10,225	225

Multivariate Kalman Filters

Multivariate versions of the Kalman filter can improve the accuracy and reliability of time-series estimates, and accommodate more realistic models of ecosystem processes. Some of these issues are briefly covered in this section. Though the complexity of a multivariate approach can be initially intimidating, the basic concepts behind complex Kalman filters are just as simple as the univariate version considered above.

Complex Tree- and Stand-level Models: Tree damage is a leading indicator of tree mortality, and tree mortality is negatively correlated with the number of live trees. This suggests a multi-response model in the Kalman filter that simultaneously considers the demographics of the statewide lodgepole pine tree population, where the average number of trees per forested acre at time t equals the tree density at $t-1$, plus the average number of ingrowth trees and minus the average numbers of new mortality and removal trees per acre between $t-1$ and t.

More accurate and detailed model predictions may be available from deterministic growth and yield models. In principle, any deterministic population or ecosystem model can be linked to the Kalman filter to improve estimates and the analysis of broad-scale trends (e.g., Van Den Brakel and Visser 1996, Williams and others 2005, and Tian and Xie 2008). Multivariate predictions from a complex nonlinear model can be made for each FIA plot or tally tree, using the most recent field measurements as initial conditions. Examples include the Forest Vegetation Simulator (Crookston and Dixon 2005, Miles 2008) and models fit to

[18] Recall that the static model for live trees fit the annual FIA panel data reasonably well in figure 13

FIA plot data by Lessard and others (2001) for updating annual forest surveys. Imputation can be viewed as an empirical multivariate prediction model, which has already been applied in the context of annual FIA surveys (Van Deusen 1997, Reams and McCollum 2000, McRoberts 2001, Gartner and Reams 2002, and Eskelson and others 2008). However, if alternative hypotheses will be compared during analysis, then deterministic models, such as the Forest Vegetation Simulator, should be used rather than purely empirical methods, such as imputation. Deterministic models can capture assumptions about future population dynamics, whereas purely empirical methods use past observations to predict future conditions.

Imputations or predictions from the model for each tree or FIA sample plot can be used to estimate the predicted state vector at the population level. This is accomplished with design-based sample survey methods, exactly like those used with actual field measurements. Those multivariate sample survey estimates are then assimilated into the Kalman filter through a multivariate transition matrix, which is analogous to the univariate φ_t in equations 23 and 24. One challenge would be estimation of the covariance matrix that describe model predictions error, analogous to the scalar variance $\sigma_w^2(t)$ in equation 24 plus any quantified estimation error associated with the transition matrix (Ni and Zhang 2008).

Such models could be formulated to compare alternative hypotheses that are related to forest health. The Forest Vegetation Simulator can model effects of disturbance agents, including insects, pathogens, and fire (Crookston and Dixon 2005). Hypotheses may be based on climate change scenarios, the consequences of which are modeled with the Forest Vegetation Simulator (*e.g.,* Malmstrom and Raffa 2000; Crookston and others 2008). Hypotheses may be formulated with a demographic model that predicts the spatial dynamics of a pest population and the associated damage to trees (*e.g.,* Logan and others 2003). Ranking the degree of agreement between model predictions and direct observations, such as annual FIA panel estimates, is briefly covered on page 39.

Standardization of Multivariate Vector Residuals: If all assumptions incorporated into a Kalman filter are approximately correct, then the standardized residuals should be approximately distributed with a zero mean, unit variance, and mutual independence over time (Maybeck 1979). If there is a convincing deviation of the standardized residuals from their expected distributions, then there is evidence of model misspecification. It can not be overemphasized that close scrutiny of residuals is essential to mitigate the risk from a model-based approach, such as the Kalman filter, while preserving the gains in efficiency that are possible with the model-based approach.

In a large and complex governmental statistical system like FIA, this level of scrutiny would have to match that already conducted to detect other sorts of non-sampling errors (*e.g.,* Pollard and others 2006). FIA database software could be

augmented with a hypothesis test that the set of standardized residuals from recent FIA panels and annual estimates fits a normal distribution with mean zero and unit variance; the Kolmogorov-Smirnov test is an example. The database could include a test that the time-series of recent residuals from the composite estimator in the Kalman filter fulfill the expectation of mutual independence over time; the non-parametric Wald-Wolfowitz runs test is an example.

If population-level estimates from the deterministic models demonstrate a good fit to the time-series of FIA panel estimates, and analyses of residuals reveal no suspicious deviations from expectations, then the accuracy of the statistical estimates with the Kalman filter will likely improve. If the fit is mediocre, at least the Kalman filter is robust against modest levels of model misspecification. Furthermore, any lack of fit, assuming it is actually discovered during an analysis, provides the opportunity to learn more about the system and improve the models. Monitoring and analysis of vector residuals from the Kalman filter could assist in this learning process.

Remotely Sensed Data: Ancillary remotely sensed data can improve the estimated area of forest, including separation of estimates into forest area with and without severe insect damage (e.g., Wulder and others 2005 2006a 2006b, Goodwin 2008). The Kalman filter can combine multi-response process models for land use, land cover and forest condition, with the time-series of annual design-based multivariate panel data, and with multivariate ancillary data from remotely sensed censuses and sample surveys (Czaplewski and others 1988, Czaplewski 1990, 1995, 1999, 2001). This approach does not require stratification of individual FIA panels based on remotely sensed data and geopolitical boundaries, such as counties (Czaplewski 2001). Therefore, the Kalman filter can avoid problems inherent with detailed stratification and when the sample size is small, which is a problem particularly acute with the Moving Average method (Patterson and Reams 2005).

In a sense, the composite estimator in the Kalman filter improves precision by "borrowing" relevant information from the past. Likewise, the Kalman filter can improve precision by "borrowing" ancillary information from remotely sensed sources. Therefore, the Kalman filter is potentially well-suited for complex monitoring systems that include multiple time-series of multivariate remotely sensed data and field data. An example is the Nevada Photo-based Inventory Pilot (NPIP), which is one attempt to implement the national FIA strategic plan (U.S. Forest Service 2007a).

Improving Accuracy of Time-Series Estimates: The mountain pine beetle example in figure 14C and figure 15C used "expert opinion" to quantify model parameters. However, the "extended" Kalman filter (Jazwinski 1970) can simultaneously estimate population attributes and model parameters. Rather than

a priori expert judgment used to model lodgepole pine damage and mortality as a 50 percent increase per year (figure 14 and figure 15), a potentially more accurate rate parameter could be estimated from the annual panel data, in addition to the estimated mean number of trees per acre. The model form (*e.g*., exponential rate of change) would need to be identified from independent external sources, but the parameter values (*e.g.,* 50 percent per year increase) for the model could be fit empirically without impairing valid inference.

Since forest populations are integrated systems, there can be strong isotropic temporal correlations among variables at proximate points in time, both past and future. More advanced versions of the Kalman filter can act as linear smoothers over multiple time increments (Jazwinski 1970). It is possible that current panel data can improve composite estimates for past conditions in addition to current conditions. This kind of extra effort seems worthwhile when addressing important analysis questions with a relatively short time-series of annual cross-sectional panel data. After all, the cost is over $500,000 for each annual datum point in Colorado.

Multivariate versions of the Kalman filter employ matrix algebra and inverse covariance matrices. Unfortunately, covariance matrices can be ill-conditioned or even singular, especially when the dimensions of the vector estimates are large. These pragmatic circumstances frequently cause numerical instability in large applications of the Kalman filter. In many cases, the numerical results from the Kalman filter will be dominated more by numerical round-off errors than random sampling and prediction errors. This vulnerability can produce disappointing, inaccurate or even infeasible results (e.g., negative variance estimates) with the Kalman filter. Aberrant numerics may not be obvious from the vector estimate alone, which is especially dangerous. Fortunately, variations of the Kalman filter are numerically robust, even with singular covariance matrices (Maybeck 1979). These numerical solutions employ various types of matrix square roots, matrix decompositions or matrix factorizations when combining vector model predictions with design-based vector panel estimates. Bierman (1977) is a particularly useful source for effective solutions to numerical hazards with large, ill-conditioned covariance matrices. While these solutions add complexity during implementation of the Kalman filter, they should not be allowed to distract from the fundamental and intuitive simplicity of the Kalman filter. If there were numerically perfect computers, then the complexity needed to solve numerical problems would not be necessary.

Comparison of FIA Annual and Periodic Surveys for Monitoring

During the 1960s to 1990s, the FIA periodic design produced relatively precise estimates for a “snapshot” in time, but these same estimates often lost much of their value well before the FIA Inventory Unit was re-measured during the following periodic cycle. Users tended to lose confidence in FIA periodic data

about 5 years after the field work was completed. Since periodic surveys were repeat only once every 10 to 20 years (AFPA 1998), FIA period surveys had limited value for half or more of their life cycle. How does the current annual FIA design compare with the previous periodic design for monitoring applications?

An annual FIA survey in the Interior West essentially uses the same field plots as the periodic survey. However, those plots are systematically sub-sampled into 10 mutually exclusive FIA panels in the western USA (Bechtold and Patterson 2005). Any single panel includes $1/10^{th}$ of all FIA field plots. The precision, as measured by the standard deviation of the sample mean, from a single panel equals $(1/\sqrt{10})=0.32$ of the precision from a full periodic survey if both were measured during the exact same time. Therefore, confidence intervals from a single panel are about $(0.32^{-1})=3.1$ times broader than those from a periodic survey conducted in the same year. While there is sufficient funding to measure a single panel in 1 year, there are not enough resources to measure the ten-fold increase in plots every year that would be required for a periodic survey. The annual design sacrifices precision to gain timeliness, although precision remains important in monitoring changes over time.

In order to assess FIA monitoring programs with the current annual design relative to the prior periodic design, hypothetically assume that a periodic FIA survey of Colorado was completed in 2002, with approximately 4,000 forested field plots measured. Therefore, the variance of this hypothetical periodic survey in 2002 would be about $1/10^{th}$ that of the actual 2002 annual survey. To complete this scenario, visualize annual FIA surveys, each of which measures about 400 forested field plots during a single year, between 2002 and 2007.

Assume that the statewide total number of lodgepole pine trees remains at an approximate steady-state between 2002 and 2007, as described by the model in equations 27 and 28 and used with the Kalman filter for figure 13. Under this model, Figure 16 shows the expected standard deviation of the hypothetical periodic survey in 2002 as an estimate for each year between 2003 and 2007. Figure 16 shows that after 3 years the data quality, as measured by the standard deviation of random estimation errors (σ), of the hypothetical 2002 periodic survey for live trees is approximately the same as the Kalman filter estimate that uses the same model with the much smaller annual surveys from 2002 to 2005. In the absence of new periodic data, and given the stated assumptions, the precision of the Kalman filter estimates for live trees with annual panel data surpasses that of the hypothetical 2002 periodic survey in years 2006 and beyond.

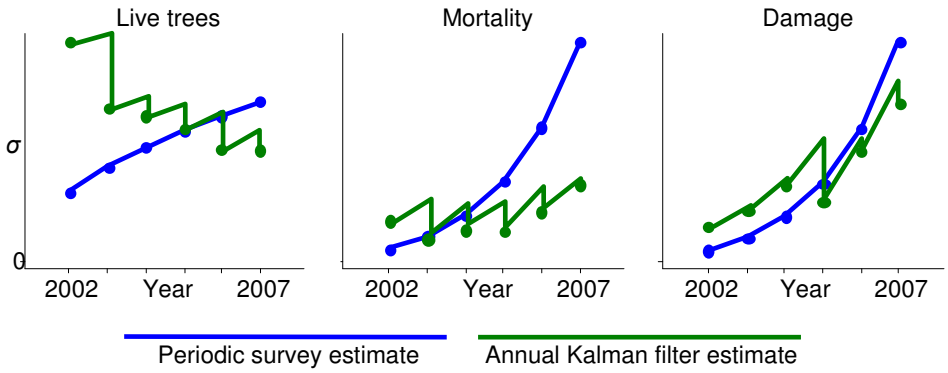


Figure 16: The standard deviation of annual estimates, each with a sample size of *n*/10, compared to that expected from a hypothetical periodic survey in 2002, with a sample size of *n*. A full periodic survey is more accurate (*i.e.,* smaller σ) immediately after all field data are measured, hypothetically in 2002. However, unobserved changes in the population can quickly accumulate after the periodic survey is completed. After 1 to 3 years, the estimated precision of the annual estimates, which are combined with the Kalman filter, equals or surpasses that which would be expected from the corresponding periodic FIA survey, which has 10-times the sample size of a single annual panel. The temporal propagation of error in the design-based periodic survey estimates assumes the same model used for the corresponding annual estimates with the Kalman filter with annual panel data. The model for live lodgepole pine trees assumes a steady-state (figure 13, equations 27 and 28). The models for lodgepole pine tree mortality and damage assumes an exponential increase (figure 14 and figure 15 equations 25 and 26, figure 5). Only the first 5 years after the hypothetical periodic survey are illustrated here. Thereafter, the precision of a periodic survey further deteriorates relative to an annual FIA survey. Likewise, the value of estimates from the periodic survey declines, at least until the next periodic survey. Historically, periodic surveys were conducted once every 10 to 20 years.

As another example, assume that statewide lodgepole pine mortality from mountain pine beetles increases exponentially 50 percent per year between 2002 and 2007. This model is expressed in equations 25 and 26, which is the same model used with the Kalman filter for figure 14C. Assuming this model is approximately true, Figure 16 shows the expected standard deviation of the hypothetical 2002 periodic survey as an estimate of tree mortality for years 2003 through 2007. An example of this same expectation is shown in more detail in figure 5. Almost immediately, the data quality from the annual FIA panels and the Kalman filter exceeds that of the hypothetical periodic survey in 2002, at least under the very dynamic circumstances caused by the mountain pine beetle epidemic in Colorado.

Assuming the same exponential rate of increase in insect damage to live trees, the annual FIA panels with the Kalman filter estimator produces more precise estimates than the hypothetical 2002 periodic survey after only 3 years (figure 16). This hypothetical example demonstrates the potential advantage of the annual FIA design, when coupled with appropriate time-series estimators, for monitoring important changes in a broadly distributed forest population.

Conclusions

The daunting migration by FIA from periodic to annual surveys has improved the timeliness of FIA forest inventory statistics. These have very been useful in strategic analyses of the current state of forests at the national, state and multi-county scales. These results are routinely produced with simple and familiar statistical estimators, namely, the Moving Average or the closely related Temporally Indifferent method. However, a successful monitoring program requires more than production of updated inventory reports (Moffat and others 2008). The Kalman filter not only can improve annual inventory updates when the forest population is undergoing rapid change, the Kalman filter can also improve the ability to monitor, quantify and interpret broad changes in the nation's forests. This is a high priority the FIA strategic plan (U.S. Forest Service 2007a). Hopefully, the descriptions and examples in our paper reveal the benefits and intuitive simplicity of Kalman filter.

The Kalman filter offers other advantages over the Moving Average and Temporally Indifferent methods. The model-based Kalman filter estimator can be more accurate for populations that are rapidly changing, especially if residuals are faithfully monitored to reveal model failures. In populations that are static or change very slowly, it appears that both the Moving Average and Kalman filter estimators yield very similar estimates of current forest inventories.

A properly implemented Kalman filter, which includes analyses of residuals, combines the statistical efficiency of a model-based estimator with the reliability of a design-based estimator. Therefore, the Kalman filter is less vulnerable to temporal "lag bias" (Patterson and Reams 2005) when population dynamics are in a relatively rapid state of flux. This is precisely the situation in which accurate monitoring is most important. This also means that the Kalman filter can be less risky than other model-based estimators when there is a chance that the model is inaccurate.

Unlike the Moving Average and Temporally Indifferent methods, alternative implementations of the Kalman filter can incorporate predictions from alternative deterministic models, which, in turn, are the manifestations of alternative sets of hypotheses. The Akaike information criterion (AIC), which quantifies the agreement between Kalman filter estimates and purely design-based FIA panel estimates, might be used to rank the fit of alternative models to FIA panel data. Therefore, the analyst can quantitatively evaluate alternative *a priori* hypotheses that are intended to explain temporal trends in forest populations. For example, does a model that includes the consequences of climate change better fit the annual FIA design-based panel estimates than a model that assumes no such affects?

The Kalman filter can incorporate deterministic models that consider the population demographics of growth, mortality, regeneration, stand succession,

and changes in land use. The Kalman filter can assimilate time-series of diverse remotely sensed data, without the burdensome constraints of post-stratification (Czaplewski 2001). These capabilities can improve precision of inventory and monitoring estimates relative to the Moving Average and Temporally Indifferent methods.

The analyst is initially interested in 3 basic questions: Is there an observable trend in population parameters over time? Does the trend make sense? Is the trend significant relative to the uncertainty in the population estimates? The Kalman filter, much like the Moving Average and Temporally Indifferent methods, can address the first question, at least through the qualitative judgment of the analyst and the temporal series of estimates (see figures 7 through 13).

The time series of estimates from the Moving Average or Temporally Indifferent methods, in concert with the analyst's professional judgment, can address the second question: Does the trend make sense? The Kalman filter is capable of the same. In addition, the Kalman filter, which can objectively rank alternative deterministic models, can progress beyond qualitative judgment, and provide quantitative evidence to answer the second question.

Quantitative evidence is needed to answer the third question: Is the trend significant relative to the uncertainty in the population estimates? The Kalman filter can include a static model in which the population is assumed to be at a steady state, where there the net change over time is zero. The relative fit of this static model compared to alternative dynamic models can be used to make inferences about the third question. The Moving Average and Temporally Indifferent methods do not provide comparable information. Therefore, the Kalman filter improves upon the Moving Average and Temporally Indifferent methods when analyzing temporal trends with FIA annual panel data.

In principle, the Kalman filter offers the opportunity to improve FIA monitoring and analyses. However, this opportunity has not been rigorously tested. Although the fundamental concepts in the Kalman filter are intuitively simple, implementation of the Kalman filter is more complex than current methods used in FIA information management systems. Complexity inescapably incurs risk.

One of the Guiding Principles of the FIA strategic plan (U.S. Forest Service 2007a) is to "*take the lead in inventorying and monitoring changes in the nation's forests, forest resources, and forested ecosystems*." This guidance suggests that additional statistical research by FIA should be directed towards time-series methods that monitor changes in the nation's forests using annual FIA data. The Kalman filter is one such method.

Acknowledgements

Dr. John D. Shaw (U.S. Forest Service, FIA, Ogden, UT) provided the necessary encouragement to further develop the original oral presentation at the 2008 FIA symposium into a full peer-reviewed article in this Proceedings. Drs. Francis A. Roesch (U.S. Forest Service, Asheville, NC), James A. Baldwin (U.S. Forest Service Albany, CA) and KaDonna Randolph (U.S. Forest Service, Knoxville, TN) generously contributed very useful peer reviews. As always, any remaining errors and omissions remain our sole responsibility.

Mr. William Ciesla (U.S. Forest Service, retired, Fort Collins, CO) donated the image of bark beetle damage in figure 11, and Jennifer Ross (U.S. Forest Service, Lakewood, CO) provided the cartographic products used in Figure 12.

Finally, this paper is dedicated to the memory of Dr. Albert R. Stage, who inspired a whole generation of scientists interested in monitoring and forecasting forest conditions in the pursuit of sustainable forestry.

Literature Cited

AFPA. 1998. The report of The Second Blue Ribbon Panel on Forest Inventory and Analysis. Washington, DC: American Forest and Paper Association. 17 p.

Akaike, H. 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 19:716-723.

Anderson, D. R. 2008. Model Based Inference in the Life Sciences: A Primer on Evidence. Springer, New York, USA. 184pp.

Bechtold W.A. and P.L. Patterson (editors). 2005 The enhanced forest inventory and analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.

Bierman, Gerald J. 1977. Factorization Methods for Discrete Sequential Estimation. Academic Press, NY. 241pp.

Brockwell, P.J.; Davis, R.A. 1996. Introduction to time series and forecasting. New York: Springer Verlag. 420 pp.

Burnham, K. P. and Anderson, D. R. 2001. Kullback-leibler information as a basis for strong inference in ecological studies. *Wildlife Research* 28:111-119.

Burnham, K. P. and Anderson, D. R. 2004. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research* 33:261-304.

Cressie, N. and Wikle, C. K. 2002. Space–time Kalman filter. *In* Encyclopedia of Environmetrics, Volume 4, pp 2045–2049. El-Shaarawi and Piegorsch (Eds.) John Wiley & Sons, Ltd, Chichester

Crookston, N. and Dixon, G. 2005. The forest vegetation simulator: A review of its structure, content, and applications. *Computers and Electronics in Agriculture* 49:60-80.

Crookston, N. L., Rehfeldt, G. E., Ferguson, D. E., and Warwell, M. 2008. FVS and global warming: A prospectus for future development. *In* R. N. Havis and N. L. Crookston (eds.), Third Forest Vegetation Simulator Conference, RMRS-P-54, pp. 7-16. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Czaplewski, R. L. 1990. Kalman filter to update forest cover estimates. In V. J. Labau and Cunia (eds.), State-of-the-art methodology of forestry inventory, GTR PNW-263, pp. 457-465. USDA Forest Service, Pacific Northwest Research Station.

Czaplewski, R. L. 1995. Continuous adaptive monitoring of status and trends in ecosystem conditions. Proceedings of the 1995 Society of American Foresters conference. Bethesda, MD: Society of American Foresters. 482 p.

Czaplewski, R. L. 1999. Multistage remote sensing: toward an annual national inventory. *Journal of Forestry*. 97(12):44-48.

Czaplewski, R. 2001 Areal control using generalized least squares as an alternative to stratification , pp. 63-65. In Proceedings of the Second Annual Forest Inventory and Analysis Symposium ; 2000 October 17-18; Salt Lake City, UT. Reams, Gregory A.; McRoberts, Ronald E.; Van Deusen, Paul C., eds. General Technical Report SRS–47. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 143 p.

Czaplewski, R. L., Alig, R. J., and Cost, N. D. 1988. Monitoring land/forest cover using the Kalman filter: a proposal. *In* Conference on Forest Growth Modelling and Prediction, GTR NC-120, pp. 1089-1096. USDA Forest Service North Central Forest Experiment Station.

Devall, M. S., Grender, J. M., and Koretz, J. 1991. Dendroecological analysis of a longleaf pine *pinus palustris* forest in Mississippi. *Plant Ecology* 93:1-8.

Dixon, B. L., and R. E. Howitt. 1979. Continuous forest inventory using a linear filter. *Forest Science* 25(4):675-689.

Eskelson, B. N. I., Hailemariam, T., and Barrett, T. M. 2008. Tree-level imputation techniques to estimate current plot-level attributes in the pacific northwest using paneled inventory data. *In* W. McWilliams, G. Moisen, and R. Czaplewski (eds.), Forest Inventory and Analysis (FIA) Symposium Proceedings, RMRS-P-56CD, pp. 1-13, Fort Collins, CO. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Gadbury, G. L.; Williams, M. S.; Schreuder, H. T. 2004. Revisiting the southern pine growth decline: Where are we 10 years later? Gen. Tech. Rep. RMRS-GTR-124. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 10 p..

Gartner, D. and Reams, G. 2002. A comparison of several techniques for imputing tree level data. *In* R. E. Mcroberts, G. A. Reams, P. C. Van Deusen, and J. W. Moser (eds.), Proceedings of the Third Annual Forest Inventory and Analysis Symposium, General Technical Report GTR-NC-230, pp. 103-108. U.S. Department of Agriculture, Forest Service, North Central Research Station.

Gao, C., Wang, H., Lakshmivarahan, S., Weng, E., Zhang, Y., and Luo, Y. 2008. Parameter estimation and uncertainty analysis of forest carbon dynamics: Ensemble Kalman filter method. *In* 93rd Annual Meeting of the Ecological Society of America: Enhancing Ecological Thought by Linking Research and Education.

Goodwin, N. R., 2008. Estimation of insect infestation dynamics using a temporal sequence of landsat data. *Remote Sensing of Environment* 112:3680-3689.

Gove, J. and Houston, D. 1996. Monitoring the growth of american beech affected by beech bark disease in Maine using the Kalman filter. *Environmental and Ecological Statistics* 3:167-187.

Gregoire, T. G. and D.K. Walters. 1988. Composite vector estimator by weighting inversely proportional to variance, *Canadian Journal of Forest Research*. 18:282–284.

Hurteau, M., H. Zald, and M. North 2007. Species-specific response to climate reconstruction in upper-elevation mixed-conifer forests of the western Sierra Nevada, California. *Canadian Journal of Forest Research* 37:1681-1691.

Jazwinski, A. H. 1970. Stochastic Processes and Filtering Theory. Academic Press, NY. 376pp

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

Johnson, E. W.; and D. Wittwer. 2006. Aerial detection surveys in the United States *In*: Aguirre-Bravo, C.; Pellicane, Patrick J.; Burns, Denver P.; and Draggan, Sidney, Eds. 2006. Monitoring Science and Technology Symposium: Unifying Knowledge for Sustainability in the Western Hemisphere Proceedings RMRS-P-42CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. pp. 809-811.

Kish, L. 1987. Statistical Design for Research. John Wiley & Sons.

Kott, P. 2005. Randomization-assisted model-based survey sampling. *Journal of Statistical Planning and Inference* 129:263-277.

Lessard, V. C., McRoberts, R. E., and Holdaway, M. R. 2001. Diameter growth models using Minnesota Forest Inventory And Analysis data. *Forest Science* 47:301-310.

Logan, J. A., Régnière, Â., and Powell, Â. A. 2003. Assessing the impacts of global warming on forest pest dynamics. *Frontiers in Ecology and the Environment* 1:130-137.

Malmstrom, C. M. and Raffa, K. F. 2000. Biotic disturbance agents in the boreal forest: considerations for vegetation change models. *Global Change Biology* 6:35-48.

Maybeck, P. S. 1979. Stochastic models, estimation, and control. Volume 1. Academic Press, NY. 423pp.

McRoberts R.E. 2001. Imputation and model-based updating techniques for annual forest inventories. Forest Science 47:322-330.

McRoberts, R. E. 2005. The enhanced forest inventory and analysis program. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 1-10

Meinhold, J. and Singpurwalla, N.D. 1983. Understanding the Kalman filter. *The American Statistician* 37:123–127.

Miles, P. D. 2008. Forest inventory and analysis data for FVS modelers. In: Havis, Robert N.; Crookston, Nicholas L., comps. 2008. Third Forest Vegetation Simulator Conference; 2007 February 13Â–15; Fort Collins, CO. Proceedings RMRS-P-54. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 125-129

Moffat, A. J., Davies, S., and Finr, L. 2008. Reporting the results of forest monitoring: an evaluation of the European forest monitoring programme. *Forestry* 81:75-90.

Ni, C. and Zhang, L. 2008. An estimator of prediction error variance for projection equations. *Forest Science* 54:569-578.

Patterson, P. L; and G. A. Reams. 2005. Combining panels for forest inventory and analysis estimation. pp. 69-74 *In*: W.A. Bechtold and P.L. Patterson, eds. The enhanced forest inventory and analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.

Pollard, J. E.; J. A. Westfall, P. L. Patterson, D. L Gartner, M. Hansen, O. Kuegler. 2006. Forest Inventory and Analysis National Data Quality Assessment Report for 2000 to 2003 Gen. Tech. Rep. RMRS-GTR-181. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 43 pp.

Reams, G. A. and McCollum, J. M. 2000. The use of multiple imputation in the southern annual forest inventory system. *In* M. Hansen and T. Burk (eds.), Integrated tools for natural resources inventories in the 21st century., General Technical Report GTR-NC-212, pp. 228-233.

Reams, G. A.; W. D Smith; M. H. Hansen; W. A. Bechtold; F. A. Roesch; and G. G. Moisen. 2005. pp. 21-36 *In*: W.A. Bechtold and P.L. Patterson, eds. The enhanced forest inventory and analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.

Roesch, F. A. 1999. Mixed estimation for a forest survey sample design, pp. 1-6. In 1999 Proceedings of the Section on Statistics and the Environment, American Statistical Association, Presented at the Joint Statistical Meetings Baltimore, MD August 8-12, 1999. 148 p.

Roesch, F. A. 2007. Compatible estimators of the components of change for a rotating panel forest inventory design. *Forest Science* 53(1):50-61.

Roesch. F. A. 2008 Spatial-Temporal Models for Improved County-Level Annual Estimates. *In*: McWilliams, Will; Moisen, Gretchen; Czaplewski, Raymond, comps. 2008. Forest Inventory and Analysis (FIA) Symposium 2008; October 21-23, 2008; Park City, UT. Proc. RMRS-P-56CD, pp. 1-20. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Roesch, F.A. and G.A. Reams, 1999. Analytical alternatives for an annual inventory system. *Journal of Forestry*. 97: 33-37.

Särndal, C., B. Swensson, and J. Wretman. 1992. Model Assisted Survey Sampling. Springer, New York. 694pp.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6:461-464.

Scott, C.T. and M. Köhl. 1993. Experiences with designing the Forest Survey of the Northeastern U.S. Proc. Statistical Methods, Mathematics and Computers. Köhl, M. and G.Z. Gertner, ed., Berlin/Eberswalde, Germany. Swiss Federal Institute for Forest, Snow and Landscape Research, Birmensdorf, Switzerland. pp. 139-148

Scott, C. T.; W. A. Bechtold; G. A. Reams; W. D. Smith; J. A. Westfall; M. H. Hansen; and G. G Moisen. 2005. Sample-based estimators used by the forest inventory and analysis national information management system. pp. 43-67 *In*: W.A. Bechtold and P.L. Patterson, eds. The enhanced forest inventory and analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.

Theil, H. 1971. Principles of econometrics. Wiley, New York

Tian, X. and Xie, Z. 2008. A land surface soil moisture data assimilation framework in consideration of the model subgrid-scale heterogeneity and soil water thawing and freezing. S*cience in China Series D: Earth Sciences* 51:992-1000.

U.S. Forest Service . 2007a. Forest Inventory and Analysis Strategic Plan: A History of Success, a Dynamic Future. FS-865. Washington, DC: U.S. Forest Service . 17 pp.

U.S. Forest Service . 2007b. Interior West Forest Inventory and Analysis Field Procedures, Version 3.01. Ogden, UT: U.S. Forest Service , Rocky Mountain Research Station.

Van Den Brakel, J. A. and Visser, H. 1996. The influence of environmental conditions on tree-ring series of Norway spruce for different canopy and vitality classes. *Forest Science* 42:206-219.

Van Deusen, P. C. 1987. Testing for stand dynamics effects on red spruce growth trends. *Canadian Journal of Forest Research* 17:1487-1495

Van Deusen, P. C. 1989. A model-based approach to tree ring analysis. *Biometrics* 45:763-779.

Van Deusen, P. C. 1997. Annual forest inventory statistical concepts with emphasis on multiple imputation. *Can. J. For. Res.* 27:379-384.

Van Deusen, P. C. 1999. Modeling trends with annual survey data. *Canadian Journal of Forest Research* 29:1824-1828.

Van Deusen, P.C. 2002. Comparison of some annual forest inventory estimators. *Canadian Journal of Forest Research*. 32: 1992-1995.

Van Deusen, P.C. 2008. Detecting trend in annual inventory data with a mixed estimator. *In*: McWilliams, Will; Moisen, Gretchen; Czaplewski, Raymond, comps. 2008. Forest Inventory and Analysis (FIA) Symposium 2008; October 21-23, 2008; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Visser, H. and Molenaar, J. 1992. Estimating trends and stochastic response functions in dendroecology with an application to fir decline. *Forest Science* pp. 221-234.

Walters, D. K., Burkhart, H. E., Reynolds, M. R., and Gregoire, T. G. 1991. A Kalman filter approach to localizing height-age equations. *Forest Science* pp. 1526-1537.

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., and Kurpius, M. R. An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology* 11:89-105

Wulder, M., Dymond, C., White, J., Leckie, D., Carroll, A. 2005. Surveying mountain pine beetle damage of forests: A review of remote sensing opportunities. *Forest Ecology and Management* 221(1–3):27-41.

Wulder, M. A., White, J. C., Bentz, B. J., Alvarez, M. F., and Coops, N. C. 2006a. Estimating the probability of mountain pine beetle red-attack damage. *Remote Sensing of Environment* 101:150-166.

Wulder, M. A., White, J. C., Bentz, B. J., and Ebata, T. 2006b. Augmenting the existing survey hierarchy for mountain pine beetle red-attack damage with satellite remotely sensed data. *Forestry Chronicle* 82:187-202.

Spatial-Temporal Models for Improved County-Level Annual Estimates

Francis A. Roesch[1]

Abstract: *The consumers of data derived from extensive forest inventories often seek annual estimates at a finer spatial scale than that which the inventory was designed to provide. This paper discusses a few model-based and model-assisted estimators to consider for county level attributes that can be applied when the sample would otherwise be inadequate for producing low-variance estimates in the smaller counties. I present and demonstrate simple spatial and/or temporal estimators that draw strength from neighboring counties and/or years in order to increase confidence in the county level annual estimates. The spatial estimators are restricted to those that do not require knowledge of exact plot locations in order to enable their use with privacy protected, publicly available data. A series of simulations is used to compare and contrast the performance of these estimators relative to position in the time series of interest under various variance prescriptions. Although none of the estimators is shown to be superior in terms of minimum mean squared error (MSE) overall, a few general conclusions are drawn. The first is that estimators that draw strength through consecutive measurements of the same set of field plots show a significant reduction in MSE under a wider variety of circumstances than those that draw strength from plots in neighboring counties. The second conclusion is that of the estimators that rely on a temporal model, a simple, centralized weight-adjusted moving average (with weights specific to time-series position) often was the most robust.*

Keywords: Small-area estimation, weighted moving average, mixed estimation, forest inventory.

Introduction

The consumers of publicly available data from extensive forest inventories, such as the one conducted by the USDA Forest Service's Forest Inventory and Analysis (FIA) program, often express a desire for annual estimates at a finer spatial scale than that which the inventory was designed to provide. For example, many users want estimates at the county level even though the sample is inadequate for producing low-variance sample-based estimates of many variables in the smaller counties. The effort can be complicated when the relative sample plot locations are masked in order to protect landowner privacy, such as they have been with FIA data. FIA developed a "fuzzing and swapping" procedure to prevent disclosure of any information that would link individual landowners to specific inventory plot information. During "fuzzing," the reported geographic locations of the plots are randomly perturbed by up to 805 m. (Lister et al., 2005). During "swapping," plot data are exchanged by location between plots of similar characteristics. These fuzzed-and-swapped locations are then published in the Forest Inventory and Analysis Database (FIADB). Because FIA cannot release the exact "swapping" rules, the procedure effectively reduces the reliability of the spatial locations of the plots to a county scale. I present and demonstrate simple spatial and/or temporal estimators that draw strength through

[1] Mathematical Statistician, USDA Forest Service, Southern Research Station, 200 WT Weaver Boulevard, Asheville, NC, 28804-3454. E-mail: FRoesch@fs.fed.us.

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

design-based models that relate to neighboring counties and/or years in order to increase confidence in the county level estimates arising from these privacy protected, publicly available data.

The paper progresses through a series of three simulations; each successive simulation delves deeper into areas suggested by the previous simulation. In Simulation 1, given the initial annual county estimates of a set of variables of interest (e.g. basal area per acre, cubic volume per acre, etc.), I start with a few simple models that could draw estimation strength from the same variable measured in neighboring counties and/or years, as well as a model relating the variable of interest to a common concomitant variable measured at the same time and place, in an attempt to increase confidence in the resulting county level estimates. In Simulation 2, I narrow the focus to circumstances affecting the relative merits of drawing strength spatially and drawing strength temporally, while at the same time introducing a model-unbiased centralized moving average into the comparison. In Simulation 3, I modify the moving average estimator in an attempt to improve estimates at the extremes of the time series and introduce a mixed estimator (e.g. see Van Deusen, 1999 or Roesch, 2007) to frame the spatial and temporal models. Conclusions are then drawn on the comparative results of the three simulations.

Initial Models

Let i=1, . . . , I index the I counties.
Let t=1, . . . , T index T discrete time points.

Let $Y_{i,t}$ denote the response variable for county i at time t. Let $\mathbf{Y}_t = \left(Y_{1,t}\ , \ldots ,\ Y_{I,t}\right)'$ denote the column vector of the response variable for all counties at time t. Further, let $X_{0,i,t} \equiv 1$, and let $X_{k,i,t}$ denote the kth independent variable at county i and time t, for k=1, . . ., K. Then let $\mathbf{X}_{i,t} = \left(X_{0,i,t}\ , \ldots ,\ X_{K,i,t}\right)$ denote the row vector of independent variables for county i at time t. Concatenate the $\mathbf{X}_{i,t}$ into a matrix $\mathbf{X}_t$ of I rows, one for each county at time t.

Naive Model:

First, I define the naive estimator of the county means based on concomitant variables measured in the county and year of interest. The estimator is naive because it does not draw strength from nearby counties or time periods. I use the single concomitant variable of basal area per acre, leading to a simple model for the county means at time t:

$\mathbf{Y}_t = \hat{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_t + \hat{\mathbf{e}}_t$.

where $\hat{\mathbf{X}}_t$ is a special case of $\mathbf{X}_t$ with two columns by I rows in which K=1, and $X_{1,i,t} = BA_{i,t}$, i.e. $\mathbf{X}_{i,t} = \left(X_{0,i,t}, BA_{i,t}\right)$, $\hat{\boldsymbol{\beta}}_t$ is a 2-row vector of estimated parameters and $\hat{\mathbf{e}}_t$ is an I-row vector of $N\left(0,\sigma^2\right)$ error terms.

Spatial Model:

For a given time *t*, assume the county response variable follows a Markov random field under the following spatial neighborhood structure:

$C_i^{(1)}$ = the set of all county indices other than county *i* such that the county centroid is within d_1 km of the centroid of county *i*.

$C_i^{(2)}$ = the set of all county indices such that the centroid of the county is greater than d_1 km from the centroid of county *i*, and less than or equal to d_2 km from the centroid of county *i*.

$C_i^{(3)}$ = the set of all county indices such that the centroid of the county is greater than d_2 km from the centroid of county *i*, and less than or equal to d_3 km from the centroid of county *i*.

Define $\overline{Y}_{C_i^m,t}$ as the mean of all $Y_{i,t}$ in which *i* is found in $C_i^{(m)}$.

Let $C_i^l = \left\{C_i^{(1)}, C_i^{(2)}, C_i^{(3)}\right\}$. Given this structure, assume that all spatial support for county *i* is represented in set C_i^l,

A simple temporally specific spatial model is then:

$$\mathbf{Y}_t = \breve{\mathbf{X}}_t\breve{\boldsymbol{\beta}}_t + \breve{\mathbf{e}}_t, \qquad [1]$$

where $\breve{\mathbf{X}}_t$ is a special case of $\mathbf{X}_t$ with 4 columns by *I* rows and $\mathbf{X}_{i,t} = \left(X_{0,i,t}, \overline{Y}_{C_i^{(1)},t}, \overline{Y}_{C_i^{(2)},t}, \overline{Y}_{C_i^{(3)},t}\right)$, $\breve{\boldsymbol{\beta}}_t$ is a 4-row vector of estimated parameters and $\breve{\mathbf{e}}_t$ is an *I*-row vector of $N\left(0,\sigma^2\right)$ error terms.

Temporal Model:

For $\left\{\mathbf{Y}_{S/2+1}, \ldots, \mathbf{Y}_{T-S/2}\right\}$, $S+1 \le T$, assume the conditional distribution of $\mathbf{Y}_t$ given other time periods depends on the nearest *S* time points. Denote this assumption as:

$p\left(\mathbf{Y}_t \middle| \mathbf{Y}_r : r = 1, \ldots, T\right) = p\left(\mathbf{Y}_t \middle| \mathbf{Y}_{t^S} : t^S = t-(S/2), \ldots, t-1, t+1, \ldots, t+(S/2)\right)$, for $t = \left(S/2+1\right), \ldots, \left(T-S/2\right)$.

A simple temporal model would then be:

$$\mathbf{Y}_t = \tilde{\mathbf{X}}_t\tilde{\boldsymbol{\beta}}_t + \tilde{\mathbf{e}}_t \text{; t=1+}S\text{/2, . . . , }T\text{-}S\text{/2, }S\text{ even.} \qquad [2]$$

where $\tilde{\mathbf{X}}_t$ is a special case of $\mathbf{X}_t$ with *S* columns by *I* rows and $\tilde{\mathbf{X}}_{i,t} = \left(Y_{i,t-S/2}, \ldots, Y_{i,t-1}, Y_{i,t+1}, \ldots, Y_{i,t+S/2}\right)$, $\tilde{\boldsymbol{\beta}}_t$ is a *S*-row vector of estimated parameters and $\tilde{\mathbf{e}}_t$ is an *I*-row vector of $N\left(0,\sigma^2\right)$ error terms.

Spatial-Temporal Model:

Simply combining the spatial [1] and temporal models [2] above leads to a spatial-temporal model for *t*=1+*S*/2, . . . ,*T*-*S*/2:

$$\mathbf{Y}_t = \left[\breve{\mathbf{X}}_t \sim \tilde{\mathbf{X}}_t \right] \begin{bmatrix} \breve{\boldsymbol{\beta}}_t \\ \tilde{\boldsymbol{\beta}}_t \end{bmatrix} + \left[\breve{\mathbf{e}}_t + \tilde{\mathbf{e}}_t \right]; \; t=1+S/2,...,T-S/2, \; S \text{ even.} \qquad [3]$$

The full model with concomitant variables:

$$\mathbf{Y}_t = \left[\breve{\mathbf{X}}_t \sim \tilde{\mathbf{X}}_t \sim \ddot{\mathbf{X}}_t \right] \begin{bmatrix} \breve{\boldsymbol{\beta}}_t \\ \tilde{\boldsymbol{\beta}}_t \\ \ddot{\boldsymbol{\beta}}_t \end{bmatrix} + \left[\breve{\mathbf{e}}_t + \tilde{\mathbf{e}}_t + \ddot{\mathbf{e}}_t \right]; \; t=1+S/2,...,T-S/2, \; S \text{ even.} \qquad [4]$$

where $\ddot{\mathbf{X}}_t$ is a special case of $\mathbf{X}_t$ with 1 column and I rows in which $\mathbf{X}_{i,t} = BA_{i,t}$, $\ddot{\boldsymbol{\beta}}_t$ is a scalar (1x1 vector) estimated parameter and $\ddot{\mathbf{e}}_t$ is an I-row vector of $N\left(0, \sigma^2\right)$ error terms.

Simulations

A series of simulation demonstrations is given using a simulated population covering a 23-year span, based on FIADB data from five states in the southeastern United States (Georgia, Alabama, Florida, North Carolina, and South Carolina.) Note that no general conclusions should be drawn for this particular five state area over these 23 years. This is simply an artificial population that is intended to approximate a real population. Specific details about the construction of the population from these data may be obtained from the author.

Simulation 1 – Initial Comparison of models on county data:

In this simulation, I acknowledge the data are a sample and use bootstrap simulations (e.g. see Efron and Tibshirani, 1998) in an exploratory comparison of the estimators. For completeness, I will augment the above models when necessary to form an estimator for all years in a particular period of interest. For all of the temporally dependent models [2] ,[3], and [4], I will set S=4, and therefore the nearest points to t would be t-2,t-1,t+1, t+2. This means that we would not have estimates for years 1, 2, T-1, and T. For all three models for years 2 and T-1, I'll set S=2. For model [3] and [4] for the first and final years, I will set S=0, virtually eliminating the temporal part of the model. In model [2] for the first and last year, I use the naive estimator.

In the spatial model [1], I used six sets of values for the spatial neighborhood structure with d_1, d_2, and d_3 defined as:

Set	d_1	d_2	d_3
1	40	60	70
2	60	90	105
3	80	120	140
4	100	150	170
5	120	180	210
6	140	210	245

I took 1000 bootstrap samples of the counties for the estimation of all five models. All of the estimators showed very small bootstrap estimates of bias, as expected. I used the mean of the squared error of 1000 bootstrap samples as an approximation to MSE. In the interest of brevity, I give the results for only three of the sets, in figures S1_1 through S1_3. The general conclusion from this simulation is that, for cubic foot volume (a very general and well-observed variable), the spatial model is not very helpful at the county level. In most instances, the estimated MSEs were slightly higher and never much lower than the naive model. However, the simple temporal model, as well as the spatial-temporal and full models, shows a significant reduction in estimated MSE over the naive model and the spatial model.

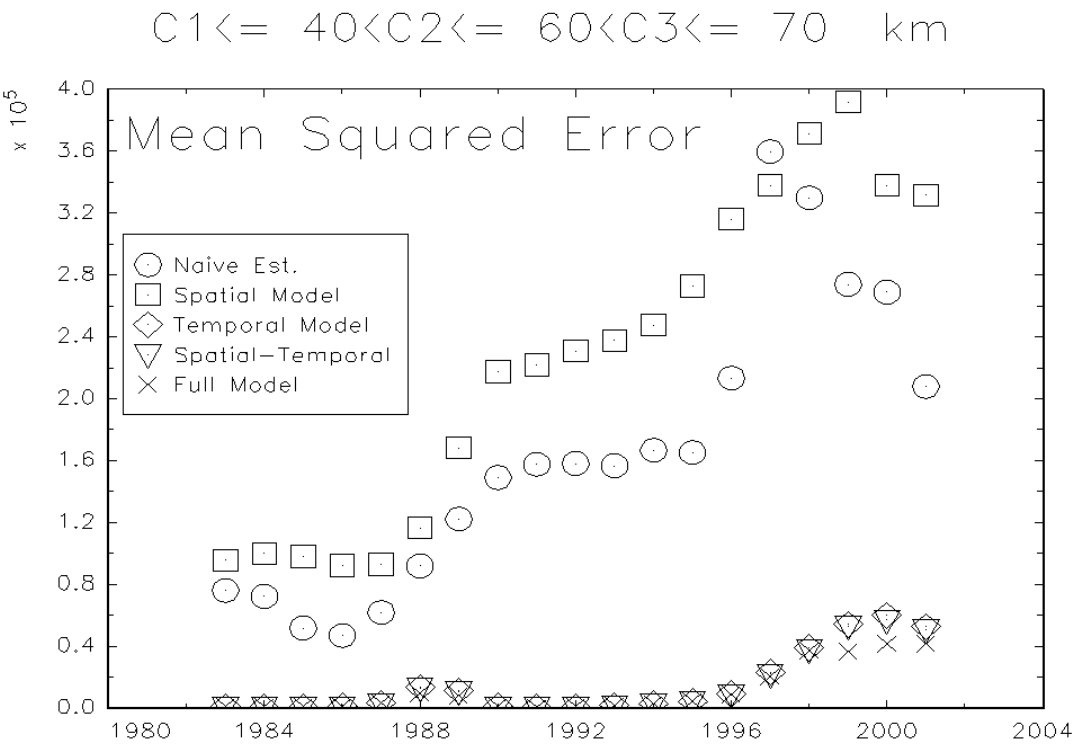


Figure S1_1: The mean squared errors $(ft^3/acre)^2$ calculated for each of the models from 1000 bootstrap samples. This graph gives the results for the spatial triple $\{d_1=40, d_2=60, d_3=70\}$ in all models that include a spatial component.

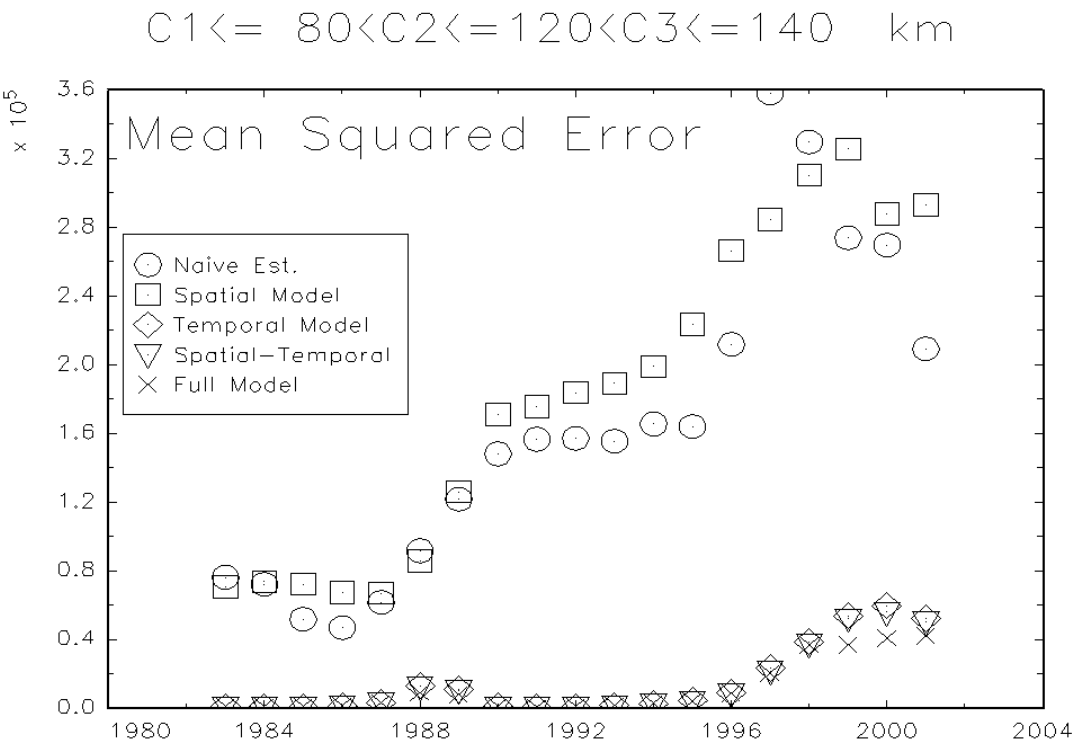


Figure S1_2: The mean squared errors $(ft^3/acre)^2$ calculated for each of the models from 1000 bootstrap samples. This graph gives the results for the spatial triple $\{d_1=80, d_2=120, d_3=140\}$ in all models that include a spatial component.

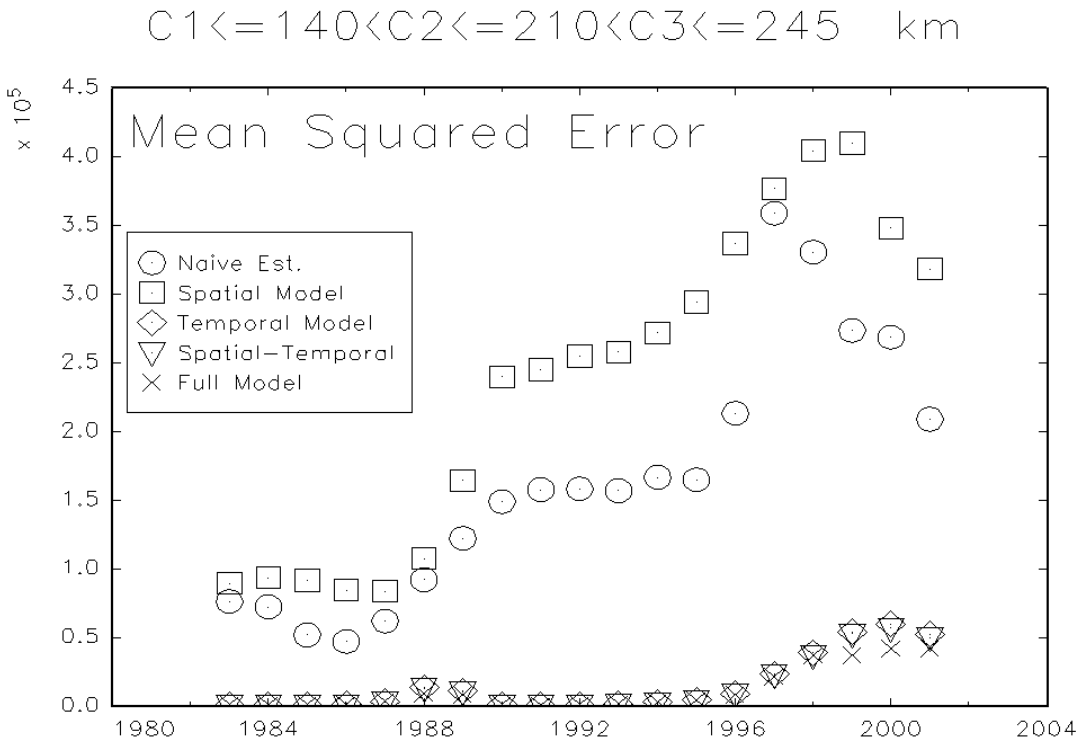


Figure S1_3: The mean squared errors $(ft^3/acre)^2$ calculated for each of the models from 1000 bootstrap samples. This graph gives the results for the spatial triple $\{d_1=140, d_2=210, d_3=245\}$ in all models that include a spatial component.

Simulation 2:

The results for Simulation 1 were demonstrative rather than revealing; cubic volume per acre at the county level should be estimated at a very low variance through this sample design. Users of FIA data often want county level estimates for variables of a much higher variance than total cubic foot volume per acre, such as county level estimates of cubic foot volume of bottomland hardwoods within 200 feet of a stream center. Therefore, I devised a second simulation that should span a realistic range of potential sampling errors for commonly desired estimates. To do this, I treat the county values of cubic foot volume per acre as a seed population observed over the 23-year span. I then simulate variables of increasing total sampling error by drawing a random standard normal deviate (N_{dev}) for each value at each year for 1000 iterations, and then scaling the random deviates by 6 factors (n_{fac}= 0.1, 0.2, 0.4, 0.6, 1.0, and 2.0). Therefore, in order to simulate variables of increasing variance, the sample value for county *i* at time *t* for each iteration $(\tilde{x}_{it})$ was generated from the population value (x_{it}) as $\tilde{x}_{it} = x_{it} + \left(N_{dev} * x_{it} * n_{fac}\right)$. Resulting negative values of $\tilde{x}_{it}$ were set to zero, thereby truncating the error distribution. For this simulation, I calculated the sample mean $(\tilde{x}_{it})$, the estimate from the spatial model, the estimate from the temporal model, the estimate from the spatial-temporal model, and a 5-year

centralized moving average. Neither the naive model nor the full model was included in this simulation, both to increase clarity and because the concomitant variable did not provide enough benefit in the first simulation to justify the increase in complexity for this simulation. For the spatial model, I used the third of the six sets of $\{d_1, d_2, d_3\}$ triples for Simulation 1 above, i.e., $\{d_1=80, d_2=120, d_3=140\}$.

The centralized moving average is assumed to give an estimate for the variable of interest during the year at the center of the interval. Therefore, under this assumption, a 5-year period cannot be used for the first two and final two years. This could lead one to alternative methods of estimating the variable for the first two and final two years. In Simulation 2, I use the 5-year moving average for years 3 to 21, a 3-year average for years 2 and 22 and the annual mean for years 1 and 23. In Simulation 3, I will describe and use an alternative approach. For a thorough understanding of the moving average and its potential interpretations, I refer the reader to Roesch et al. (2003).

Results of Simulation 2: Because all of these estimators have little or no theoretical bias, the simulation results are given solely in terms of MSE in Figures S2_1, S2_2, and S2_3, in which n_{fac} equals 0.2, 0.4, and 2.0, respectively. Results for the other values of n_{fac} are omitted, for the sake of brevity. The graphs show that an advantage of the spatial model, which uses the strength from adjacent counties, increases with increasing simulated sampling error. In Figure S2_1, with little sampling error, the temporal estimators (including the moving average) show MSEs that are slightly better than the simple county mean, and much better than the spatial model. In Figure S2_2, the temporal estimators show MSEs that are much lower than the simple county mean and the spatial model. The value of gathering strength from adjacent counties steadily increases as the sampling error increases and we progress through the graphs to S2_3. I also note that the simple centralized moving average performs quite well except at the extremes of the temporal series, where fewer years are contributing to the estimator. I will show the results of a simple adjustment to the moving average estimator for estimates of the extreme temporal values in Simulation 3.

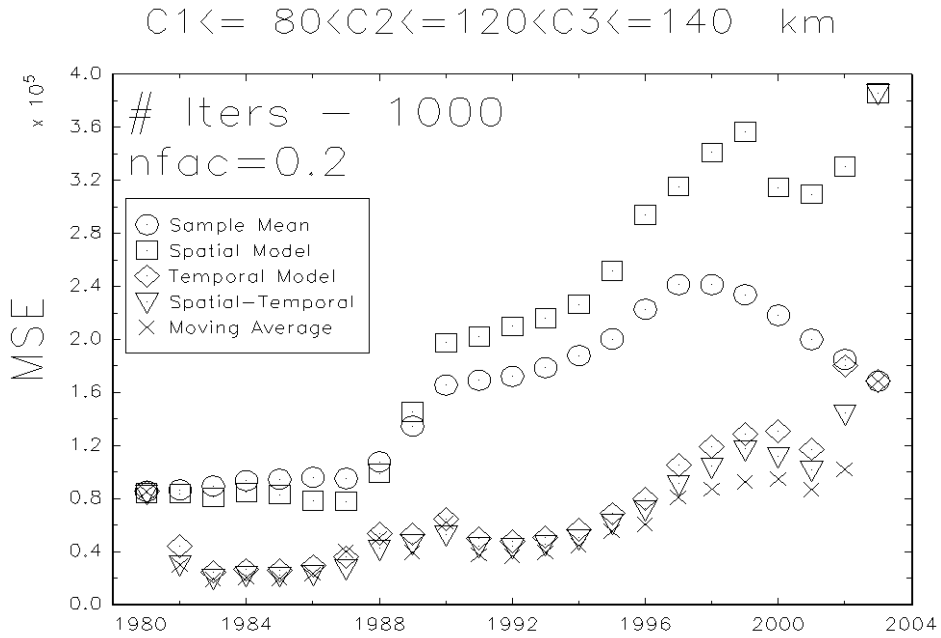


Figure S2_1: The annual Mean Squared Errors $(ft^3/acre)^2$ for each of the models defined in the text, achieved after a 1000-iteration simulation with n_{fac} equal to 0.2.

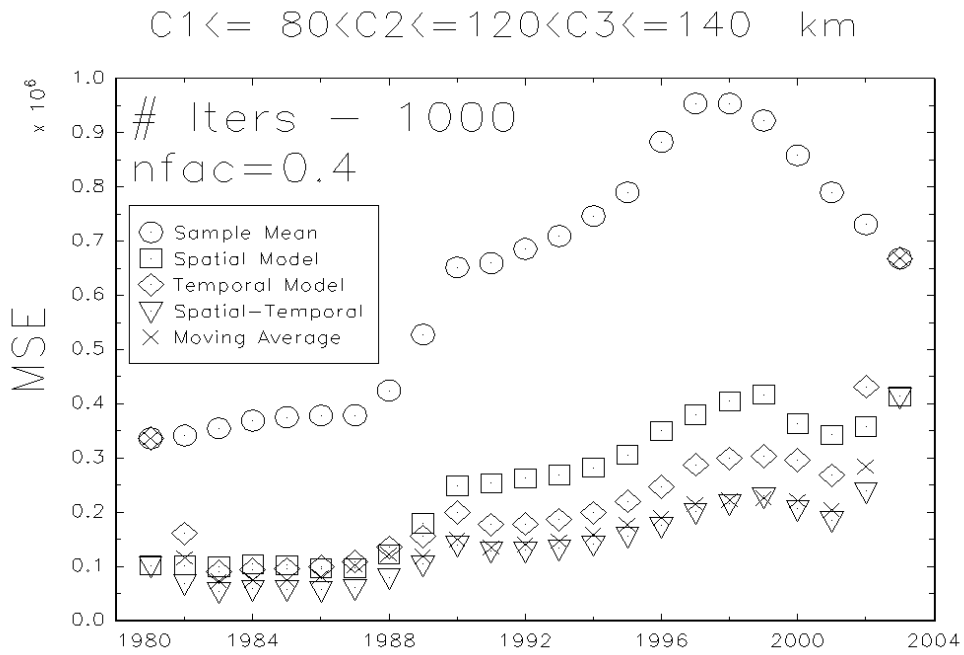


Figure S2_2: The annual Mean Squared Errors $(ft^3/acre)^2$ for each of the models defined in the text, achieved after a 1000-iteration simulation with n_{fac} equal to 0.4.

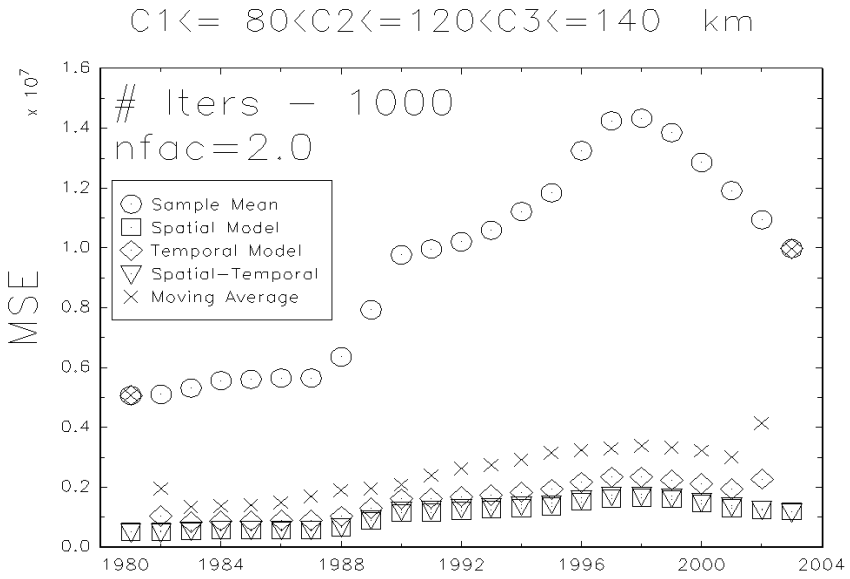


Figure S2_3: The annual Mean Squared Errors $(ft^3/acre)^2$ for each of the models defined in the text, achieved after a 1000-iteration simulation with n_{fac} equal to 2.0.

Simulation 3:

Recall that in Simulation 2, I addressed the desire of users of FIA data for county level estimates of variables that might have a high variance, due, in part, to the infrequency of their occurrence. Of course, even an estimate of total cubic volume can become a high variance variable in rarely occurring condition classes. I devised a third simulation to bring this problem further into context and show some potential mitigation approaches. That simulation is similar to Simulation 2 except that estimates were made within subsets of the population corresponding to six commonly recognized condition classes: publicly owned softwoods (CCA), publicly owned hardwoods (CCB), privately owned hardwoods (CCC), privately owned naturally regenerated southern yellow pine (CCD), privately owned plantations of southern yellow pine (CCE), and privately owned softwoods (CCF). Table S3_1 shows the number of plot locations within the population's 438 counties that were classified within each of the six condition classes during each year.

Note that these condition classes are not mutually exclusive; rather they are intended to represent varying levels of difficulty in estimation due to their frequency within this particular population. Because the population was derived directly from observed data, I assume that I have captured a fair representation of the spatial aggregation of these condition classes. Again, the intention is to span a realistic range of potential sampling errors by treating the county values of cubic foot volume per acre as a seed population observed over the 23-year span. As in Simulation 2, I drew a random standard normal deviate (N_{dev}) for each value at each year for 1000 iterations, and then scaled the deviates by 2 of the 6 factors used previously (n_{fac}= 0.2, and 2.0). Again, the sample value for county i at time t for each iteration $\left(\tilde{x}_{it}\right)$ was generated from the population value $\left(x_{it}\right)$ as: $\tilde{x}_{it} = x_{it} + \left(N_{dev} * x_{it} * n_{fac}\right)$. Resulting negative values of $\tilde{x}_{it}$ were set to zero, thereby truncating the error distribution.

Note that setting the sampling simulation error as high as (n_{fac}= 2.0) may seem onerous; however, it represents a very real phenomenon in broad-scale inventories, resulting from the occasional observation of a very high value in conjunction with an increased rarity of occurrence. This effect is represented here by a normal distribution truncated on the left at zero.

Table S3_1: The number of plot locations within the population's 438 counties classified by each of the 6 condition classes during each year defined in Simulation 3.

Year	CCA	CCB	CCC	CCD	CCE	CCF
1	1159	1183	12549	4994	4132	9184
2	1201	1237	12832	5013	4300	9371
3	1196	1378	13244	5034	4291	9413
4	1197	1312	13028	4955	4323	9361
5	1267	1402	13423	4905	4815	9805
6	1356	1569	13667	4790	5185	10060
7	1368	1586	13551	4544	5529	10158
8	1369	1706	13923	4381	5522	9990
9	1316	1786	13539	3978	5504	9574
10	1290	1664	12514	3700	5290	9056
11	1324	1690	12628	3730	5473	9271
12	1250	1614	11915	3501	5291	8852
13	1068	1487	11235	3304	4845	8209
14	897	1399	10646	3098	4255	7413
15	825	1266	9935	2780	3966	6806
16	790	1254	9641	2658	3607	6327
17	773	1251	8774	2425	3414	5902
18	705	1129	7897	2205	3186	5452
19	632	1048	7299	2002	2935	5002
20	576	1000	6740	1872	2752	4697
21	549	855	6263	1693	2567	4323
22	460	666	5129	1461	2184	3697
23	339	518	3781	1105	1637	2783

For each iteration, I calculated the sample mean, a weighted 5-year moving average, a temporal mixed estimator under two constraint models, and a spatial-temporal mixed estimator under the same two constraint models. A brief description of each estimator follows.

Weighted 5-year Moving Average: The weighted 5-year moving average (wMA5) is the same as the centralized 5-year moving average for years 3 to T-2, that is, the weights are equal to 1 for each of the five years in those estimates. It differs in that rather than reducing the number of years contributing to the estimate for the initial two and final two annual estimates in the series, the annual contributions to the initial and final 5-year averages are weighted somewhat arbitrarily in an attempt to compensate for the relative imbalance in these estimates as I apply them to "off center" years. That is indexing wMA5 as $wMA5_t$ for year *t*:

$$wMA5_t = \begin{cases} (3.0x_1 + 1.2\,x_2 + 0.4\,x_3 + 0.2\,x_4 + 0.2\,x_5)/5.0 & \text{for } t = 1, \\ (1.4\,x_1 + 1.6\,x_2 + 1.0\,x_3 + 0.6\,x_4 + 0.4\,x_5)/5.0 & \text{for } t = 2, \\ (x_{t\text{-}2} + x_{t\text{-}1} + x_t + x_{t+1} + x_{t+2})/5.0 & \text{for } t = 3:T\text{-}2, \\ (0.4\,x_{T\text{-}4} + 0.6\,x_{T\text{-}3} + 1.0\,x_{T\text{-}2} + 1.6\,x_{T\text{-}1} + 1.4\,x_T)/5.0 & \text{for } t = T\text{-}1, \quad \text{and} \\ (0.2\,x_{T\text{-}4} + 0.2\,x_{T\text{-}3} + 0.4\,x_{T\text{-}2} + 1.2\,x_{T\text{-}1} + 3.0\,x_T)/5.0. & \text{for } t = T. \end{cases}$$

Mixed Estimation: The mixed estimator can also be used to draw strength from temporally adjacent measurements for annual forest inventory designs. The reader unfamiliar with the use of mixed estimation for annual inventories is referred to Van Deusen (1996, 1999, and 2000) and Roesch (2007). A simple interpretation of the mixed estimator in this context is that one is mixing the strength of one's belief in a set of constraints on a population with the strength of one's belief in a model for the relationship of a set of observations to the population. Here I use mixed estimation to estimate the trends in county means. I apply the mixed estimator to both the temporally specific county sample means and to (temporally specific) spatial "moving-window" county estimates. In both cases, assume an observation model for time *t*:

$$x_t = \beta_t + e_t,$$

where x_t is the county mean of interest, β_t is an unknown random coefficient, and e_t is a zero-mean error term with variance σ_t^2 / n_t. This model is combined with a model describing constraints on the temporally ordered vector of the β_ts ($\boldsymbol{\beta}$). Here, for t=5,6,7,…,T, I use the constrained transition model:

$$\beta_t - 3\beta_{t-1} + 4\beta_{t-2} - 3\beta_{t-3} + \beta_{t-4} = v_t.$$

where v_t is an error term for time *t*. Collect the ordered v_ts into a zero-mean vector **v**, which has variance $p\boldsymbol{\Omega}$. Also, form the temporally ordered vectors $\mathbf{X}$, from the x_ts, and $\mathbf{e}$ from the e_ts. Assume **v** to be a partition of **e**. Our observation model then becomes:

$$\mathbf{X} = \boldsymbol{\beta} + \mathbf{e}$$

Represent the covariance matrix of $\mathbf{X}$ with $\boldsymbol{\Sigma}$. The constraints can be expressed as:

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{v}$$

where **R** is the matrix of constraints for the transition model with (T-4) rows by T columns, in this case.

Combining the models results in the solution set for a mixed estimator:

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{R} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{e} \\ \mathbf{v} \end{bmatrix}.$$

Applying the constraints with strictness moderated by the parameter *p*, the mixed estimator is:

$\hat{\boldsymbol{\beta}} = \left[\boldsymbol{\Sigma}^{-1} + p^{-1}\mathbf{R}'\boldsymbol{\Omega}^{-1}\mathbf{R} \right]^{-1} \boldsymbol{\Sigma}^{-1}\mathbf{X}$ (Van Deusen 1999).

Van Deusen (1999) shows how to optimally select the parameter *p* and a model using maximum likelihood and information criteria. For this simulation, those selections would be difficult to track as well as computer-intensive, so I use a single model and assumed values of *p*. The temporal model used here is a trivial extension of the series of constraint matrices found in equations 9 through 11 in Van Deusen (1999).

As noted above, rather than estimate the parameter *p*, I initially set it equal to 0.1, which corresponds to a high degree of confidence in the transition model. Also, the practice of adjusting the degree of strictness in constraints by adjusting *p* applies the same level of constraint strictness to the entire vector of estimates, at a level chosen by the data. Suppose rather that one is relatively satisfied with the model for most of the annual estimates but has some foreknowledge of problematic occurrences in the population at specific points in time. In that case, it might be better to eliminate constraint rows that span the affected times, rather than re-optimize over both affected and unaffected time periods. I incorporate an example here and dub it the ME-reduced approach. In this population, the constraint matrix for the full model consisted of 19 rows. Examination of the temporal trends in the population revealed constraints that appeared to be inconsistent with those trends. This is equivalent to knowing for instance that certain changes in harvesting pressure had occurred in the past, and those changes are likely to have affected the volume trend in particular ways that differ from the temporal model. For the ME-reduced run, I kept rows 1-5, 10, 14, and 19 to achieve the reduced constraint matrix $\mathbf{R}_{\mathbf{red}}$ **:**

$$\mathbf{R}_{\mathbf{red}} = \begin{bmatrix} 1 & -3 & 4 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -3 & 4 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -3 & 4 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -3 & 4 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -3 & 4 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -3 & 4 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -3 & 4 & -3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -3 & 4 & -3 & 1 \end{bmatrix}.$$

Naturally, the corresponding rows and columns of $\boldsymbol{\Omega}$ must also be eliminated in order to arrive at $\boldsymbol{\Omega}_{\mathbf{red}}$.

A Temporal Mixed Estimator Model: A Temporal Mixed Estimator Model results when the temporally specific county means comprise the **X** matrix in the above development. Both the fully constrained model and the model using the reduced constraint set were evaluated at each iteration.

A Simple Spatial Temporal Mixed Estimator Model: For this simulation, the spatial model was simplified, relative to previous simulations, to a weighted moving-window model. That is, the temporally specific estimate for county *i* is first calculated as the mean of all plots in all counties with county centroids less than or equal to a distance of 140 km from the centroid of county *i*. These means are then the initial estimates, i.e., those comprising the **X** matrix, for the Mixed Estimator described above, resulting in a simple spatial-temporal mixed estimator. As for the temporal model, both the fully constrained model and the model using the reduced constraint set were evaluated at each iteration.

Results of Simulation 3: The area-weighted mean and MSE, respectively, for total cubic foot volume per acre for all land and for each of the condition classes, are shown in S3_1 and S3_2, with n_{fac} set equal to 0.2, and in S3_3 and S3_4 with n_{fac} set equal to 2.0.

The somewhat arbitrary weighting used for the wMA appears to have worked well, allowing that estimator to compare favorably to the ME in all four figures for this simulation.
The most striking result of the output in the first two figures is the relationship between the behaviors of the full-constraint versus the reduced-constraint temporal models. In Figure S3_1, note that the fully constrained temporal model is displaying more bias than the other estimators, although the subsequent figure will show that the fully constrained model had very low MSEs, relative to the other ME models.

In figure S3_2, the wMA estimator is usually lowest or tied for lowest in MSE, across the condition classes, as one might suspect, given the low simulation variance and the more homogenous populations than the full population in the upper right graph of the figure. In that upper right graph, representing the entire land base, the fully constrained temporal mixed estimator more often outperforms the wMA and other ME models. The ME temporal models are generally outperforming the spatial temporal models under these circumstances in terms of MSE. Additionally, the fully constrained models are usually lower in MSE than their reduced constraint counterparts. Therefore, reducing the constraints did serve to encourage less bias in the estimates at the cost of increased variance. This effect is present but less striking in the spatial-temporal models.

The last two figures (S3_3 and S3_4) plot the area-weighted means and MSEs, respectively, for total cubic foot volume per acre for all land and for each of the condition classes, with n_{fac} set equal to 2.0. The realized sample means are shown to be greater than the population means in Figure S3_3. The fully constrained temporal model appears to be the least biased of the estimators, however that appears to be an artifact of the stiffness of the constraints. Figure S3_4 gives the MSEs and shows the advantage of the estimators that reach out to adjacent areas and temporal periods to obtain improved estimates, with the temporal models again appearing to contribute the most to the reduction in MSE.

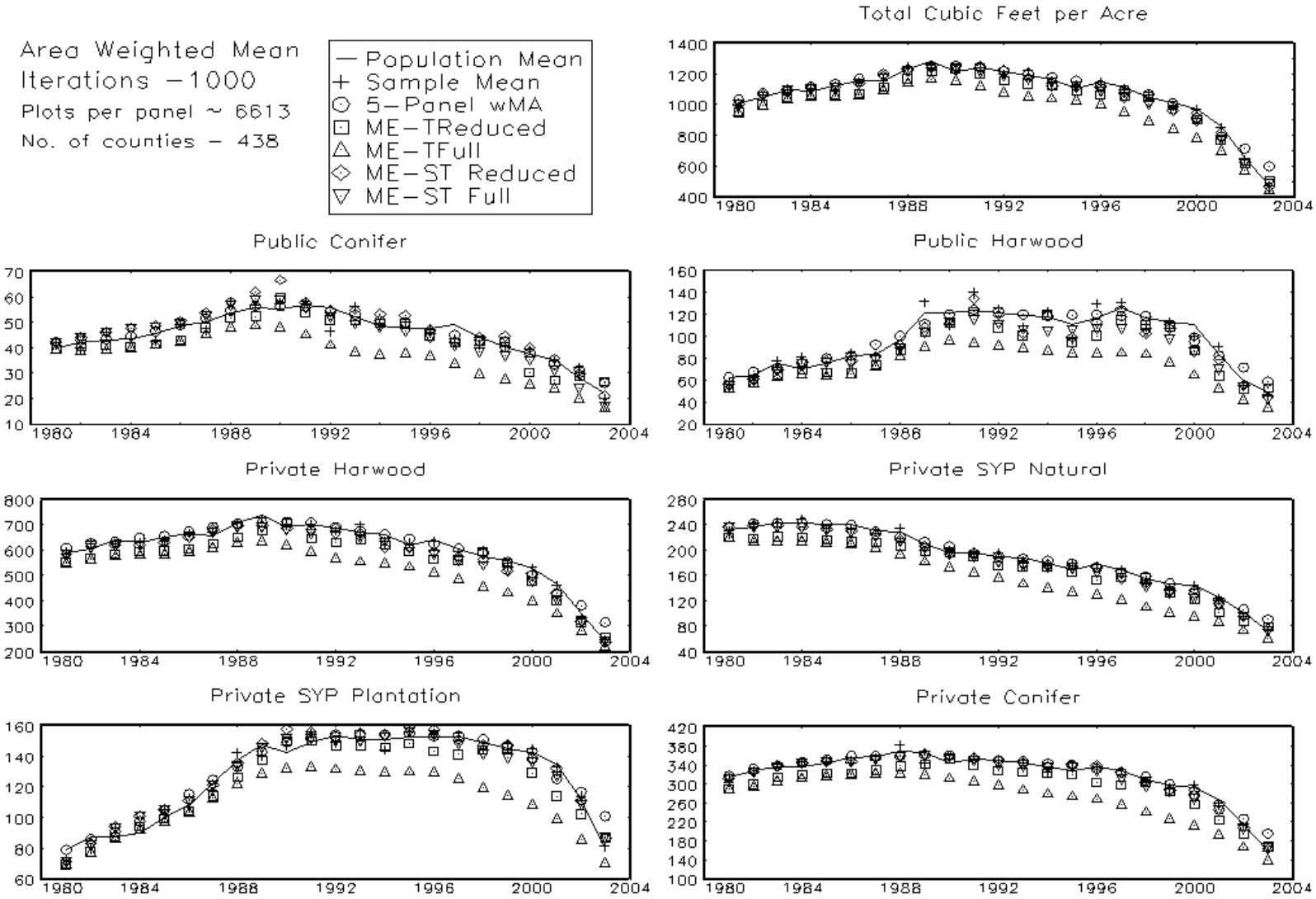


Figure S3_1: The annual area weighted Mean(ft^3/acre) for each of the estimators defined for simulation 3 in the text, achieved after a 1000 iteration simulation with n_{fac} equal to 0.2.

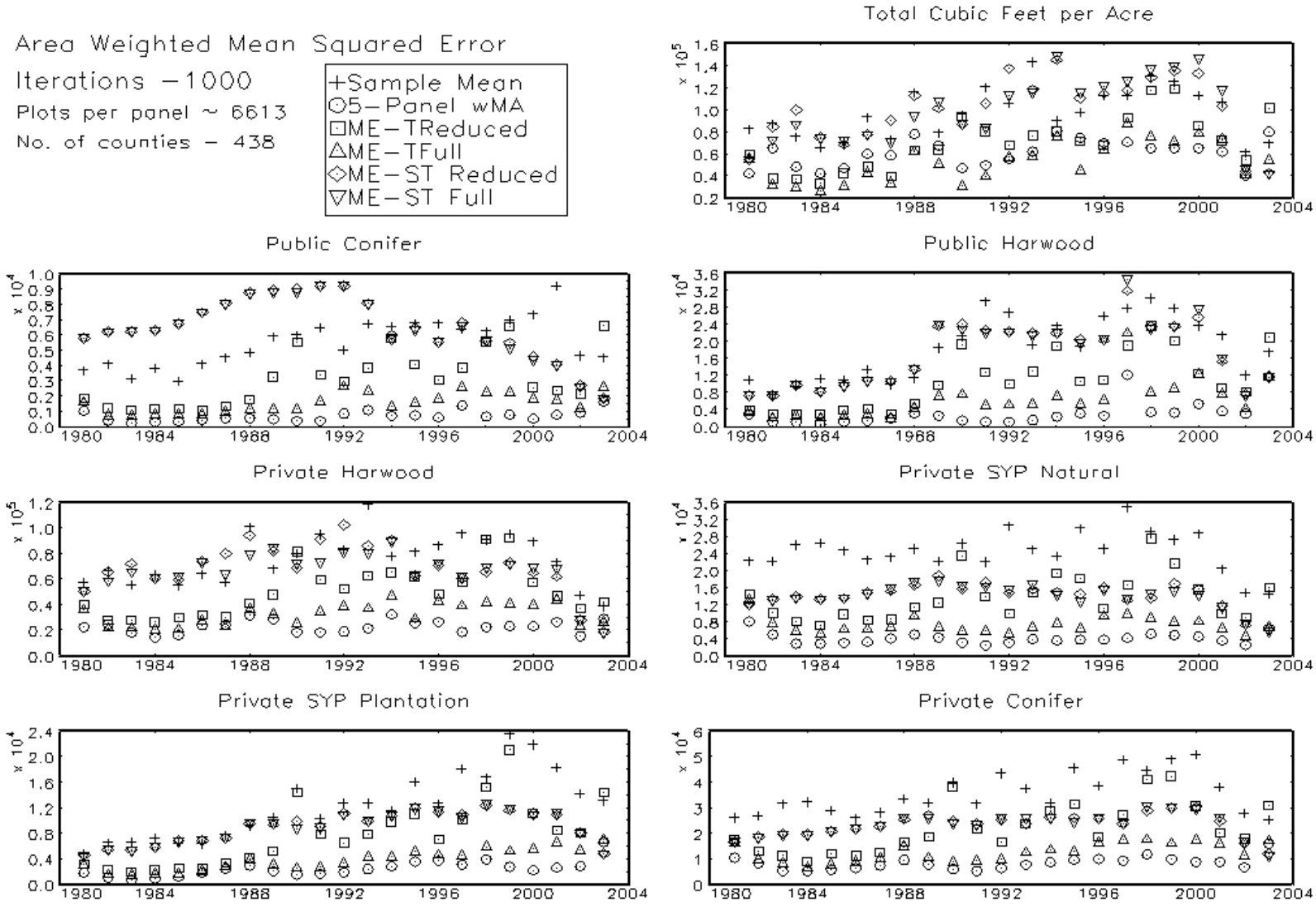


Figure S3_2: The annual area weighted Mean squared Error $(ft^3/acre)^2$ for each of the estimators defined for simulation 3 in the text, achieved after a 1000 iteration simulation with n_{fac} equal to 0.2.

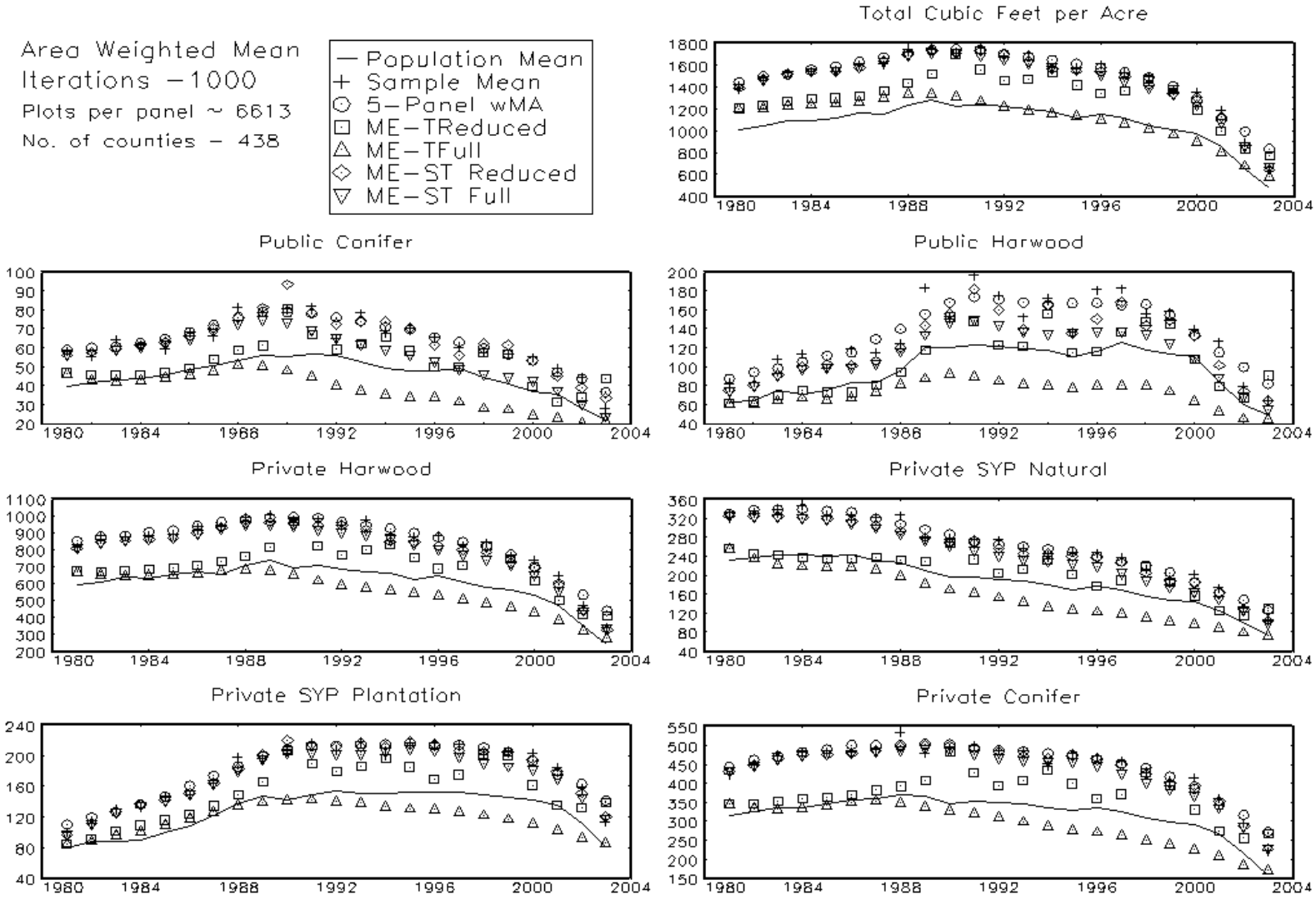


Figure S3_3: The annual area weighted Mean (ft^3/acre) for each of the estimators defined for simulation 3 in the text, achieved after a 1000 iteration simulation with n_{fac} equal to 2.0.

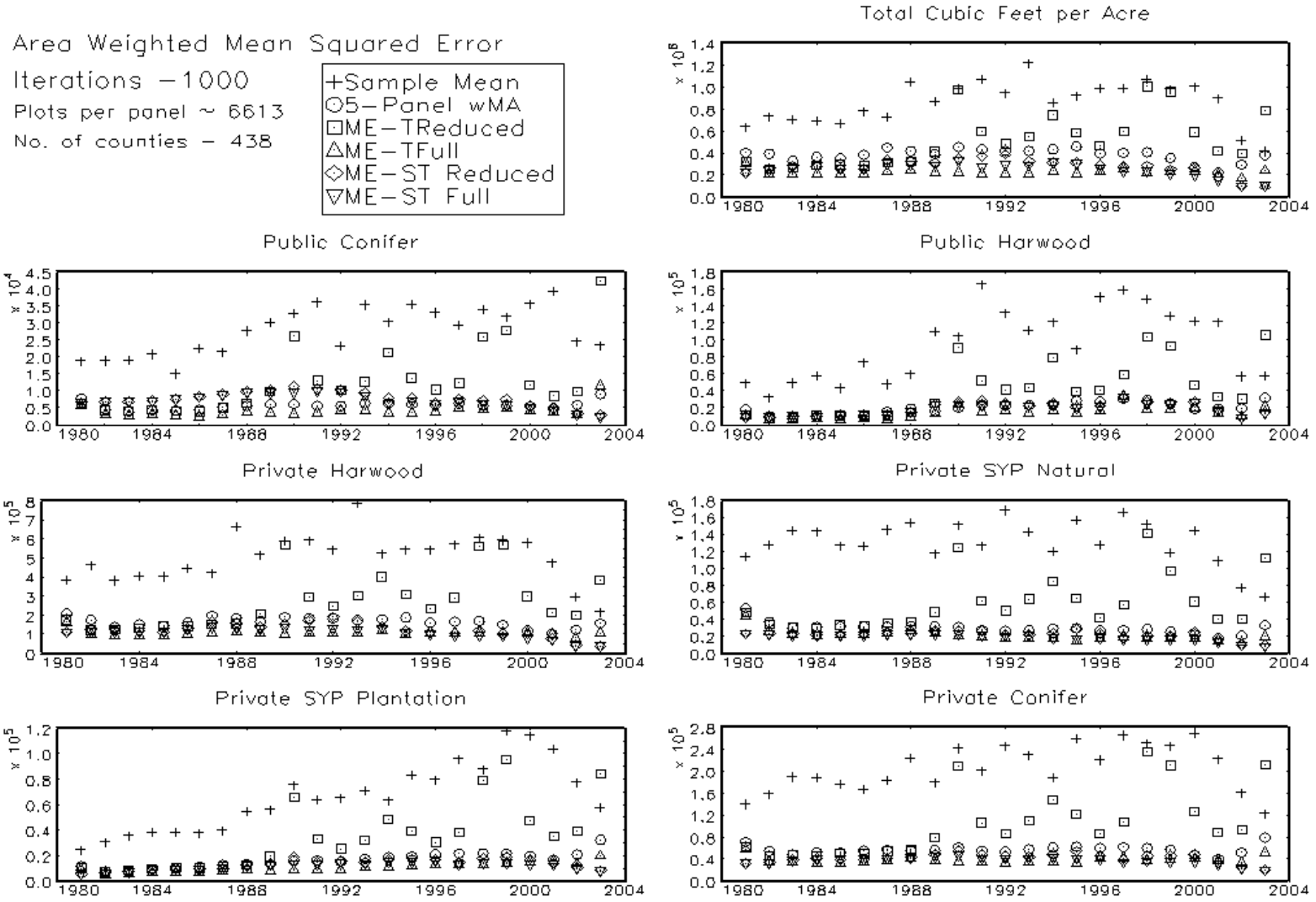


Figure S3_4: The annual area weighted Mean squared Error $(ft^3/acre)^2$ for each of the estimators defined for simulation 3 in the text, achieved after a 1000 iteration simulation with n_{fac} equal to 2.0.

Conclusions

Initially, this study was intentionally general, starting from aggregated county estimates as opposed to individual plot estimates to avoid obfuscating the question being investigated: Under what circumstances does it become advantageous to draw strength by gathering information from areas larger than a county or periods longer than a year to make county level annual estimates? These simulations have shown quite clearly that except for variables that enjoy an extremely low sampling variance, such as general cubic foot volume per acre, estimates of county level variables can benefit greatly from "outside" information. For the variables with a sampling error in the middle of the range investigated, the most benefit seems to come from adjacent years, however, the spatial model using adjacent county data increased in favor relative to the moving average and the other temporal models with increasing sampling error. Additionally, this study did not uncover any cost to using the models that incorporate both space and time, that is the spatial-temporal model and the full model were always low in MSE.

There have been quite a few papers on approaches to incorporating temporal trend for FIA data under the current design, most notably those utilizing a mixed estimator (i.e. Van Deusen 1996, Van Deusen 1999, Van Deusen 2000, Roesch 1999, Roesch 2006, and Roesch 2007). This work lends further credence to those approaches while also suggesting that they should be expanded to include a spatial component for estimating variables with an especially egregious variance structure. The advantage of the mixed estimator over the centralized moving average is its formal treatment of improvement in estimates of the initial and final values in the time series. The wMA also gives improved estimates for these endpoint values. The latter estimator is quite simple, but less formal than the ME. Although one could try to optimize the weights in the wMA, the result would be similar to a specific case of the ME. The methods investigated were restricted to those that could be applied to the privacy protected, publicly available data in FIADB and the spatial model did show an advantage when being applied in high variance circumstances. It is possible, however, that spatial models developed without this restriction would perform better than the coarse-scale models used here.

Acknowledgements:
The final version of this manuscript benefited from helpful reviews by Paul C. Van Deusen, Stanley Zarnoch, and William Burkman. Paul Smith and Kathryne Roesch contributed valuable editorial comments.

Literature Cited

Efron, B. and R.J. Tibshirani, 1998. *An Introduction to the Bootstrap.* New York. Chapman and Hall/CRC. xvi + 436 p.

Lister, A., C. Scott, S. King, M. Hoppus, B. Butler, and D. Griffith. 2005. Strategies for Preserving Owner Privacy in the National Information Management System of the USDA Forest Service's Forest Inventory and Analysis Unit, pp. 163-166 In: McRoberts, Ronald E.; Reams, Gregory A.;

Van Deusen, Paul C.; McWilliams, William H.; Cieszewski, Chris J., eds. Proceedings of the fourth annual forest inventory and analysis symposium; Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 257 p.

Roesch, F. A. 1999. Mixed Estimation for a Forest Survey Sample Design, pp. 1-6. In *1999 Proceedings of the Section on Statistics and the Environmen*, American Statistical Association, Presented at the Joint Statistical Meetings Baltimore, MD August 8-12, 1999. 148 p.

Roesch, F.A., Steinman, J.R., and Thompson, M.T. 2003. Annual Forest Inventory Estimates Based on the Moving Average, pp. 21-30. In: McRoberts, R.E., Reams, G.A., Van Deusen, P.C., Moser, J.W. Proceedings of the third annual forest inventory and analysis symposium; 2001 October 17-19; Traverse City, Michigan. Gen. Tech. Rep. NC-230. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 208 p.

Roesch, F.A. 2006. Continuous Inventories and the Components of Change, pp. 355-362. In: Second International Conference on Forest Measurements and Quantitative Methods and Management & the 2004 Southern Mensurationists Meeting. Cieszweski, C. J. and M. Strub, eds. Proceedings of a conference held June 15-18, 2004. Hot Springs, AR. USA. 412 p.

Roesch, F. A. 2007. Compatible Estimators of the Components of Change for a Rotating Panel Forest Inventory Design. *For. Sci.* 53(1):50-61.

Van Deusen, P.C. 1996. Incorporating predictions into an annual forest inventory. *Can. J. For. Res.* 26:1709-1713.

Van Deusen, P.C. 1999. Modeling trends with annual survey data. *Can. J. For. Res.* 29(12):1824-1828.

Van Deusen, P.C. 2000. Alternative sampling designs and estimators for annual surveys, pp. 192-196. In: *Integrated Tools For Natural Resources Inventories In The 21st Century,* USDA For. Ser. Gen. Tech. Rep. NC-212, Hansen, M. and T. Burk (eds). 744 p.

Predictive Geo-Spatial Modeling



Using Landsat Imagery and FIA Data to Examine Wood Supply Uncertainty

Curtis A. Collins[1] and Ruth C. Seawell[2]

Abstract: *As members of the forest products industry continue to reduce their landholdings, monitoring reliable future timber supplies becomes an increasingly important issue. This issue requires both spatial and forest inventory information to meet the strategic planning needs of these entities. Increased depth in the archival span of imagery available from the Landsat program in conjunction with timber estimates from FIA data have been used by Larson and McGowin, Inc. to fill this information gap in the southern U.S. by further refining timber supply projections.*

In generating this information for various clients, Landsat MSS and TM (including ETM+) data have been used to create forest age and type maps with temporal origins as early as 1972 to as recent as 2007. By using forest type and age layers, spatial products can be created that have matching attributes to those presented by FIA. In doing so, FIA volumes can be adjusted to generate expected total volumes for regions by using the remotely sensed/image processed type and age acreages. Once volumes by type and age classes are established, the SubRegional Timber Supply (SRTS) model can be use to analyze supply/demand relationships between multiple products, endogenous land use, varying growth and yield components and regional and local demand estimates.

Results from a sampling of these projects will be presented in such a way as to share image processed accuracies and overall project results. Special attention will be given to the nature of the results and how our clients have found them helpful. Examples will also be presented showing how the integration of GIS operations can be utilized along with the spatial data product to glean further results that are more widely applicable.

Keywords: Landsat, change detection, SRTS, supply, remote sensing.

Introduction

With the large-scale shift of private forest ownership in the U.S. ongoing, many changes are occurring as public and private entities adapt to answer evolving problems. One obvious problem with the ownership shift is noted in the

[1]Larson and McGowin Forestry Consultants, P.O. Box 230, Watkinsville, Georgia 30677, ccollins@larsonmcgowin.com
[2]Larson and McGowin Forestry Consultants, P.O. Box 2143, Mobile, Alabama 36652, rseawell@larsonmcgowin.com

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

reduction of integrated land holding entities within organizations that consume wood and timber. With no more company lands, procurement personal are quickly adapting to rely heavily on outside sources of material that, in some cases, were marginalized in the past. One developing tool of importance to many of these procurement personnel involves wood basin analysis where regional mill consumption basin(s) are identified and analyzed to yield information for strategic planning purposes.

Use of these techniques and resulting information is not new. While many mills are having reliable company land bases removed from the equation, some have been successfully operating without these holdings for some time. In doing so, these firms relied on forest analysts to monitor supply/demand trends at the multiple scales, not the least of which being the basin from which a mill of interest relied upon for raw material. Uses of FIA data and reports were paramount in performing these analyses, but many users, knowing the limitations of FIA, desired additional information and at finer scales, temporally and spatially, than FIA intended.

One way where FIA data accentuation was explored and achieved involved the most successful space-borne remotes sensing program, the Landsat program, which gives the public access to data from 1972 to the present. Using this large bank of data, FIA results could be updated in longer lag periods between field measurement periods. This data could also be used to accentuate or even replace age class estimates as FIA is a sampling approach that, in smaller areas in particular, may not yield as accurate results as Landsat-based products. Similarly, Landsat data can be used to map broad forest type classes which, again, could be used to crudely update or replace FIA forest type area estimates.

With FIA and Landsat estimates in hand, various approaches can be taken to yield viable supply/demand/consumption information that can be used to analyze regional timber supply situations as well as inform existing timber supply models. The widely adopted SubRegional Timber Supply (SRTS) model is one such model that was developed to examine southern timber markets (Abt *et al.*, 2000), and has been applied both in the southern and northern U.S. to examine timber supply and prices. Over the last five years, SRTS' capability has expanded to include the analysis of multiple products, as well as endogenous land use and the integration of growth and yield components and regional/local demand estimates. The model provides ending inventories, annual removals, and price indices for the analysis of future timber supply scenarios.

Methodology

Landsat Data and Analyses

With the launch of Landsat 1 on July 23, 1972, the U.S. embarked on a program that has and continues to yield continuous and comprehensive earth surface science information for various mapping projects. The data yielded from

this ongoing program includes Multispectral Scanner System (MSS) data with ~60m resolution data in visible green, visible red, near-infrared (NIR), and mid-infrared (MIR) bands; Thematic Mapper (TM) data with ~30m resolution data in visible blue, visible green, visible red, near-infrared (NIR), and two mid-infrared (MIR) bands; and Enhanced Thematic Mapper+ (ETM+) data with the same capabilities as TM data plus a ~15m panchromatic layer covering 0.52-0.9 μm. For the purpose of this paper, these data were used in two analyses: forest age mapping and forest type mapping.

The forest age mapping approach used by Larson and McGowin currently utilizes a post-classification comparison approach (similar to Collins *et al.*, 2005). In this methodology, a current forest/non-forest map, usually derived from unsupervised classification techniques and interpretation, is manipulated in descending years through a temporal dataset of derived "cut" layers until each forested pixel is tagged with a year of origin, unless this origin is beyond the scope of Landsat (earlier than 1972). Similarly, a harvest layer was built by ascending, again using cut layers, from a forest/non-forest layer that was ~10 years old, using the forest class, to the current forest/non-forest layer, using the non-forest class.

The age determination process is designed to detect clearcut areas, not partial harvest or thinning operations. However, if these partial harvests or thins are heavy enough they may trigger an origin set inappropriately. With regards to temporal resolution, datasets are targeted to be acquired from every 1 to 5 years usually throughout the span of Landsat data (usually 1972-1973, depending on data quality) depending on site and growing conditions. It is also worth noting that while leaf-off data, data collected after deciduous leaf senescence has occurred, can be useful in this process; leaf-on data is targeted as it typically better discriminates forest from non-forest. The creation of cut layers from individual Landsat images from the temporal dataset is usually done through independent unsupervised classification or applying image thresholds using a variety of image transforms incorporating individual Landsat bands and/or Normalized Difference Moisture Index (NDMI) and/or Normalized Difference Vegetation Index (NDVI).

The forest type mapping approach uses the same current forest/non-forest layer to mask out non-forest pixels. In the remaining forested areas, a current Landsat leaf-off dataset is acquired. This dataset is next either passed through an unsupervised classification and interpretation routine or interpreted for and classes via NDVI thresholds in order to derive deciduous (usually hardwood), evergreen (usually softwood), and mixed deciduous-evergreen classes.

Timber Supply Analyses with SRTS

Traditionally, USFS FIA periodic inventories have been the data source for the SRTS model. However, because of the increasing age between FIA surveys and

dramatic shifts in harvest patterns, remote sensing-based classification and change detection techniques for forest timber types have been used as a means to update these estimates as well as to refine FIA area assessments. Typically, the area estimates by broad timber types and five-year age classes are used in conjunction with FIA per-acre estimates to derive timber inventories.

While the inventories in the study basins may not necessarily be too old for timber supply assessment, there could exist some conditions that merit an update using satellite imagery. Dramatic shifts in harvest patterns since the FIA survey or damage activities like hurricanes, ice storms, insect infestation, etc., can be detected in a satellite update. Also, permanent losses in timberland due to urbanization can be measured and trends observed. Finally, the satellite image area assessment is oftentimes a more accurate estimate of the area than the FIA estimates.

In the previous versions of the SRTS model, harvests were held constant, sometimes creating unrealistic price projections. In this "harvest" mode the annual harvest levels input by the user are assumed constant and harvest does not react to higher or lower prices. Because of that the model can only adjust the price needed to achieve the desired harvest given the inventory levels thus creating what appear at times unrealistic price trends.

The latest version, MP SRTS, now allows users to specify a demand curve to find the equilibrium solution between supply and price. Demand is modeled at the aggregate level; however, through the inventory shifts by product, region and owner, a solution for equilibrium price simultaneously determines harvest shifts across regions and owners. Demand can either be held constant as in a harvest run or increased through time. The goal program then harvests across management types and age classes to get the projected target mix, while harvesting consistently with historical harvests for the region. The difference is that harvest and price react to the change in demand. The effect of an increasing demand on harvest and price depends on the supply, i.e. inventory. Other things being equal, an increase in demand will raise price and harvest; but harvest will not increase proportionately since the price increase dampens some of the harvest.

Results and Discussion

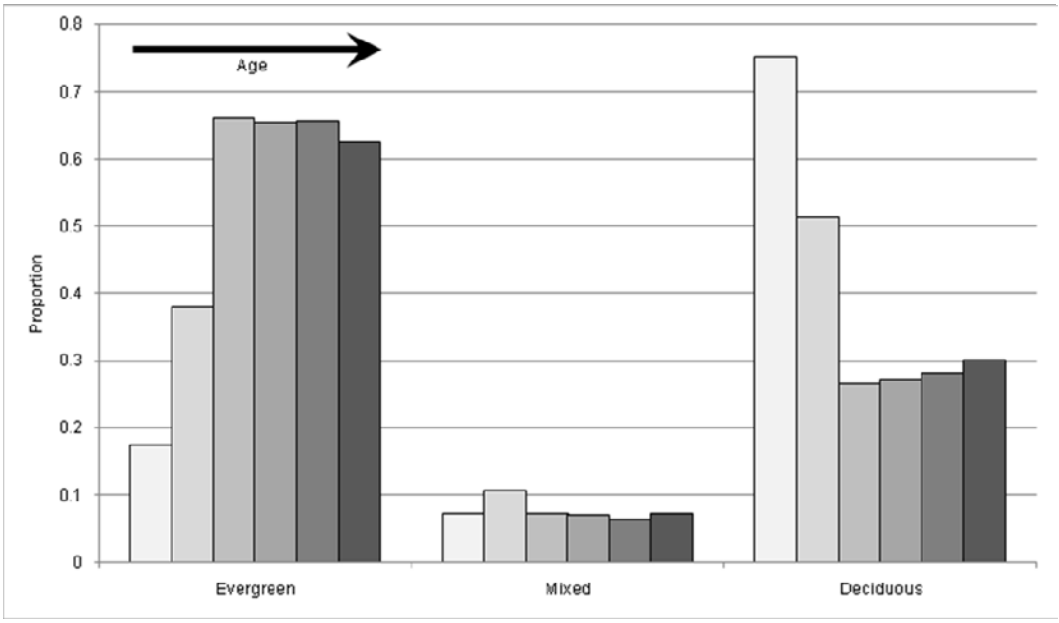
Landsat Data and Analyses

One set of resulting Landsat classifications were compared to client provided stand data in order to get a notion, albeit not definitive, with regard to accuracy for a specific project. The data and client were situated in the southeastern U.S. The client provided stand data came in the form of a polygon vector layer with each delineated stand attributed with management species composition (usually a target management composition, not necessary what was present), age (actually year of establishment), and size (acres). The data was last updated in June 2006

and covered approximately 400,000 acres of land being actively managed for loblolly pine (*Pinus taeda*). After receiving the data, it was cleaned to better match existing GIS stand types, mainly from the management composition field, with the Landsat derived classifications.

In cleaning the data, the management classes were re-classed to better match the deciduous/hardwood, evergreen/pine, and mixed forest classes as well as open (there were some fields and wildlife foodplots) and harvest non-forest classes. In addition to these classes, this cleanup also used observed histogram data to note that rather consistent forest type given age distributions were heavily, and inappropriately, favoring the deciduous type in the youngest age classes (Figure 1). For this reason, young forested areas were stripped of a class designation and assigned as a "regeneration" class. These data were next overlaid on overlapping forest age and type layers, created with current data from 2007, with zonal attributes being calculated so as to indicate the majority type and age of each stand with respect to the two classified products. The observed age attributed to the stands and the zonal extracted ages were then rescaled to 5 year age classes to better match an FIA matching scenario.

Figure 1. The distribution of proportion of forest type within six younger age classes shown in forest type bundles.



With types and ages scaled to similar classes, the acreages in each stand were assigned to the appropriate age and type error matrix cells as indicated in Tables 1 and 2, respectively. The resulting two error matrices illustrate results from a process that, using these test data, performs well in the age determination aspect and below expected in the forest type classification aspect.

Table 1: Age comparison (error matrix) by stand polygons (acres), making sure only to compare ages in stands noted as forested stand layer

Reference Data \ Classified Landsat Data	0-5	6-10	11-15	16-20	21-25	26-30	31-35	> 35	
0-5	**56,707**	9,248	53.9	52.7	449.1	274.3	251.5	709	67,744
6-10	16,708	**53,933**	1,744	8.4	9.6	77.8	1146.1	2,219	75,847
11-15	8.4	14,544	**20,608**	262.3	8.4	33.5	6	2,859	38,329
16-20	316.2	673.1	8,532	**16,768**	796.4	28.7	34.7	4,278	31,425
21-25	2,238	1,814	16.8	3,353	**9,513**	162.9	58.7	1,628	18,786
26-30	1,943	1,569	2.4	25.1	725.7	**2,529**	326.9	2,562	9,681
31-35	789.2	959.3	13.2	13.2	2.4	232.3	**1,266**	942.5	4,219
> 35	6,224	7,921	1,720	1,402	1,965	3,965	2,905	**106,807**	132,910
	84,933	90,661	32,690	21,886	13,471	7,304	5,996	122,002	**378,941**

	Percent	psKHAT
Overall %	71%	63%

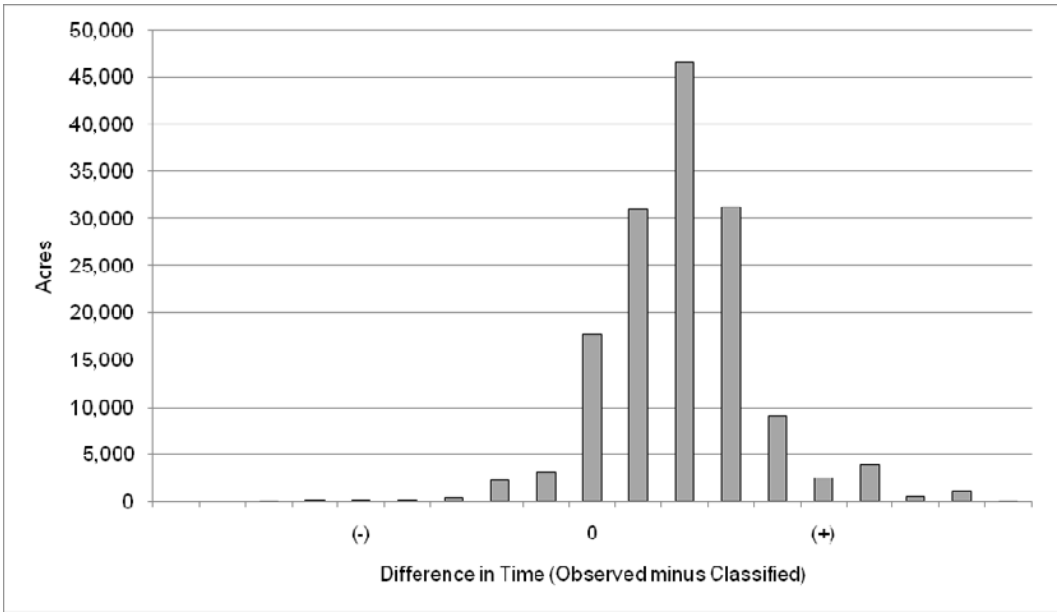
Table 2: Type comparison (error matrix) by stand polygons (acres)

Reference Data \ Classified Landsat Data	Harvest	Regen	Open	Pine	Mixed	Hdwd	
Harvest	**3,853**	5,702	5,497	4,465	4,970	13,344	37,830
Regen	31,432	**78,338**	10,158	1,529	713.8	1,635	123,806
Open	35.9	704.2	**3,917**	257.5	470.7	1039.5	6,424
Pine	480.2	18,431	6,149	**75,162**	5,739	2,638	108,599
Mixed	445.5	2,047	268.3	4,361	**23,674**	396.4	31,193
Hdwd	358.1	5,927	1,309	7,020	9,078	**66,713**	90,406
	36,606	111,150	27,298	92,793	44,646	85,767	**398,258**

	Percent	psKHAT
Overall %	63%	53%

Note that the age results in Table 1 show a fairly good overall agreement at 71% with a pseudo-Kappa-hat of 63% (this is a pseudo measure as penalizing chance agreement, which is the goal of the Kappa statistic, assumes certain sample allocation practices that were not employed in this loblolly management-focused forestland ownership test data). In further analyzing the data, it was noted that the age classification process seemed to have a bias so that regenerated forests were being assigned an age of zero even after they had already been regenerated (Figure 2). This was first noted in the elevated off diagonal cells in the Table 1 error matrix (highlighted in gray) where large area amounts were noted in younger classified classes than observed.

Figure 2. The distribution areas, in acres, of age bias (observed minus classified), in one year age bias classes, for the forested stands in the polygon vector test dataset.



With regard to the type classification (Table 2), the results were below expected at 63% overall agreement with a pseudo-Kappa-hat of 53%. These poor results, however, can be refined if the regeneration and harvest classes, which are very close in meaning, are merged (new diagonal matrix value derived from gray shaded area), yielding a refined overall agreement proportion of 73%. Further refinement involving only looking at the forested stands (the error matrix region with individual cells bordered) reveals an overall agreement of 72%. Aside from these results, it is worth echoing the previous statement involving the nature of the stand data itself. This problematic nature is primarily focused on the management species attribute field, which again, is sometimes a management target as opposed to what is actually present, and the one year age difference between the stand data (last updated in June 2006) and the classification (based off of leaf-on Landsat data from 2007). Also, this data, again, is biased toward actively managed loblolly pine plantations, thus, it is not representative of the wood basin as a whole.

Timber Supply Analyses with SRTS

Because most regional southern timber supply analyses focus significantly on pine plantation availability, the forest type accuracy bias toward pine in the satellite classification processes is not as restrictive as it might be if supply was equally influenced by other forest types. Likewise timely and improved updates of regeneration and harvest classes can provide more critical information than improved allocation of five year age classes through time in the SRTS model. In other words it is more important to know how many acres are cut and planted in a given period than the exact age class distribution of them through time when projecting future supply.

However, because of the difficultly in distinguishing between "regenerating" or "harvested" classes these two forest types were combined in their age class group and the average forest type ratios between pine, hardwood and mix for the age classes 5-15 were used to allocate their "type". In essence this assumes that the reforestation trends of the last fifteen years are still applicable. This may or may not be the case. This includes acres that were "changed" or "harvested" but that are not designated as timberland yet either. Similarly, because of the age of the Landsat data, the oldest age class able to be determined by the satellite classification is 30-35. Because the SRTS model utilizes the FIA inventory five year age classes up to 50, the FIA age class proportions are used to allocate the satellite acres accordingly for the 35 and older group.

The SRTS model uses FIA's five broad management types as well as two ownership groupings: corporate and private. Because the satellite classification can only distinguish "forest cover types", assumptions must be made to allocate these three forest cover types (pine, mix and hardwood) into the five FIA/SRTS management types (pine plantation, natural pine, mix/hardwood, upland hardwood, and bottomland hardwood). In the above client dataset the satellite classification estimated 3% more overall timberland acres than was provided in the SRTS inventory data files. The allocation of the satellite acres after adjustment for the age class 0-5 was 49% pine, 11% mix, and 39% hardwood. This compares to a SRTS allocation of 44% pine, 11% mix and 44% hardwood. The SRTS percentages between planted and natural pine and upland and bottomland hardwood were used to further allocate the satellite pine and hardwood acres within their age class groupings (Table 3).

Table 3: A comparison of the original FIA/SRTS acreage allocation to adjusted (using FIA/SRTS ratios/proportions) classified satellite acreages.

FIA/SRTS Type and Age Area Estimates

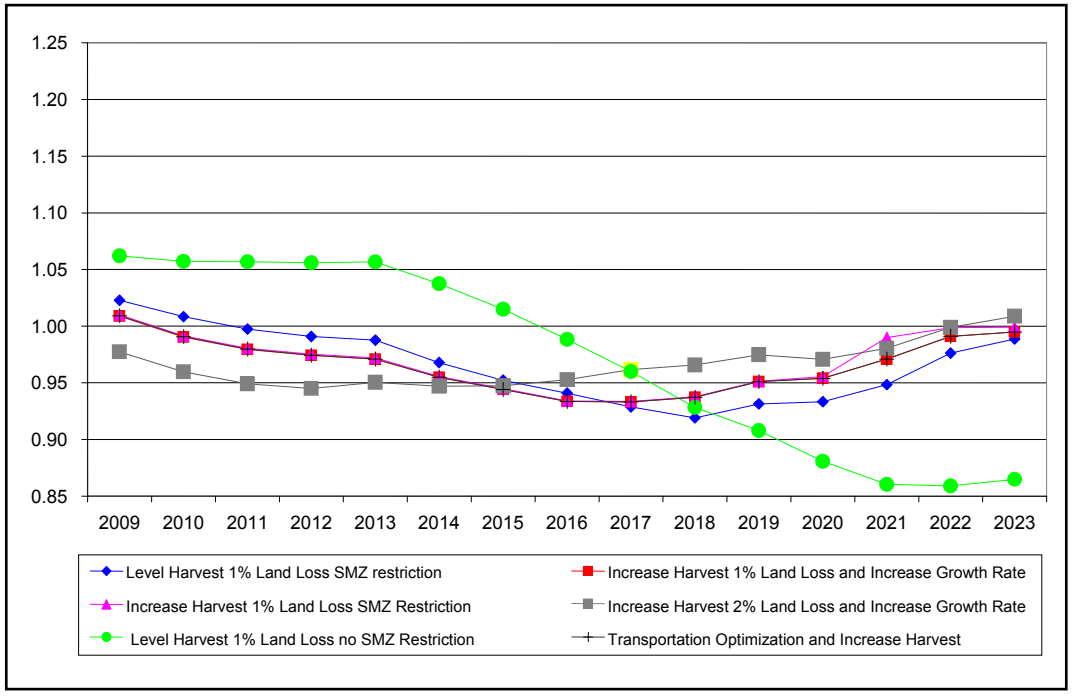
Age Class	Planted Pine	Natural Pine	Mix	Upland Hardwood	Bottom Hardwood	Grand Total
0-5	62,860	89,796	46,119	280,600	39,229	638,364
6-10	220,595	91,612	70,710	174,317	22,301	579,535
11-15	200,117	129,937	66,582	135,369	30,184	562,189
16-20	251,314	99,511	50,647	82,960	14,699	499,132
21-25	299,795	108,334	60,949	94,566	17,123	580,768
26-30	181,636	133,447	51,426	81,831	34,849	483,189
31-35	74,105	90,762	67,053	89,498	18,976	340,394
36-40	26,366	66,588	54,538	129,572	51,416	328,480
41-45	17,055	103,187	60,024	118,759	26,138	325,163
46-50	0	75,502	42,169	177,677	14,919	310,266
>50	0	219,328	92,723	842,904	223,950	1,378,905
Total	1,453,605	1,208,004	662,939	2,208,055	493,783	6,026,386
% Total	24%	20%	11%	37%	8%	

Adjusted Satellite Classified Type and Age Area Estimates

Age Class	Planted Pine	Natural Pine	Mix	Upland Hardwood	Bottom Hardwood	Grand Total
0-5	192,684	94,745	46,085	110,605	15,462	459,581
6-10	360,629	149,768	83,286	198,435	25,386	817,503
11-15	250,703	162,782	60,705	139,119	31,020	644,329
16-20	267,847	106,057	65,866	148,978	26,396	615,144
21-25	301,960	109,116	36,765	82,437	14,927	545,206
26-30	89,441	65,711	44,299	107,317	45,703	352,472
31-35	42,965	52,622	31,829	116,709	24,746	268,871
36-40	43,462	109,764	65,677	113,660	45,102	377,666
41-45	28,114	170,095	72,283	104,175	22,928	397,594
46-50	0	124,459	50,781	155,857	13,086	344,183
>50	0	361,544	111,661	739,390	196,448	1,409,042
Total	1,577,805	1,506,662	669,236	2,016,681	461,205	6,231,589
% Total	25%	24%	11%	32%	7%	

Using the FIA per acre averages with the new satellite acres a revised inventory can then be used in the SRTS model runs. Likewise additional spatial constraints such as SMZ or slope restrictions, urban growth projections, transportation limitations can be derived to be used in conjunction with availability or sensitivity analysis. Figure 3 compares the growth/removal ratios for such a sensitivity analysis looking at harvest increase, land loss assumptions and increased federal land supply.

Figure 3: Growth and removal rates for a test wood basin of interest using various scenarios for sensitivity analysis purposes.



Conclusions

With regard to overall project performance there is still, as always, room for improvement. When the practices of application and science meet, compromises are usually made, as was the case here. Had more time been available, the results could have been further refined and, as one might expect, the accuracy of the products would have been improved. With this said, from a remote sensing aspect, lessons from this process include thrusts to improve efficiency; reduce costs (which is tied to efficiency of course); address bias in classified age; examine the classification of forest types at varying ages, removing the trend to over-classify the deciduous type in younger classes; and generally work to better align classified products with FIA data. These improved classification techniques in conjunction with improved application of per acre inventory averages and growth rates should help to provide inventory estimates that are more reflective of current market conditions for SRTS model runs. Likewise, because spatial analysis is available using the classification and its classed acreage distribution, sensitivity analysis can also be more reflective of future market scenarios improving strategic forest planning opportunities.

References

Collins, C.A., D.W. Wilkinson, D.L. Evans. 2005. Multi-temporal analysis of Landsat data to determine forest age classes for the Mississippi statewide forest inventory-preliminary results, pp. 10-14. In Proceedings of the 2005 International Workshop on the Analysis of Multi-Temporal Remote Sensing Images. [CD-ROM]; 2005 May 16-18; Biloxi, MS.

Abt, R.C., F.W. Cubbage, G. Pacheco. 2000. Southern Forest Resource Assessment Using the SubRegional Timber Supply (SRTS) Model. Forest Products Journal. 50(4).

Mapping Upland Hardwood Site Quality and Productivity with GIS and FIA in the Blue Ridge of North Carolina

Claudia A. Cotton[1], Stephen R. Prisley[2], and Thomas R. Fox[2]

Abstract: *The forested ecosystems of the southern Appalachians are some of the most diverse in North America due to the variability in climate, soils, and geologic parent material coupled with the complex topography found throughout the region. These same characteristics cause stands of upland hardwoods to be extremely variable with regard to site quality and productivity. Site index has been the tool most commonly used to measure existing site quality and productivity, but measured site index may not accurately quantify potential site quality and productivity, largely due to ubiquitous disturbance and variable land-use history. Because of this, environmental factors may hold merit in predicting the quality of a forested site in the southern Appalachians. To assess the accuracy of existing methods, three indices developed within the region were used to predict the site quality of the upland hardwood forests throughout a six-county study area in the Blue Ridge physiographic province of western North Carolina. We hypothesized that predictions of site quality generated by the indices would correlate with similar estimates from Forest Inventory and Analysis (FIA) plots. We also predicted that the indices that included multiple types of information would produce higher correlations with FIA estimates. Finally, we felt we would be able to reasonably predict site index, but not basal area or volume increment. The environment of the study area was derived from a layered GIS that depicted variables related to water availability. FIA data and actual plot locations were compared to the predictions. Results indicated a moderate correlation between one index and site quality. The index with multiple layers of information did not produce a higher correlation, and there was no relationship among any of the indices to basal area or volume increment. Future research will include finer-scaled estimates of soil information and estimates of water inputs as well as usage.*

Keywords: Upland hardwood forests, site quality, site productivity, GIS, Blue Ridge.

Introduction

Quantifying site quality and upland hardwood productivity is a major challenge for accurate growth and yield modeling in the upland hardwood forests of the southern Appalachians. These ecosystems are some of the most complex

[1] Research Associate and [2] Associate Professors; Virginia Polytechnic Institute and State University; Department of Forestry; 305 Julian Cheatham Hall; Blacksburg, VA 24060; email: ccotton9@vt.edu

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

and diverse in North America due to the many combinations of topography, geology, climate, and landform found throughout the region. These attributes lead to a wide range of site quality and productive capacity from the stand to the region, driven largely by topographic controls on water availability. Site index (Frothingham 1917) has been the standard for measuring site quality and productivity, but accuracy is affected by disturbance, variable land-use history, the presence of suitable trees to measure, and measurement error.

Traditionally, site classification has addressed the landscape by analyzing the vegetation, measuring site index, and inferring relationships of water and nutrient availability based on the current vegetation. However, due to problems with site index accuracy, this methodology often does not capture the true site quality and consequently, estimates of stand growth and yield may be inaccurate. A more precise approach would be to identify the primary environmental variables found to significantly influence the availability of water and nutrients in upland hardwood forests, determine how the variables influence the measured site quality, and then use these relationships to develop a model to predict the quality and productivity of a site.

Site quality is assumed here to be a measure of potential productivity. Site, in silviculture terms, can be expressed qualitatively through the local climate, soil, and vegetation present, and quantitatively through the local productivity, or potential wood production per unit land area per unit of time (Helms 1998; Johnson et al. 2002). Site quality and productivity can be predicted with reasonable accuracy in small areas using measured site index, but the prediction is costly and requires specific vegetation to be present. Field-measured site index is not feasible for use across the entire southern Appalachians. However, previous research has shown that site quality can be interpreted through topographic, edaphic, geologic, and vegetative characteristics for discrete areas (Elliott et al. 1997; Fralish 1994; McNab et al. 2004; Smalley 1984, 1986; Whittaker 1956; Williard et al. 2005). More recently, advances in digital terrain modeling and mapping have allowed us to combine and analyze the landscape with higher accuracy and precision (Bolstad et al. 1998; Host et al. 1996; Iverson et al. 1997; Kelley et al. 2005; Simon et al. 2005).

By increasing the accuracy of our site quality estimations, we may identify sites with the highest potential productivity, focus our management efforts onto these sites, and reduce the land base used for commercial wood production. As our forests are becoming increasingly important, this work could improve the efficiency with which we utilize our forested resource. In order to manage a forest for any reason, it is imperative to have an accurate inventory.

The objective of this study was to determine if environment-based geospatial estimates of site quality and productivity within a six-county study area in the Blue Ridge of western North Carolina were correlated with estimates from the Forest Inventory Analysis (FIA) data from that area. We hypothesized that site

quality predictions generated from three recent indices based on topography and soils would be correlated with estimates taken by the USFS Forest Inventory Analysis plots. A second hypothesis predicted that multiple layers of data (topography and soils) would produce a higher correlation with the validation data. Finally, we felt we would be able to reasonably predict height-driven productivity estimates (site index), but not those derived from diameter (basal area and volume increment).

Methods

Study Area

The six counties chosen for the study included Buncombe, Haywood, Jackson, Madison, Swain, and Yancey, and are located in the mountainous western area of North Carolina known as the Blue Ridge (Figure 1). The terrain is heavily dissected, resulting in maximum elevations around 2000 meters (m). The six counties cover approximately 780,000 hectares, or 34% of the Blue Ridge within North Carolina.

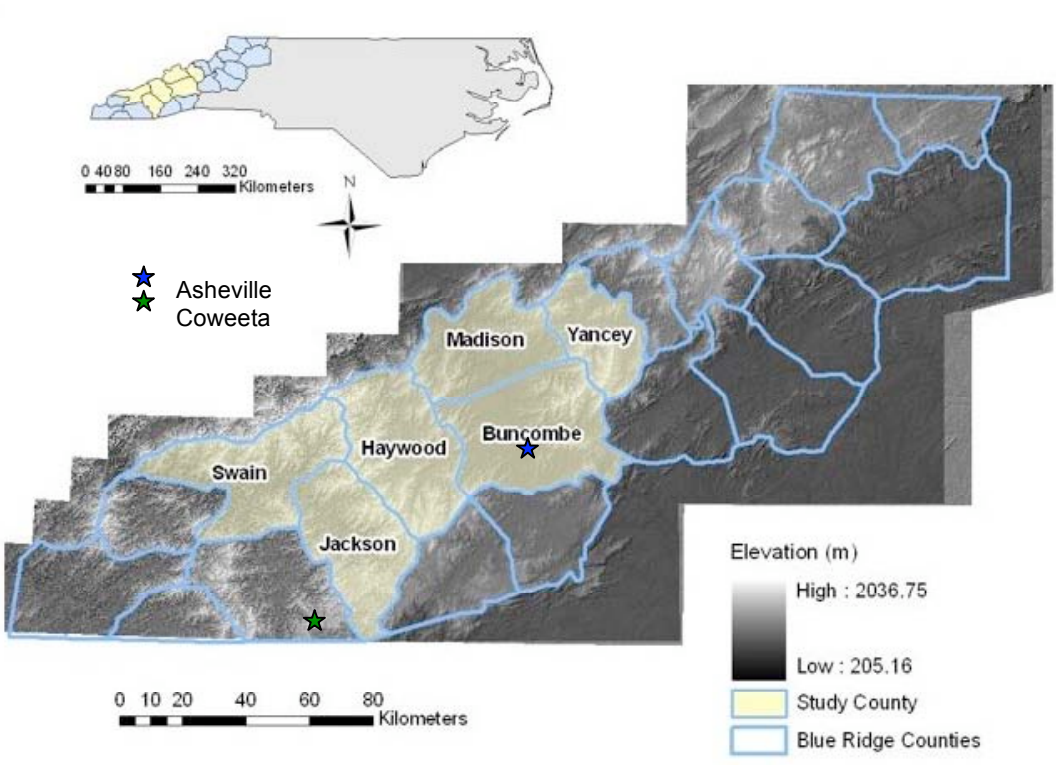


Figure 1: The six counties included in the study area located within the mountains of western North Carolina.

The mean annual temperature is 8 to 16°C and ranges from 3.3°C in January to 24°C in July (USDA NRCS 2006). Precipitation varies widely in the study area

and is controlled for plant uptake by the topography. For instance, Asheville, NC, located in the Asheville Basin, receives as little as 900mm of annual rainfall because it is in a rain shadow created by the surrounding mountains (Figure 1). On the other hand, Coweeta Hydrological Laboratory, less than 100 km from Asheville, can receive over 2000mm of annual precipitation, one of the highest rates in the eastern United States. A growing season of 135 to 235 days prevails (USDA NRCS 2006). From the many combinations of topography and landform in the area, aspect significantly affects microclimate, which in turn influences the type and vigor of the vegetation communities. South and west facing slopes are warmer and drier than those found on north and east aspects, as well as those that are shaded by neighboring landforms (USDA NRCS 2006).

General Approach

Several layers of geospatial maps, based on widely available spatial data, were generated. Second, three existing indices were used to predict site quality as a function of the environment for specific locations where there was available FIA data. Finally, the predictions were compared to FIA estimates of productivity.

GIS

An elevation mosaic for the study area was created from 10-meter (m) Digital Elevation Models (DEM's) that were obtained from the USGS National Map Seamless Server. Basic surface analyses, including aspect, slope percent and degree, curvature, and hillshade, were derived from the DEM.

In order to get a measure of slope position (summit, backslope, footslope, etc.) we used the hydrological characteristics of the area. We filled sinks in the DEM and derived flow direction and flow accumulation, in succession. The flow accumulation grid was classified into classes of ½ standard deviation, and a threshold was established to mark the beginning of the streams. The result was then reclassified to create a mask that separated water (NoData) from land (1). The masked flow accumulation was then multiplied by the original flow direction grid, which created a masked flow direction. This was done to allow the water flow to go to the stream edge and stop instead of traveling onto the stream outlet. In essence, we wanted slope position of the land and not the water. Two flow length grids were calculated for downhill and uphill flow length using the masked flow direction. The two length rasters (downhill and uphill) were then used in the formula:

Downhill flow length / (Uphill flow length + Downhill flow length)

This calculation allowed for an estimate of slope position as a percent of the slope, where 0% slope position represented the bottom of the slope at stream edge and 100% was at the ridge top.

Indices Used to Predict Site Quality

Three indices were used to predict the site quality of the study area. They were chosen because they represented site quality and productivity based on water availability, which has been cited as the most limiting factor to tree growth in the southern Appalachians (Smith 1994).

Terrain Shape Index (McNab 1989): The simplest algorithm for determining upland hardwood productivity, the TSI provided a measure of the concavity or convexity of the land that surrounded the plot center. It is defined as the mean slope gradient of the plot boundary as viewed from the plot center (McNab 1989). High values indicated maximum convexity (ridges, spur ridges, nose slopes), and minimum values indicated maximum concavity (creek beds, coves, bottoms). The original development of the model related the TSI to yellow-poplar site index in the Blue Ridge. The formula of the TSI is as follows:

TSI = mean elevation of the plot boundary / plot radius

In ArcGIS, a grid of the TSI function was calculated in raster calculator as:

TSI = [DEM-focalmean (DEM, annulus, 3,4)] / 35

The focalmean is a neighborhood function, which allows a cell-to-cell comparison to determine the answer to the command. The "3,4" portion specified that the plot boundary was to be determined between 30-40m from the plot center and would produce a radius of 35m.

Forest Site Quality Index (FSQI) (Meiners et al. 1984): Meiners and others (1984) examined the effect of topography on available water in the Ridge and Valley of southwestern Virginia and the result was the FSQI. It combined slope position, slope percent, and aspect to determine topographic position as it affected water availability for forest growth. Once the index number was determined it was then correlated to the upland oak site index. This was the only index that categorized all variables and outputs into specific productivity classes.

In ArcGIS, the aspect, slope percent, and slope position grids were reclassified to reflect the assigned FSQI score (Table 1). Adding the three grids together in raster calculator produced the final index grid, which had a potential value range of 3 - 16. A high index value indicated high site quality.

Table 1: Values assigned to the input variables in the FSQI productivity model.

FSQI Value	Aspect	% Slope	Slope Position
1	196-260	>=60	Shoulder
2	166-195; 261-280	45 - 59	Backslope
3	146-165; 281-340	30 - 44	Summit
4	0-20; 341-360	15 - 29	Footslope
5	81-145	0 - 14	Toe Terrace Floodplain
6	21-80		

Integrated Moisture Index (IMI) (Iverson et al. 1997): The most complex index of the three, the IMI combined topographic and soils data to get an index of site quality. It was developed in the Alleghany Plateau in Ohio, which should be considered when assigning weights to the input variables. It is the only index that was originally created in a digital environment. From a 10m DEM, we derived hillslope, curvature, and flow accumulation, and from a 1:250,000 STATSGO map, we derived whole-profile available soil water (cm). The four variables were weighted and added in raster calculator to produce a final map of IMI. The authors were able to correlate the index to upland oak site index, and high index numbers indicated high site quality.

FIA Database

Productivity estimates from the North Carolina FIA database were compared to predictions of site quality in the study area (Figure 2). We obtained, for the entire state of North Carolina, the tree, plot, and condition tables for the years of 1982 (Cycle 5), 1990 (Cycle 6), and 2002 (Cycle 7). As part of a Privacy Policy Study Group Memorandum of Understanding, the Southern Research Station FIA unit provided the actual plot locations.

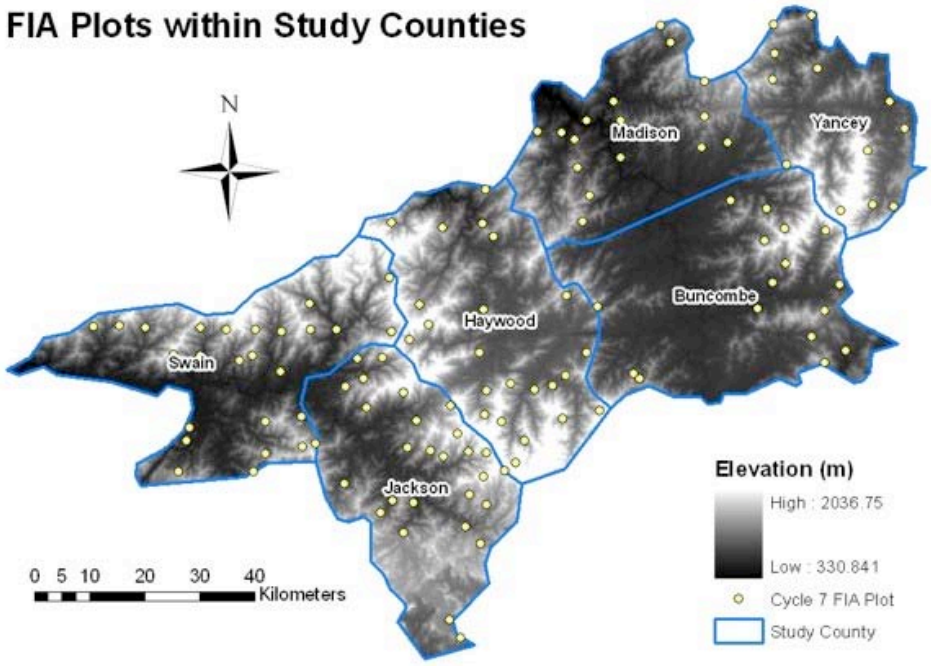


Figure 2: FIA plots used to test the accuracy of the site quality predictions generated by the TSI, FSQI, and the IMI.

FIA productivity estimates used to test the accuracy of the predictions included site index and basal area increment from Cycle 7, and volume increment from a

six-year difference between Cycles 5 and 6. A larger increment was preferred but could not be obtained due to a change in survey methodology between Cycles 6 and 7 (from periodic to annual). This resulted in an inability to match trees and plots between the two cycles.

Some stratification of the database was done to ensure that the FIA plots were forested, mostly hardwood, and minimally disturbed (Table 2). Note that Table 2 reflects the number of usable FIA plots within the entire Blue Ridge portion of North Carolina; for the plot sample size that covered the six-county study area, please refer to Table 3. According to the FIA database, stand age is subject to large error because it is difficult to quantify (Forest Inventory and Analysis Program 2008). For this reason, age was not used as a rigid exclusion criterion.

Table 2: FIA database screening criterion and details for forested plots in the Blue Ridge physiographic province for cycles 5, 6, 7, and Merged 5 & 6.

Cycle 5 Criterion	Initial Plot *n*	Plot *n* after screening	Difference (# Plots Deleted)	Percent of initial plot *n* kept
forested land in Bl. Ridge	1025	804	221	
water or balance plot	804	733	71	
>20% conifer BA in plot	733	546	187	
removed tree in plot	546	**495**	51	48.3

Cycle 6 Criterion	Initial Plot *n*	Plot *n* after screening	Difference (# Plots Deleted)	Percent of initial plot *n* kept
forested land in Bl. Ridge	1022	793	229	
water or balance plot	793	735	58	
>20% conifer BA in plot	735	553	182	
removed tree in plot	553	**499**	54	48.9

Cycle 7 Criterion	Initial Plot *n*	Plot *n* after screening	Difference (# Plots Deleted)	Percent of initial plot *n* kept
forested land in Bl. Ridge	930	767	163	
water or balance plot	767	767	0	
one condition only	767	507	260	
no lat/lon recorded	507	482	25	
>20% conifer BA in plot	482	370	112	
removed tree in plot	370	355	15	
stand treatment	355	**347**	8	37.3

Merge 5 & 6 Criterion	Total initial merged plot *n*	Plot *n* after Screening[a]	% of Total initial merged plot *n* kept	No. Counties
Merge Cyc 5 & 6	575	**376**	65.391	23

[a]Screening criteria include unmatching plots and age difference of 6 years, after individual cycle screening criteria were applied

Statistics

Pearson's correlation coefficient *(r)* was calculated to determine the relationship between the FIA productivity estimates (site index, standing basal area, and volume increment) and the predicted site quality estimates which were generated by the TSI, FSQI, and IMI.

Results and Discussion

Digital Imagery of the Productivity Models

The first step of the analysis involved predicting site quality using the TSI, FSQI, and IMI. Maps of predicted site quality were generated in ArcGIS using each of the indices (Figures 3, 4, and 5).

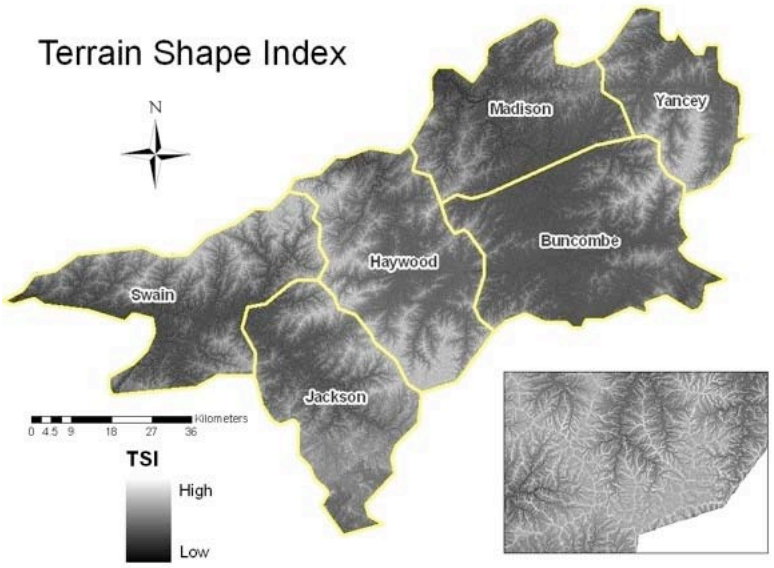


Figure 3: The digital representation of the TSI across the study area.

The TSI model was best represented as a continuous black-to-white grid where high, or light, values indicated maximum convexity, or areas where water would not be stored. Dark, or low values represented maximum concavity, or areas where water would be more accessible for tree uptake and growth. This could translate to a continuous grid of site quality, where high TSI values served as a proxy for low site quality, and low TSI values for high site quality.

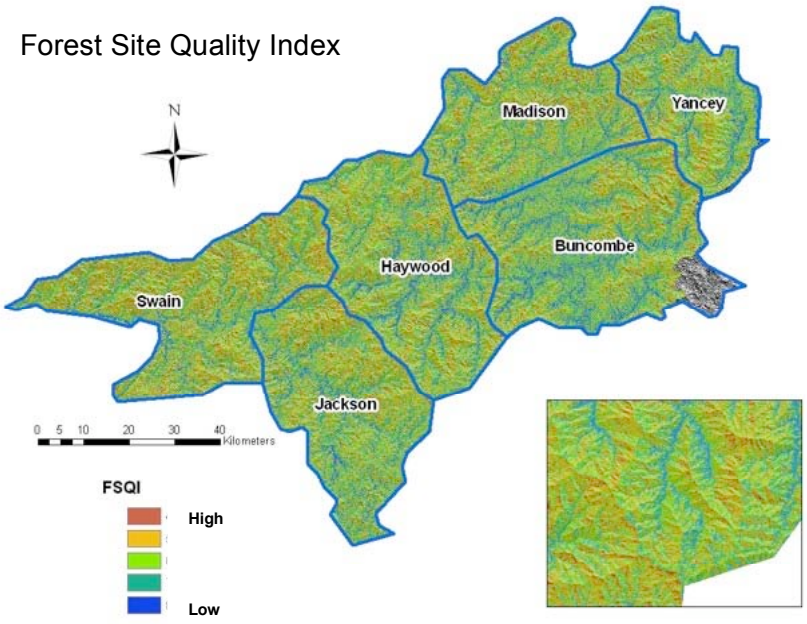


Figure 4: The digital representation of the FSQI across the study area.

The FSQI model was represented as a categorical grid to mirror the final index scores, with cooler colors (blue and green) reflecting higher site quality and warmer colors (red and orange) indicative of lower site quality. This map represents the combination of aspect, slope percent and slope position as they affected water availability, and in turn, site quality.

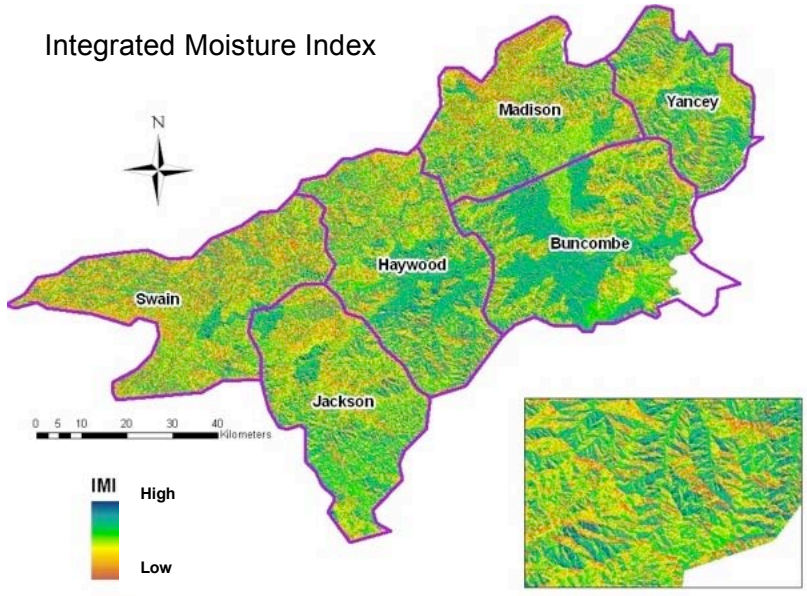


Figure 5: The digital representation of the IMI across the study area.

The IMI model was represented as a continuous color map to show the wide variation of site qualities across the study area. The color scheme mirrored that of the FSQI, where high index scores (demonstrating to high site quality) were depicted in blue and low index scores (low site quality) were in red. The final model grid was a weighted combination of flow accumulation, curvature, whole-profile available water holding capacity, and hillshade. We felt the IMI would have the most power in reflecting the true site quality of the study area; however, it predicted high site quality in the Asheville Basin, which has low site quality due to its landscape position within a rain shadow. Questions emerged at this juncture, in particular, what else was affecting the site quality besides topographic and edaphic conditions?

Correlation Analysis

Of the three FIA estimates of productivity (site index, standing basal area, and volume increment), site index alone was somewhat predictable (Table 3). This was expected since any measure of productivity that is diameter-based will be highly influenced by stand conditions, such as density, stage of succession, management practices, and disturbance history. Comparative analysis through correlation revealed a moderate relationship between the FSQI and FIA site index (Figure 6).

Table 3: Summary of Pearson's correlation coefficients and sample sizes (in parentheses) for predicted versus actual site productivity.

	Predictive Index		
FIA Productivity Estimate	**TSI**	**FSQI**	**IMI**
Site Index	-0.25 (127)	0.38 (143)	0.09 (125)
Standing Basal Area	0.05 (127)	-0.01 (143)	0.18 (125)
Volume Increment	-0.03 (137)	-0.04 (138)	0.05 (148)

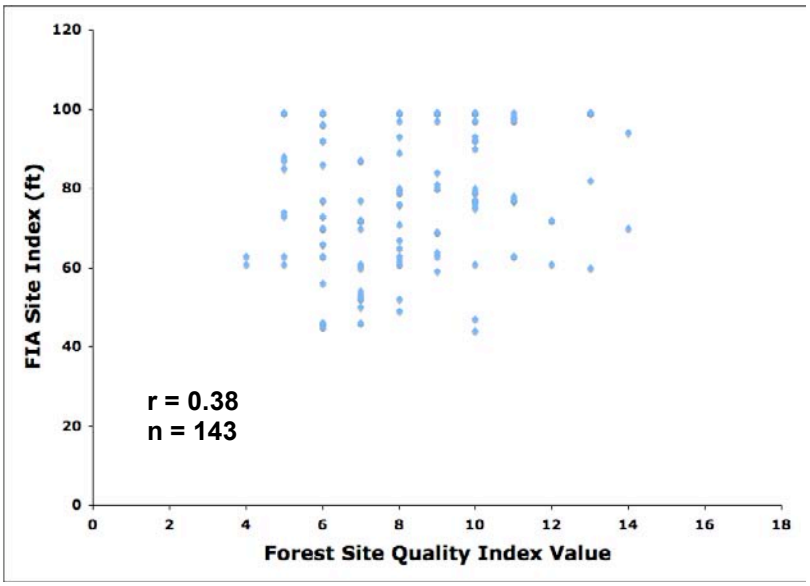


Figure 6: Scatter plot showing the correlation between FIA Site Index and the FSQI.

The correlation between FIA site index and the FSQI could have been higher had the FIA site index not been capped at 99 feet (Figure 6). Discussions with FIA personnel indicated that at some point site index values had been allocated only two numeric digits, limiting site index to a maximum of 99 (pers. comm., Ray Sheffield, USFS SRS FIA). Further, considering that we were working with a wide range of hardwood species as well as age classes, we felt that 0.38 was a reasonable correlation.

The TSI did not correlate with any of the FIA estimates of productivity (Table 3). We could not use this model to predict the actual site index for the study area because of the larger area, more extreme terrain values, and issues with replicating the index digitally, which produced more extreme values than what McNab originally derived in his model development. Because of this, a decision was made to use the calculated index values versus the model calculated site index for all of the correlations to maintain consistency.

Unexpectedly, the addition of soils information did not improve the predictive ability of the IMI (Table 3). This was likely a result of the scale of the STATSGO data layer (1:250:000). This particular index may have had more power if finer-scale data, such as SSURGO, (scale of 1:25,000) could better reveal differences that may exist among the mapping units. Additionally, it is necessary to account not only for the capacity of a soil for water storage, but for usage and inputs to that system as well, as was exemplified with high IMI value for the rain-shadowed Asheville Basin. We feel by incorporating all aspects of the water cycle (inputs of climate, usage in the form of evapotranspiration, and storage as defined by the environment and soils), we will come closer to predicting the true site quality of these forested ecosystems.

Overcoming Issues to Progress Forward

Productivity and site quality research shows promise using FIA data but there are some issues that need to be resolved in order for this database to be truly workable in such an environment. Specifically, the inability to match trees between the periodic and annual surveys restricts the calculation of any reasonable growth increment to capture productivity. The site index estimates as taken in the field are subject to considerable error, but they are the best we have for this study. Changed plot numbers between the two survey methods also hinders analyses between cycles. Finally, a consistent way to capture plot age would be helpful when dealing with stands of mixed-species and mixed-aged hardwoods. This is a subjective task at best, but perhaps if a protocol were established by the FIA, our site quality predictions for upland hardwoods would be more robust.

Conclusions

As hypothesized, site quality predictions generated by the FSQI were reasonably correlated with the FIA data. Second, we were able to predict site index but not basal area or volume increment. Finally, additional information about soils did not improve the predictive ability of the Integrated Moisture Index, mainly due to the coarse scale of the data.

These findings indicate that more work is needed to predict site quality as a function of environment in the upland hardwood forests of the southern Appalachians. Future efforts to capture this model include the incorporation of water usage and inputs in the form of evapotranspiration and climate, respectively. Further, the use of more detailed and different types of soils information, such as SSURGO, may allow for better explanation of how water is stored in the forest, and how it affects site quality. This data may also refine the predictions by providing a measure of nutrient availability through the characteristics of soil organic matter content, texture, and depth. Along these lines, geofertility classes based on parent material are being considered as well.

Site quality and productivity work in the southern Appalachians has been continuous for at least the past one hundred years. The significance of this work is even more important now as our forests are valued for multiple products, such as a timber, clean water, wildlife habitat, carbon stores, and recreation. As our estimates of site quality increase, we will be able to more effectively partition the forest into efficient management units so all usage objectives may be met for the long term. It is imperative we sustain our forests for future generations, and this work will directly affect that goal.

Literature Cited

Bolstad, P., W. Swank, and J. Vose. 1998. Predicting southern Appalachian overstory vegetation with digital terrain data. Landscape Ecology 13:271-283.

Elliott, K., L. Boring, W. Swank. 1999. Long-term patterns in vegetation-site relationships in a southern Appalachian forest. Journal of the Torrey Botanical Society 126:320-334.

Forest Inventory and Analysis Program. 2008. The forest inventory and analysis database: database description and users manual version 3.0 for Phase 2, revision 1. USDA Forest Service, Washington Office.

Fralish, J. 1994. The effect of site environment on forest productivity in the Illinois Shawnee Hills. Ecological Applications 4:134-143.

Frothingham, E. 1917. Height growth as a key to site. Journal of Forestry 16:754-760.

Helms, M. (ed.). 1998. The Dictionary of Forestry. Society of American Foresters, Bethesda, MD. 224 p.

Host, G., P. Polzer, D. Mladenoff, M. White, and T. Crow. 1996. A quantitative approach to developing regional ecosystem classification. Ecological Applications 2:608-618.

Iverson, L., M. Dale, C. Scott, and A. Prasad. 1997. A GIS-integrated moisture index to predict forest composition and productivity of Ohio forests. Landscape Ecology 12:331-348.

Johnson, P., S. Schifley, and R. Rogers (eds.). 2002. Site Productivity. Pp. 168-189 *in* The Ecology and Silviculture of Oaks. Biddles Limited, King's Lynn, England. 503 p.

Kelley, A., D. Powell, and R. Riggs. 2005. Predicting potential natural vegetation in an interior northwest landscape using classification tree modeling and a GIS. Western Journal of Applied Forestry 20:117-127.

McNab, W. 1989. Terrain shape index: quantifying the effect of minor landforms on tree height. Forest Science 35:91-104.

McNab, W., D. Loftis, and C. Schweitzer. 2004. A pilot test of indicator species to assess uniqueness of oak-dominated ecoregions of central Tennessee. Pp. 88-93 *in* Upland Oak Ecology Symposium: history, current conditions, and sustainability, Speitch, M. (ed.). USDA Forest Service Southern Research Station. GTR-SRS-73. 311 p.

Meiners, T.M., D.W. Smith, T.L. Sharik, and DE Beck. 1984. Soil and plant water stress in an Appalachian oak forest in relation to topography and stand age. *Plant and Soil* 80:171-179.

Schmidt, J. and A. Hewitt. 2004. Fuzzy land element classification from DTM's based on geometry and terrain position. *Geoderma* 121:243-256.

Simon, S., T. Collins, G. Kauffman, W. McNab, and C. Ulrye. 2005. Ecological zones in the southern Appalachians: first approximation. USDA Forest Service Southern Research Station. RP-SRS-41. 41 p.

Smalley, G. 1984. Classification and evaluation of forest sites in the Cumberland Mountains. USDA Forest Service Southern Forest Experiment Station. GTR-SO-50. 84 p.

Smalley, G. 1986. Classification and evaluation of forest sites on the northern Cumberland Plateau. USDA Forest Service Southern Forest Experiment Station. GTR-SO-60. 76 p.

Smith, D.W. 1994. The southern Appalachian hardwood region. Pp. 173-225 *in* Regional silviculture of the United States, 3rd ed. JW Barrett (ed.). J Wiley and Sons, New York. 551 p.

USDA NRCS. 2006. Land resource regions and major land resource areas of the United States, Caribbean, and the Pacific Basin. USDA Handbook 296.

Whittaker, R. 1956. Vegetation of the Great Smoky Mountains. Ecological Monographs 26:1-80.

Williard, K., D. Dewalle, and P. Edwards. 2005. Influence of bedrock geology and tree species composition on stream nitrate concentrations in mid-Appalachian forested watersheds. Water, Air, and Stream Pollution 160:55-76.

Testing Ecoregions in Kentucky and Tennessee with Satellite Imagery and Forest Inventory Data

W. Henry McNab and F. Thomas Lloyd (deceased)[1]

***Abstract**: Ecoregions are large mapped areas of hypothesized ecological uniformity that are delineated subjectively based on multiple physical and biological components. Ecoregion maps are seldom evaluated because suitable data sets are often lacking. Landsat imagery is a readily available, low-cost source of archived data that can be used to calculate the normalized difference vegetation index (NDVI), which is associated with seasonal and annual vegetation conditions. This paper reports on a study designed to test the use of NDVI as an integrating variable for detecting differences among mapped ecoregions and determine if NDVI was associated with biological components within ecoregions using USDA Forest Service Forest Inventory and Analysis (FIA) field data. Published 1.1 km resolution georeferenced NDVI imagery was obtained at the beginning (June 22) and end (September 23) of the summer growing season for 11 years (1989-1999). Public domain GIS software (WinDisp4) was used to determine NDVI values at 5,399 georeferenced, forested plot locations in Kentucky and Tennessee that are periodically inventoried by FIA. Plots were grouped by ecoregions of various scale and tested for significant differences. Analysis of variance revealed significant ($P<0.001$) differences in mean NDVI between the two macro-scale ecoregions (divisions) and among most of the next lower ecological units (provinces). Regression analysis indicated that NDVI was associated ($P<0.01$) with season of sampling, elevation, and forest stand basal area of the inventory plots. At the ecoregion analysis scale, NDVI values consistently increased with higher elevation and forest basal area at the plot locations. We concluded that NDVI determined at FIA plot locations has potential for testing differences among ecoregions. Satellite imagery has often been used to determine classification accuracy within small mapped ecosystems; results from our study suggest that imagery may also be used to test for hypothesized overall differences between large mapped ecosystems.*

Keywords: AVHRR, ecoregionalization, NDVI, Palmer drought severity index, remote sensing.

---

[1] Research Foresters; United States Department of Agriculture, Forest Service; Southern Research Station; Upland Hardwood Ecology and Management; Bent Creek Experimental Forest; 1577 Brevard Road; Asheville, NC 28806 USA; hmcnab@fs.fed.us.

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

Ecoregions are "any large portion of the Earth's surface over which the ecosystems have characteristics in common" (Bailey 1998). When mapped, they can be considered as hypotheses for testing and refinement (Rowe and Sheard 1981). Ecoregionalization is the process of knowledgeable integration of spatial environmental physical factors to form regions of hypothesized ecological uniqueness (Bailey 1998).Recently, this process has been used by conservation and land-management agencies in the planning, analysis, and assessment of resource-related issues (Griffith and others 1999). Maps of ecoregions have been produced by state (Homoya and others 1985, Hargrave 1993) and federal (Keys and others 1995, Griffith and others 1997) agencies, and by conservation organizations (The Nature Conservancy 2001). Ecoregional delineations have been used for many purposes, including development of forest growth models (Huang 1999) and ecosystem-based management plans (Haynes and others 1998), selection of research areas (Snyder and others 1999), assessment of forest resources (Rudis 1999), and analysis of water quality (Griffith and others 1999). Although questions have been raised about the appropriateness of ecoregional delineations for these purposes (Wright and others 1998), little attention has been given to validation of them (Rowe and Sheard 1981, Bailey 1984, Omernik 1995, Smith and Carpenter 1996, Brown and others 1998).

Validation of ecoregional delineations differs in several important ways from the process of testing conventional classifications such as soil maps or maps of vegetation cover types (Edwards and others 1998). First, the "real" taxonomic identities of ecological units on many maps are often unknown because these units are characterized subjectively by consensus during delineation and review. However, objective, quantitative methods have been investigated (Host and others 1996, Hargrove and Hoffman 1999) and criticized as inefficient (Rowe and Sheard 1981). Second, ecoregional delineations are generalizations and syntheses of many environmental components (Bailey 1998) and are not delineated for a specific purpose that provides a convenient basis for testing. Finally, and perhaps most importantly, obtaining a suitable dataset for an appropriate response factor is difficult and costly because ecoregions extend over large areas ranging from hundreds of thousands to millions of km^2 (Bailey 1984). Vegetation is typically used as a response variable for testing ecoregional delineations (Bailey 1984, Edwards and others 1998, Schreuder and others 1999), and water quality has been employed in this way (Bailey 1984). Remote sensing is an economical means of acquiring data bases and satisfies some of the identified problems of testing and validation of ecoregional delineations (Loveland and others1991, Brown and others 1998).

Remote sensing has often been used to develop classifications of vegetation types, but has seldom been employed to test for differences between hypothesized ecosystems. Ramsey and others (1995) found that normalized difference vegetation index (NDVI) was associated significantly with five ecoregions in

Utah mapped by Omernik (1987), each of which occupied a distinct elevation zone and was dominated by a characteristic type and abundance of xerophytic vegetation. The NDVI, a ratio of the red and near-infrared spectral bands, is correlated with the amount of actively photosynthesizing biomass. Since being devised in the 1970s, the use of NDVI has largely dominated vegetation analysis by remote sensing (Loveland and others 1991), although a number of other indices have been evaluated (Gutman 1991). NDVI has been associated with a number of climatic and vegetation parameters, including drought (Singh and others 2003), wildland fire risk (Maselli and others 2003), land degradation (Wessels and others 2004), forage production (Hill and others 2004), leaf area index (Wang and others 2005), stand age and density (Sivanpillai and others 2006), continental-scale vegetation type classification (Loveland and others 1995), and net primary production (Handcock 2000, Meng and others 2007). It is reasonable to think that NDVI data obtained by remote sensing could be used to test for vegetational differences among ecoregions.

Forest Inventory and Analysis (FIA) plot data constitutes another data base that has characteristics desirable for use in the validation of ecoregional delineations. Schroeder and others (1997) found that FIA plots provided suitable data for regional estimates of biomass production. Schreuder and others (1999) reported that inventory data provides a suitable basis for a system of forest monitoring. Plot and tree level data are readily available for inventory plots, field locations are known with reasonably acceptable accuracy, and the network covers a broad geographic area (Hansen and others 1992).

Testing of ecoregional delineations by means of remote sensing has not been reported for the eastern U.S. The eastern U.S. has more precipitation, less topographic relief, and less contrast between forest cover types than the western landscape studied by Ramsey and others (1995). We think it probable that remotely sensed NDVI integrates the quantity, composition, and condition of vegetation in eastern forests in response to environmental relationships as it does in Utah. Effects of exposed soil on NDVI (Huete 1988) should be considerably less in the eastern U.S. because canopy cover is almost complete over forested sites. This paper reports on a study designed to test the use of NDVI as an integrating variable for detecting differences among mapped ecoregions in the relatively humid eastern U.S. We addressed two questions: (1) Does mean NDVI vary among ecoregions within a hierarchical framework of ecological units and (2) How does NDVI vary in relation to environmental and biological factors? We used software, data sets, and techniques that would be readily available to ecological researchers having an understanding of remote sensing but who lack extensive technical training in the discipline.

Materials and Methods

NDVI Data

Imagery, which originated from the AVHRR (advanced very high resolution radiometer) sensor on the NOAA-11 satellite, is retrieved daily and processed by the U.S. Geological Survey (Loveland and others1991, Burgan and Hartford 1993). The 1.1 km resolution imagery was georeferenced and made into weekly or biweekly composite images of generally cloud-free data bases of several spectral wavelength bands (Holben 1986). Data from two of the five spectral channels are used for calculation of NDVI:

NDVI = (Channel 2 - Channel 1)/(Channel 2 + Channel 1)

where Channel 1 is the red, visible portion of the electromagnetic spectrum (0.58 to 0.68 microns) and Channel 2 is the infrared portion of the electromagnetic spectrum (0.725 to 1.10 microns). The underlying premise of the NDVI is that incoming radiation in the visible spectrum (0.58-0.68) is absorbed by chlorophyll in vegetation and that infrared wavelengths are reflected. Thus, NDVI is directly correlated with the quantity of green, photosynthesizing vegetation. Values of NDVI, which can range from -1.0 to 1.0, were recoded from 0 to 255. Values < 100 indicate the presence of clouds, snow, bare soil (i.e. fallow agricultural fields), and land surfaces naturally lacking vegetation. Values ≥100 indicate the presence of varying quantities of photosynthesizing vegetation biomass. Water bodies were coded as 255. We used 11 years (1989 - 1999) of stock, "off-the-shelf," NDVI imagery that had been retrieved from the EROS data center by the USDA Forest Service and published as annual archives primarily for use in fire danger rating (Burgan and Chase 1997a-1997h, Burgan and Chase 1998, Burgan and others 1999, USDA Forest Service 2000). The georeferenced imagery consisted of weekly or biweekly composites of cloud-free pixels of the highest NDVI value. None of the NDVI imagery had been corrected for atmospheric effects (Song and others 2001). We used public domain image display and analysis software, WinDisp4 (FAO 1999), to extract and merge blocks of NDVI images from the data sources and to determine median NDVI within ecoregions. Because our analysis did not involve classification, change detection, time series analysis, or other highly specialized use of imagery, we did not consider that data manipulation or transformation, such as tasseled cap or Fourier, was necessary to achieve our objectives.

Area Studied

Vegetation: We utilized data for ecoregions within the southeastern portion of the upland oak-hickory (*Quercus-Carya*) forest type of the Eastern U.S, also known as the Central Hardwood region (Braun 1950). This is the most extensive forest type of the conterminous U.S., occupying about 46 million ha (Burns 1983). The oak-hickory type occurs in the east-central U.S., from the prairie

borders in Texas northward to the Dakotas and eastward to the Appalachian Mountains from Georgia to southern New England. It is most extensive in Kentucky, Tennessee, and the Ozark and Ouachita Highlands. The type is characterized by deciduous hardwoods that form a nearly closed canopy consisting of four or more layers on mesic sites: a tall (25-30 m) tree overstory, an open midstory of immature canopy and shade-tolerant tree species, a low (1-5 m) shrub layer of ericaceous species, and an herb layer. Northern red oak (*Q. rubra* L.), black oak (*Q. velutina* Lam.), and white oak (*Q. alba* L.) occur throughout the range of the type; other species of local importance include chestnut oak (*Q. prinus* L.), post oak (*Q. stellata* Wangenh.), and scarlet oak (*Q. coccinea* Muenchh.). Hickory species include pignut (*C. glabra* (Mill.) Sweet) and mockernut (*C. tomentosa* (Poir)Nutt.). Species in the midstory include sourwood (*Oxydendrum arboreum* (L.) DC.), eastern redbud (*Cercis canadensis* L.), and flowering dogwood (*Cornus florida* L.). The shrub layer consists of rosebay rhododendron (*Rhododendron maximum* L.) and mountain laurel (*Kalmia latifolia* L.). The oak-hickory type predominates on almost all landscape positions throughout its range. In mountainous areas, however, lower slopes and coves often have a greater proportion of mesophytic species, including yellow-poplar (*Liriodendron tulipifera* L.) and red maple (*Acer rubrum* L.). The type grades into northern hardwoods in the higher elevation and northern parts of its range and the oak-pine type in its southern range.

Climatic regime of the oak-hickory region is humid continental with generally short, cool winters and long, warm summers (Burns 1983). In the central part of the area occupied by the type, mean daily temperature averages about 12.8 ^{o}C and the frost-free period is around 180 days. Annual precipitation in the central part averages from 75 to 100 cm, about half of which occurs during the growing season. Periods of moisture deficit lasting from 2 to 6 weeks are common during the late summer. Annual snowfall averages about 25 cm.

The topography of the area consists mostly of low, open hills, although areas of steep relief occur in the Appalachian, Cumberland, and Ozark Mountains. The oak-hickory type is restricted to elevations below about 1,650 m on exposed slopes and ridges in the Southern Appalachians. Geologic parent material consists mostly of glacial material, residual sandstone, shales, and limestone, although gneisses and schists also occur. Soils range from cool-moist Spodosols to warm-dry Alfisols.

The data utilized in our study are for the central portion of the oak-hickory type in Kentucky and Tennessee, an area of about 213,000 km^2. The oak-hickory type in these states consists of areas where the plurality of tree species are upland oaks or hickory, and where associated species include yellow-poplar, elm (*Ulmus* L spp.), maple, and black walnut (*Juglans nigra* L.).

Ecoregions: This area includes ecoregions delineated at the upper three levels (i.e. domain, division, province) of the USDA Forest Service hierarchical

framework of ecological units (Cleland and others 1997) (figure 1). In this framework, domains occupy millions of hectares, divisions within a domain occupy hundreds of thousands of hectares, and provinces in a division occupy tens of thousands of hectares. Ecoregion delineation and nomenclature follow Keys and others (1995).

The Humid Temperate Domain, 200, includes all of the study area. This domain is characterized by sufficient precipitation to support forests of broadleaf deciduous and needleleaf evergreen trees, which constitute the dominant vegetation. This domain is subdivided into divisions based on characteristics of the climate during the winter (Bailey 1995). The study area contains two divisions: 220 and 230. Divisions delineate major changes in temperature regimes (Bailey 1995) and tend to be oriented in an east-west direction. The Hot Continental Division, 220, delineates an area of hot summers and cool winters and supports winter deciduous forest vegetation characterized by tall broadleaf trees. The climate of the Subtropical Division, 230, which lies south of 220, is characterized by lack of cold winters and by high humidity during the summer and also by a higher proportion of conifers (primarily *Pinus L.spp.*) in the forest vegetation. The dividing line between the 220 and 230 divisions represents an isotherm of about 22 ^{o}C for the warmest month. Most of the study area consists of division 220, which includes three provinces.

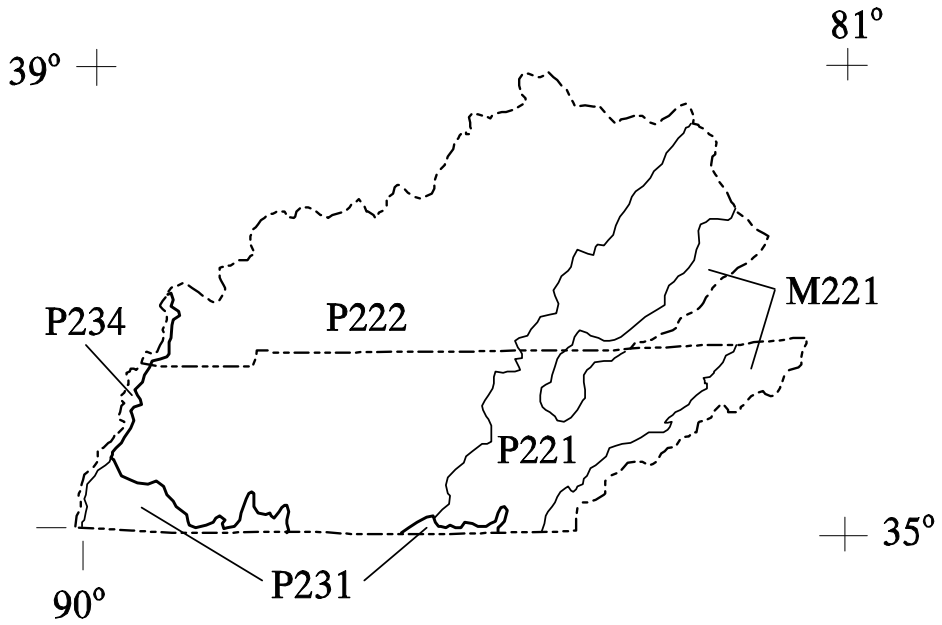


Figure 1: Ecoregions at the domain, division, and province hierarchical levels in the study area, which consisted of the states of Kentucky and Tennessee. Only ecoregions at the province level are identified. In the hierarchical structure of the Forest Service classification framework (Cleland and others 1997), domain 200 and divisions 220 and 230 are implicitly represented by the taxonomic structure of the numbering system. Thus, the entire area is contained in domain 200 and the wide boundary line along the southern and western parts of Tennessee separates divisions 220 and 230.

Provinces are components of a division that correspond to broad vegetation regions, and that conform to climatic sub-zones controlled primarily by continental weather characteristics such as length of dry season and duration of cold temperatures (Bailey 1995). Provinces are also characterized by similar soil orders and potential natural vegetation. Five provinces are represented in the study area, one mountainous (M221) and four with planar or plateau-type surfaces (P221, P222, P231, and P234).

Province M221 extends across the western slope of the Appalachian Mountains in Tennessee and includes a small part of the Cumberland Mountains in Kentucky. A landscape of high relief is characteristic of this province. Geologic formations generally consist of Precambrian metamorphosed gneisses and schists, although the Cumberland Mountains were formed by differential erosion of level-bedded sandstones. Vegetation is dominated by species of *Quercus*, although mesophytic species predominate in mountain valleys. Annual precipitation is greater here than in other provinces, primarily as a result of orographic effects. Nine percent of the study area is contained in M221. Except for the heavily populated broad intermountain basins, most of the province is forested, particularly the steep, mountainous slopes.

Province P221, Eastern Broadleaf Forest (Oceanic), is in the eastern part of the study area and makes up about 22 percent of the study area. Precipitation here is less than in M221, but greater than in the adjacent province to the west. Geology is variable, ranging from level-bedded sandstones on the Cumberland Plateau to the carbonate rocks of the Ridge and Valley geologic province. Parts of P221 have been weathered and dissected to present mountainous terrain. Altitude ranges from about 200 to 800 m. Vegetation consists of arborescent species dominated by oaks. Forests have been cleared from 30 to 40 percent of the land area for urban, pasture, and agricultural land uses.

Province 222, Eastern Broadleaf Forest (Continental), is the largest of the ecoregions under consideration, occupying 63 percent of the study area. Its geology consists of Paleozoic level-bedded sandstones and siltstones that have weathered to form topography of low, rolling hills and dissected plateaus. Vegetation is dominated by oaks and hickories. Forest cover ranges from about 45 percent in the eastern and central part of this province to 70 percent in the west.

A small part of P231 (Southeastern Mixed Forest) occurs along the southern boundary of Tennessee, and constitutes 4 percent of the study area. Geologic formations are irregularly bedded Quaternary and Cenozoic sands and clays. Forest canopy vegetation consists mostly of oaks and hickories; pines (*Pinus* L.spp.) are more prevalent here than in neighboring provinces. Land use is mostly forestry and agriculture.

Province 234 (Lower Mississippi Riverine Forest), the smallest of the five ecoregions included in the study (2 percent of the study area), borders the

Mississippi River and is characterized by flat, alluvial plains that are flooded periodically. Areas of late Pleistocene, wind-deposited loess occur and form a distinctive boundary with P222. A large portion of this province was formerly utilized for agricultural purposes, particularly areas that had more favorable soil moisture relations, but reforestation is now occurring (Schweitzer and Stanturf 1997). Forest vegetation includes a number of mesophytic species that can withstand periodic flooding, such as water tupelo (*Nyssa aquatica* L.), blackgum (*N. sylvatica* Marsh.), sweetgum (*Liquidambar styraciflua* L.), some oaks, and baldcypress (*Taxodium distichum* (L.) Rich.). The physical characteristics and vegetation of P234 are quite distinct from those of the other provinces.

In summary, the natural vegetation of the ecoregions studied consists primarily of forests with an upper and mid-canopy cover of arborescent species dominated by oaks. Divisions are most clearly distinguished by physiography and differences in vegetative composition. Vegetational differences among provinces within a division are less distinct. Bailey (1995), Griffith and others (1997), and Delcourt and Delcourt (2000) provide more information on ecological relationships in the study area. Ecoregions delineated at the division and province scales, unlike ecological units at lower hierarchical levels, are not repeated across the landscape. Approximately half of the two-state study area is forested.

Experimental Design

We designed our study with the assumption that the ecoregion (e.g. division, province) is the experimental unit, or the smallest portion of the study that receives a treatment. Treatments are the environmental conditions and biological processes used to define each ecoregion. Treatments are defined by physical factors -- climate (e.g. temperature, precipitation), geology, landform, and soils. Forest vegetation is the relatively stable biological characteristic that responds to the physical environmental treatment. NDVI was employed to quantify the biological response. The environmental treatment of each ecoregion was assumed to be a random effect.

Climate varies both with location within the study area and over time. Therefore, because NDVI quantifies the phenological state of vegetation, which we hypothesize varies among ecoregions as a function of climate, it was measured at two fixed times during the growing season.

Replication of the experimental units for statistical testing is problematic because ecoregions at the division and province hierarchical levels are unique and thus do not repeat across the landscape. We used years as replications and considered their effects to be random. Experimental error for testing treatments consisted of unexplained variation in NDVI resulting from phenological differences in vegetation at the measurement dates, unmeasured atmospheric conditions, and other variation that was not accounted for by treatments.

Ecoregion Analyses

The population of NDVI values within each ecoregion was sampled as values for the areas around the FIA inventory plot locations (Hansen and others 1992). The FIA plots were utilized to identify forested sample sites for determination of NDVI and to exclude landcover associated with nonforest uses (e.g. metropolitan, agricultural). A total of 5,399 forested plots were available for which the geographic coordinates had been determined to the nearest 100 seconds of latitude and longitude. One hundred seconds of latitude equals approximately 1.6 km. The median of nine NDVI values was determined from a 3x3 grid (3.3 x 3.3 km) centered at each FIA inventory plot to ensure inclusion of the plot in the sample area.

We determined NDVI at the FIA plot locations for 11 years. Each year, NDVI was sampled at the summer solstice (June 22) and the fall equinox (September 22). In the study area, these dates generally correspond to full expansion of foliage at the highest altitudes in late spring and also before initiation of leaf abscission, near the end of the growing season. In addition, the June date is representative of the annual period before soil moisture generally becomes limiting, and September is the period when moisture deficits may become maximized. The Palmer drought severity index (PDSI) was used to account for variation in annual and seasonal NDVI associated with meteorological drought (Palmer 1965). Published values of PDSI that had been calculated for each of the four climatic divisions in each state were used in the analysis.

We used analysis of variance to determine if variation in the response variable (NDVI) was affected by differing treatments, which were represented by ecoregions. We restricted tests of significance to ecoregions within the hierarchical structure of the classification. For example, we first used all data to compare divisions, and then conducted separate analyses of provinces within each division (i.e. two data sets were formed by subsetting on division). Orthogonal contrast tests were used to determine all differences among divisions and provinces.

NDVI Modeling

We used multiple regression on the linked FIA and NDVI data bases to investigate the effects of selected topographic, geographic, and vegetational variables on NDVI. We followed Hansen and others (1992) in selecting altitude and stand basal area as vegetational and topographic variables to characterize each inventory plot. Plot latitude and longitude were included as surrogate variables to account for variation in climate associated with geographic location within ecoregions. The response variable was NDVI averaged by season over the 11 years studied. Other variables in the FIA data set (i.e. aspect and slope gradient) were excluded from analysis because of the broad area represented by each sample plot (10.1 km^2).

Results

Inventory Plot Characteristics

A total of 5,399 FIA inventory plots (3,049 in Kentucky and 2,350 in Tennessee) were used in the analysis (table 1). Plots were well distributed throughout each ecoregion. Mean topographic elevations were greatest (563 m) in M221, the mountainous province, and least (90 m) in P234, near the Mississippi River. Elevations in P222 and P231 were similar, but variation was greater in P231. Stand basal area was about 2.8 m^2/ha higher for division 230 than for division 220. Among provinces, stand basal area was greatest (20.3 m^2/ha) in M221 and least (13.2 m^2/ha) in P222.

NDVI and Ecoregions

Mean summer (June) NDVI averaged 155.8 among provinces for the 11-year period studied (figure 2). The range in average summer NDVI across all ecoregions varied from about 167 in 1990 to 127 in 1994. M221 had the highest summer NDVI in most years, and P234 had the lowest summer NDVI every year. Annual direction and magnitude of change in NDVI was generally consistent among provinces for most years of the study, although there were notable exceptions in 1996 and 1997. The mean summer PDSI averaged across all provinces ranged from -0.6 to 4.4 during the study period. The PDSI pattern agreed with the pattern of NDVI for about half of the years, but differed visibly for other years (e.g. 1989, 1994, and 1997).

Ecoregions accounted for highly significant differences in average NDVI between divisions and provinces within divisions for the summer growing season (table 2). Similar patterns were apparent in the fall (September) data, but the relationships were not as strong. The multiple coefficient of determination (R^2) values of the analyses of variance for the summer and fall seasons were R^2 =

Table 1: Mean (±SD) topographic, geographic, and vegetation characteristics of forest inventory and analysis sample plots by ecoregion hierarchical level.

Ecoregion	Plots	Elevation	Latitude	Longitude	Basal area
	number	meters	- - - - - degrees - - - - -		m^2/ha
Domain 200	5399	299(161)	36.8(1.0)	85.5(1.9)	15.7(10.7)
Divisions in domain 200					
220	5222	302(161)	36.8(0.9)	85.4(1.8)	15.6(10.8)
230	177	195(123)	35.3(0.5)	88.4(1.4)	18.4(8.4)
Provinces in division 220					
M221	624	563(212)	36.4(0.7)	83.2(0.9)	20.3(10.0)
P221	1479	380(105)	36.7(0.9)	84.0(0.9)	18.6(9.7)
P222	3119	214(70)	36.9(1.0)	86.4(1.5)	13.2(10.8)
Provinces in division 230					
P231	145	217(124)	35.1(0.1)	88.1(1.4)	18.8(7.2)
P234	32	90(17)	36.3(0.6)	89.4(0.3)	16.5(12.3)

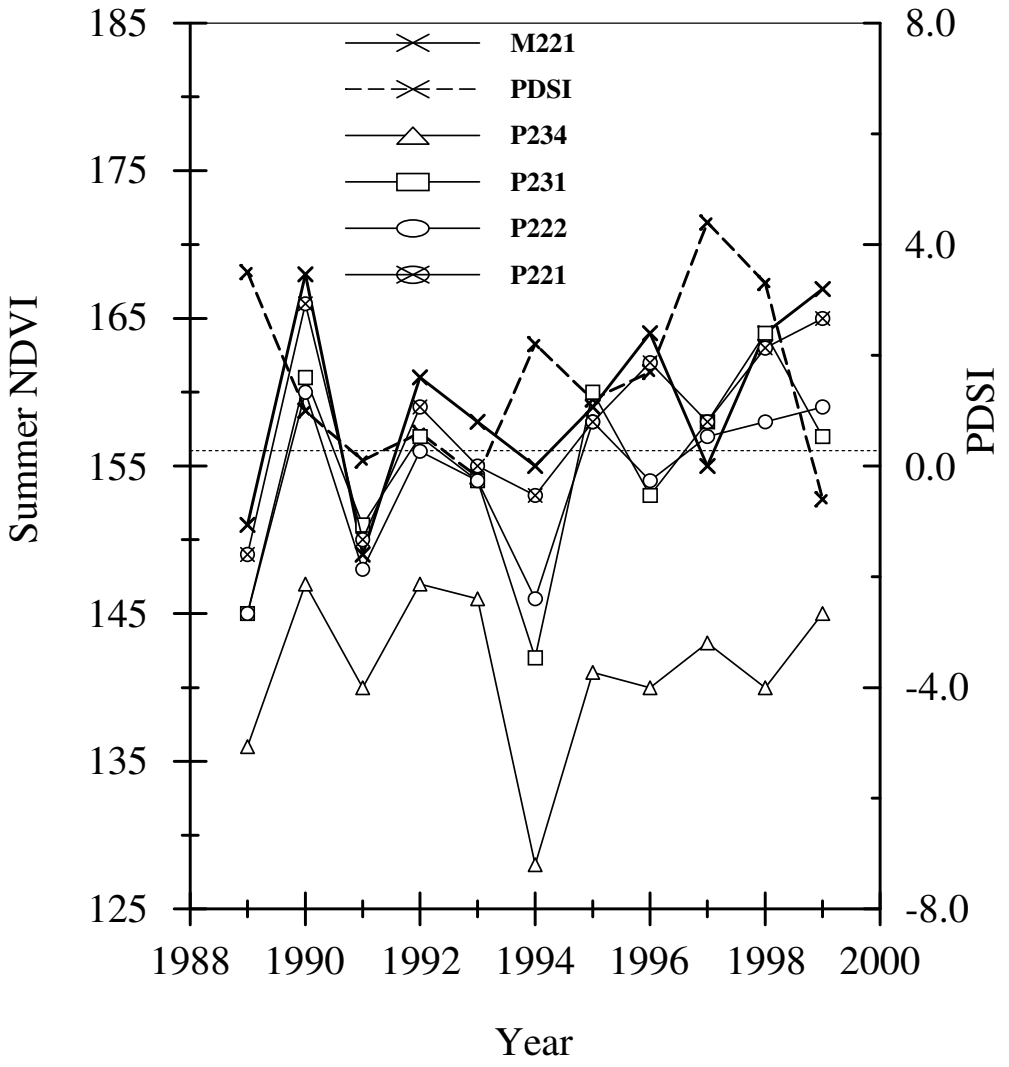


Figure 2: Mean normalized difference vegetation index (NDVI) and regional Palmer drought severity index (PDSI) at the summer solstice (June 22) by ecological province over the 11 years of the study. The long-term (11 years) mean summer NDVI is indicated by the horizontal dashed line.

0.9068 and $R^2 = 0.7436$ respectively. The mean summer division-to-division difference was approximately 10 NDVI units, which was significant at the P<0.0001 level. The three provinces in division 220 were M221, P221, and P222. Mean summer NDVI of M221 did not differ from that of P221 (P=0.1560). However, NDVI of P222 was significantly different from that of M221 (P<0.001) and that of P221 (P=0.002). The two provinces in division 230 were P231 and P234. Mean summer NDVI of P231 (154.7) was significantly different (P<0.0001) from that of P234 (141.1). NDVI for P231 and NDVI for P234 were not significantly different (P=0.7511) for the fall season.

Inclusion of PDSI as a covariable in the analysis of variance for the summer season explained a small but significant (P=0.0032) portion of the variation in

Table 2: Seasonal mean (±SD) NDVI and orthogonal contrasts among hierarchical ecoregions. [1]

Ecoregion	Summer (June 21)	Fall (September 21)
Divisions in domain 200		
220	157.2(6.1)a	153.8(6.2)a
230	147.0(9.2)b	151.6(5.6)b
Provinces in division 220		
M221	159.4(6.4)a	155.6(7.3)a
P221	158.1(5.7)a	153.8(6.1)ab
P222	154.1(5.4)b	151.9(4.8)b
Provinces in division 230		
P231	154.7(6.7)a	151.8(6.1)a
P234	141.1(5.6)b	151.3(5.3)a

[1]Within hierarchical ecoregion groups, means followed by the same letter are not significantly different (P<0.001).

mean NDVI within and between provinces. The coefficient for PDSI was 1.9998, and the value of R^2 increased from 0.907 to 0.926. PDSI was not significant (P=0.1689) when included as a covariable in an analysis of variance model for fall NDVI, and had a coefficient of -0.9708.

NDVI and Environmental Data

Regression analysis indicated that phenological, topographic, vegetative, and geographic variables were highly significant (P<0.001) sources of variation in NDVI at all levels of the hierarchical framework (table 3). Among provinces, variation in NDVI explained ranged from R^2=0.206 for P222 to R^2=0.553 for M221. Among the independent variables, time of measurement during the growing season (i.e. June 22 or September 22) explained the largest proportion of variation, followed by elevation and stand basal area. Effects of season, elevation, and stand basal area were generally consistent in magnitude and sign of the variable coefficients: NDVI increased with increasing elevation and stand basal area primarily as a result of greater quantities of green vegetation.

Season of determination was not significant for division 230 because of an interaction between the two provinces. As indicated by the signs of the coefficients for season, NDVI decreased from summer to fall for P231, but increased for P234. Effect of season was consistent among the other ecoregions.

Discussion

NDVI and Ecoregions

This study demonstrated that ecoregions explained significant variation in NDVI at several levels of a hierarchical classification framework in the humid eastern US, where landcover consists of forests dominated by relatively uniform arborescent vegetation. Our results parallel those reported by Ramsey and others (1995), who found that nonhierarchical ecoregions of the arid environments of

Table 3: Regression coefficients of season, topographic, vegetation, and geographic variables and fit statistics (R^2, coefficient of correlation; $S_{Y \cdot X}$ standard error of mean) in models of NDVI by hierarchical ecoregion.[1]

Ecoregion	Season	Elevation	B.A.	Latitude	Longitude	R^2	$S_{Y \cdot X}$
		meters	m^2/ha	- - - - degrees - - - - -			
Domain 200	-2.901*	0.010*	0.091*	-0.231*	0.306*	0.313	4.212
Divisions in domain 200							
220	-2.972*	0.009*	0.095*	-0.220*	-0.276*	0.319	4.134
230	-0.714	0.031*	0.016*	-1.917*	-0.537	0.320	5.562
Provinces in division 220							
M221	-3.745*	0.010*	0.046*	1.723*	-2.186*	0.482	2.889
P221	-4.241*	0.013*	0.039*	2.681*	-1.678*	0.474	2.924
P222	-2.215*	0.025*	0.115*	-0.618*	-0.066	0.265	4.388
Provinces in division 230							
P231	-2.928*	0.029*	-0.032	-5.515*	-0.750*	0.430	4.166
P234	10.438*	0.181*	0.123*	-15.680*	23.661*	0.468	7.156

[1]Level of significance: *=P<0.01.
[2]B.A.: Basal area of stand.

Utah were significantly related to NDVI. Mean NDVI was higher and more variable in our study than in the Utah study. NDVI in the oak-hickory type averaged about 156(±9s.d.) for summer and 153(±8s.d.) for fall whereas NDVI in the Utah study was about 118(±5s.d.) (Ramsey and others 1995). Minimum NDVI was similar for both areas, about 100, indicating a lack of photosynthesizing vegetation. Maximum NDVI, however, ranged from 169 to 178 across the deciduous forested ecoregions and from 115 to 148 in the ecoregions in Utah. Because NDVI is a surrogate variable for on-site quantification of vegetation (Loveland and others 1995), our results suggest that there are ecological differences among most ecoregions in our study area.

One might draw useful conclusions from a single year's results if by-province NDVI rankings did not change from year to year. Ramsey and others (1995) found little annual variation in NDVI, but we found substantial switching in year-to-year rankings of ecoregional NDVI means. This variation served as the experimental error in our tests of significance. We hypothesize that year-to-year differences in NDVI at the season date of measurement could result from a combination of varying annual temperature and precipitation regimes. Spring temperatures that are cooler or warmer than normal would affect development of foliage by the June 22 measurement date, resulting in less or more green photosynthesizing vegetation than average for imaging based on NDVI. In a similar manner, annual variation in the soil moisture regime affects the quantity of vegetation present.

Atmospheric effects are a possible cause of annual variation in NDVI values within and among ecoregions. The NDVI imagery used in our study was uncorrected for atmospheric effects of water vapor, ozone, and Rayleigh scattering. NDVI band 2 (0.75-1.1μm) is particularly sensitive to increased levels of atmospheric water vapor, which tends to reduce estimated NDVI values. The use of biased values of NDVI can reduce the accuracy of estimates of vegetation

moisture content and other predictions. NDVI was not used for predictive purposes in our study. Also, the use of biweekly NDVI composites tends to reduce atmospheric effects because the greenest pixel is used in such composites, which suggests that the image was made when there was the least amount of water vapor in the atmosphere. Clearly, however, use of atmospherically corrected NDVI values is desirable.

Our use of PDSI to quantify meteorological drought accounted for small but significant variation in annual NDVI. Walsh (1987) also found that NDVI was correlated with PDSI. A measure of air temperature for each season by year would likely account for additional variation in NDVI. Annual variation in soil moisture availability may account for some of the variation in NDVI in our study. The PDSI indicated a surplus of soil moisture (PDSI>0.5) for broad, state climatic divisions in 6 years and a deficit (PDSI<-0.5) in 3 years. Availability of refined values of PDSI from the weather station nearest each FIA plot would likely have shown stronger relationships in the analysis of covariance.

Because FIA plots are established in forested locations, we were able to use the locations of FIA plots to subsample NDVI values within ecoregions. However, the positional accuracy of these plots to only 0.01 seconds is an important limitation in their usefulness. This lack of accuracy caused us to expand our sample area to the eight neighboring cells surrounding the cell in which the FIA plot supposedly fell. Despite the relative lack of accuracy of inventory plot location and the greater likelihood of including nonforested areas for NDVI determination in the expanded sample area, we consider this method as superior to the total polygon method. That method uses all cells in a polygon for analysis, which could include many that are nonforested, such as urban areas and agricultural fields, thereby potentially increasing variation. Marked improvement in results should be possible with increased location accuracy of plots. One possible drawback to this rationale, however, is exclusion of NDVI values associated with land uses other than forests, a classification factor included in some ecoregion map delineations (Omernik 1995).

NDVI and Environmental Variables

Our study demonstrates a significant association between NDVI and characteristics of field inventory plots. Stand basal area consistently and significantly accounted for small amounts of variation in NDVI. We assume that the relationship would have been stronger if sampling had been restricted to a single 1.1 km cell, instead of the 3.3 km^2 grid that was used. Our findings agree with those of Kremer and Running (1993) who found that characteristics of vegetation on ground-sampled plots were correlated with 1.1 km NDVI imagery.

The explanation of the significance of the other topographic and geologic variables is logical. Elevation was directly associated with NDVI in all of the ecoregions we studied, and is associated with length of growing season, initiation

of spring growth, and onset of winter dormancy, as have been reported in other studies (Ramsey and others 1995, Burgan and Hartford 1993, Loveland and others 1995). Latitude is likely related to variation in temperature within ecoregions and to variation in NDVI. The effect of longitude on NDVI variations is probably related to precipitation, which tends to increase from west to east. In our study area, longitude is probably more closely associated with precipitation than with temperature.

Conclusions

Remote sensing data can be used to test the relatively uniform, humid environments of the eastern U.S. for uniqueness in some vegetational characteristics. Our test was designed to determine if ecoregions differed based on measurements of NDVI, which quantifies the amount of photosynthesizing vegetation present. Other characterization criteria of ecoregions, including plant species composition and water quality, have proven useful for ground-based validation testing, but NDVI is one of the few variables that can be applied using remote sensing. We found that one difficulty associated with the use of NDVI in the forested eastern U.S. is annual variation in its magnitude. This variation could result from subtle year-to-year climatic effects on vegetation attributable to annual variation in temperature and moisture regimes, or to varying atmospheric conditions associated with a humid climate. Multiple years of NDVI data are needed both to avoid possible bias arising from analysis of atypical data and as a source of experimental error. Results of our study in deciduous hardwoods confirm and extend those of other studies in which NDVI was used to test ecoregion maps in arid environments.

The use of FIA plot locations provides data of sufficient sensitivity for testing and reducing variation between ecological map units. Other studies have used polygons in which all cells are utilized for computation of a single average value of NDVI for a particular time period. This total polygon method adds variation in NDVI that comes from nonforest conditions. However, the large number of sample sites based on inventory plot locations provided data for testing of the within-polygon relationship NDVI has with environmental factors. FIA plots also provided a large amount of useful ancillary information.

NDVI should not be the sole criterion for testing ecoregion hypotheses. However, NDVI evaluation could be used to identify parts of ecoregions warranting further investigation or review. NDVI should be considered a tool for preliminary screening. NDVI and inventory plot analysis is a relatively inexpensive and readily available tool that could be used by persons who do not have knowledge of GIS techniques and expensive GIS software. Although we used relatively simple methods, additional more sophisticated techniques are available, such as correcting for variation in atmospheric conditions.

In conclusion, our study provides methodology for using inventory data and remote sensing imagery for testing hypotheses about ecoregion map units. The use of FIA inventory plots to provide specific locations for sampling and information on the vegetation present, in conjunction with NDVI data obtained by remote sensing, provides a method suited to testing of ecoregion maps. The purpose of this investigation was not to develop or refine mapped ecological units, but to examine the potentials of NDVI data as a tool for use in the future development and testing of map units. Our results suggest that NDVI may be well suited for this purpose.

Acknowledgments

We gratefully acknowledge the contribution of Roberta Bartlett, Forester, USDA Forest Service, Missoula, MT who provided archived NDVI imagery and suggested the use of WinDisp3 and WinDisp4 software for image processing and data summarization. We acknowledge also the contribution of Tracy Roof, Lead Forestry Technician, USDA Forest Service, Asheville, NC who provided ancillary data sets required for the analyses, typically on short notice.

References

Bailey, R.G. 1984. Testing an ecosystem regionalization. Journal of Environmental Management. 19:239-248.

Bailey, R.G. 1995. Description of the ecoregions of the United States. 2d ed. rev. and expanded (1st ed. 1980). Misc. Publ. No. 1391 (rev). Washington, DC: USDA Forest Service. 108 p. with separate map (scale 1:7,500,000).

Bailey, R.G. 1998. Ecoregions. New York: Springer-Verlag. 176 pp.

Braun, E.L. 1950. Deciduous forests of eastern North America. Philadelphia, PA: The Blakiston Company. 596 pp.

Brown, D.E.; Reichenbacher, F.; Franson, S.E. 1998. A classification of North American biotic communities. The University of Utah Press. Salt Lake City. 141 pp.

Burgan, R.E.; Chase, C.H. 1997a. NDVI and derived image data: Data archives 1989. Gen. Tech. Rep. INT-GTR-354CD. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Burgan, R.E.; Chase, C.H. 1997b. NDVI and derived image data: Data archives 1990. Gen. Tech. Rep. INT-GTR-355CD. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Burgan, R.E.; Chase, C.H. 1997c. NDVI and derived image data: Data archives 1991. Gen. Tech. Rep. INT-GTR-356CD. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Burgan, R.E.; Chase, C.H. 1997d. NDVI and derived image data: Data archives 1992. Gen. Tech. Rep. INT-GTR-357CD. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Burgan, R.E.; Chase, C.H. 1997e. NDVI and derived image data: Data archives 1993. Gen. Tech. Rep. INT-GTR-358CD. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Burgan, R.E.; Chase, C.H. 1997f. NDVI and derived image data: Data archives 1994. Gen. Tech. Rep. INT-GTR-359CD. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Burgan, R.E.; Chase, C.H. 1997g. NDVI and derived image data: Data archives 1995. Gen. Tech. Rep. INT-GTR-360CD. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Burgan, R.E.; Chase, C.H. 1997h. NDVI and derived image data: Data archives 1996. Gen. Tech. Rep. INT-GTR-361CD. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Burgan, R.E.; Chase, C.H. 1998. NDVI and derived image data: Data archives 1997. Gen. Tech. Rep. RMRS-GTR-9-CD. Ft. Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Burgan, R.E.; Chase, C.H.; Bartlette, R.A. 1999. NDVI and derived image data: Data archives 1998. Gen. Tech. Rep. RMRS-GTR-27-CD. Ft. Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Burgan R.E.; Hartford, R.A. 1993. Monitoring vegetation greenness with satellite data. Gen. Tech. Rep. INT-GTR-297. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 13 p.

Burns, R.M. 1983. Silvicultural systems for the major forest types of the United States. Agriculture Handbook 445. Washington, DC: USDA Forest Service. 191 p.

Cleland, D.T., and others. 1997. National hierarchical framework of ecological units. In.Boyce, Mark S.; Haney, Alan, eds. Ecosystem management: applications for sustainable forest and wildlife resources. New Haven and London: Yale University Press: 181-200.

Delcourt, H.R.; Delcourt, P.A. 2000. Eastern deciduous forests. In: Barbour, M.G.; Billings, W.D., eds. North American terrestrial vegetation. 2nd edition. Cambridge, U.K.: Cambridge University Press: 357-395.

Edwards, Jr., T.C.; Moisen, G.G.; Cutler, D.R. 1998. Assessing map accuracy in a remotely sensed, ecoregion-scale cover map. Remote Sensing of Environment 63:73-83.

FAO 1999. WinDisp4 Map and Image Display and Analysis Software, User's Manual. Technical Report ES:GCP/INT/619/EC. Food and Agriculture Organization of the United Nations. Rome. 103 p.

Gutman, G.G. 1991. Vegetation indices from AVHRR data: an update and future prospects. Remote Sensing of Environment 35:121-136.

Griffith, G.E.; Omernik, J.M.; Azevedo, S.H. 1997. Ecoregions of Tennessee. EPA/600/R-97/022. U.S. Environmental Protection Agency, Environmental Research Laboratory: Corvallis, OR. 49 p.

Griffith, G.E.; Omernik, J.M.; Woods, A.J. 1999. Ecoregions, watersheds, basins, and HUCs: How state and federal agencies frame water quality. Journal of Soil and Water Conservation. 54:666-677.

Handcock, R.N. 2000. Spatio-temporal analysis of net primary production across Ontario using an ecoregionalization. Ph.D. Thesis. University of Toronto. 153 pp.

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

Hargrave, B. 1993. The upper levels of an ecological classification system for Minnesota. Department of Natural Resources. Minneapolis, MN. 27 pp.

Hargrove, W.W.; Hoffman, F.M. 1999. Using multivariate clustering to characterize ecoregion borders. Computing in Science & Engineering 1(4):18-25.

Haynes, R.W.; Graham, R.T.; Quigley, T.M. 1998. A framework for ecosystem management in the interior Columbia basin. Journal of Forestry 96(10):4-9.

Hill, M.; Donald, G.E.; Hyder, M.W.; Smith, R.C.G. 2004. Estimation of pasture growth rate in the south west of Western Australia from AVHRR NDVI and climate data. Remote Sensing of Environment 93:528-545.

Holben, B.N. 1986. Characteristics of maximum-value composite images from temporal AVHRR data. International Journal of Remote Sensing 7:1417-1434.

Homoya, M.A.; Abrell, D.B.; Aldrich, J.R.; Post, T.W. 1985. The natural regions of Indiana. Indiana Academy of Science 94:245-268.

Host, G.E.; Polzer, P.L.; Mladenoff, ; D.J. White, M.A.; Crow, T.R.. 1996. A quantitative approach to developing regional ecosystem classifications. Ecological Applications 6:608-618.

Huete, A.R. 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment 25:295-309.

Huang, S. 1999. Ecoregion-based individual tree height-diameter models for lodgepole pine in Alberta. Western Journal of Applied Forestry. 14(4):186-193.

Keys, Jr., J., and others. 1995. Ecological units of the eastern United States - first approximation (map and booklet of map unit tables). Atlanta, GA. USDA Forest Service, presentation scale 1:3,500,000; colored.

Kremer, R.G., and S.W. Running. 1993. Community type differentiation using NOAA/AVHRR data within a sagebrush-steppe ecosystem. Remote Sensing of Environment 46:311-318.

Loveland, T.R.; Merchant, J.W.; Ohlen, D.O.; Brown, J.F. 1991. Development of a land-cover characteristics database for the conterminous United States. Photogrammetric Engineering and Remote Sensing 57(11):1453-63.

Loveland, T.R.; Merchant, J.W.; Brown, J.F.; Ohlen, D.O.; Reed, B.C. Olson, P. J. Hutchinson, J. 1995. Seasonal land-cover regions of the Unites States. Annals of the Association of American Geographers 85(2):339-355.

Maselli, F.; Romanelli, S; Bottai, L.; Zipoli, G.. 2003. Use of NOAA_AVHRR NDVI images for the estimation of dynamic fire risk in Mediterranean areas. Remote Sensing of Environment 86:187-197.

Meng, Q., Cieszewski, C.J., Madden, M., Borders, B. 2007. A linear mixed-effects model of biomass and volume of trees using Landsat ETM+ images. Forest Ecology and Management 244:93-101.

Omernik, J.M. 1987. Ecorgions of the conterminous United States. Annals of the Association of American Geographers 77:118-125.

Omernik, J.M. 1995. Ecoregions: A spatial framework for environmental management. In: Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making, W.Davis and T. Simon (Eds.). Lewis Publishers, Boca Raton, Florida, pp. 49-62.

Palmer, W.C. 1965. Meteorological drought. Research Paper No. 45. U.S. Department of Commerce, Washington, D.C. 58 p.

Ramsey, R.D.; Falconer, A.; Jensen, J.R.,. 1995. The relationship between NOAA-AVHRR NDVI and ecoregions in Utah. Remote Sensing of Environment 53:188-195.

Rowe, J.S.; Sheard, J.W. 1981. Ecological land classification: a survey approach. Environmental Management 5:451-464.

Rudis, V.A. 1999. Ecological subregion codes by county, conterminous United States. Gen. Tech. Rep. SRS-36. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 95 p.

Schreuder, H.T.; Czaplewski, R.; Bailey R.G. 1999. Combining mapped and statistical data in forest ecological inventory and monitoring-supplementing an existing system. Environmental Monitoring and Assessment 56:269-291.

Schroeder, P.; Brown, S.; Mo, J. and others. 1997. Biomass estimation for temperate broadleaf forest of the United States using inventory data. Forest Science 43:424-434.

Schweitzer, C.J., Stanturf, J.A. 1997. From okra to oak: reforestation of abandoned agricultural fields in the lower Mississippi alluvial valley. In: Meyer, Dan A., ed. Proceedings of the twenty-fifth annual hardwood symposium; 25 years of hardwood silvicluture: a look back and a look ahead; 1997 May 7-10; Cashiers, NC. Memphis: National Hardwood Lumber Association: 131-138.

Singh, R.P.; Roy, S.; Kogan F. 2003. Vegetation and temperture condition indices from NOAA AVHRR data for drought monitoring over India. International Journal of Remote Sensing 24:4393-4402.

Sivanpillai, R., Smith, C.T., Srinivasan, R., Messina, M.G., Wu, X.B. 2006. Estimation of managed loblolly pine stand age and density with Landsat ETM+ data. Forest Ecology and Management 223:247-254.

Smith, M-L.; Carpenter, C. 1996. Application of the USDA Forest Service national hierarchical framework of ecological units at the sub-regional level: the New England-New York example. Environmental Monitoring and Assessment 39:187-198.

Snyder, S.A.; Tyrrell, L.E.; Haight, R.G. 1999. An optimization approach to selecting research natural areas in national forests. Forest Science 45(3):458-469.

Song, C.; Woodcock, C.E.; Seto, K.C.; Lenney, M.P.; Macomber, S.A. 2001. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects. Remote Sensing of Environment 75:230-244.

The Nature Conservancy. 2001. Ecoregions of the United States of America. Arlington, VA: The Nature Conservancy. Map.

USDA Forest Service. 2000. NDVI and derived image data: Data archives 1999. Boise, ID: U.S. Department of Agriculture, Forest Service, National Interagency Fire Center. [CD-ROM format available from National Interagency Fire Center, 3833 S. Development Avenue, Boise, ID 83705.]

Walsh, S.J. 1987. Comparison of NOAA AVHRR data to meteorologic drought indices. Photogrammetric Engineering and Remote Sensing 53:1069-1074.

Wang, Q., Adiku, S.; Tenhunen, J.; Granier, A. 2005. On the relationship of NDVI with leaf area index in a deciduous forest site. Remote Sensing of Environment 94:244-255.

Wessels, K.J.; Prince, S.D.; Frost, P.E.; van Zyl, D. 2004. Assessing the effects of human-induced land degradation in the former homelands of northern South Africa with a 1 km AVHRR NDVI time-series. Remote Sensing of Environment 91:47-67.

Wright, G.R.; Murray, M.P.; Merrill, T. 1998. Ecoregions as a level of ecological analysis. Biological Conservation 86:207-213.

Biomass Estimation



Improving North American Forest Biomass Estimates from Literature Synthesis and Meta-Analysis of Existing Biomass Equations

David C. Chojnacky[1]; Jennifer C. Jenkins[2], and Amanda K. Holland[2]

***Abstract**: Thousands of published equations purport to estimate biomass of individual trees. These equations are often based on very small samples, however, and can provide widely different estimates for trees of the same species. We addressed this issue in a previous study by devising 10 new equations that estimated total aboveground biomass for all species in North America (Jenkins et al. 2003). We selected 318 total biomass equations (out of 2,626 identified in the literature), based on applicability for estimating biomass from only diameter measurements, and used a modified meta-analysis to develop new equations. This was done by using regression analysis on data generated from those 318 equations for tree sizes within the diameter bounds of the original data. We also included two sets of ratio equations, for hardwood and softwood species, to separate out biomass of different tree components—foliage, branches, bark, and roots.*

The Joint Fire Science Program funded this work to create more generalized biomass equations for regional fire-fuels managers by extending our literature synthesis. We are updating our work with literature published through 2008. We will devise new equations to allow for more differences within North American regions and provide greater accuracy for local use. The new analysis will use equations from the literature that include height as well as diameter, but will result in biomass equations that do not require height as an input variable.

The results of our analyses to date suggest that allometric scaling theory may be applied in future studies to more accurately estimate tree biomass from diameter and whole-tree density measurements.

Keywords: allometric equations, tree biomass, biomass estimation, tree carbon

[1]Virginia Tech, Falls Church, VA; Visiting Professor, University of Vermont, Burlington, VT
[2] University of Vermont, Burlington, VT

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

We previously developed a series of national-scale tree biomass estimation equations that can be used with inventory data for producing forest carbon budgets. We are updating this work for the Joint Fire Science Program, in an effort to develop generalized estimators for regional land management use. In the future, we plan to use allometric scaling theory to resolve issues associated with estimation of forest biomass.

Previous Study

Equation Compilation

In our previous work (Jenkins et al. 2003), we compiled biomass equations from literature published through 1999. Our resulting database included 2,626 biomass equations for either total tree biomass or tree components including bole, bark, branches, foliage, and roots. From the equation database, we determined 318 diameter-based equations for over 100 species from 104 published sources were useful for our analysis for estimating total aboveground biomass. In a modified meta-analysis, we generated data for each published equation at 5-cm intervals within the diameter ranges of the original equations—resulting in what we called "pseudodata" (following the concept pioneered by John Tukey on jackknife estimation [Mosteller and Tukey 1977]). We then fit these pseudodata into 10 species-group-specific equations by using regression and log transformation.

Modeling pseudodata

Using a simple log form (Eq. 1) we modeled pseudodata with 5 equations for conifer species groups (cedar and larch, Douglas fir, pine, spruce, and true firs and hemlocks), 4 equations for hardwood groups (aspen, alder, and poplar; soft maple and birch; maple, oak, and hickory; and mixed hardwood), and a woodland equation including both conifer and hardwood dryland forest species.

$$\ln BM = a + b \ln dbh \qquad (1)$$

Where

BM = total biomass,
dbh = stem diameter at breast height

Pine Pseudodata

Data from the pine species group illustrate our rationale for combining species into broad groups that corresponded to overlapping ranges within the pseudodata. The widely planted loblolly pine in the Southeastern United States has a biomass range that includes that of the much different pinyon and lodgepole pines found in the Western United States (Figure 1). This observed variation does not seem to be explained by other factors such as tree growth form, site, or climate; three distinct patterns would have been expected.

Because collecting new biomass samples across the full continental range of species and regions would be impractical, we concluded that fitting new equations using species-grouped pseudodata represented the most reasonable option for producing a consistent set of biomass equations for all North American species. We sorted through all pseudodata and distinguished 10 species groups for final equation development.

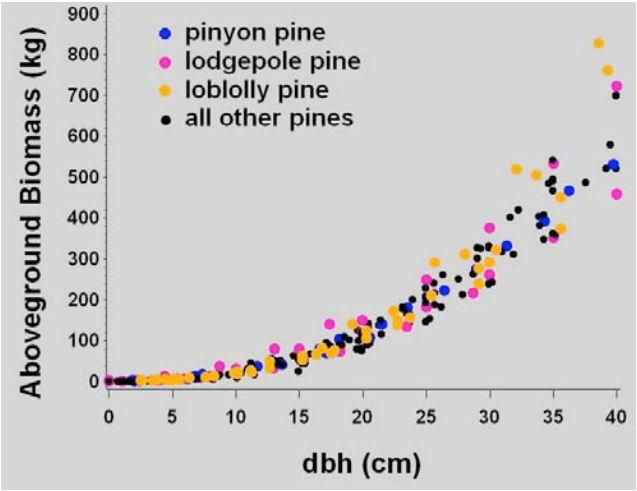


Figure 1: Modeled pseudodata depicting aboveground biomass of pinyon, lodgepole, and loblolly pines show no distinct pattern for separating species (biomass expressed in kg as a function of stem diameter at breast height; from Jenkins et al. [2003]).

Component ratio equations

We also developed generalized hardwood and conifer ratio equations to partition the estimates into foliage, stem bark, stem wood, and coarse root components (Figure 2). The hardwood and conifer models are similar except the conifer model shows less branch variation with diameter and more foliage variation.

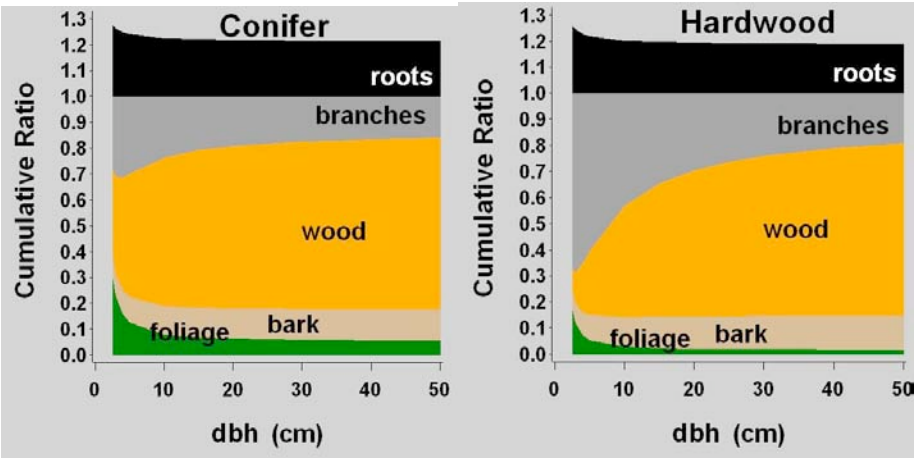


Figure 2: Biomass component ratios for conifer and hardwood species as a function of stem diameter at breast height, from equations in Jenkins et al. (2003).

Because pseudodata for biomass components were much more variable than for total biomass, and because there were few studies that measured biomass components in a consistent fashion, all pseudodata were grouped into two general models (hardwood and conifer). The resulting ratio equations serve as a first approximation until more data are available for detailed modeling of component biomass. The woodland species group, which included both conifer and hardwood species, was omitted from this analysis; there were not enough component equations in the literature for analysis with these species.

FIA Data Comparison

We compared tree-level data results from the ten equations against county-scale biomass calculations from the old Eastwide FIA database from the 1990s (Hansen et al. 1992) by subtracting Hansen et al. (1992) biomass estimates from the Jenkins et al. (2003) biomass estimates, and then dividing by the latter and multiplying by 100 to express the results as a percentage (Figure 3).

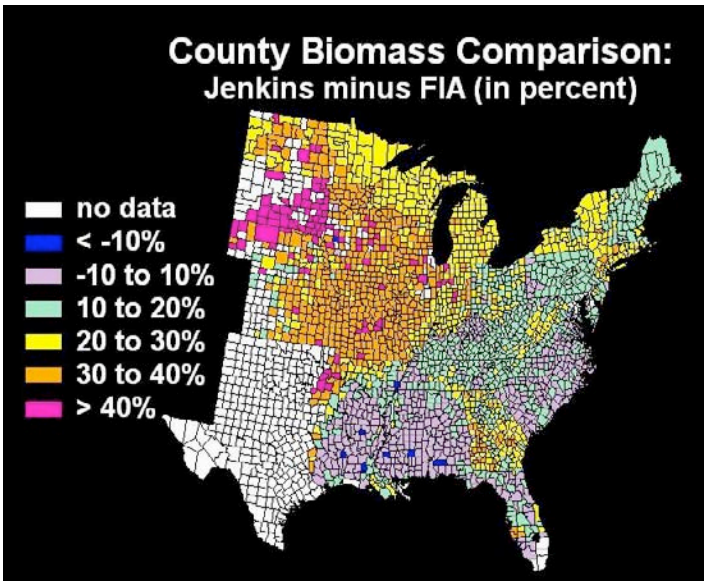


Figure 3: Amount Jenkins et al. (2003) biomass estimates generally exceed those of FIA (Hansen et al. 1992), expressed as a percent of Jenkins et al. (2003) estimates.

It is difficult to say which equations are better. Jenkins et al. (2003) equations used consistent methodology throughout, which is a plus, but unidentified, unquantified biases for certain species may exist. In the FIA approach, each FIA unit uses the best methods for a particular state or region. However, this is less consistent, particularly when equations change at state boundaries. For example, the map patterns in Figure 3 showing straight-line boundaries—particularly in the southeastern United States—suggest variation introduced by biomass equation choice. Also, FIA estimates should theoretically be a little smaller than those of Jenkins et al. (2003) because they do not include foliage, but foliage should account for only about 5%.

Current Work

Our current objective is more focused toward regional biomass estimation for fire fuels managers, and will likely result in more than 10 equations. We are updating the literature synthesis to include data published through 2008.

We are also going back to the earlier literature to include previously omitted equations that use both diameter and height as predictor variables. Our initial plan is to generate pseudodata as before—based on diameter only—but to use a height-to-diameter equation for inclusion of height variation. We plan to use FIA data to develop a series of simple height-to-diameter equations for this purpose. This will allow use of biomass equations that include height but will also allow us to fit final regression equations from pseudodata from diameter only. Equations for predicting shrub biomass may be added to our analyses if resources permit.

Allometric Scaling Theory Applied To Biomass Estimation

Allometric Scaling Theory

Allometric scaling theory may hold promise for biomass estimation. This theory explains why dimensions of biological organisms are scaled in proportion to one another (Enquist et al. 1998, 1999; West et al. 1997, 1999a, 1999b). An analysis based on this theory could greatly improve our meta-analysis work. Tree biomass and volume are proportional to dbh, as are a host of other tree measures, but to date there is no coherent theory explaining exactly why this might be true. Allometric scaling theory uses fractal dimensions of tree architecture and physics of fluid transport up a tree to offer an explanation in terms of a generalized biomass model. Parameterization of this model is postulated to hold regardless of tree species and regardless of tree size.

Generalized Tree Biomass Equation

The biomass equation described by allometric scaling theory (Eq. 2) suggests that the exponent on dbh is not a parameter to be estimated; rather, it is fixed at eight-thirds or 2.67.

$$\text{Biomass} = (C\ \rho)\ \text{dbh}^{2.67} \qquad (2)$$

Where

C = proportionality constant
ρ = component-integrated tree specific gravity

The measure of specific gravity included here is not specific gravity of bole wood commonly reported, but a measure of specific gravity integrated over the entire tree and including bark, wood, branches, and foliage. We refer to this as "component-integrated tree specific gravity," or "CIT specific gravity." It would be difficult to obtain CIT specific gravity for every tree species in the United States. However, as the scaling theory is postulated to hold for small and large trees, CIT specific gravity might be measured for saplings of each species rather

easily and cost-effectively. The proportionality constant C is postulated to be a constant independent of species.

Allometric Theory Equation Applied

The 10 equations from Jenkins et al. (2003) can be expressed as a constant (α) times dbh raised to β, where β is the scaling coefficient in the allometric scaling model (Table 1). In 9 of our 10 biomass equations, the scaling parameter estimated from pseudodata is fairly close to the theoretical eight-thirds or 2.67 for the allometric scaling model. Only the woodland equation differs markedly; this is likely due to inconsistent biomass measurements for a diverse group of evergreen hardwood and some softwood species that grow in Southwestern U.S. dryland forests. Thus, anecdotal evidence indicates that allometric scaling theory may provide a sound theoretical basis for tree biomass estimation, but field measurements of biomass and whole-tree specific gravity of some trees are needed in order to apply it.

Table 1: Estimated β-values for Jenkins et al. (2003) equations expressed in the form of the allometric scaling equation: Biomass = $\alpha\ dbh^{\beta}$, where dbh = stem diameter at breast height. In the generic allometric scaling model equation, β = 8/3 or 2.67.

Conifer		Hardwood	
Species Group	**β**	**Species Group**	**β**
Cedar and Larch	2.3	Aspen and Alder	2.4
Douglas Fir	2.4	Soft Maple and Birch	2.4
Pine	2.4	Maple, Oak, and Hickory	2.4
Spruce	2.3	Mixed Hardwood	2.5
True Fir and Hemlock	2.5		
		Woodland (both conifer and hardwood)	1.7

Summary

The goal of our work is to improve forest biomass estimates. To that end, we are pursuing the following:

- An update of our earlier database and synthesis of individual tree biomass equations for North America
- Development of regional-scale biomass equations that will include both diameter and height variation
- Future research considering allometric scaling theory as the basis for biomass estimation

Acknowledgments

We are grateful to the Joint Fire Science Program for fully funding our study. Scientific Notations LLC provided valuable writing and editing services.

References

Enquist, B.J.; Brown, J.H.; West, G.B. 1998. Allometric scaling of plant energetics and population density. Nature 395:163–166.

Enquist, B.J.; West, G.B.; Charnov, E.L.; Brown, J.H. 1999. Allometric scaling of production and life history variation in vascular plants. Nature 401:907–911.

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

Jenkins, J.C.; Chojnacky, D.C.; Heath, L.S.; Birdsley, R.A. 2003. National-scale biomass estimators for United States tree species. Forest Science 49(1):12–30.

Mosteller, F.; Tukey, J.W. 1977. Data analysis and regression: a second course in statistics. Addison-Wesley Series in Behavioral Science: Quantitative Methods. Reading, Mass.: Addison-Wesley; p. 133–164.

West, G.B.; Brown, J.H.; Enquist, B.J. 1997. A general model for the origin of allometric scaling laws in biology. Science 276:122–126.

West, G.B.; Brown, J.H.; Enquist, B.J. 1999a. The fourth dimension of life: Fractal geometry and allometric scaling of organisms. Science 284:1677–1679.

West, G.B.; Brown, J.H.; Enquist, B.J. 1999b. A general model for the structure and allometry of plant vascular systems. Nature 400:664–667.

Landscape-Level Analyses I



Relationships between forest structure, composition, site, and spruce beetle occurrence in the Intermountain West

DeRose, R. Justin[1,2]

Long, James N.[3]

Shaw, John D.[4]

Abstract: *Engelmann spruce forests are structurally and compositionally diverse, occur across a wide range of physiographic conditions, and are the result of varying disturbance histories such as fire, wind and spruce beetle. The spruce beetle is a natural disturbance agent of spruce forests and has population levels that fluctuate from endemic to epidemic. Conceptually, structural and compositional diversity of Engelmann spruce forests are thought to provide resistance to spruce beetle activity, and have traditionally been used to assess stand risk to the beetle. We used FIA data from nine states in the Intermountain West to test the effect of stand structure, composition, and site attributes on the occurrence of spruce beetle. A suite of independent variables from FIA data were used to predict the presence or absence of spruce beetle. Results suggested structural variables such as trees per acre and stand density index ratio were influential in the model. Similarly, compositional variables such as percent stand density index of aspen, lodgepole pine, Douglas-fir, and subalpine fir were also influential. Co-occurrence of multiple species with Engelmann spruce helped explain some of the patterns of spruce beetle occurrence. Ubiquitously occurring species such as subalpine fir and Douglas-fir explained less of the variation in model predictions. Elevation and latitude were significant predictors of spruce beetle occurrence; however, co-variation made direct interpretation of these independent variables difficult. Although statistically weak, the model results were ecologically interpretable and provided a range-wide context in explaining spruce beetle activity at some sites in the Intermountain West.*

[1] Corresponding author (R. Justin DeRose)

[2] Department of Wildland Resources and Ecology Center, Utah State University, 5230 Old Main Hill, Logan, UT 84322-5230, Phone: 435-797-3219, Facsimile: 435-797-3796, Email: rjustinderose@gmail.com

[3] Department of Wildland Resources and Ecology Center, Utah State University, 5230 Old Main Hill, Logan, UT 84322-5230

[4] Forest Inventory and Analysis, Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Keywords: Biogeography, *Dendoctronus rufipennis*, *Picea engelmannii*, stand density index ratio.

Introduction

Engelmann spruce (*Picea Engelmannii* Parry es. Engelm.) forests are widespread in the Intermountain West at higher elevations and are effectively southern extensions of the boreal forest. They occur from approximately 105° to 118° west in longitude and from approximately 33° to 50° north in latitude and range in elevation from less than 3,000 ft in northern Idaho to almost 12,000 ft on the Colorado Plateau. These forests occur across a wide range of physiographic conditions and are typically structurally and compositionally diverse (Peet 2000). Structurally, Engelmann spruce forests are commonly multi-aged, exhibiting a wide range of tree sizes (Aplet and others 1988). Compositionally, many other species co-occur with spruce, its most common associates are subalpine fir (*Abies lasiocarpa* (Hook.) Nutt.) and Douglas-fir (*Pseudotsuga menziesii* var. *glauca* (Beissn.) Franco), and more regionally common species such as aspen (*Populus tremuloides* Michx.), limber pine (*Pinus flexilis* James), whitebark pine (*Pinus albicaulis* Engelm.), and lodgepole pine (*Pinus contorta* Dougl. ex. Loud.) (see Figure 1).

Both low- and high-severity disturbances influence the structure and composition of Engelmann spruce forests (Veblen and others 1994). Windthrow, root and butt rots, and endemic spruce beetle (*Dendoctronus rufipennis* Kirby) all act locally to create low-severity canopy gaps in spruce forests. In contrast, crown fires, wind storms, and spruce beetle outbreaks, while infrequent, are high-severity disturbances that occur at much larger scales.

The spruce beetle is a host-specific bark beetle native to Engelmann spruce forests across the Intermountain West (Holsten and others 1999). Endemic populations of the beetle can build to epidemic (outbreak) levels in recently downed material such as logging slash, wind throw, or avalanche debris (Schmid 1981). Transition from endemic to epidemic population levels can be facilitated by warmer-than-average summer temperatures allowing univoltine beetles to reproduce successfully (Hansen and others 2001a) and, in combination with semivoltine beetles, overwhelm their host.

It is commonly though that composition and structure of Engelmann spruce-dominated stands influence the potential for spruce beetle activity. For example, a commonly used stand-level spruce beetle risk-rating (Schmid and Frye 1976) suggests dense (high basal area), pure (greater than 65 percent Engelmann spruce), old (large diameter) stands on well-drained creek bottoms (high site productivity potential) are most likely to be attacked by the spruce beetle. The Schmid and Frye (1976) risk-rating in effect quantifies the view that

compositionally and structurally diverse stands have less potential for spruce beetle activity, including the shift from endemic to epidemic population-levels.

We use FIA data to explore the relationships between stand structure, composition, site attributes, and spruce beetle occurrence in Engelmann spruce forests across the Intermountain West. We examine the putative relationship between compositional and structural diversity and spruce beetle occurrence and explore whether stand and site variables in fact have potential for use in the construction of predictive models of spruce stands susceptible to the spruce beetle.

Methods

Forest Inventory and Analysis (FIA) data from nine Intermountain West states: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, eastern Washington, and Wyoming (Table 1) were used to explore the relationship between structure, composition, site attributes, and spruce beetle occurrence. The nationwide hexagonal sampling strategy implemented in 2000 provided most of the data for this analysis while the remaining data came from previous periodic plots. Each hexagon is 5937 acres or, 1 plot per approximately 6000 acres, national sampling precision. Each plot will be visited on a 5 (east) or 10 (west) year return interval. (FIA Fact Sheet Series, online http://fia.fs.fed.us/library/fact-sheets/default.asp). Because the FIA data is unbiased and systematically collected it is appropriate for analyzing spruce beetle occurrence patterns in the Intermountain West.

Data preparation

Phase 2 data (inventory data) from plots which had the presence of Engelmann spruce were used in this analysis. Data indicating the occurrence of spruce beetle in FIA came from two sources: the tree table and the condition table. If spruce beetle was found on a measurement tree the recorder assigned that tree an Agent Code of 10. Similarly, if spruce beetle activity was found on or near the FIA plot, Disturbance Code = 10 was designated in the conditions table (USDA 2007). Therefore reference to spruce beetle activity or occurrence in this paper is interpreted as Engelmann spruce activity and/or mortality as a result of the spruce beetle. For the analysis we assumed both designations indicated beetle activity and collectively analyzed them as spruce beetle affected stands. Final data preparation included filtering so that re-measurement plots were only represented once (most recent) and only plots with Condition Proportion = 1.0.

A suite of independent variables characterizing structural, compositional, and site-level attributes for each plot (Table 2) were calculated from the filtered data. Although a total of 41 species co-occurred with Engelmann spruce across the

Intermountain West we limited calculation of compositional variables to the 13 species that shared at least 100 plots with spruce (Figure 1). As an example, individual species contribution to total stand basal area and stand density index as a percent were calculated to describe stand composition. To characterize structural attributes we calculated trees per acre, quadratic mean diameter, basal area, and the stand density index ratio (SDIr). SDIr is the ratio of stand density index calculated using the summation method (Shaw 2000) to stand density index calculated using the average method (Reineke 1933). SDIr is interpreted as an index of structural diversity from least (greater than 0.9) to most diverse (less than 0.78) (Shaw and Long 2007). With the exception of aspect value (Roberts and cooper 1989), site variables were taken straight from the FIA data.

Data analysis

The presence/absence nature of the dependent variable, spruce beetle occurrence, lent itself to the logistic regression approach to model fitting. We initially predicted the presence/absence of spruce beetle in Engelmann spruce stands using compositional, structural, and site-level variables individually and their interactions (Table 2). To accurately reflect pre-beetle conditions reconstruction of stand attributes was done for plots that had beetle-caused mortality. One shortcoming with the binary nature of the dependent variable (spruce beetle occurrence) was that it could indicate either endemic or epidemic beetle populations. Therefore the results are interpreted as spruce beetle activity or occurrence in Engelmann spruce stands, and differentiation between population levels was not made. Models were evaluated for goodness-of-fit using pseudo-R^2, or deviance (D^2). Pearson's correlations (r^2) were used to further evaluate the significant independent variables.

Results

Both Engelmann spruce plots and plots with spruce beetle were wide ranging across the Intermountain West (Figure 2). Therefore it was not surprising that generally the logistic models explained very little of the overall variance in spruce beetle occurrence, with a range of D^2 from 6 percent to 19 percent. However, the models did indicate independent variables with statistical significance and we evaluated these further.

Of the 13 species that co-occurred with Engelmann spruce, four were influential in predicting spruce beetle activity: basal area of aspen, basal area of lodgepole pine, and percent stand density index of aspen, lodgepole pine, Douglas-fir, and subalpine fir. Aspen and lodgepole pine basal areas were negatively related to spruce beetle occurrence, whereas aspen and lodgepole percent stand density index were positively related. The contradiction in signs was difficult to explain; however, the fact that both these species perform the same

role ecologically, although primarily at different latitudes, suggested this pattern was not spurious. Both Douglas-fir and subalpine fir had a positive relationship with spruce beetle occurrence, occurred nearly ubiquitously with Engelmann spruce across the Intermountain West, and were weakly inter-correlated (r^2 = 0.23). Subalpine fir, because of its near complementary relationship with corkbark fir, was weakly correlated with latitude (r^2 = 0.28) which likely drove its significance in the model.

Structural variables significant in the prediction of spruce beetle occurrence were: trees per acre, stand basal area, stand density index, and SDIr. Trees per acre and basal area were positively related to beetle occurrence whereas stand density index and SDIr had negative relationships. Trees per acre and basal area were highly correlated (r^2 = 0.60) and probably co-varied in the model. However, the difference in sign between trees per acre and stand density index was suspect since they were strongly positively correlated (r^2 = 0.74).

Site-level variables significant in the prediction of spruce beetle occurrence were elevation and latitude. Unfortunately latitude, longitude, and elevation were highly inter-correlated (range of r^2 = 0.79 – 0.81) in this study so isolating the effect of one from the other was difficult. Model results indicated either latitude or elevation was significant individually, but not simultaneously. Furthermore, although they appeared to proxy one another their signs differed. The difference in sign between these two variables in the model (elevation was positive, latitude was negative) can be explained by the inherent negative relationship between latitude and elevation. Individually, their patterns with spruce beetle occurrence were ecologically explainable. The number of spruce beetle plots decreased with decreasing latitude (Figure 3) but did not occur disproportionately to the total number of spruce plots (data not shown). In contrast, the number of spruce beetle plots generally increased with increasing elevation but for the 9,000 – 11,000 foot elevation range spruce beetle plots occurred disproportionately more than the number of spruce plots (Figure 4).

Discussion

The large variation in compositional, structural, and site-level attributes for Engelmann spruce forests across the Intermountain West proved only marginally useful in predicting the occurrence of spruce beetle activity. Regardless, ecologically meaningful patterns between the independent variables and observed spruce beetle activity were revealed by the data. The weakness in statistical results generally provides support for the complex, non-linear nature of spruce beetle populations which exhibit shifts from endemic to epidemic levels that are catalyzed by specific events and therefore not reliably predictable. The geographically unbiased FIA data revealed the spruce beetle generally occurred across the entire range of Engelmann spruce (Figure 2), suggesting it is a normal

part of spruce forests but also complicated efforts to identify useful predictors of future outbreaks.

Understanding the relationships between co-occurring species and spruce beetle in Engelmann spruce forests is important because mixed stands are much more common than pure stands. For example, when the FIA data were filtered for nearly pure Engelmann spruce composition (greater than 80 percent stand density index), approximately 13 percent of the spruce plots met the criteria. All significant composition variables were early successional species with the exception of subalpine fir. Perhaps more interesting was that the stocking of all four species was positively related to spruce beetle occurrence. This is counterintuitive to our expectation that compositionally simple stands of relatively pure spruce are more susceptible to beetle activity. The biogeographical patterns of aspen, lodgepole, and subalpine fir were interesting and somewhat predictable, for example, corkbark fir gave way to subalpine fir at approximately 38° latitude. Similarly, aspen shifts from common to rare at approximately 44° latitude where lodgepole pine becomes common and remains so to the Canadian border (Figure 2). This is consistent with the observations along the Colorado Front Range made by Peet (1978) who suggested aspen yielded to lodgepole as the early successional player in spruce stands as one moved north in latitude. Biogeographical patterns for the other common species that co-occurred with spruce such as limber pine, whitebark pine, and grand fir (Figure 1) were also predictable from site and structural variables but were not influential in the spruce beetle models.

Structural variables such as spruce basal area and spruce mean diameter greater than 10 inches are commonly used to interpret stand risk to a beetle outbreak. However, only stand-level variables that ignored the spruce component were indicated in our analysis. Based on previous research the positive relationship between basal area and spruce beetle occurrence was expected (Schmid and Frye 1977). What was less intuitive was the significance of tree density on beetle occurrence, except that it likely co-varied with basal area. One explanation is that stand basal area was highly correlated with spruce basal area ($r^2 = 0.53$) but only weakly correlated with trees per acre ($r^2 = 0.19$). Furthermore, to reconcile the difference in signs between stand density index and trees per acre it helps to understand stand density index and spruce basal area were correlated ($r^2 = 0.48$). An index of structural diversity, SDIr, did not co-vary with any of the other structural indices. Lower SDIr indicates stands that are irregular in structure, for instance, stands with an index of greater than 0.9 have a unimodal diameter distribution, stands of approximately 0.75 exhibit a reverse-J distribution, and stands with an index below 0.7 exhibit increasingly bimodal or irregular structures. The negative relationship of SDIr with spruce beetle occurrence found here is counterintuitive to our expectation that stands of simple structure would have more beetle activity. This result is likely driven by the non-linear nature of the distribution of SDIr across the Intermountain West range of spruce and the

possible effects of elevation on the relationship due to the correlation between these variables on the spruce beetle plots (r^2 = 0.28, see below).

Other variables such as stand age and site quality (xeric versus mesic) have traditionally been used in combination with composition and structure to predict the risk of spruce beetle populations shifting from endemic to epidemic (Schmid and Frye 1976). While some compositional and structural variables were significant in predicting spruce beetle occurrence in this study, the strongest pattern that emerged was the disproportionate number of spruce beetle plots at high elevation (Figure 4). Generally the proportion of spruce beetle plots increased with increasing elevation with the exception of the 11,000 – 12,000 foot elevation bin (Figure 4). This increasing relationship can be partially explained by the relatively strong relationship between elevation and Engelmann spruce stand density index (r^2 = 0.41) which is not inconsistent with the idea that increasing spruce composition increases spruce beetle activity. However, this does not fully explain the spike in spruce beetle observations in the 9,000 – 11,000 foot elevation range. As for the 11,000 – 12,000 foot elevation range perhaps the spruce that occur here are very small, for example krumholz stands, and climatic extremes are such that it is relatively inhospitable for spruce beetle populations.

Climatic and weather variables have been found to exert a heavy influence on insect populations including the family Scolytidae (Logan and Bentz 1999; Hansen and others 2001b). If changes in climate were manifest at varying levels across an elevation gradient we would have reason to suspect the higher than expected numbers of spruce beetle in the 9,000 – 11,000 foot range would be influenced. We suggest the unpredictability of spruce beetle activity from structural and compositional attributes in this study is likely due to the influence of an overriding factor, climate. Indeed, climate, and specifically temperature, has surfaced as the controlling factor in spruce beetle populations in Alaska (Berg and others 2006) and the recent explosive mountain pine beetle (*Dendroctonus ponderosae* Hopkins) outbreak in Colorado (Kulakowski and Veblen, unpublished paper). Since temperature controls insect populations, we hypothesize climatically sensitive areas, such as those found at high elevations, will exhibit the most pronounced response to spruce beetle activity. Compounding this problem is the observation that strong temperature-precipitation interactions control the range limit for Engelmann spruce, and climate modeling under various global climate change scenarios revealed the bioclimatic window for spruce is likely to close for populations at high elevations and lower latitudes (Rehfeldt and others 2006). Since both beetle populations and the realized niche for Engelmann spruce are controlled by temperature, perhaps increased spruce beetle activity at higher elevations is exacerbating the predicted shift in spruce range limits.

Acknowledgements

We wish to thank the T.W. Daniel Endowment, Ecology Center, and the Forest Health Monitoring Analysis and Reporting Program, agreement number 08-CA-11046000-610, for financial support of this project.

References

Aplet, G. H.; Laven, R.D.; Smith, F.W. 1988. Patterns of community dynamics in Colorado Engelmann spruce-subalpine fir forests. Ecology. 69: 312-319.

Berg, E.E.; Henry, D.J.; Fastie, C.L.; De Volder, A.D.; Matsuoka, S.M. 2006. Spruce beetle outbreaks on the Kenai Peninsula, Alaska, and Kluane National Park and Reserve, Yukon Territory: relationship to summer temperatures and regional differences in disturbance regimes. Forest Ecology and Management. 227: 219-232.

Hansen, E.M.; Bentz, B.J.; Turner, D.L. 2001a. Physiological basis for flexible voltinism in the spruce beetle (Coleoptera: Scolytidae). The Canadian Entomologist. 133: 805-817.

Hansen, E.M.; Bentz, B.J.; Turner, D.L. 2001b. Temperature-based model for predicting univoltine brood proportions in spruce beetle (Coleoptera: Scolytidae). The Canadian Entomologist. 133: 827-841.

Holsten, E.H.; Their, R.W.; Munson, A.S., Gibson, K.E. 1999. The spruce beetle. Forest Insect and Disease Leaflet 127. U.S. Department of Agriculture, Forest Service, 12 p.

Logan, J.A.; Bentz, B.J. 1999. Model analysis of mountain pine beetle (Coleoptera: Scolytidae) seasonality. Environmental Entomology. 28: 924-934.

Long, J.N.; Shaw, J.D. 2005. A density management diagram for even-aged ponderosa pine stands. Western Journal of Applied Forestry. 20: 1-11.

Peet, R.K. 1978. Latitudinal variation in southern Rocky Mountain forests. Journal of Biogeography. 5: 275-289.

Peet, Robert K. 2000. Forests and meadows of the Rocky Mountains. In: Barbour, Michael G.; Billings, William Dwight, eds. North American Terrestrial Vegetation. Second edition. Cambridge University Press. p 75-121.

Rehfeldt, G.E.; Crookston, N.L.; Warwell, M.V.; Evans, J.S. 2006. Empirical analyses of plant-climate relationships for the western United States. International Journal of Plant Science. 167: 1123-1150.

Reineke, L.H. 1933. Perfecting a stand-density index for even-aged forests. Journal of Agricultural Research. 46: 627-638.

Roberts, D.W.; Cooper, S.V. 1989. Concepts and techniques of vegetation mapping. Gen. Tech. Rep. INT-257. U.S. Department of Agriculture, Forest Service, 90-96 p.

Schmid, J.M. 1981. Spruce beetles in blowdown. Research Note RM-141, U.S.

Department of Agriculture, Forest Service, 4 p.

Schmid, J.M.; Frye, R.H. 1976. Stand ratings for spruce beetles. Research Note RM-309 U.S. Department of Agriculture, Forest Service, 4 p.

Schmid, J.M. and Frye, R.H. 1977. Spruce beetle in the Rockies. Gen. Tech. Rep. RM-49 U.S. Department of Agriculture, Forest Service, Report, Rocky Mountain Forest and Range Experiment Station: 38 p.

Shaw, J.D. 2000. Application of stand density index to irregularly structured stands. Western Journal of Applied Forestry. 15: 40-42.

Shaw, J.D.; Long, J.N. 2007. A density management diagram for longleaf pine stands with application to red-cockaded woodpecker habitat. Southern Journal of Applied Forestry. 31: 28-38.

USDA. 2007. Interior West Forest Inventory and Analysis field procedures. Version 3.01. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 420 p.

Veblen, T.T.; Hadley, K.S.; Nel, E.M.; Kitzberger, T.; Reid, M.; Villalba, R. 1994. Disturbance regime and disturbance interactions in a Rocky Mountain subalpine forest. Journal of Ecology. 82: 125-135.

Table 1: Forest Inventory and Analysis data from Engelmann spruce plots by state.

State	# plots	Species present[a]	Trees per acre	Basal area (ft^2)	Stand density index	Stand density index ratio
				- - - ranges - - -		
Arizona	105	WF, CF, SF, LM, PP, DF, AS	36 – 705	54 – 338	54 – 575	0.39 – 0.98
Colorado	587	WF, CF, SF, LP, LM, PP, DF, AS	12 – 860	6 – 367	12 – 640	0.06 – 0.99
Idaho	548	GF, SF, WL, WBP, LP, LM, WWP, PP, DF, WRC, WH, AS	12 – 625	5 – 348	8 – 452	0.06 – 0.98
Montana	819	GF, SF, WL, WBP, LP, LM, WWP, PP, DF, WRC, WH, AS	3 – 821	7 – 327	10 – 528	0.04 – 1.0
Nevada	4	WF, LM, DF, AS	88 – 412	63 – 223	99 – 346	0.25 – 0.55
New Mexico	101	WF, CF, SF, LM, PP, DF, AS	30 – 876	18 – 354	32 – 648	0.35 – 0.98
Utah	345	WF, SF, LP, LM, PP, DF, AS	7 – 950	8 – 280	13 – 498	0.06 – 1.0
Washington	6	GF, SF, WL, LP, WWP, DF, WRC, WH,	199 – 542	77 – 138	132 – 274	0.48 – 0.83
Wyoming	388	SF, WBP, LP, LM, PP, DF, AS	12 – 762	11 – 343	16 – 550	0.04 – 0.98

[a] Species codes: ES – Engelmann spruce, SF – subalpine fir, CF – corkbark fir (*Abies lasiocarpa* (Hook.) Nutt. var. *arizonica* (Merriam) Lemmon AS – aspen, WBP – whitebark pine, LM – limber pine, LP – lodgepole pine, DF – Douglas-fir, WL – western larch (*Larix occidentalis* Nutt.), WF – white fir (*Abies concolor* (Gord. And Glend.) Hildebr.), GF – grand fir (*Abies grandis* (Dougl. ex D. Don) Lindl.), PP – ponderosa pine (*Pinus ponderosa* Dougl. ex Laws.), WH – western hemlock (*Tsuga heterophylla* (Raf.) Sarg.), WRC – western redcedar (Thuja plicata Donn ex D. Don).

Table 2: Independent variables describing structure, composition, and site used in logistic regression models

Independent variable	Description
Structural	
Trees per acre	Absolute density
Stand density index (summation)	Relative measure of stand density (Shaw 2000)
Stand density index ratio	Ratio of summation stand density index to Reineke stand (Long and Shaw 2005) density index (Reineke 1933), index of structural diversity
Stand basal area	Measure of stocking
Quadratic mean diameter	Geometric mean of stand diameter
Compositional	
Species basal area	Individual species basal area
Percent stand density index by species	Percent composition of each species present on the spruce plot
QMD of Engelmann spruce	Geometric mean diameter for spruce component only
QMD of Engelmann spruce greater than 10 inches	Mean of stand diameter for all Engelmann spruce greater than 10 inches
Site	
Elevation	Height above sea level
Site index	Measure of site potential productivity based on tree height and age
Aspect value	Moisture index*
Slope	Steepness of the plot
Stand age	Age of the plot
Latitude	North
Longitude	West

*Aspect value = [Cos((aspect-30)/180*3.14)+1]/2 (Roberts and Cooper 1989)

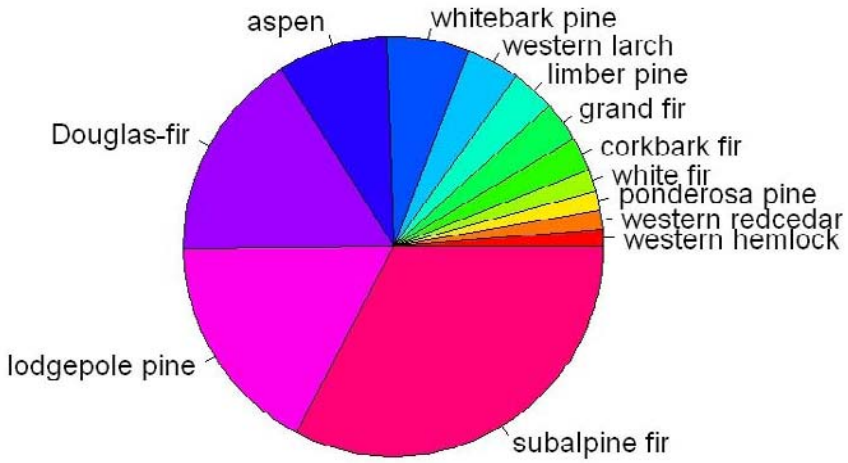


Figure 1: Representation of species which co-occur with Engelmann spruce on at least 100 of the FIA spruce plots (total spruce plots, *n* = 2093). For species identification see Table 1.

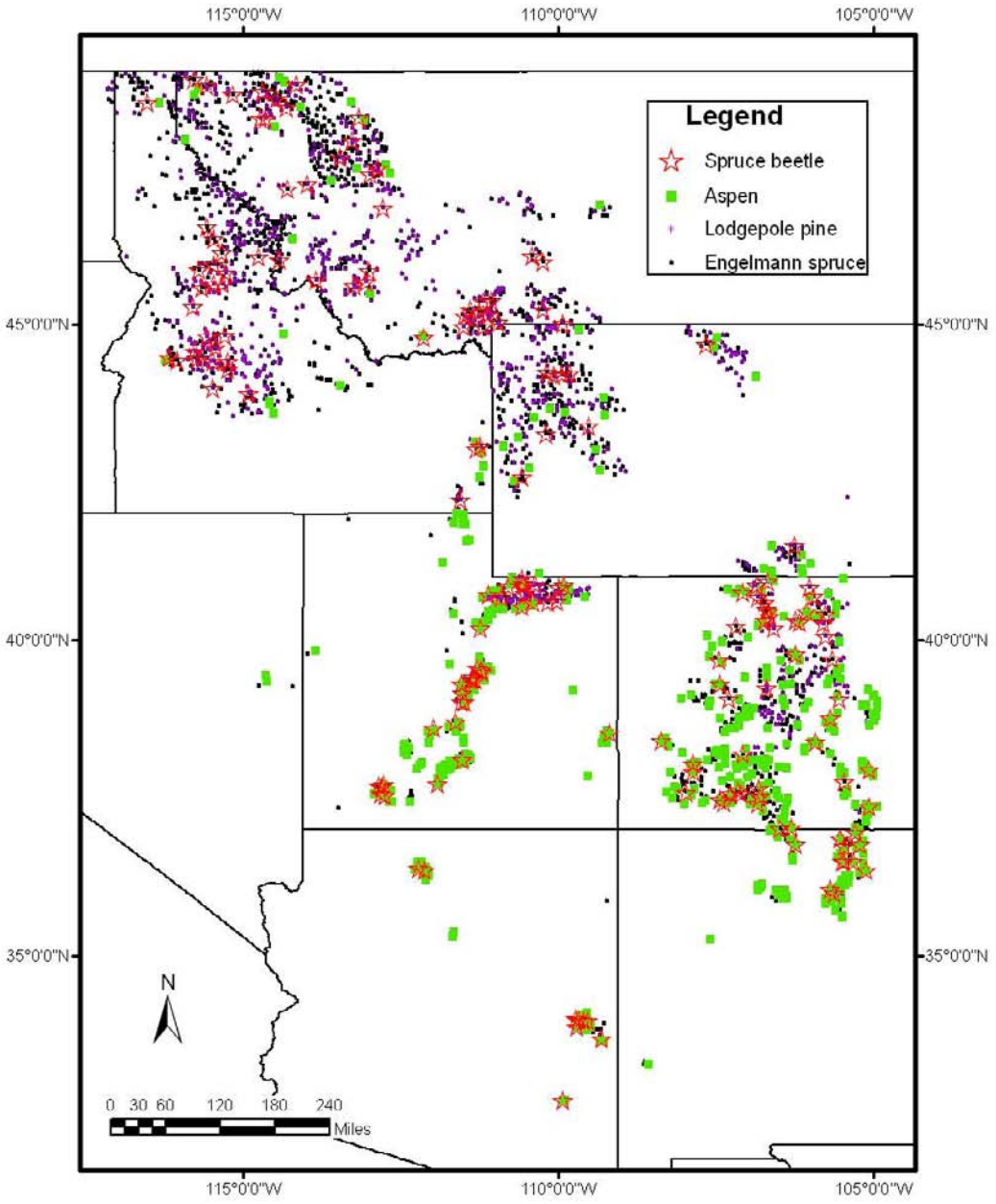


Figure 2: Map of the distribution of Engelmann spruce in the Intermountain West based on FIA data (*n* = 2093). Plots with spruce beetle, aspen, and lodgepole pine are also shown.

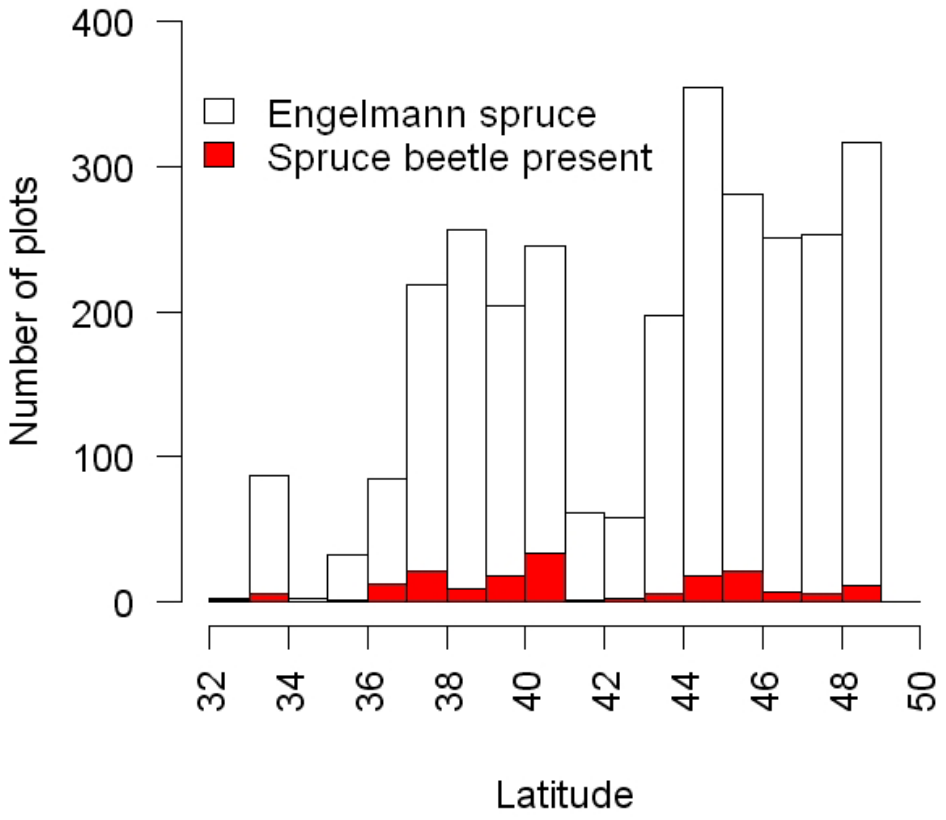


Figure 3: Frequency distribution of Engelmann spruce plots (*n* = 2093) juxtaposed with Engelmann spruce plots that had spruce beetle (*n* = 175) over latitude.

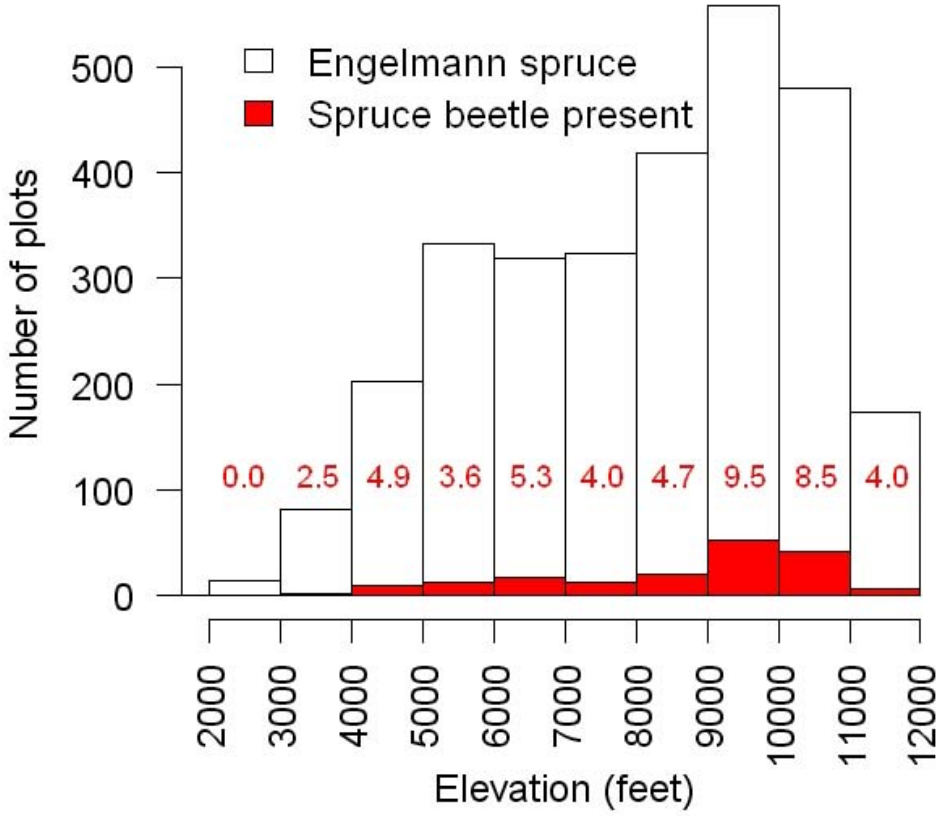


Figure 4: Frequency distribution of Engelmann spruce plots (*n* = 2093) juxtaposed with Engelmann spruce plots that had spruce beetle (*n* = 175) over thousand foot elevation bins. Spruce beetle plots as a percent of total Engelmann spruce plots by elevation bin in red showing the disproportionate number of spruce beetle plots in the 9000 – 11000 elevation range.

Evaluating Cypress Sustainability—FIA Data "In the Hot Seat"

Mark J. Brown[1]

Abstract: *The use of cypress (Taxodium species) for mulch boomed during the 1990s, and its growth in popularity created concerns about the sustainability of cypress forests in the Southern United States. A combination of factors, including Hurricane Katrina, cypress harvesting practices, and the unique requirements for successful regeneration of cypress drew media attention and fueled debates over sustainability of cypress forests. The wood industry, Federal, State, and local governments, academia, environmental and conservation organizations, and retailers all became entangled in the issue. This eventually culminated in proposed legislation to place a moratorium on cypress harvesting and the decision by at least one large retailer to cease selling cypress mulch that was produced in Louisiana. These factors launched an intensive search for viable data about the cypress resource. Requests for Forest Inventory and Analysis (FIA) cypress data escalated. This paper describes the challenges FIA encountered in meeting the need to quantify trends in the cypress resource despite changes in inventory methods, forest-type definitions, plot design, processing algorithms, sample intensity, etc. The paper explains the constraints that were overcome to report region-wide cypress data and to develop a Southwide cypress factsheet. It also shows how FIA data was used to provide State level cypress trends, focusing on Florida as an example.*

Keywords: Cypress-tupelo forest type, cypress volume, factsheets, FIA, forums, mulch, sustainability.

Introduction—Background on Cypress Mulch

Traditionally, cypress has been used for lumber. Popular uses included siding, decking, and trim work. Entire houses and even boats were built with cypress lumber. More recently, the use of cypress as a mulch has increased. Originally composed of cypress edgings, shavings, and mill residues, its increased popularity, due to its perceived durability, increased demand. This led to the use of low quality saw log and small diameter whole tree utilization for mulch. Ironically, old growth cypress heartwood lumber earned the decay-resistant reputation, yet evidence exists that second growth wood does not possess these qualities to the same degree (U.S. Department of Agriculture Forest Service Forest Products Laboratory). Cypressene is the natural oil in the wood that resists

[1] United States Forest Service; Southern Research Station; Forest Inventory and Analysis; 4700 Old Kingston Pike; Knoxville, TN 37919 USA; mbrown03@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

insects and decay (Campbell and Clark 1960) and is the basis of the heartwood reputation. Decomposition rates suggest that much of the cypress mulch was actually little better than some other types of mulch, such as pine bark (Duryea and others 1999). Nevertheless, the cypress mulch industry boomed and demand for cypress mulch persists.

The Controversy—Sustainability

Opposition to the use of cypress for mulch has swelled among environmental organizations since the 1990s. Media attention was drawn to the issue and focused on the perceived threat to survival of the forest type. This concern emanates from two basic premises surrounding the species. First, there is a perception that cypress stands are not being regenerated after harvest. This is fueled by the unique silvics of cypress and the sites on which they occur. Cypress is one of the few softwoods capable of stump sprouting, but not all stumps resprout. Resprouting ranges from one-quarter to one-half of the former stand (Randall and others 2005). Natural regeneration from seed varies greatly and is keyed to unpredictable water level fluctuations (Brandt and Ewel 1989). Although guidelines exist for planting cypress (Vince and Duryea 2004), there is little evidence that it occurs on a large scale.

Second, the facts that cypress typically grows on wetland sites and that many of these sites fall under federal jurisdiction (Section 404 of the Clean Water Act), complicate the issue and highlight environmental sensitivities.

The quest for accurate information on Southwide trends in the cypress resource developed from all sides of the controversy. Immediately apparent was the limited data available to track cypress forests and a paucity of research beyond the species' physiology, wood value, and regenerative capabilities. This gap in knowledge was unlike that for other softwoods in the South, especially the southern yellow pines (loblolly, slash, and longleaf). Hanging in the balance and threatening the cypress mulch industry, were important decisions by major retailers involving the sale of cypress mulch and by some State governments concerning moratoriums on cypress harvesting. Any suspension of sales or harvesting would likely lead to the demise of the cypress mulch industry.

Demands for data to establish the facts about the extent of and trends in the cypress resource were significant enough to cause the U.S. Forest Service (USFS) Southern region to seek a rapid assessment of the cypress resource.

Analysis of Constraints—Solutions

The USFS Forest Inventory and Analysis (FIA) program was the main and obvious source for broadscale data related to forests. Therefore, the Southern region requested FIA data regarding cypress sustainability for use in the rapid

assessment. However, only limited information was available in the FIA data. One reason this occurred is that FIA historically collected cypress acreage information based on the oak-gum-cypress forest-type category. Use of this category always overestimates actual cypress forest acreage because it includes acres of all bottomland hardwood stands, gum stands, cypress stands, and various other mixtures. FIA developed a cypress-tupelo forest-type definition, but it included pure tupelo stands along with pure cypress and any mixed stands. The cypress-tupelo forest type essentially captures all stands where cypress or tupelo equal 25 percent or more of the stocking (U.S. Department of Agriculture 2007).

In order to address issues identified with the burgeoning cypress mulch industry, a preliminary definition of the cypress type first materialized in FIA with the 1995 Florida's Forests report (Brown 1999). Stocking levels were used to establish a threshold beyond which a stand would be called a cypress type. This format was similar to the threshold used to change an oak-pine type to one of the individual pine forest types and was based on 50 percent and greater stocking of the particular species. The new cypress forest-type data was limited to Florida. Pre-1995 cypress data for Florida was developed solely for the 1995 report and did not exist previously. As a result, recent requests for trends in area of cypress have been limited even in Florida and difficult to address across the remaining southern States. Eventually, interest in cypress led FIA to implement a permanent cypress forest-type definition in the early 2000s. However, many States were either between surveys or in transition to annual inventories and all did not implement the type simultaneously. Thus, the solution for the rapid assessment was to use the cypress-tupelo type for a general area trend analysis of the extent of cypress in the South. At least the cypress-tupelo type would represent all the acres in a cypress forest type plus a (typically) small percentage in tupelo. In addition, a map was produced depicting the distribution of all sample plots with one or more cypress trees recorded (figure 1) in the 13 States of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, and Virginia. Except for west Texas, the distribution emulated the native range of cypress (Burns and Honkala 1990).

Volume of cypress could be ascertained because FIA had the capability of tracking volume of any individual species by compiling estimates from single trees throughout the sample base. Therefore, trends in volume were obtainable and sourced from the FIA database (FIADB) through the FIA mapmaker Web application (Miles 2008). The former lack of a cypress forest type is probably why volume was used in a similar analysis of the status of southern cypress in the 1970s (Sternitzke 1972). At that time, evidence suggested the cypress resource was increasing after bottoming out during the housing boom shortly after World War II. Sternitzke touched on growth and removals for baldcypress, highlighting changes in Florida and Louisiana, while combining that for the remaining States. At the time of his analysis, baldcypress growth Southwide was 2.5 times that of removals. In 2008, FIA data showed that growth of all cypress in the South was

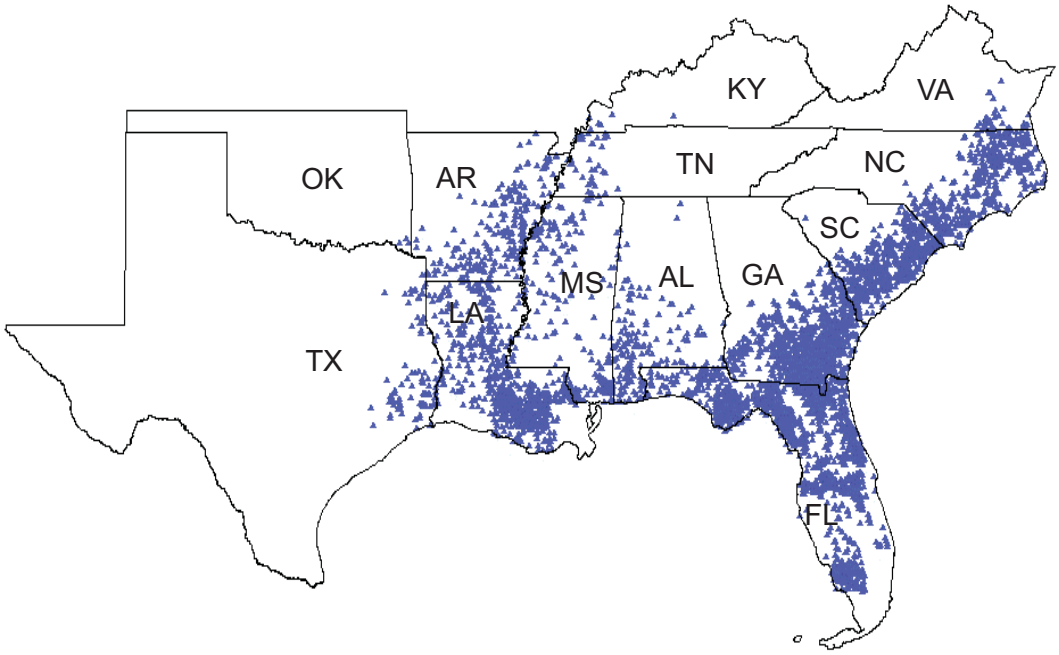


Figure 1: Distribution of FIA sample plots where one or more cypress trees were recorded.

2.1 times that of removals. There was substantial variation among the States, as evidenced by Florida's cypress growth at 1.5 times that of removals versus 0.8 in North Carolina, 1.7 in Georgia, 2.3 in South Carolina, and 12.4 in Alabama. However, even State level data disguised large differences between subregions of a State. For instance, in the four regions of Florida, cypress growth ranged from 4.4 times removals in the northwest panhandle, 2.9 in central Florida, 1.3 in northeast Florida, and just 0.6 times removals in south Florida. These are growing-stock numbers that can be compared with Sternitzke's use of growing stock. In order to avoid any subtle inaccuracies in applying growing-stock standards, FIA's all-live volumes were used in the rapid assessment for current reporting purposes.

Bald and pondcypress were combined in the assessment to avoid two potential pitfalls in the data. First, this precludes misidentification between the bald and pondcypress species, particularly in light of some intergrades between the two. Second, Florida contains sizable populations of pondcypress compared with other southern States.

Immediately obvious were the differences between the States in year of inventory data available for use in the assessment. On an individual State basis, analysis of the cypress resource within a State was easier and mostly limited by the timeframe of the surveys which determined if any trends in area could be established. Whether considering area or volume, the validity of individual State analysis was dependent on the degree of prevalence of cypress within that State. If cypress occurrence were minimal, as was the case in Oklahoma,

Kentucky, Virginia, and Tennessee, then sampling errors around the data were too high for statistically reliable estimates to stand alone. This justified the Southwide approach to reporting on the cypress resource. However, since Florida and Louisiana contain more than one-half of the South's cypress resource, both in area and in volume, either of these States could be evaluated individually with some degree of reliability. The disparity in date of surveys among the 13 States precluded declaration of Southwide trends between two specific years. It required stating any Southwide trends between an aggregation of the 13 States' **"latest"** surveys to an aggregation of the 13 States' **"previous"** surveys' data. This problem and solution was not limited to area, but involved the otherwise preferable volume comparison method as well.

It was difficult to ascertain trends in area because of the historical lack of a cypress forest-type definition and the varying dates per State at which a definition was finally enacted in the data collection. The randomness of the timing and date of the previous periodic surveys dictated whether or not a particular State would have a valid area trend for cypress. Interest was higher among data seekers and lay people for area information, thus inclusion of "best we could" area estimates and trends into the assessment output despite all the anomalies involving the use of area (figure 2). Confidence intervals were withheld for area trends due to variances in obtaining the older data.

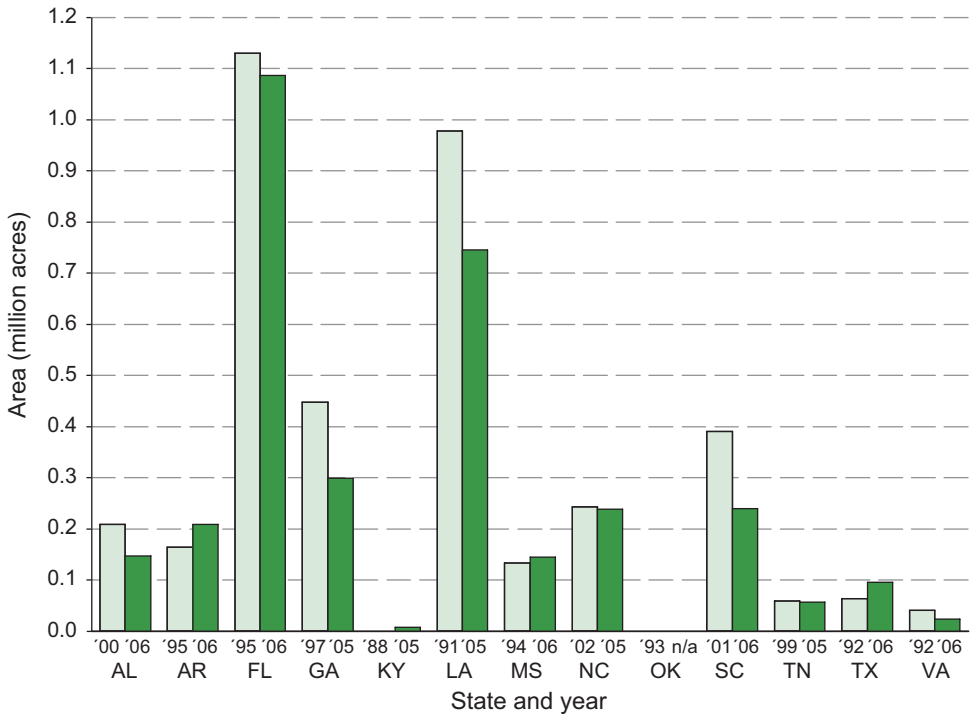


Figure 2: Area trends of cypress-tupelo timberland by State (based on available data) and year.

Trends were more easily obtained for cypress volume, which is why the rapid assessment relied on volume to best indicate sustainability trends. Confidence intervals were included in the portrayal of volume information (figure 3). To reiterate however, State level volume data often subdued differences between subregions of a State. As in the case of Florida, where more of the cypress volume occurred in the Northeast and Central regions, between the latest two surveys it increased slightly in the Northwest region while remaining almost stable in the Central part of the State and decreasing in the Northeast and South regions (figure 4).

Throughout FIA reporting and with the FIADB, clients are cautioned against using small subsets of the data. Ironically, analysis of cypress was a subset of FIA survey data in that the cypress-tupelo forest type comprised <2 percent of the total timberland in 11 of the 13 southern States. Even in the two States where it was most prevalent, it comprised <7 percent of the total timberland present.

In the initial analysis, we planned to assess the growth and removal situation for cypress at localized levels. However, these further subsets of data were too troublesome to contend with for a rapid broadscale general assessment of the resource. These factors are part of the reason why the analysis was limited to area of cypress-tupelo forest type and volume of cypress species and ownership. Incorporating growth and removals further reduced the sample to a smaller and less reliable subset of data. As it were, just with volume and area, several States declined in cypress-tupelo area but their volume of cypress rose or remained almost steady. These trends were inexplicable or defied logic. The many differences between the periodic surveys and the annual surveys complicated or

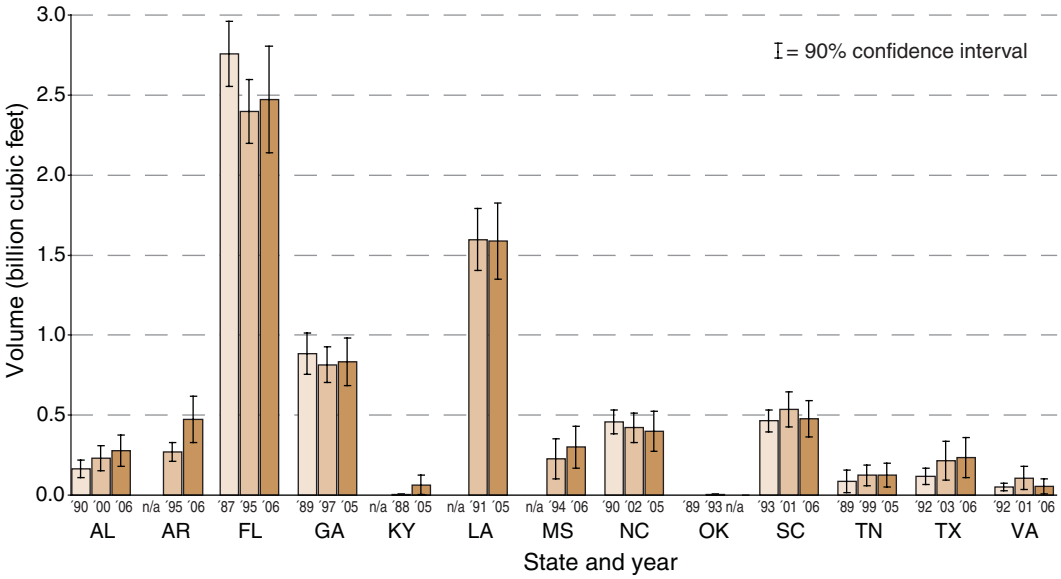


Figure 3: Volume trends of cypress by State (based on available data) and year.

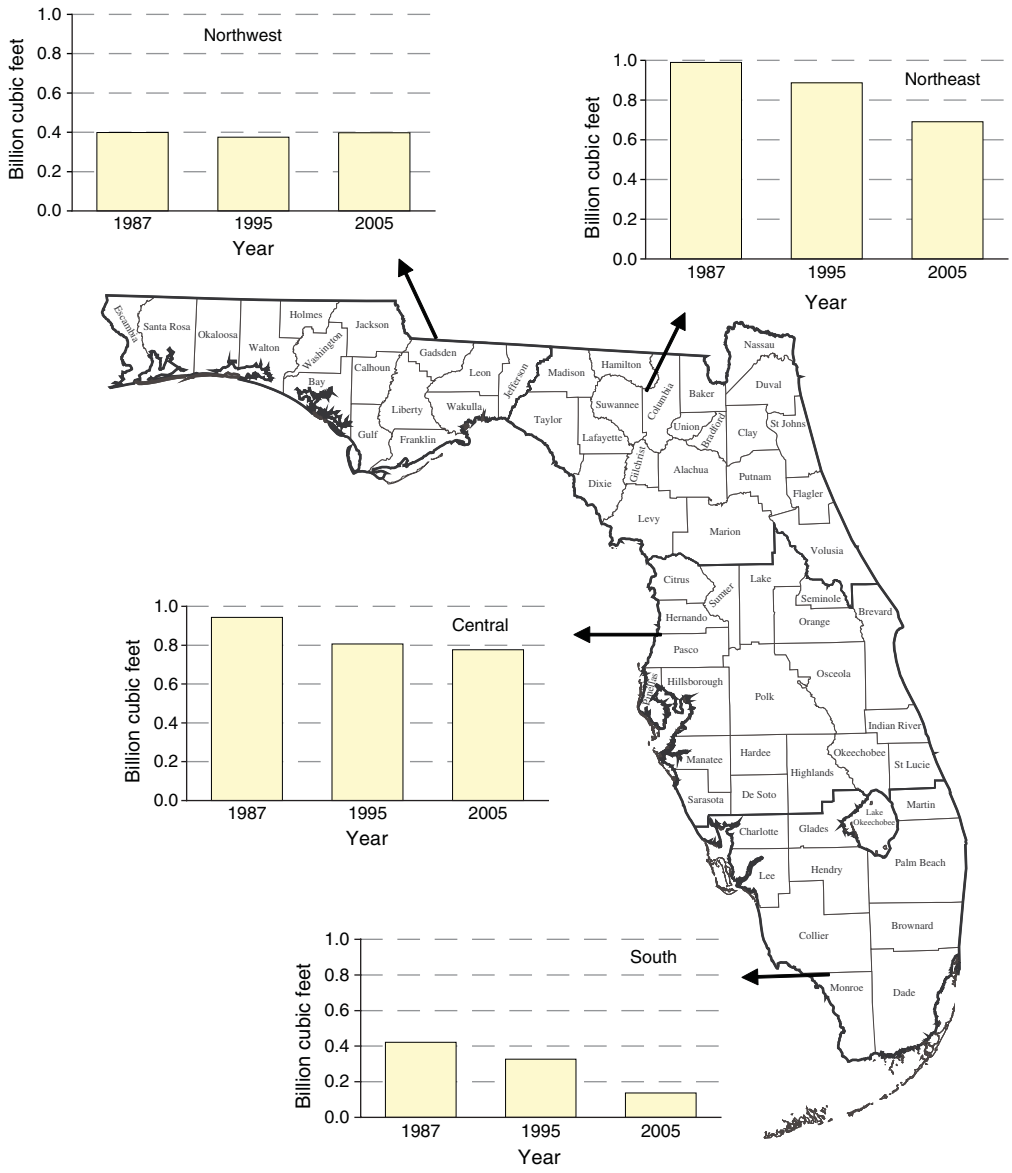


Figure 4: Volume of cypress in Florida, by region and year.

hindered accurate analysis of survey data. For example, in contrast to the fixed radius plots of the newest surveys, samples in many of the older surveys were based upon variable radius plots and a 37.5 factor prism for inventory. In addition, in 5 of the 13 southern States administered by the old Southeastern Forest Experiment Station, sample intensity was considerably higher. Instead of the current 5,700 acres per plot sample intensity implemented in annual inventories across the South, samples in Coastal Plain regions of the five southeastern States occurred on every 2,800 acres.

Conclusions

Between the latest two surveys, area of cypress forest type declined in 8 of 13 States, rose in 3, and there was insufficient data for 2 (Kentucky and Oklahoma). Volume of cypress was up in 8 of 13 States, down in 4, and there was no data in 1 (Oklahoma). Broadscale findings such as these, often are unable to illuminate possible localized countervailing trends in both area and volume. Increases in a few areas of the South offset losses in others and ameliorated totals at the regional or State level. For instance, growth exceeded removals at the State level in Florida, but in south Florida removals were in excess of growth, which poses some concern for continued sustainability of the resource at the former levels of utilization in that part of the State.

Southwide, the cypress resource appears stable. Ultimately, the sustainability of the cypress resource will largely be driven by what happens to it in Florida and Louisiana. Together, these two States contain more than one-half of all cypress acreage and volume in the South. Southwide, one-fourth of the cypress-tupelo area is publicly owned and somewhat less subject to change. Most of the sustainability issues will be related to what happens on the three-fourths of the cypress-tupelo acreage and the cypress volume that is privately owned. The broadscale regional and even State level data do mask declines in some sub-State areas. Even in Florida and Louisiana, sub-State evaluations quickly became weaker due to diminishing sample sizes. There also are some indications that stability of the resource wanes in some sub-State areas.

Ultimately, the constraints of weaknesses at small scales or subsets along with inconsistent availability of surveys with cypress forest type defined, required that the assessment be limited to trends in area and volume and adhere to the more broad State level and Southwide facts regarding the cypress resource. Although growth and removals relationships are in demand, to increase reliability, we focused on trends in total volume, and used area changes as a way of grasping the extent to which changes occur to the cypress resource. In lieu of a publication *per se*, a more concise and highly graphic factsheet format (Greis and Brown 2008) was chosen to quickly portray the status of the South's cypress resource.

References

Brandt, K.; Ewel, K.C. 1989. Ecology and management of cypress swamps: A review. Gainesville, FL. University of Florida, Institute of Food and Agricultural Sciences Bulletin 252. 19 p.

Brown, M.J. 1999. Florida's forests, 1995. Resour. Bull. SRS–48. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 83 p.

Burns, R.M.; Honkala, B.H., tech. cords. 1990. Silvics of North America: 1. Conifers. Agric. Handb. 654. Washington, DC: U.S. Department of Agriculture, Forest Service, Vol. 1. 675 p.

Campbell, R.N.; Clark, J.W. 1960. Decay resistance of bald cypress heartwood. Forest Products Journal. May: 250-253.

Duryea, M.L.; English, R.J.; Hermansen, L.A. 1999. A comparison of landscape mulches: chemical, allelopathic, and decomposition properties. Journal of Arboriculture. 25(2): 88-96.

Greis, J.G.; Brown, M.J. 2008. Cypress facts for the South. [Brochure]. On file with: U.S. Department of Agriculture Forest Service, Southern Research Station, Knoxville, TN. [Not paged]. http://srsfia2.fs.fed.us/states/state_information.shtml.

Miles, P.D. 2008. Forest inventory mapmaker web application version 3.0. [Online]. http://www.ncrs2.fs.fed.us/4801/fiadb/fim30/wcfim30.asp. [March 31, 2008].

Randall, C.K.; Duryea, M.L.; Vince, S.W.; English, R.J. 2005. Factors influencing stump sprouting by pond cypress (*Taxodium distichum* var. *nutans* (Ait.) Sweet). New Forests. 29(3): 245-260.

Sternitzke, H.S. 1972. Bald cypress: endangered or expanding species? Economic Botany. 26(2): 130-134.

U.S. Department of Agriculture Forest Service, Forest Products Laboratory. (N.d.). Center for Wood Anatomy Research. Technology transfer fact sheet bald cypress. [Online]. http://www2.fpl.fs.fed.us/TechSheets/SoftwoodNA/htmlDocs/taxodiumdisticum.html. [September 4, 2008].

U.S. Department of Agriculture Forest Service. 2007. Forest inventory and analysis national core field guide: field data collection procedures for phase 2 plots. Version 4.0. Washington, DC. 225 p. Vol. I. Internal report. On file with: U.S. Department of Agriculture Forest Service, Forest Inventory and Analysis, 201 14th St., Washington, DC 20250.

Vince, S.W.; Duryea, M.L. 2004. Planting cypress. Circular 1458; University of Florida, Gainesville, FL: University of Florida, School of Forest Resources and Conservation, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences. 9 p.

Use of Damage Surveys and Field Inventories to Evaluate Oak and Sugar Maple Health in the Northern United States

Randall S. Morin[1]
Chris Woodall[2]
Jim Steinman[3]
Charles Perry[4]

***Abstract**: Oak species (Quercus spp.) and sugar maple (Acer saccharum) are substantial components of the forest ecosystems in the 24-state region spanning the northern U.S. During recent decades, both damage surveys and forest inventories have documented declines of sugar maple and oak health. In order to more fully assess the status of oak and sugar maple health, we examined correlations between damage detected by aerial survey data, a soil dryness index, and field-based forest conditions. Study results indicated that aerial damage surveys were correlated with an overstory attribute: percent standing dead basal area. Additionally, we present a state-level analysis as an example of how this study could be replicated for inclusion in a Forest Inventory and Analysis state report.*

Keywords: FIA, forest health, aerial damage surveys, oak decline, sugar maple decline, *Acer saccharum*, *Quercus* spp.

Introduction

Oak species (*Quercus* spp.) and sugar maple (*Acer saccharum*) are substantial components of the forest ecosystems in the 24-state region spanning the northern U.S. During recent decades, both forest inventories and aerial damage surveys have documented declines in sugar maple and oak health.

It has been suggested that North America's oak forests may be in an extended period of poor growth and susceptibility to invasive pests and droughts (Kessler 1992), a situation that has been a national forest health problem since 1960 (Thomas and Boza 1984). Oak decline results from the interaction of predisposing stress factors (defoliating insects, drought, frost/ice damage, poor

[1] Research Forester, U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA, 610-557-4054, 610-557-4250 FAX, rsmorin@fs.fed.us
[2] Research Forester, U.S. Department of Agriculture, Forest Service, Northern Research Station, St. Paul, MN, 651-649-5141, 651-649-5140 FAX, cwoodall@fs.fed.us
[3] Forest Health Monitoring Coordinator, U.S. Department of Agriculture, Forest Service, Northeastern Area State & Private Forestry, Newtown Square, PA, 610-557-4158, 610-557-4250 FAX, jsteinman@fs.fed.us
[4] Research Soil Scientist, U.S. Department of Agriculture, Forest Service, Northern Research Station, St. Paul, MN, 651-649-5191, 651-649-5140 FAX, charleshperry@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

site quality, and advanced tree age) and secondary disease and insect pests (root fungi, canker fungi, and insect borers) (Lawrence et al. 2002, Manion 1991, Starkey and Oak 1989). These many stresses eventually weaken oak trees, resulting in sparse foliage, thin crowns, crown dieback, reduced radial growth, and eventually death (Lawrence et al. 2002). Oak decline has been observed throughout most of the range of oak in the eastern United States, but has been most problematic in midwestern states (e.g., Missouri: Dwyer et al. 1995). Because oak decline is a complex etiological combination of predisposing, inciting, and contributing factors (Manion 1991, Oak et al. 1996), there is need for baseline data, long-term studies, and new analytical procedures (Kessler 1989, Nebeker et al. 1992, Oak et al. 1996).

Similarly, numerous reports of sugar maple decline or dieback have been recorded over the last 50 years. Sugar maple decline has been associated with insect defoliation, drought, unbalanced soil nutrition (particularly a lack of calcium, magnesium, and potassium), stand density, and midwinter thaw/freeze events (Horsley et al. 2000, Houston 1999, Long et al. 1997). Sugar maple trees in decline are characterized by a slow loss of crown vigor, dieback of fine twigs, and reduced radial increment over a period of years, often ending in death (Horsley et al. 2002). Houston (1999) found that crown dieback and death result when at least one predisposing stress event reduces resistance to invasion by opportunistic, secondary organisms that kill tissues. Episodes of sugar maple decline have been observed across the northern U.S.: Wisconsin in the 1950s, Massachusetts in the 1960s, New York and Vermont in the 1980s, and Pennsylvania in the 1980s and 1990s (Horsley et al. 2002).

Two inciting factors that have been attributed to oak and sugar maple decline are insect defoliation and drought. At least one past study integrated forest inventory data with aerial damage survey data to assess tree health (Morin et al. 2004) but never at the landscape-level geographic scale. Likewise, previous studies have cited drought as a factor affecting sugar maple (Drohan et al. 2002, Horsley et al. 2000, Payette et al. 1996) and oak health (Dwyer et al. 1995, Starkey and Oak 1989, Stringer et al. 1989, Tainter et al. 1984) in localized areas.

The goal of this study was to compare aerial damage surveys and a soil dryness index with field inventories of sugar maple and oak forests in the 24-state region of the northern U.S. The study had four objectives: 1) to assess the current status of oak and sugar maple health across the study area, 2) to examine correlations between the number/frequency damages identified through aerial and ground surveys with the ratio of standing dead basal area of oak and sugar maple (based on forest inventories), 3) to examine correlations between soil dryness index and the ratio of standing dead basal area of oak and sugar maple (based on forest inventories), and 4) to suggest opportunities for supplementing Forest Inventory and Analysis (FIA) state and regional reports with forest health assessments using damage survey data.

Methods

Forest Inventory Data

The FIA program of the U.S. Department of Agriculture, Forest Service, the only congressionally mandated national inventory of U.S. forests, conducts a three-phase inventory of forest attributes of the country (Bechtold and Patterson 2005). The FIA sampling design is based on a tessellation of the United States into hexagons approximately 2,428 ha in size with at least one permanent plot established in each hexagon. In phase 1, the population of interest is stratified and plots are assigned to strata to increase the precision of estimates. In phase 2, tree and site attributes are measured for forested plots established in each hexagon. Phase 2 plots consist of four 24-ft fixed-radius subplots on which standing trees are inventoried.

For assessment of current forest attributes, inventory data collected from 2002 to 2006 were used with a total of 16,689 inventory plots included in the oak analysis and inventory data collected from 2001 to 2006 were used with a total of 8,722 inventory plots included in the sugar maple analysis. This study's 24-state study region includes CT, DE, IL, IN, IA, KS, ME, MD, MA, MI, MN, MO, NE, NH, NJ, NY, ND, OH, PA, RI, SD, VT, WV, and WI. Plots were included in the analyses if at least three trees of the species of interest greater than 1-inch diameter at breast height (d.b.h.) were measured.

Aerial and Ground Survey Data

The national Forest Health Monitoring (FHM) program was initiated by the U.S. Forest Service in 1990 to monitor, assess, and report the status of and trends in forest health across the Nation. The survey component of FHM detection monitoring consists of aerial and ground surveys to detect damage in the form of tree defoliation, mortality, and damage as associated with the occurrence of damaging insects, diseases, windthrow, and other biotic and abiotic forest disturbances (Conkling et al. 2005).

Aerial surveys supply a landscape-level overview of forest health conditions at a relatively low cost (McConnell et al. 2000). Forest defoliation usually is documented by a remote-sensing technique known as sketch-mapping. A sketch-map is created while flying in an aircraft and observing damage and outlining its location on topographic maps. Sketch-mapping is an acquired and difficult skill that is somewhat subjective because human observers must rely on their judgment in identifying and delineating damaged areas.

All available aerial survey data for the 24-state region were acquired for 1997-2005. The aerial survey damage polygons were limited to include only types of damage that would be expected to affect oak and sugar maple growth and/or mortality; most of the recorded damages were from defoliators (gypsy moth, *Lymantria dispar* L., for oak and forest tent caterpillar, *Malacosoma disstria* Hubner, for sugar maple). Using a GIS, the FIA plots were overlayed with the aerial survey damage polygons to assign each plot with the number of times it received damage during the 9-year period.

Due to differing levels of effort and spending among states, the detail and quality of the survey data vary across the 24-state study area. Therefore, a subset of states deemed to have high quality survey data were selected for the study (Table 1). Pennsylvania was also chosen for an independent analysis as an example for FIA state reporting.

Table 1: List of states with high quality survey data.

Oak Analysis	Sugar Maple Analysis
Massachussetts	
Minnesota	Michigan
Missouri	Minnesota
New Hampshire	Missouri
New Jersey	New Hampshire
Pennsylvania	New York
Vermont	Vermont
West Virginia	

Soil Dryness Index

Patterns related to soil water content are often a primary factor related to tree stress and thus to insects and diseases (Elliott and Swank 1994, He and Richard 2000). The soil dryness index (DI), originally named the "natural soil wetness index" (Schaetzl 1986), is a measure of the long-term wetness of a soil. The DI concept was first initiated by Hole (1978) and Hole and Campbell (1985). It indicates the amount of water that a soil contains and makes available to plants under normal climatic conditions. The DI ranges from 0 to 99. The higher the DI, the more water the soil can supply to plants. Sites with a DI of 99 are, essentially, open water. A soil with a DI of 1 is thin and dry enough to be almost bare bedrock. This layer was obtained and intersected with the FIA plots (R. Schaetzl, 2007, pers. commun.). The DI values were binned into three categories (0-33, 34-66, 67-99) for analysis (Fig. 1).

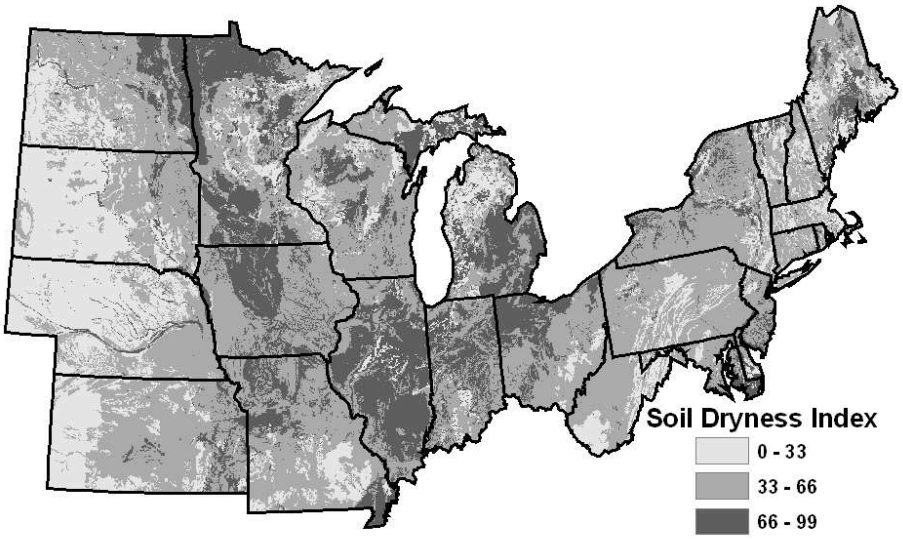


Figure 1: Map of soil dryness index.

Analysis

Spatially interpolated maps of host density and percent standing dead were produced on a 5-km by 5-km raster via a moving-window average of FIA plot values and were visually compared to the damage survey maps and soil dryness index map. One-way analysis of variance (ANOVA) was used to test the relationship of percent standing dead host material on number of damages and soil dryness index (PROC ANOVA, SAS Institute Inc. 2004). Additionally, a general linear model (PROC GLM, SAS Institute Inc. 2004) was employed to evaluate percent standing dead host material as a function of the number of damages, the soil dryness index, and the interaction between the two. Visual inspections and statistical analyses were conducted for the high quality damage survey states and for Pennsylvania separately.

Results

Oak

Oak is a critical component in forest ecosystems across the study area (Fig. 2 and Table 2). Mortality events, many presumably attributed to oak decline, over the past two decades (Lawrence et al. 2002, Manion 1991, Starkey and Oak., 1989) have led to pockets of elevated levels of standing dead oak biomass (Fig. 3 and Table 2). Based on visual inspection there appears to be some general coincidence of areas with elevated standing dead oak (Fig. 3) and recorded damage (Fig. 4); this is especially evident in Missouri, New Jersey, Pennsylvania, West Virginia, and Wisconsin.

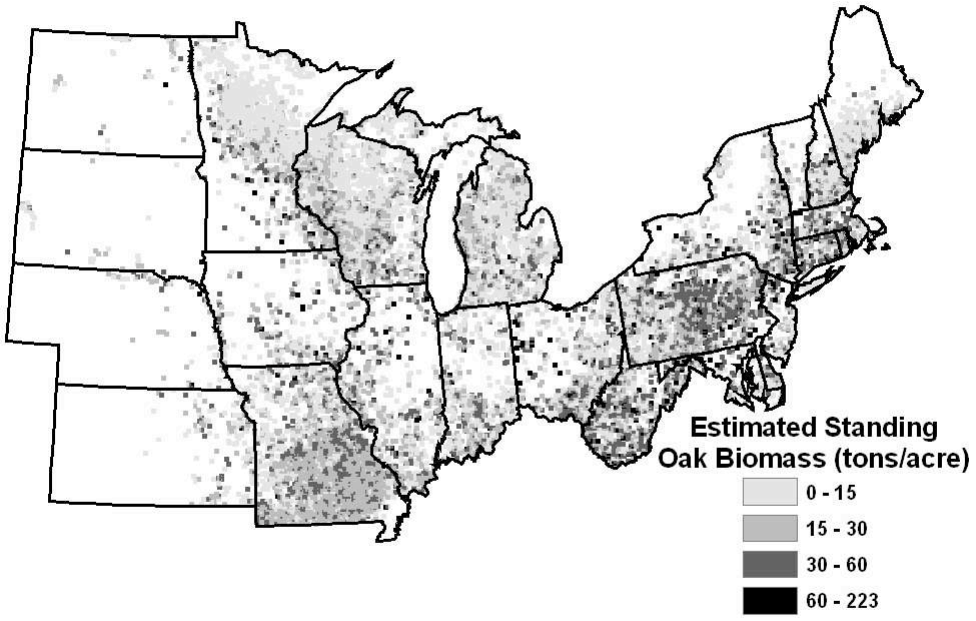


Figure 2: Map of estimated oak biomass across the 24-state study area (FIA data from 2002 to 2006).

Table 2: Mean and standard error (SE) of oak biomass and percent standing dead oak biomass by state

State	N	Oak Biomass (tons/acre)	SE (%)	Percent Standing Dead Oak Biomass	SE (%)
Delaware	47	19.9	3.572	7.2	2.190
Missouri	3189	23.7	0.304	3.8	0.165
Wisconsin	2932	15.4	0.316	3.4	0.207
Rhode Island	55	27.7	3.308	3.2	1.191
Maryland	98	25.2	2.583	3.0	0.997
Illinois	660	21.5	0.794	3.0	0.400
Connecticut	113	31.1	2.253	2.7	0.748
Iowa	295	20.7	1.057	2.6	0.648
Michigan	2187	15.6	0.383	2.2	0.179
Pennsylvania	1805	28.4	0.613	2.2	0.208
Kansas	154	15.5	1.162	2.0	0.730
Indiana	622	19.3	0.805	1.8	0.357
Minnesota	1663	12.7	0.398	1.8	0.194
New Jersey	84	26.2	2.865	1.7	0.789
New Hampshire	242	18.7	1.617	1.6	0.465
Massachussetts	178	26.5	1.774	1.6	0.410
West Virginia	502	29.9	1.136	1.6	0.280
Maine	444	11.8	0.691	1.4	0.335
Ohio	675	22.5	1.023	1.0	0.197
New York	536	23.0	1.104	0.9	0.226
Vermont	42	17.1	2.441	0.8	0.448
Nebraska	65	20.1	1.944	0.0	0.000
North Dakota	60	15.2	1.764	0.0	0.000
South Dakota	41	14.1	2.401	0.0	0.000

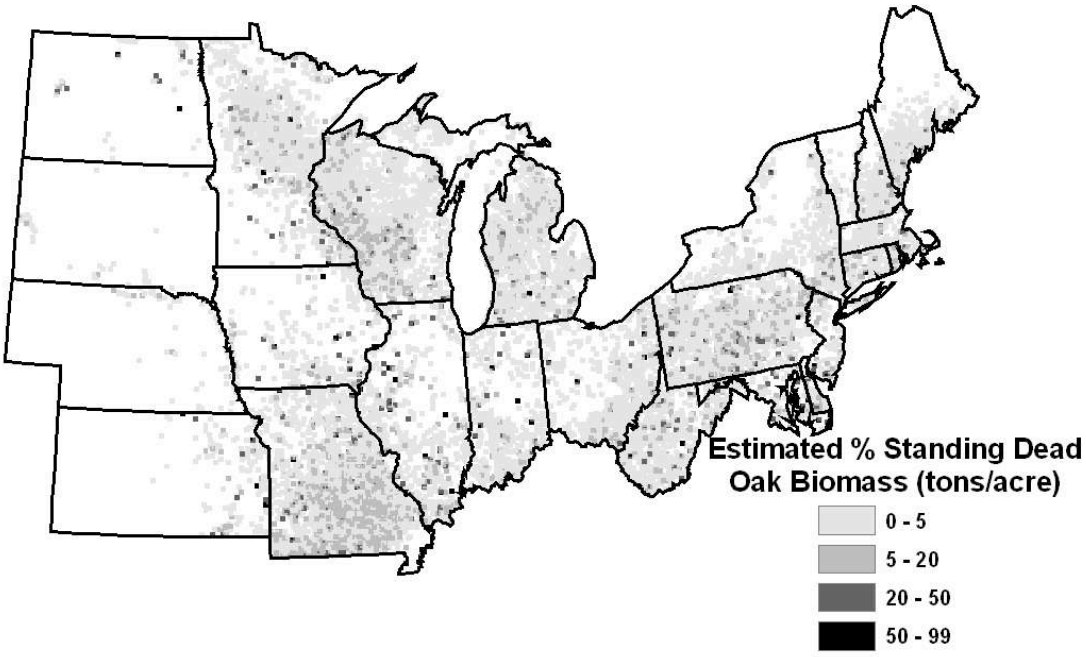


Figure 3: Map of estimated percent standing dead oak biomass across the 24-state study area (FIA data from 2002 to 2006).

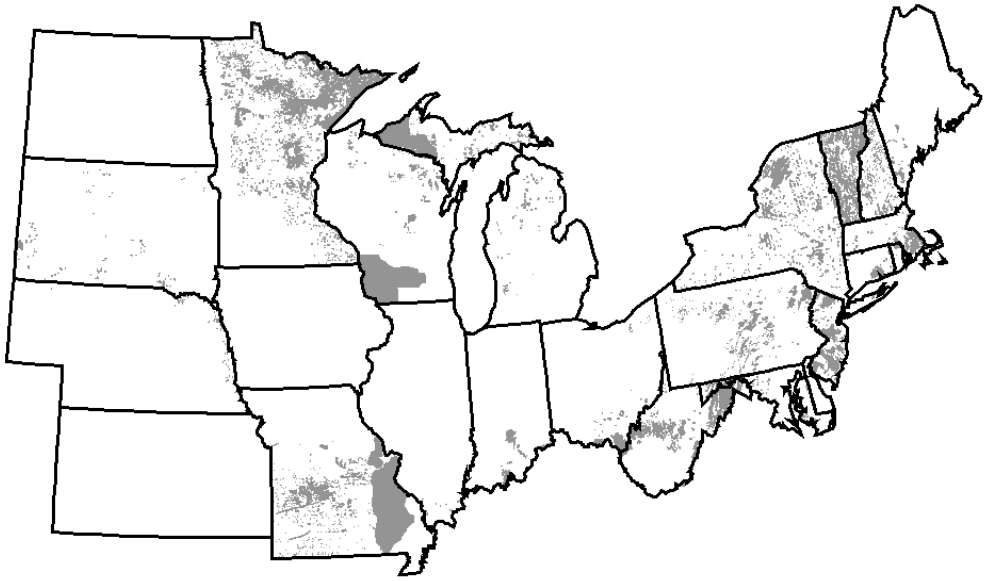


Figure 4: Map of damages expected to affect oak across the 24-state study area (survey data from 1997 to 2005).

One-way ANOVA models between percent standing dead oak biomass and years of damage (p=0.0001; Fig. 5) and between percent standing dead oak biomass and DI category (p=0.0001; Fig. 6) were statistically significant and linear in nature. Additionally, in a general linear model, years of damage (p=0.0025) and DI (p=0.0016) were significant independent predictors of percent standing dead oak biomass; the interaction between the two was not significant.

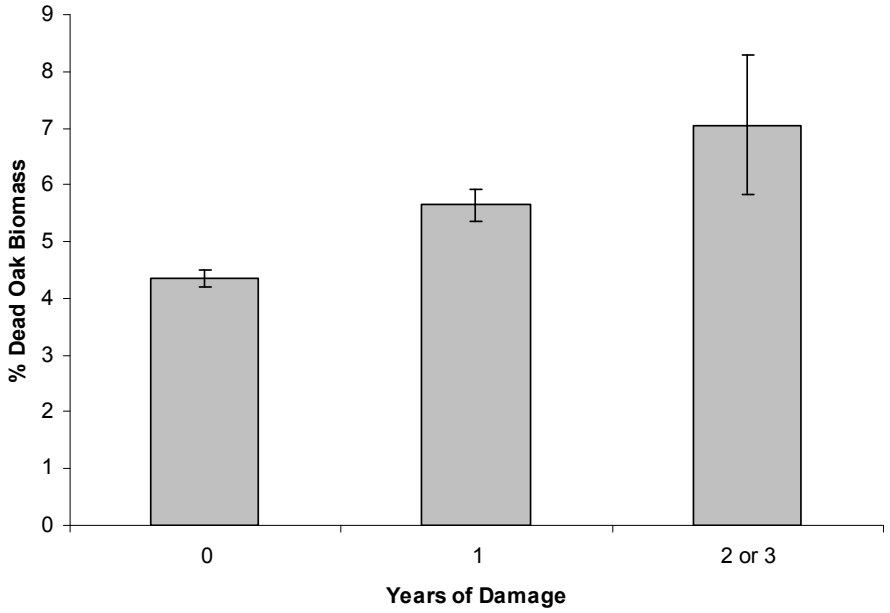


Figure 5: Percent standing dead oak biomass against years of damage across the 24-state study area.

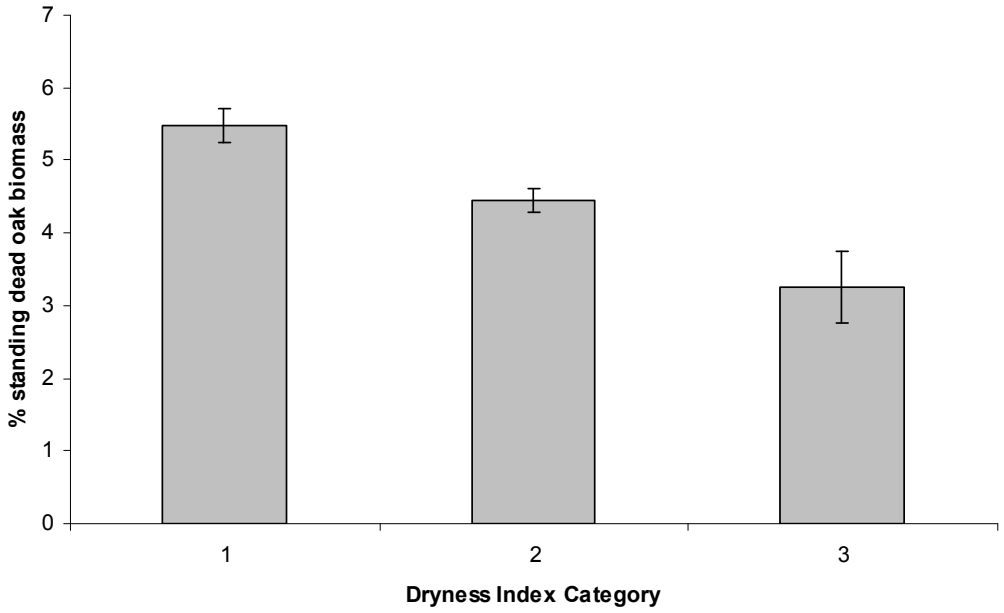


Figure 6: Percent standing dead oak biomass against dryness index category across the 24-state study area.

Sugar Maple

Sugar maple is a vital species in the forests across the study area (Fig. 7 and Table 2). Areas of elevated levels of standing dead sugar maple basal area (Fig. 8 and Table 2) are evident. Much of this mortality over the past five decades is seemingly related to sugar maple decline (Horsley et al. 2002). A visual inspection reveals a general coincidence of areas with elevated standing dead sugar maple (Fig. 8) and recorded damage (Fig. 9); this is especially evident in New York, Vermont, New Hampshire, and Massachusetts.

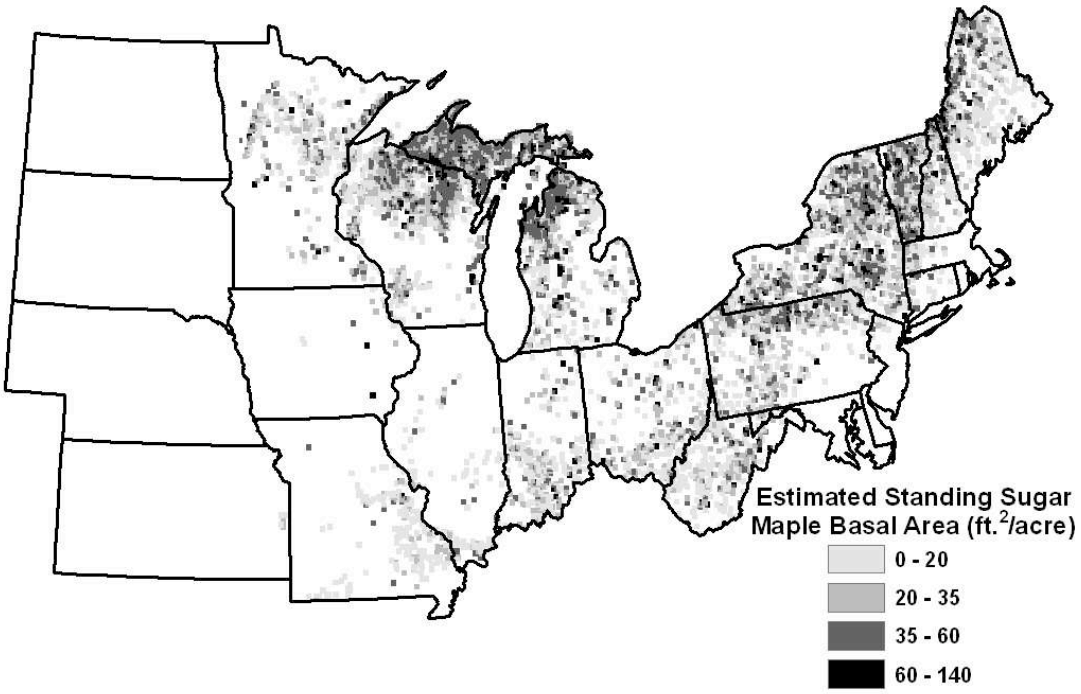


Figure 7: Map of estimated sugar maple basal area across the 24-state study area (FIA data from 2002 to 2006).

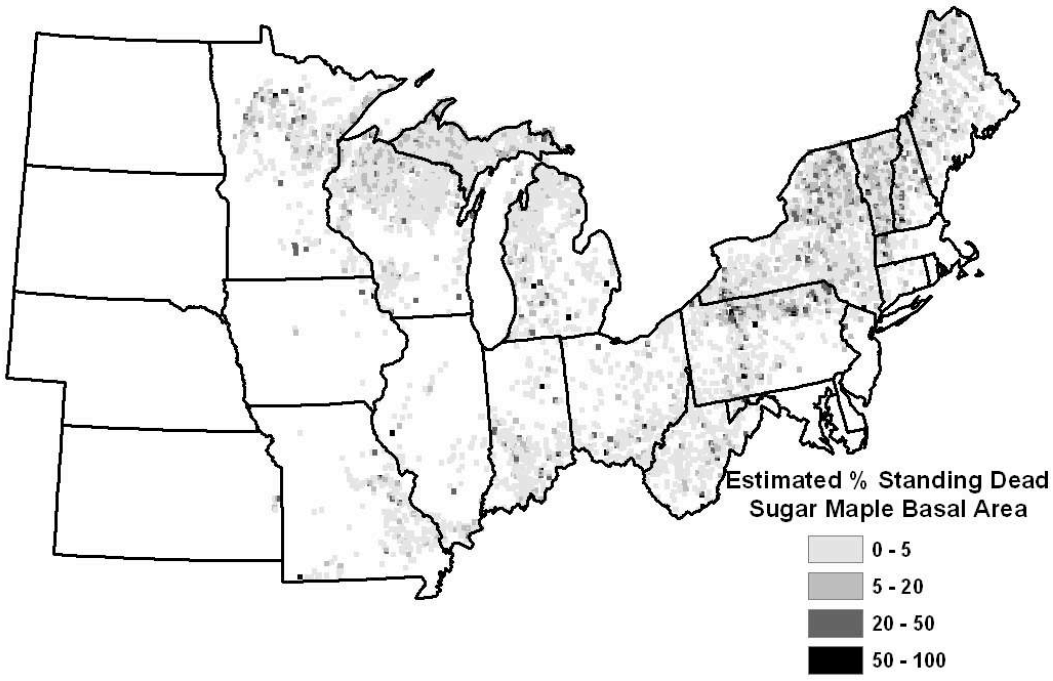


Figure 8: Map of estimated percent standing dead sugar maple basal area across the 24-state study area (FIA data from 2002 to 2006).

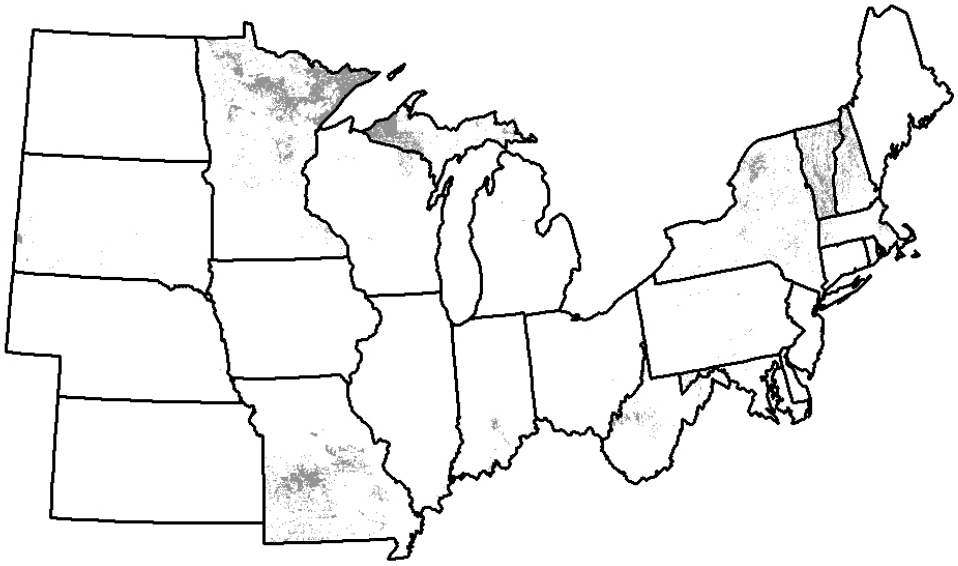


Figure 9: Map of damages expected to affect sugar maple across the 24-state study area (survey data from 1997 to 2005).

A one-way ANOVA model between percent standing dead sugar maple basal area and years of damage (p=0.0019; Fig. 10) was statistically significant and linear. A one-way ANOVA between percent standing dead sugar maple basal area and DI category (p=0.0007; Fig. 11) was also statistically significant although the relationship was not linear. In a general linear model, years of damage (p=0.0006) was a significant independent predictor of percent standing dead sugar maple basal area; DI and the interaction between years of damage and DI were not significant.

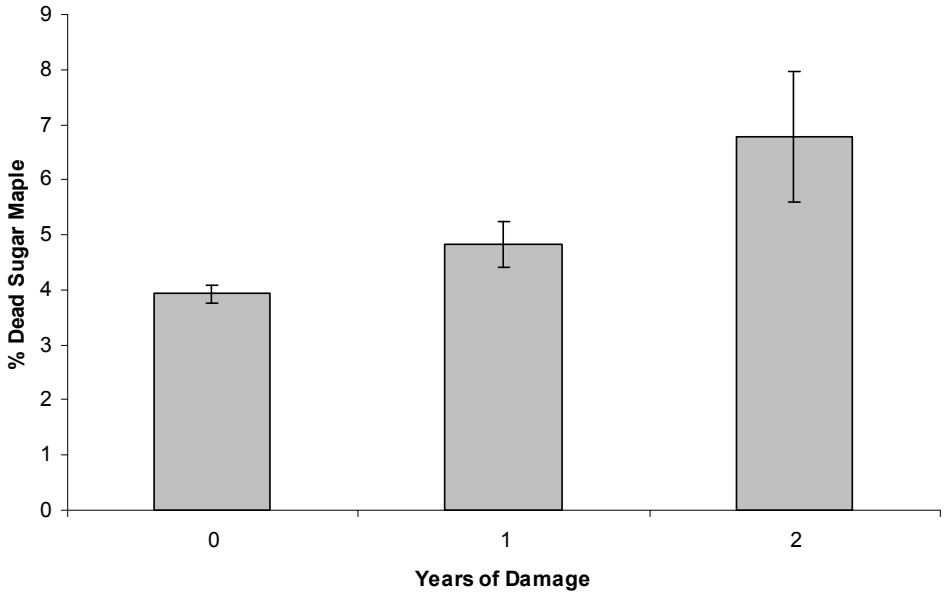


Figure 10: Percent standing dead sugar maple basal area against years of damage across the 24-state study area.

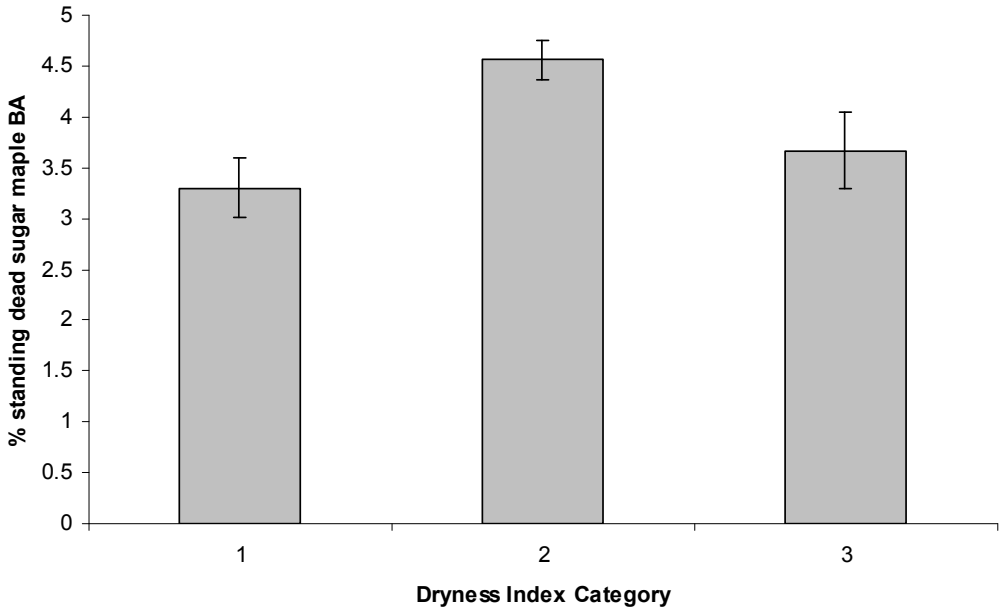


Figure 11: Percent standing dead sugar maple basal area against dryness index category across the 24-state study area.

Pennsylvania – An Example of an Analysis for an FIA State Report

Oak is the most abundant genus in Pennsylvania and is very important from an ecological and economic standpoint. Elevated levels of standing dead oak biomass (Fig. 12) exist in specific areas across the state. There appears to be some general coincidence between these areas and recorded damage (Fig. 13). A one-way ANOVA model between percent standing dead oak biomass and years of damage (p=0.0001; Fig. 14) was statistically significant and linear. A one-way ANOVA between percent standing dead oak biomass and DI category (p=0.0007; Fig. 15) was also statistically significant although the relationship was not linear. In a general linear model, years of damage (p=0.0136) and DI category (p=0.0026) were significant independent predictors of percent standing dead oak biomass; the interaction between years of damage and DI was not significant.

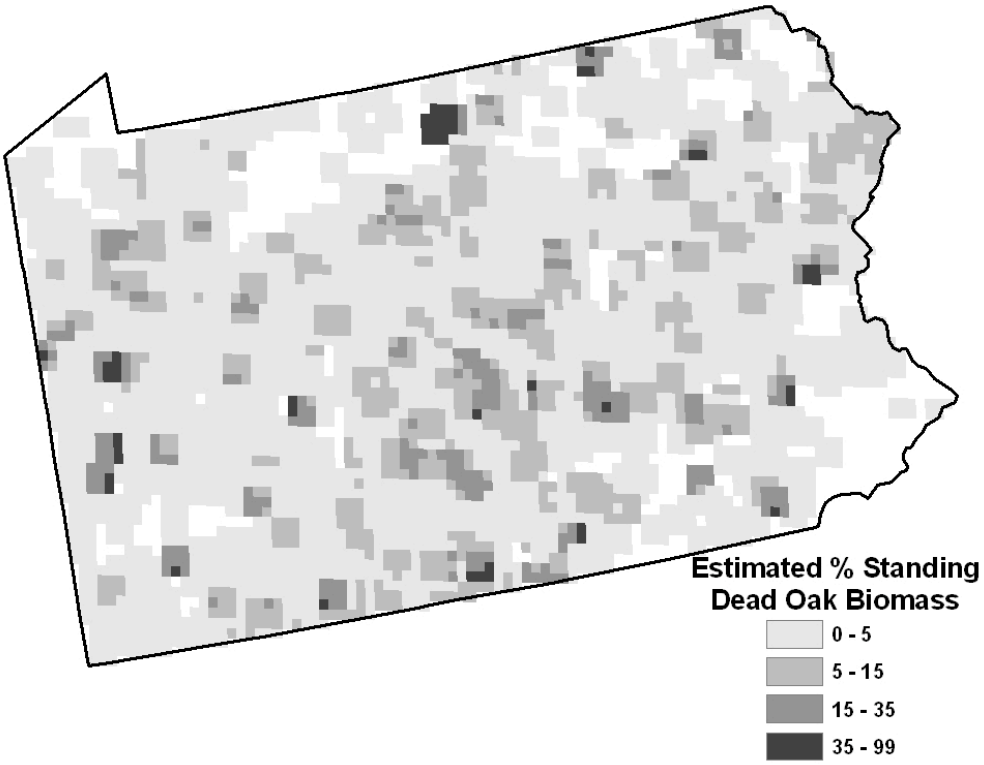


Figure 12: Map of estimated percent standing dead sugar maple basal area in Pennsylvania (FIA data from 2002 to 2006).



Figure 13: Map of damages expected to affect oak in Pennsylvania (survey data from 1997 to 2005).

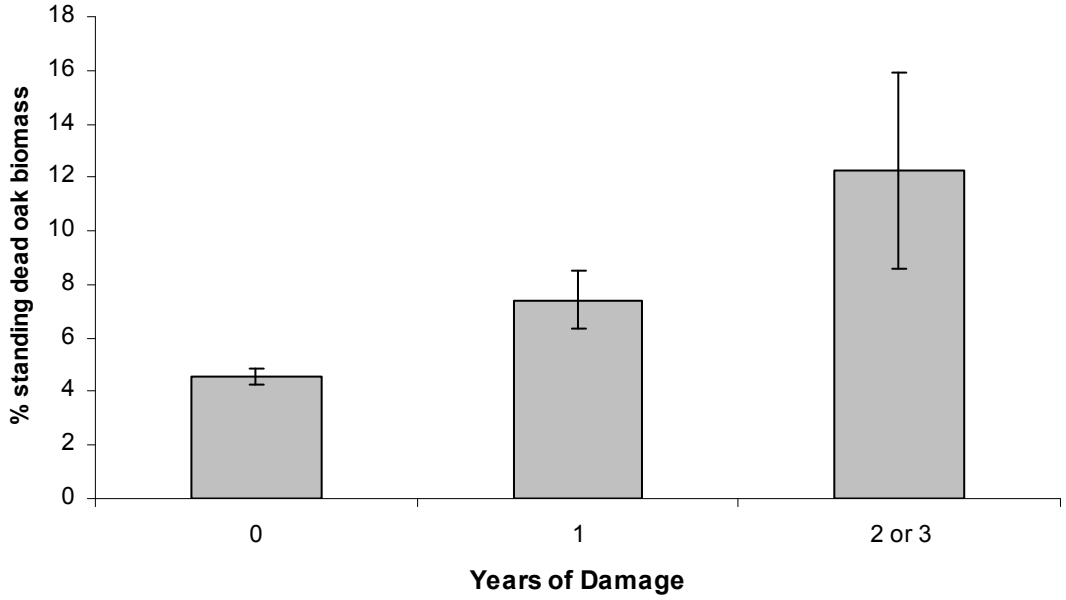


Figure 14: Percent standing dead oak biomass against years of damage in Pennsylvania.

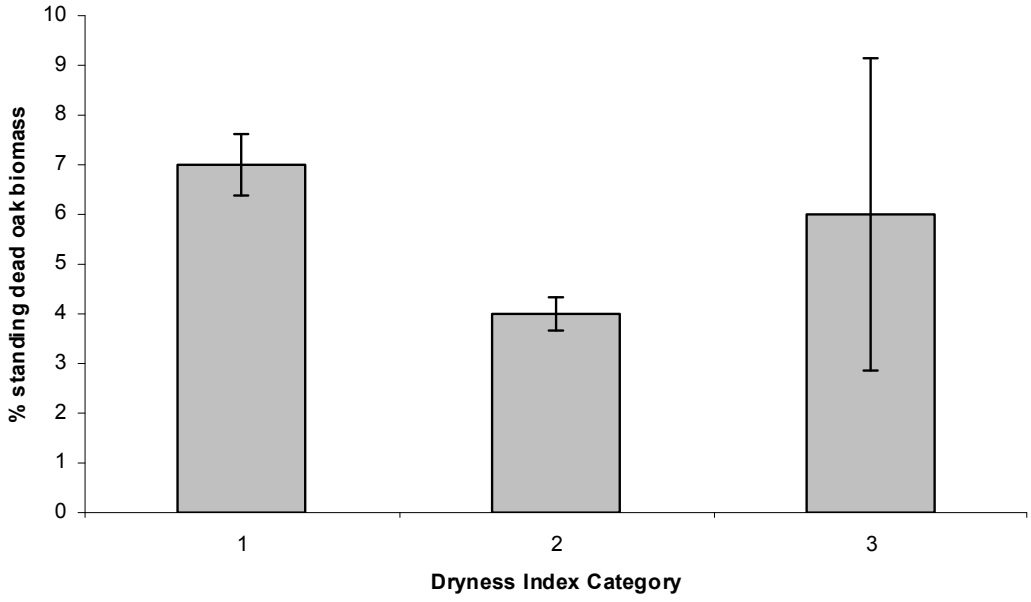


Figure 15: Percent standing dead oak biomass against dryness index category in Pennsylvania.

Discussion and Future Direction

Many FIA national, regional, and state reports include maps of various characteristics of the forest resource. Maps of mortality and/or standing dead trees are often presented with little more than anecdotal evidence to support spatial patterns. The spatial data used in this study provide an opportunity to relate levels of mortality or standing dead to potential causal factors. Quantification of these relationships is valuable from a land management perspective because it will allow managers to predict the impact of defoliation and/or significant drought events on important forest species. The potential is there for FIA analysts to use the damage survey datasets as an ancillary source for resource analysis, but caution is recommended due to the inconsistency of data collection between states.

Future work is underway to expand the scope of this study to include FIA phase 3 variables including crown dieback, crown density, and soil chemistry measures. Adding the crown measures as dependent variables will provide further insight into the relationship among tree health, damage, and soil dryness. Furthermore, soil nutrition has been cited as an important factor in sugar maple decline (Horsley et al. 2000, Long et al. 1997). Measures of available calcium and magnesium in soil samples from phase 3 plots will be related to sugar maple crown health and percent standing dead basal area. Although the soil dryness index provided a measure of drought potential, it did not account for actual levels of precipitation. Therefore, precipitation is another important factor that we plan to add to future analyses. Finally, we plan to compare spatial patterns of standing dead and/or mortality from FIA plots to areas of predicted mortality highlighted by the U.S. Forest Service's Forest Health Protection national risk map (Krist et al. 2007).

Acknowledgments

We would like to thank the Forest Health Monitoring program for providing funding for this study.

Literature Cited

Bechtold, W.A.; Patterson, P.L., eds. 2005. Forest Inventory and Analysis National Sample Design and Estimation Procedures. Gen. Tech. Rep. SRS-GTR-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Conkling, B.L.; Coulston, J.W.; Ambrose, M.J. eds. 2005. Forest health monitoring: 2001 national technical report. Gen. Tech. Rep. SRS-81. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 204 p.

Drohan, P.J.; Stout, S.L.; Petersen, G.W. 2002. Sugar maple (*Acer saccharum* Marsh.) decline during 1979-1989 in northern Pennsylvania. Forest Ecology and Management. 170(1): 1-17.

Dwyer, J.P.; Cutter, B.E.; Wetteroff, J.J. 1995. A dendrochronological study of black and scarlet oak decline in the Missouri Ozarks. Forest Ecology and Management. 75: 69-75.

Elliott, K.J.; Swank, W.T. 1994. Impacts of drought on tree mortality and growth in a mixed hardwood forest. Journal of Vegetation Science. 5(2): 229-236.

He, F.; Richard, D.P. 2000. Density-dependent effects on tree survival in an old growth Douglas fir forest. Journal of Ecology. 88(4): 676-688.

Hole, F.D. 1978. An approach to landscape analysis with emphasis on soils. Geoderma. 21: 1-13.

Hole, F.D.; Campbell, J.B. 1985. Soil landscape analysis. Totowa, NJ: Rowman and Allanheld. 196 p.

Horsley, S.B.; Long, R.P.; Bailey, S.W.; [et al.]. 2000. Factors associated with the decline disease of sugar maple on the Allegheny plateau. Canadian Journal of Forest Research. 30(9): 1-14.

Horsley, S. B.; Long, R. P.; Bailey, S. W.; [et al.]. 2002. Health of eastern north American sugar maple forests and factors affecting decline. Northern Journal of Applied Forestry. 19(1): 34-44.

Houston, D. R. 1999. History of sugar maple decline. In: Horsley, S.B.; Long, R.P., eds. Sugar maple ecology and health: Proc. international symposium; 1998 June 2-4; Warren, PA. Gen. Tech. Rep. GTR-NE-261. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station: 19-26.

Kessler, K.J., Jr. 1989. Some perspectives on oak decline in the '80's'. In: Rink, G.; Budelsky, C. A., eds. Proc. 7th central hardwood conference; 1989 March 5-8; Carbondale, IL. Gen. Tech. Rep. NC-GTR 132. St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station: 25-29.

Kessler, K. J., Jr. 1992. Oak decline on public lands in the Central Forest Region. Res. Note RN-NC-362. St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station. 4 p.

Krist, F.J., Jr.; Sapio, F.J.; Tkacz, B.M. 2007. Mapping risk from forest insects and diseases. FHTET 2007-06. Fort Collins, CO: U.S.Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team. 115 p.

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

Long, R.P.; Horsley, S.B.; Lilja, P.R. 1997. Impact of forest liming on growth and crown vigor of sugar maple and associated hardwoods. Canadian Journal of Forest Research. 27: 1560-1573.

Manion, P. D. 1991. Tree disease concepts, 2nd ed. Upper Saddle River, NJ: Prentice Hall Career & Technology. 416 p.

McConnell, T.J.; Johnson, E.W.; Burns, B. 2000. A guide to conducting aerial sketchmapping surveys. FHTET 00-01. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team. 88 p.

Morin, R.S., Jr.; Liebhold, A.M.; Gottschalk, K.W. 2004. Area-wide analysis of hardwood defoliator effects on tree conditions in the Allegheny plateau. Northern Journal of Applied Forestry. 21(1): 31-39.

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.

Oak. S.; Tainter, F.; Williams, J.; Starkey, D. 1996. Oak decline risk rating for the southeastern United States. Annals of Forest Science. 53: 721-730.

Payette, S.; Fortin, M.J.; Morneau, C. 1996. The recent sugar maple decline in southern Quebec: probable causes deduced from tree rings. Canadian Journal of Forest Research. 26: 1069-1078.

SAS Institute Inc. 2004. SAS OnlineDoc® 9.1.2**.** Cary, NC: SAS Institute.

Schaetzl, R.J. 1986. A soilscape analysis of contrasting glacial terrains in Wisconsin. Annals of the Association of American Geographers. 76: 414-25.

Starkey, D.A.; Oak, S.W. 1989. Site factors and stand conditions associated with oak decline in southern upland hardwood forests. In: Rink, G.; Budelsky, C.A., eds. Proc. 7th central hardwood conference; 1989 March 5-8; Carbondale, IL. Gen. Tech. Rep. NC-GTR 132. St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station: 95-102.

Stringer, J.W.; Kimmerer, T.W.; Overstreet, J.C.; [et al.]. 1989. Oak mortality in eastern Kentucky. Southern Journal of Applied Forestry. 13: 86-91.

Tainter, F.H.; Fraedrich, S.W.; Benson, D.M. 1984. The effect of climate on growth, decline, and death of northern red oaks in the western North Carolina Nantahala Mountains. Castanea. 49(3): 127-137.

Thomas, W.D.; Boza, C.A. 1984. The oak decline complex. Journal of Arboriculture. 10: 170-177.

Biomass Applications



A real-time, web-based optimal Biomass Site Assessment Tool (BioSAT)

Module 1. An Economic Assessment of Mill Residues for the Southern U.S.

Timothy M. Young[1]
James H. Perdue
Andy Hartsell
Robert C. Abt
Donald G. Hodges
Timothy G. Rials

ABSTRACT: *Optimal locations for biomass facilities that use mill residues are identified for 13 southern U.S. states. The Biomass Site Assessment Tool (BioSAT) model is used to identify the top 20 locations for 13 southern U.S. states. The trucking cost model of BioSAT is used with Timber Mart South 2009 price data to estimate the total cost, average cost, and marginal costs for biomass facilities that use mill residues for up to 1.5 million dry tons of annual consumption. Demand locations are based on the U.S. Census Bureau zip code tabulation areas (ZCTA). There are 9,353 zip code tabulation areas (ZCTA) in the 13-state study region. Demand point location based on a ZCTA offers an improvement in truck cost estimates when compared to demand point location based on a county centroid. The top 20 ZCTAs in the study region are located in south Mississippi, southeast Georgia, southeast Oklahoma, southwest Alabama, and east Texas. Costs in these areas range from $25 to $38 per dry ton for up to 1.5 million annual dry tons.*

Additional research on BioSAT is forthcoming for 33 eastern U.S. states. These studies will include more types of woody and agricultural biomass (e.g., logging residues, pulpwood, corn stover, etc.). Additional cost models for transportation such as truck combinations with rail and barge will be components of BioSAT.

KEYWORDS: Biomass, economic availability, siting model, BioSAT, mill residues.

---

[1]Research Associate Professor (Ph.D.), Forest Products Center, University of Tennessee, 2506 Jacob Drive, Knoxville, TN 37996-4570, phone: 865.946.1119, e-mail: tmyoung1@utk.edu
Biomass and Bioenergy Liaison, Director's Office, USDA Forest Service, Southern Research Station. Office Location @ University of Tennessee, Office of Bioenergy Programs, Room 112, 2506 Jacob Drive, Knoxville, TN 37996-4570, phone: 865-946-1123, e-mail: jperdue@fs.fed.us
Research Forester, USDA Forest Service, Forest Inventory and Analysis, 4700 Old Kingston Pike, Knoxville, TN 37919, phone : 865.862.2032, e-mail: ahartsell@fs.fed.us
Professor and Director (Ph.D.), Department of Forestry and Environmental Resources, North Carolina State University, 3126 Jordan Hall Campus Box 8008, Raleigh, NC 27695-8008, phone: 919. 515.7791, e-mail:bob_abt@ncsu.edu
Professor and Director (Ph.D.), Department of Forestry, Wildlife and Fisheries, University of Tennessee, 274 Ellington Plant Sciences, Knoxville, TN 37996, phone: 865.974.7126, e-mail: dhodges2@utk.edu
Professor and Directors (Ph.D.), Office of Bioenergy Programs, Forest Products Center, University of Tennessee, 2506 Jacob Drive, Knoxville, TN 37996-4570, phone: 865.946.1129, e-mail: trials@utk.edu

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

INTRODUCTION

The 20th century was marked by rapid growth and an increase in prosperity throughout the world. Even though there is a current global economic recession with declines in oil and other fossil fuel prices, fossil fuel supply chains remain sensitive to disruption from unanticipated geo-political events. A resurgence of emerging economies in the future will most likely increase the future demand for fossil fuels that exist in complex geopolitical areas. By 2030, some experts predict the world's energy consumption will be 50 percent higher than it is today (International Energy Outlook 2008). As noted in the USDA Forest Service Woody Biomass Utilization Strategy, using woody biomass for renewable energy contributes to the Nation's energy independence (USDA Forest Service 2008). The woody biomass removed during ecological restoration, wildfire risk reduction, and conventional silvicultural activities can become a source of energy that are renewable and contribute to U.S. energy independence (USDA Forest Service 2008).

The forest products industry is an established user of wood wastes and residues for energy generation and is the major supplier of bioenergy in most of the developed world (Sedjo 1997). The proximity of wood wastes to forest products mills makes it a sensible solution for energy production. The emerging bioenergy industry can learn from the forest products industry in the procurement of forest-based biomass. Economic benefits may arise from a synergistic coexistence between the established forest products industry and the emerging bioenergy industry.

The development of any new industry involves the establishment of many relationships (Altman and Johnson 2008). Assessing the economic capability and stability of the bioenergy supply infrastructure is essential for market organization of this emerging industry and is addressed by this study. Perlack et al. (2005) indicate that the nation's forests represent a strategic asset in meeting the national goal of replacing 30 percent of the domestic petroleum consumption by 2030. Even though research has been conducted which estimates the economic availability of biomass (Young and Ostermeier 1989, Young et al. 1991; Lunnan 1997; Walsh 1998, 2000; DiPardo 2000; Ugarte et al. 2000, 2006, 2007; Western Governors Association 2008; Biomass Research and Development Board 2008), additional research on the economics of biomass energy in the context of web-based user tools would benefit bioenergy research and provide practitioners with useful information. An emphasis on the development of web-based information tools for policy makers, planners, and investors is essential for facilitating market organization.

The study presented in this paper develops estimates of the marginal costs curves and supply curves for mill residues, and identifies optimal sites for mill residue

using facilities for 13 southern U.S. states.[2] Future studies will be expanded for 20 additional eastern U.S. states.[3] The resolution of the study is the "zip code tabulation area" or ZCTA (U.S. Census Bureau 2000).

The study is on-going and all objectives are not complete. The objectives of the overall study are: 1) develop a Microsoft SQL database of resource data (forest, mill residue, and agricultural feedstocks); 2) develop resource costs for the database; 3) develop a transportation cost model for database; 4) develop a harvesting cost model for database; 5) develop a web-based software of the system (e.g., www.BioSAT.net); and 6) update of key input data, e.g., diesel prices, mill residue prices, FIA data, etc., as necessitated.

METHODS

A model for siting biomass processing facilities ("BioSAT") is developed in the study (Figures 1A, 2A, and 3A in Appendix). The model has three cost components (i.e., resource, harvesting, and transportation) which are discussed in more detail this section. Forest resource data are obtained from the USDA Forest Service current FIA inventory data (U.S. Department of Agriculture, Forest Service 2008). Agricultural resource data development is on-going and will be obtained from state and federal reporting agencies.

Forest Resource Data

County level estimates of all-live total biomass, as well as average annual growth, removals, and mortality are obtained from the Forest Inventory and Analysis Database (FIADB) version 3.0 (U.S. Department of Agriculture, Forest Service, 2008). The latest complete cycle of data for each state is used (Table 1). New FIA data, when available, will be updated in the BioSAT model. Estimates of mill residues, urban waste, logging residuals, thinnings, and other removals are obtained from the Billion Ton 2 (BT2) study (Perlack et al. 2005). All data in green tons are converted to dry tons in the analyses.

County level estimates are allocated to "zip code tabulation areas" (ZCTA's) based on area proportionality, e.g., if a ZCTA accounts for ten percent of a county, ten percent of the county's data are assigned to that ZCTA. If a ZCTA boundary crosses multiple counties, proportions for each county are summed.

ZCTA's are based on the 2000 census definition and are obtained from the U.S. Census Bureau (U.S. Census Bureau 2000). Area proportionality is performed using ArcGIS (http://www.esri.com/software/arcgis/ Accessed January 5, 2009),

[2] Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, and Virginia.
[3] Connecticut, Delaware, Illinois, Indiana, Iowa, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Hampshire, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, Vermont, West Virginia, and Wisconsin

which produces a file containing ZCTAs, county Federal Information and Processing Standard (FIPS) codes, and the percentage each county has in the ZCTAs. An ORACLE™ database (http://www.oracle.com/database/index.html Accessed January 5, 2009) is created for this file of FIA and BT2 county level data. ZCTA level estimates are derived from the information in this database.

TABLE 1. State and year of USFS FIA inventory data.

State	Year	State	Year
Alabama	2007	Nebraska	2006
Arkansas	2007	New Hampshire	2006
Connecticut	2005	New Jersey	2006
Delaware	2006	New York	2006
Florida	2006	North Carolina	2006
Georgia	2007	North Dakota	2007
Illinois	2006	Ohio	2006
Indiana	2007	Oklahoma	1993
Iowa	2006	Pennsylvania	2006
Kansas	2006	Rhode Island	2006
Kentucky	2006	South Carolina	2006
Louisiana	2005	South Dakota	2007
Maine	2006	Tennessee	2006
Maryland	2006	Texas	2007
Massachusetts	2006	Vermont	2006
Michigan	2007	Virginia	2007
Minnesota	2007	West Virginia	2006
Mississippi	2006	Wisconsin	2007
Missouri	2006		

As ZCTAs do not account for all zip codes, a file containing all possible zip codes as of January 31, 2008 is used from zip-codes.com (http://www.zip-codes.com/ Accessed January 5, 2009*)*. This file contains the zip code, latitude, and longitude of the mail office associated with each zip code. These points are then assigned to the corresponding ZCTA. Users can query using any zip code, although the results are based on ZCTAs. i.e., there are 33,568 zip codes and 24,795 ZCTAs in the 33-state study region.

Confidence bounds of individual county level FIA data can be wide. Therefore, estimates of individual ZCTAs are not used in this study, but ZCTAs are aggregated together into larger groupings or "biosheds" where confidence bounds may be comparable to aggregate county groupings. Confidence bounds of the resource supply in any given bioshed which is a grouping of ZCTAs do not offer any improvement over existing studies which aggregate county-level resource supply data. However, using the ZCTA of a demand point offers improvement of confidence bounds of cost estimates when compared to existing studies which rely on estimates using the county centroid as a demand point. Counties can be large and have geographic barriers that impact transportation time and distance (e.g., bridges over large waterways, mountains, large metropolitan areas, etc.).

Using the ZCTA of a demand point with the transportation network from a grouping of supply ZCTAs will improve the estimate of costs relative to using the centroid of the county as the demand point.

Land use of both counties and ZCTAs are not considered in the current study, i.e., a zip code may be predominately urban, but it will receive the amount of resources assigned to it based solely on area proportion. Future studies and enhancements of the BioSAT model will attempt to recognize county and zip code land type, and hence the allocation of resources to a ZCTA. All areas classified as water and unproductive lands are removed from all datasets before the area proportionality process is performed.

Resource Costs

Resources cost data (e.g., stumpage, mill residue prices, log prices, etc.) for the southeastern U.S. are obtained from Timber Mart South (TMS), see http://www.tmart-south.com/tmart/ (Accessed January 12, 2009). TMS mill residue price data (e.g., hardwood sawdust, pine sawdust, and pine shavings) for a state are allocated equally to all ZCTAs. There are currently no estimates for logging residue stumpage in the model.

Resource cost data in the northern U.S. are obtained from multiple sources (Table 2). A significant constraint for resource cost data in these northern regions is the absence of a regional price reporting system similar to TMS. This research is on-going and results for the northern regions are not reported in this paper.

Transportation Costs

Transportation Network:

Microsoft© MapPoint® 2006 (http://www.microsoft.com/MapPoint/en-us/default.aspx Accessed January 5, 2009) is used in BioSAT to provide the shortest travel time routes and distances between ZCTAs. Road networks in MapPoint® are a combination of the Geographic Data Technology, Inc. (GDT) and Navteq data. GDT data are used for rural areas and small to medium size cities. Navteq data are used for major metropolitan areas.

The GDT data are based on "Tele Atlas Dynamap Streets" which is designed for address level geocoding (http://www.teleatlas.com/index.htm Accessed January 12 2009). When an address level geocode is not available the GDT data set uses cascading accuracy at the ZIP+4, ZIP+2, and ZIP Code centroid to return the highest level of geocode for the address. ZIP code boundary data are based on the Dynamap/5-Digit ZIP code Boundary data from Tele Atlas North America. It is designed to identify the boundaries of United States Postal Service ZIP Codes.

Navteq maps provide a highly accurate representation of the detailed road

network including up to 260 attributes like turn restrictions, physical barriers and gates, one-way streets, restricted access, and relative road heights http://www.navteq.com/about/whatis.html Accessed January 12 2009).

TABLE 2. Price reporting sources in the northern 20 U.S. state region.

State/Region	Report	Data	Frequency	weblink
Connecticut	Southern New England Stumpage Price	Stumpage: Sawtimber (multiple species), fuelwood, pulpwood, biomass	Quarterly	http://forest.fnr.umass.edu/snespsr/reports/all%20reports.htm
Illinois	Illinois Timber Prices	Stumpage, FOB: Sawtimber (multiple species), pulpwood	Bi-annual	http://web.extension.uiuc.edu/forestry/il_timber_prices/index.html
Indiana	Indiana Forest Products Price Report	Delivered: Sawtimber (multiple species), pulpwood	Annual	http://www.fnr.purdue.edu/extension/pricereports.shtml
Kentucky	Kentucky Delivered Prices	Delivered: Sawtimber (multiple species), pulpwood	Quarterly	http://www.forestry.ky.gov/programs/utilize/Kentuckys+Growing+Gold+Bulletin.html
Maine	Maine Annual Price Report	Stumpage: Sawtimber (multiple species), fuelwood, pulpwood, biomass	Quarterly	http://www.state.me.us/doc/mfs/pubs/annpubs.htm#stump
Massachusetts	Southern New England Stumpage Price	Stumpage: Sawtimber (multiple species), fuelwood, pulpwood, biomass	Quarterly	http://forest.fnr.umass.edu/snespsr/reports/all%20reports.htm
Michigan	Michigan Stumpage Price Report	Stumpage: Sawtimber (multiple species), fuelwood, pulpwood, biomass	Quarterly	http://www.michigan.gov/dnr/0,1607,7-153-10368_22594-81536--,00.html http://www.michigandnr.com/ftp/forestry/tsreports/StumpagePriceReports/12_Month_Stumpage_Price_Reports/
Minnesota	Minnesota Forest Resources Report	Stumpage: Sawtimber (multiple species), pulpwood	Annual	http://www.dnr.state.mn.us/forestry/um/index.html
Missouri	Missouri Timber Price Trends	Stumpage: Sawtimber (multiple species)	Quarterly	http://mdc4.mdc.mo.gov/applications/MDCLibrary/MDCLibrary2.aspx?NodeID=854
New Hampshire	New Hampshire Average Stumpage	Stumpage: Sawtimber (multiple species), pulpwood	Bi-annual	http://www.nh.gov/revenue/munc_prop/avgstumpval.htm
Ohio	Ohio Timber Prices	Delivered: Sawtimber (multiple species)	Bi-annual	http://www.oardc.ohio-state.edu/ohiowood/
Pennsylvania	Pennsylvania Timber Market Report	Stumpage, Delivered: Sawtimber (multiple species)	Quarterly	http://www.sfr.cas.psu.edu/TMR/TMR.htm
Vermont	Vermont Forest Quarterly	Stumpage: Sawtimber (multiple species), fuelwood, pulpwood, biomass	Quarterly	http://stumpage.uvm.edu/stumpage.php
West Virginia	West Virginia Timber Market Report	Stumpage: Sawtimber (multiple species)	Quarterly	http://ahc.caf.wvu.edu/index.php?option=com_wrapper&Itemid=116

Trucking Costs:

The current transportation cost model estimates trucking costs. The analysis in this paper assumes dry-van storage trailers for trucking given that mill residues are the biomass type being hauled. Other trailer types are planned in the trucking model for different biomass types, e.g., pulpwood, corn stover, switchgrass, etc. The final study will include additional truck/rail, rail, and barge cost models.

The trucking cost model given in equations (1), (2), (3), and (4) is an adaptation of the model by Berwack et al. (2003). Diesel fuel cost efficiencies; tire variable costs; tax and license fees; and management and overhead costs of the Berwack et al. (2003) model are modified for the BioSAT model. Modifications to Berwack et al. (2003) model are made from a review of the model in October 2008 by three trucking companies[4] and one wood-using company that requested anonymity. The trucking cost model is assumed to be for contract carriers of the biomass consuming company. In most cases, contract carriers are the least cost form of truck transportation for a biomass consuming facility (personal communication: see footnote 2).

Trucking costs are a function of: variables costs which are dependent on haul time; variable costs which are dependent on haul distance; fixed costs which are dependent on haul distance; and the quantity demanded at a ZCTA demand point. The following equations are presented for the cost model:

$$\text{Total Truck Cost} = \sum_{r=1}^{z} \left(\left(V_{t,s(i,j)} + V_{d,s(i,j)} + F_{d,s(i,j)} \right) \times B_{(i,j)} \right) \qquad (1)$$

where, $V_{t,s(i,j)}$ = variable cost for t for s of (i,j),
$V_{d,s(i,j)}$ = variable cost for d for s of (i,j),
$F_{d,s(i,j)}$ = fixed cost for d for s of (i,j),
$B_{(i,j)} = Q_i / C_s$ = total hauls for each (i,j) for all routes r, $r = 1 \ldots z$,
C_s = legal trailer capacity for s,
Q_i = annual capacity of demand ZCTA, i ($i = 1 \ldots m$),
d = round-trip travel distance (i,j),
i = demand ZCTA, $i = 1 \ldots m$,
j = supply ZCTA, $j = 1 \ldots n$,
m = total number of biomass supply ZCTAs,
n = total number of biomass supply ZCTAs,
r = route (i, j), $r = 1 \ldots z$,
s = U.S. state, $q = 1 \ldots 33$,
t = round-trip travel time (i,j).

[4] Pemberton Truck Lines, Inc. (Knoxville, TN); Skyline Transportation, Inc. (Knoxville, TN); and Mason Dixon, Inc. (Scottsboro, AL).

$$V_{d,s(i,j)} = D_{d,s(i,j)} + M_{d(i,j)} + T_{d(i,j)} \quad (2)$$

where, $D_{d,s(i,j)}$ = diesel fuel cost for *d* for *s* of (*i, j*),
$M_{d(i,j)}$ = maintenance and repair cost for *d* for (*i, j*),
$T_{d(i,j)}$ = tire cost for *d* for (*i, j*).

$$V_{t,s(i,j)} = L_{t,s(i,j)} \quad (3)$$

where, $L_{t(i,j)}$ = labor cost for *t* for *s* of (*i,j*),

$$F_{d,s(i,j)} = E_{d(i,j)} + S_{s(i,j)} + N_{s(i,j)} + O_{d(i,j)} + I_{s(i,j)} \quad (4)$$

where, $E_{d(i,j)}$ = equipment cost for *d* for (*i,j*),
$S_{s(i,j)}$ = tax for *s* for (*i,j*),
$N_{s(i,j)}$ = license fee for *s* for (*i,j*),
$O_{d(i,j)}$ = management and overhead cost for *d* for (*i,j*),
$I_{s(i,j)}$ = insurance cost for *s* for (*i,j*).

The variable cost inputs for the trucking model (e.g., diesel fuel, labor wages, etc.) are updated bi-monthly in the BioSAT model. Minimum transportation travel times and distances between ZCTAs in a bioshed are estimated from Microsoft© MapPoint® 2006 (http://www.microsoft.com/MapPoint/en-us/default.aspx Accessed January 5, 2009).

Trucking costs of the BioSAT model are estimated using equations (1), (2), (3), and (4) between each supply $ZCTA_{(j)}$ and demand $ZCTA_{(i)}$ within a bioshed Q_i. Trucking costs are sorted by least cost between each supply $ZCTA_{(j)}$ and demand $ZCTA_{(i)}$. Trucking variable costs are a function of travel time between ZCTAs and trucking fixed costs are a function of travel distance between ZCTAs. The least cost set of supply ZCTAs to meet a demand quantity are generally dependent on shortest travel time between a supply $ZCTA_{(j)}$ and demand $ZCTA_{(i)}$.

Harvesting Costs

Logging Residues:

The BioSAT model uses the Subregional Timber Supply (SRTS) model to estimate and project logging residues in the southeastern U.S. SRTS uses U.S. Forest Service FIA data to project timber supply trends based on current conditions and the economic responses in timber markets (Abt et al. 2000). Abt et al. (2000) note SRTS is a partial equilibrium market simulation model that can be used to analyze various forest resource and timber supply situations. It uses a

biological inventory projection model and a conventional supply/demand framework to project future timber prices and inventories given exogenous assumptions about land area and demand.
SRTS was developed initially to provide an economic overlay to traditional timber inventory models, e.g., ATLAS (Mill and Kincaid 1992), and to develop a consistent methodology for disaggregating the impacts of national and global models, e.g., TAMM (Adams and Haynes 1996), that treated the South as a homogenous supply region (Abt et al. 2000). Timber market and inventory modules are the two major SRTS model components**.** Market parameters are first used to solve for equilibrium price changes, where the market is defined by all of the included subregions. Price and supply shift information from the individual regions are used to calculate harvest change by subregion.

The internal inventory module in SRTS is based on the GRITS model (Cubbage et al. 1990). GRITS extrapolated forest inventories based on USDA Forest Service FIA estimates of timberland area, timber inventory, timber growth rates, and timber removals. GRITS classifies data into 10-year age class groups by broad species group (softwoods and hardwoods) and forest management type (planted pine, natural pine, oak-pine, upland hardwood, and lowland hardwood). FIA data by species group, forest management type, and 10-year age class are summarized for each relevant region in the analysis. Land area trends by forest management type are exogenous to the model. Within a management type, the model can allocate harvest across age classes based on starting harvest proportions, current inventory proportions, or oldest age class first (Abt et al. 2000).

Logging Residue Costs:

Even though logging residue estimates are not presented in this paper, the Fuel Reduction Cost Simulator (FRCS) as modified for the Billion Ton Study (Perlack et al. 2005) by Dykrsta (2008) will be used in future estimates from BioSAT to estimate the cost of harvesting logging residues (Fight et al. 2006; Stokes 1992). The original FRCS model was designed to simulate fuel-reduction treatments in the Interior West, where wildfire is a significant problem (Dykstra 2008). The FRCS was substantially revised by Dykstra (2008) including the development of new procedures to simulate harvests in the North (North Central and Northeast), the South, and the coastal West as well as the Interior West.

In the modified FRCS model the following harvesting operations are considered to collect biomass (Dykstra 2008):

- Manual felling and whole-tree extraction, either with conventional skidders or with cable systems; the simulator uses cable systems if the average ground slope is 40% or more;
- Mechanized felling and whole-tree skidding where mechanized felling is not used with cable yarding.

For ground-based logging, the FRCS model calculates the production rates and costs for both of the possible alternatives (manual felling and mechanized felling). The model then selects the lower-cost alternative for use in deriving the supply curve for the Billion Ton Study which is the same approach that will be used in the BioSAT model.

The variable cost inputs for the FRCS model (e.g., diesel fuel, labor wages, etc.) are updated bi-monthly in the BioSAT model. Forest resource input data is obtained from the USDA Forest Service current FIA inventory data and logging residue estimates are obtained from the SRTS model (Abt et al. 2000).

RESULTS

Mill Residue Economic Supply

The physical supply of mill residues from U.S. Forest Service data indicates distinct regions in the 13-state study region that have high densities of mill residue for potential biomass using facilities (Figure 1). BioSAT estimates the economic availability of such residues and is discussed in this paper.

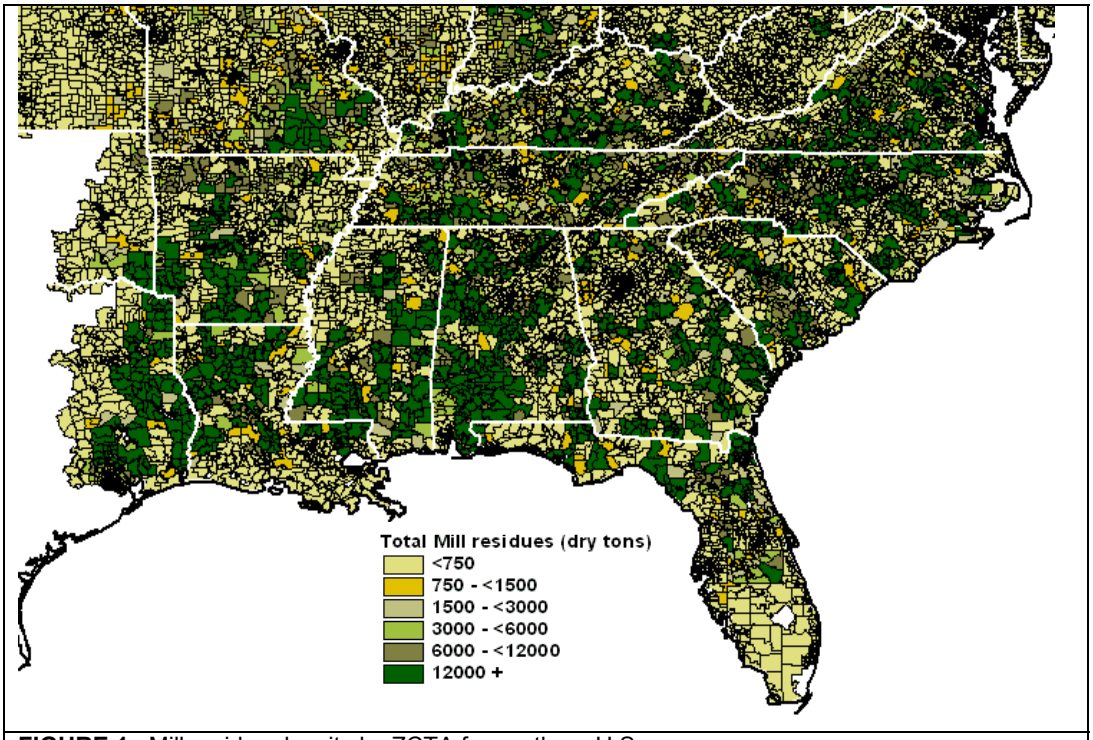


FIGURE 1. Mill residue density by ZCTA for southern U.S.

The trucking cost model component of the BioSAT model is used to generate total cost, average total cost, and marginal costs for up to 1.5 million dry tons of annual consumption of mill residues for 9,353 ZCTA demand points within the

13-state study region. Marginal cost curves for trucking costs only are also generated from the model by ZCTA (Figures 2 and 3).

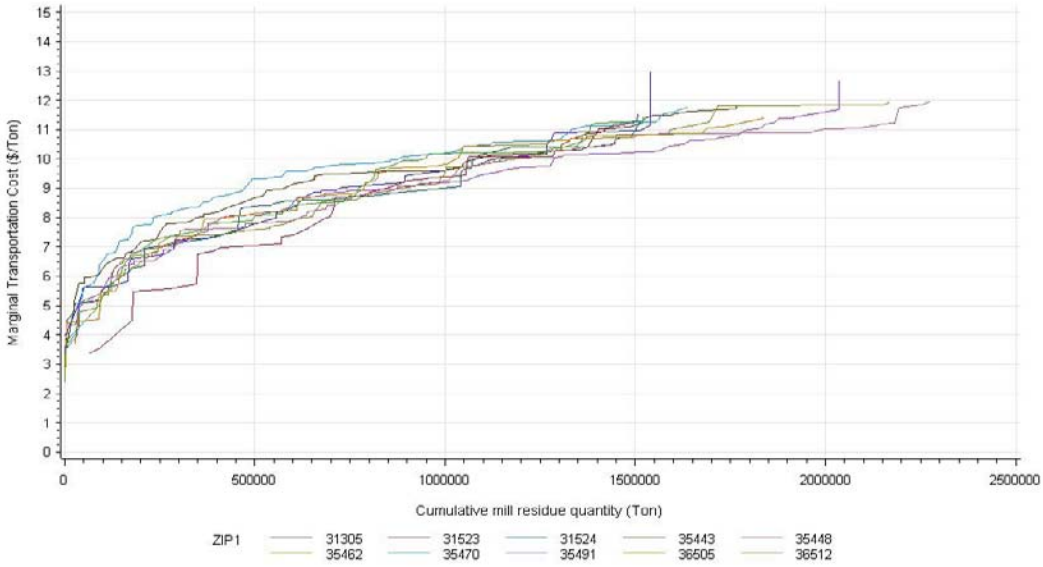


FIGURE 2. Marginal cost curves for trucking of mill residues by ZCTA in Alabama and Georgia for up to 1.5 million dry tons of consumption per year.

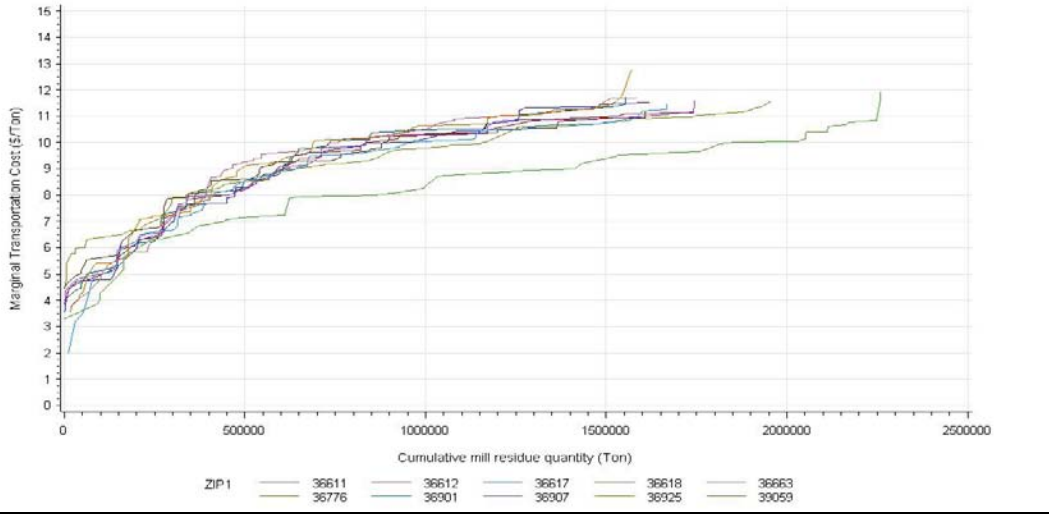


FIGURE 3. Marginal cost curves for trucking of mill residues by ZCTA in Mississippi and Louisiana for up to 1.5 million dry tons of consumption per year.

A comparison of the marginal curves in Figures 2 and 3 for trucking costs indicate significant differences in marginal costs for some demand ZCTAs due to trucking expenses. A key assumption of the BioSAT model is that transportation costs are a key cost factor in the location of biomass consuming facilities.

This is illustrated in Figure 4 in the marginal cost curves for ZCTA 31305 (Darien, GA) that includes trucking costs and TMS 2009 mill residue prices for that region. For an annual consumption of up to 1.5 million dry tons of mill

residues the cost of procuring mill residues declined from $29.65 per dry ton in September 2008 (diesel price of $4.068 per gallon) to $28.67 per dry ton in January 2009 (diesel price of $2.228 per gallon). This differential of $0.98 per dry ton may be a significant cost for a biomass facility in a highly competitive market. The cost per ton for ZCTA 31305 (Darien, GA) for consumption between 500,000 and 1.5 million dry tons of consumption per year is given in Table 3. Note that ZCTA 31305 (Darien, GA) is ranked in the 13-state study region as a low cost demand point for mill residues.

There is more than 1.5 million dry tons of available mill residues in the bioshed for ZCTA 31305 (Darien, GA) but for illustration purposes only 1.5 million dry tons are presented in Figure 4. The distinct shifts in the marginal cost curve in Figure 4 occur from larger travel times between the demand $ZCTA_{(i)}$ and supply $ZCTA_{(j)}$ after mill residues are procured from the preceding lower cost supply $ZCTA_{(j)}$. It is assumed that mill residues are economically available if the demand $ZCTA_{(i)}$ buyer is willing to pay an additional price from the supplier. Given the 2008-2009 economic recession and the many mill curtailments and shutdowns of residue suppliers, the next version of the BioSAT model will allow the user to select a percentage of mill residues physically available before starting a cost estimation and search by ZCTA for a bioshed (e.g., 100%, 80%, 75%, 50%, etc.).

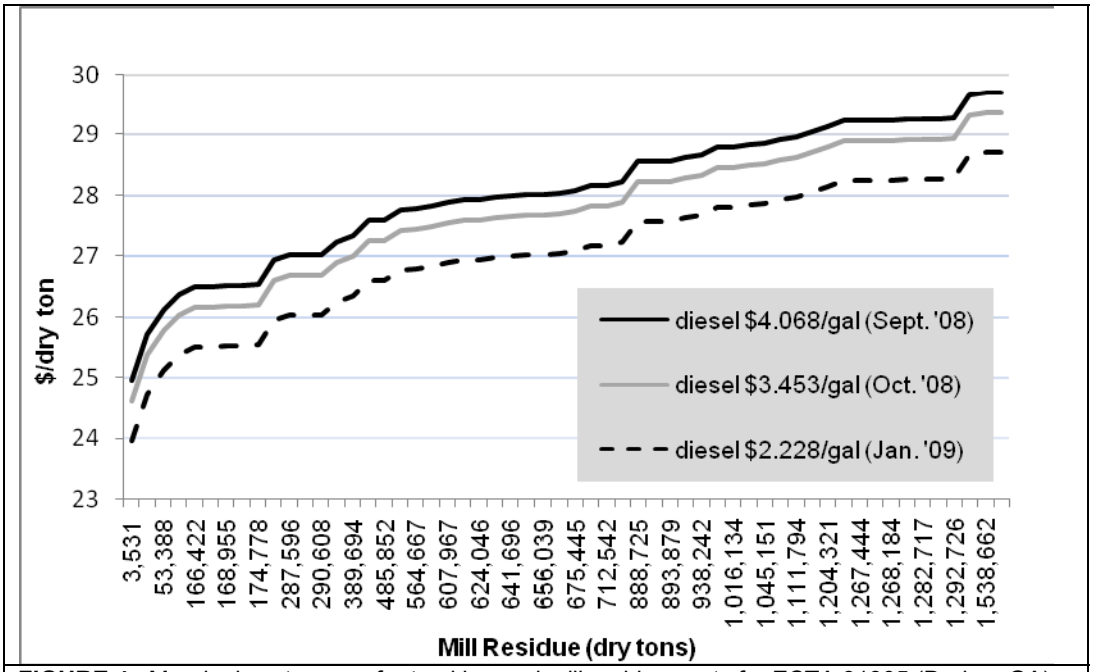


FIGURE 4. Marginal cost curves for trucking and mill residue costs for ZCTA 31305 (Darien, GA) for consumption up to 1.5 million tons per year.

TABLE 3. Cost of mill residues in dry tons for ZCTA 31305 (Darien, GA).

Dry Tons	diesel $4.068/gal (Sept. '08)	diesel $3.453/gal (Oct. '08)	diesel $2.228/gal (Jan. '09)
557141	$27.76	$27.43	$26.77
564667	$27.78	$27.45	$26.79
576221	$27.81	$27.48	$26.82
607967	$27.89	$27.55	$26.90
620670	$27.92	$27.59	$26.93
624046	$27.93	$27.60	$26.94
640828	$27.97	$27.64	$26.99
641696	$27.98	$27.65	$26.99
653744	$28.01	$27.68	$27.02
656039	$28.02	$27.69	$27.03
658074	$28.02	$27.69	$27.03
675445	$28.07	$27.74	$27.08
712447	$28.16	$27.83	$27.17
712542	$28.16	$27.83	$27.18
730901	$28.21	$27.88	$27.22
888725	$28.55	$28.22	$27.56
893656	$28.56	$28.23	$27.57
893879	$28.56	$28.23	$27.57
926886	$28.63	$28.30	$27.64
938242	$28.65	$28.32	$27.66
1009023	$28.78	$28.45	$27.80
1016134	$28.80	$28.47	$27.81
1029843	$28.82	$28.49	$27.83
1045151	$28.85	$28.52	$27.86
1080692	$28.91	$28.58	$27.92
1111794	$28.97	$28.64	$27.98
1151927	$29.04	$28.71	$28.05
1204321	$29.12	$28.79	$28.14
1266766	$29.22	$28.89	$28.23
1267444	$29.22	$28.89	$28.24
1268084	$29.22	$28.89	$28.24
1268184	$29.22	$28.89	$28.24
1278278	$29.24	$28.91	$28.26
1282717	$29.25	$28.92	$28.26
1286112	$29.26	$28.93	$28.27
1292726	$29.27	$28.94	$28.28

1512672	$29.65	$29.32	$28.67
1538662	$29.70	$29.37	$28.71
1539276	$29.70	$29.37	$28.71

Optimal Sites for Biomass Facilities that use Mill Residues

Twenty low cost demand ZCTAs of mill residues for the 9,353 ZCTAs in the 13-state study region are presented in Figure 5. Southern Mississippi is a low cost region for a facility that consumes mill residues up to 1.5 million dry tons per year (ZCTAs 39653, 39436, and 39059). There are also low cost ZCTAs for mill residue consumption of 1.5 million dry tons per year in southeast Georgia (ZCTA 31305), southeast Oklahoma (ZCTA 74737), and southwest Alabama (ZCTA 35448). East Texas has two low cost ZCTA locations (ZCTAs 75534 and 77657) for mill residue consumption up to 1.5 million dry tons per year.

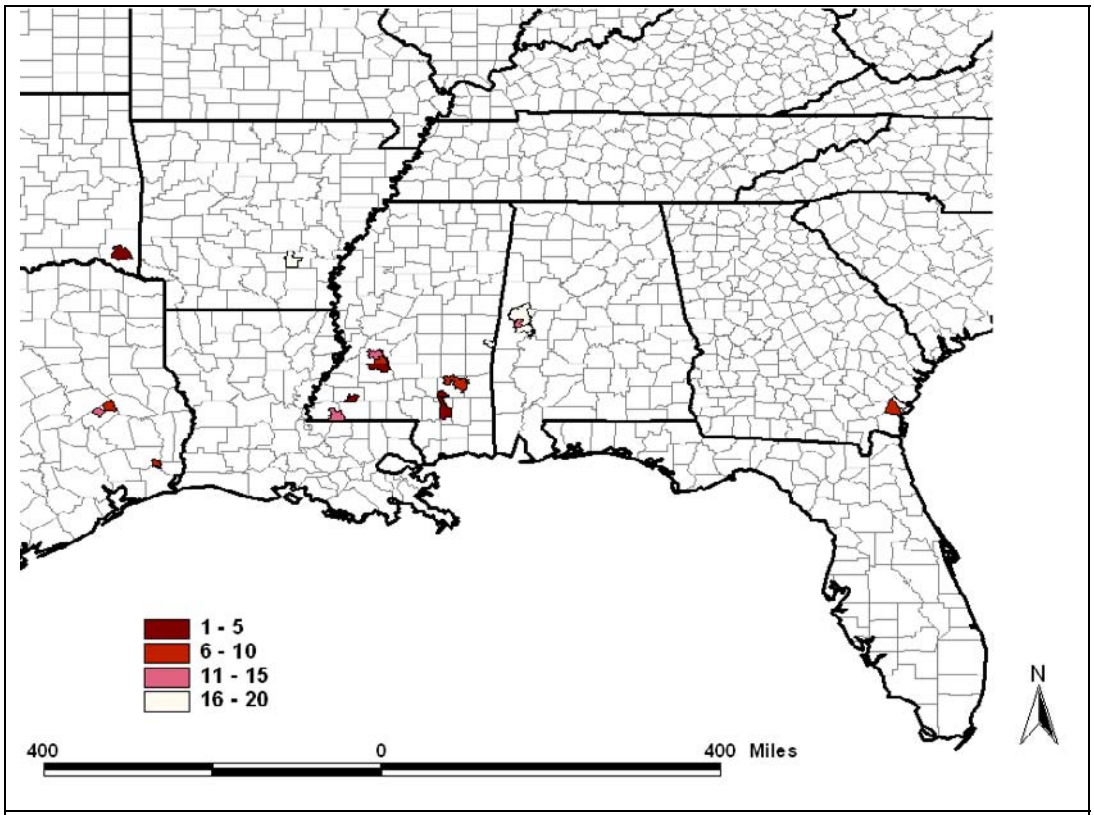


FIGURE 5. Top 20 ZCTA locations with county boundaries for mill residue consumption up to 1.5 million tons consumption per year in the 13-state study region.

A strength of the BioSAT model is that it estimates trucking costs as a function of the MapPoint® road network. This is illustrated in Figure 6 for ZCTA 31305 (Darien, GA) for annual consumption of mill residues of up to 1.5 million dry tons. Some ZCTAs are located in very close proximity (gray color) to the east of the demand ZCTA (highlighted in blue) but are excluded from the model given

lack of mill residues and/or long travel-time road networks that incur high trucking costs.

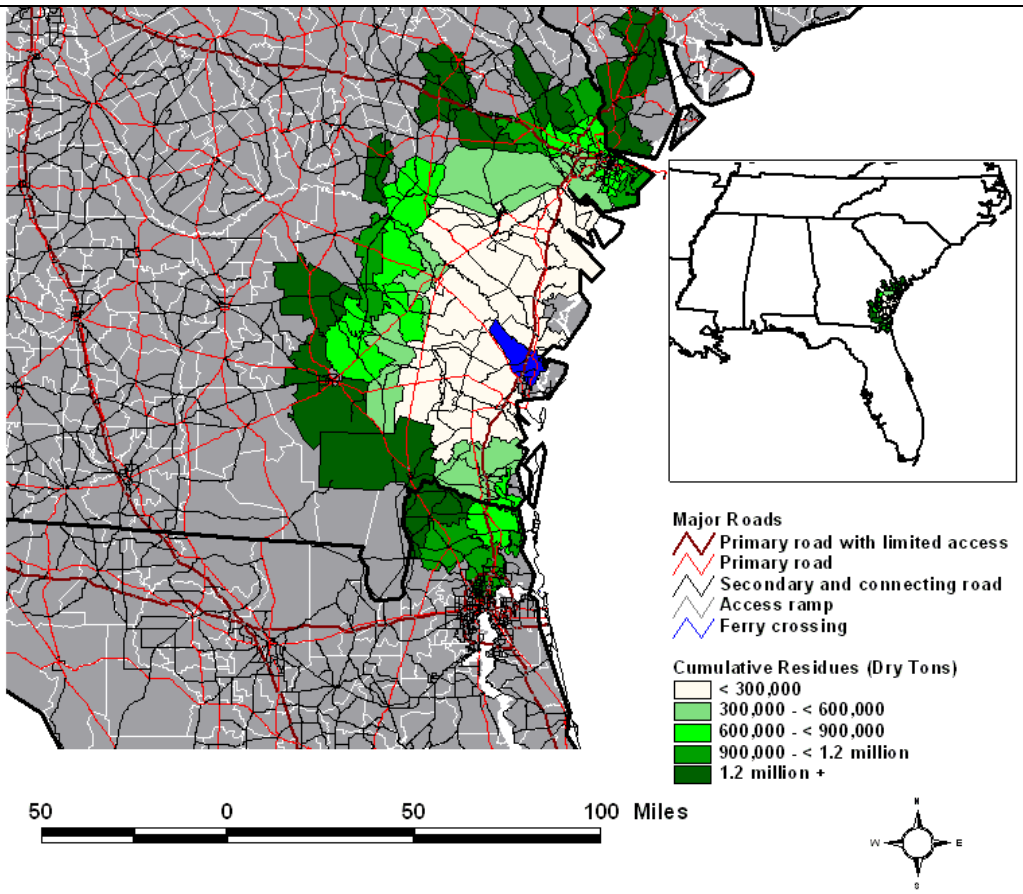


FIGURE 6. Bioshed for ZCTA 31305 (Darien, GA) with road network for consumption of mill residues up to 1.5 million tons per year.

CONCLUSIONS

A study using the Biomass Site Assessment Tool (BioSAT) model for procurement of mill residues for 13 southern U.S. states is presented in this paper. The BioSAT model for mill residues assumes truck transportation with dry-van storage for a maximum one-way haul distance of five hours. Mill residue prices are obtained from Timber Mart South.

BioSAT has a trucking cost model that estimates costs as a function of the road network provided by MapPoint®. Road networks in MapPoint® are a combination of the Geographic Data Technology, Inc. (GDT) and Navteq data. County level estimates of all-live total biomass, as well as average annual growth, removals, and mortality are obtained from the Forest Inventory and Analysis Database

(FIADB) version 3.0. The latest complete cycle of data for each state are used. Data in BioSAT are organized by 24,975 zip code tabulation areas (ZCTA) in 33 eastern U.S. states. ZCTAs are based on the U.S. Census Bureau 2000 census definition.
Confidence bounds of the resource supply in any given grouping of ZCTAs ("bioshed") does not offer improvement over existing studies which aggregate county-level resource supply data. However, using the ZCTA as a demand point may offer improvement in cost estimates when compared to studies which use the county centroid as a demand point. Counties can be large and have geographic/economic barriers that impact road networks (e.g., bridges over large waterways, mountains, large metropolitan areas, etc.). Such geographic/economic barriers can increase the travel time and costs for transportation.

Twenty low-cost ZCTA demand-points for annual consumption of mill residues up to 1.5 million dry tons are located in southern Mississippi, southeast Georgia, southeast Oklahoma, southwest Alabama, and east Texas. Costs for these ZCTAs range from $25 per dry ton to $38 per dry ton for up to 1.5 million dry tons of annual consumption.

Research on BioSAT is on-going and studies are forthcoming for 33 eastern U.S. states. These studies will include different types of woody and agricultural biomass (e.g., logging residues, pulpwood, corn stover, etc.). Additional cost models for transportation by truck with rail and barge intra-modal transfer will also be forthcoming.

REFERENCES

Abt, Robert C.; Cubbage, Fred W. and Pacheco, G. 2000. Southern forest resource assessment using the subregional timber supply (SRTS) model. Forest Products Journal. 50(4):25-33.

Adams, D. and Haynes, R.W. 1996. The 1993 Timber Assessment Market Model: Structure, projections, and policy simulations. PNW-GTR-368. USDA Forest Serv., Pacific Northwest Forest and Range Expt. Sta., Portland, OR. 58 p.

Altman, I. and Johnson, T. 2008. The choice of organizational form as a non-technical barrier to agro-bioenergy industry development. Biomass and Bioenergy. 32:28–34.

Biomass Research and Development Board. 2008. Increasing feedstock production for biofuels economic drivers, environmental implications, and the role of research. http://www.brdisolutions.com/default.aspx. (Accessed January 5, 2009).

Cubbage, Fred W.; Hogg, D.W.; Harris, T.G. and Alig, R.J. 1990. Inventory projection with the Georgia Regional Timber Supply (GRITS) Model. Southern J. of Appl. Forestry. 14(3):137-142.

Berwick, M. and Farooq, M. 2003. Trucking cost model for transportation managers.

Upper Great Plains Transportation Institute, North Dakota State University. Fargo, ND. 61p.

DiPardo, J. 2000. Outlook for biomass ethanol production and demand. U.S. Energy Information Administration. www.eia.doe.gov/oiaf/analysispaper/pdf/biomass.pdf. (Accessed January 5, 2009).

Dykstra, D.P. 2008. Subject: estimating biomass collection costs for the "Billion-Ton Study" update. Memo: Estimating Forest Biomass Collection Costs for the Billion-Ton Study Update (BTS2). Dykstra, April 25, 2008.

Graham, Robin L.; English Burt C. and Noon, Charles E. 2000. A Geographic Information System-based modeling system for evaluating the cost of delivered energy crop feedstock. Biomass and Bioenergy. 18(4):309-329.

Fight, R.D.; Hartsough, B.R. and Noordijk, P. 2006. Users guide for FRCS: Fuel Reduction Cost Simulator software. Gen. Tech. Rep. PNW-GTR- 668. Portland, OR. Pacific Northwest Research Station, Forest Service, US Department of Agriculture. 23 p. http://www.fs.fed.us/pnw/pubs/pnw_gtr668.pdf . (Accessed January 5, 2009).

Energy Information Administration. 2008. International Energy Outlook. Report #:DOE/EIA-0484(2008), Release Date: June 2008 http://www.eia.doe.gov/oiaf/ieo/world.html**.** (Accessed January 5, 2009).

Lunnan, A. 1997. Agriculture-based biomass energy supply-a survey of economic issues. Energy Policy. 25(6):573-582.

Mill, J.R, and Kincaid, J.C. 1992. The Aggregate Timberland Assessment System-ATLAS: A comprehensive timber projection model. Gen. Tech. Rept. PNWGTR-281. USDA Forest Serv.. Pacific Northwest Res. Sta., Portland, OR. 168 p. http://www.treesearch.fs.fed.us/pubs/9007. (Accessed January 5, 2009).

Perlack, Robert D.; Wright, Lynn L.; Turhollow, Anthony F.; Graham, Robin L.; Stokes, Bryce J. and Erbach, D.C. 2005. Biomass as feedstock for a bioenergy and bioproducts Industry: the technical feasibility of a billion-ton annual supply. US Department of Energy, Oak Ridge National Laboratory, Oak Ridge, TN. DOE Publication DOE/GO-102005-2135, ORNL/TM-2005/66. 59 p. Available online at http://feedstockreview.ornl.gov/pdf/billion_ton_vision.pdf. (Accessed January 5, 2009).

Sedjo, R.A. 1997. The economics of forest-based biomass supply. Energy Policy. 25(6):559-566.

Stokes, Bryce J. 1992. Harvesting small trees and forest residues. Biomass and Bioenergy. 2(1):131-147.

Ugarte, Daniel L.T.; Daniel, G. and Ray, D.E. 2000. Biomass and bioenergy applications of the POLYSYS modeling framework. Biomass and Bioenergy.

4(3):1-18.

Ugarte, Daniel L.T.; English, Burt C.; Menard, R.J. and Walsh, Marie. 2006. Conditions that influence the economic viability of ethanol from corn stover in the midwest the USA. International Sugar Journal. 108(1287):152-156.

Ugarte, Daniel L.T.; English, Burt C.; and Jensen, Kim. 2007. Sixty billion gallons by 2030: Economic and agricultural impacts of ethanol and biodiesel expansion. American Journal of Agricultural Economics. 89(5):1290-1295.

Walsh, Marie E. 1998. U.S. bioenergy crop economic analyses: status and needs. Biomass and Bioenergy. 14(4):341-350.

Walsh, Marie E. 2000. Method to estimate bioenergy crop feedstock supply curves. Biomass and Bioenergy. 18:283-289.

U.S. Census Bureau. 2000. ZIP Code Tabulation Area (ZCTA) for Census 2000. http://www.census.gov/geo/ZCTA/zcta.html. (Accessed January 5, 2009).

U.S. Department of Agriculture, Forest Service. 2008. U.S. Forest Service research and development strategic plan, 2008-2012. Washington, DC. 32p.

U.S. Department of Agriculture, Forest Service. 2008. Woody biomass utilization strategy. U.S. Department of Agriculture, FS-899. GPO: Washington, D.C. 33p.

Western Governor's Association. 2008. Strategic assessment of bioenergy development in the west – spatial analysis and supply curve development. University of California, Davis. 86p. http://www.westgov.org/wga/initiatives/transfuels/Task%203.pdf. (Accessed January 5, 2009).

Young, Timothy M. and Ostermeier, David M. 1989. IFCHIPSS – The Industrial Fuel Chip Supply Simulation Model. Final Report for Contract with Southeastern Regional Biomass Energy Program as administered by the Tennessee Valley Authority. 141p.

Young, Timothy M.; Ostermeier, David M.; Thomas, J.Daniel and Brooks, Robert T. 1991. The economic availability of woody biomass for the Southeastern United States. Bioresources Technology. 37(1):7-16.

APPENDIX

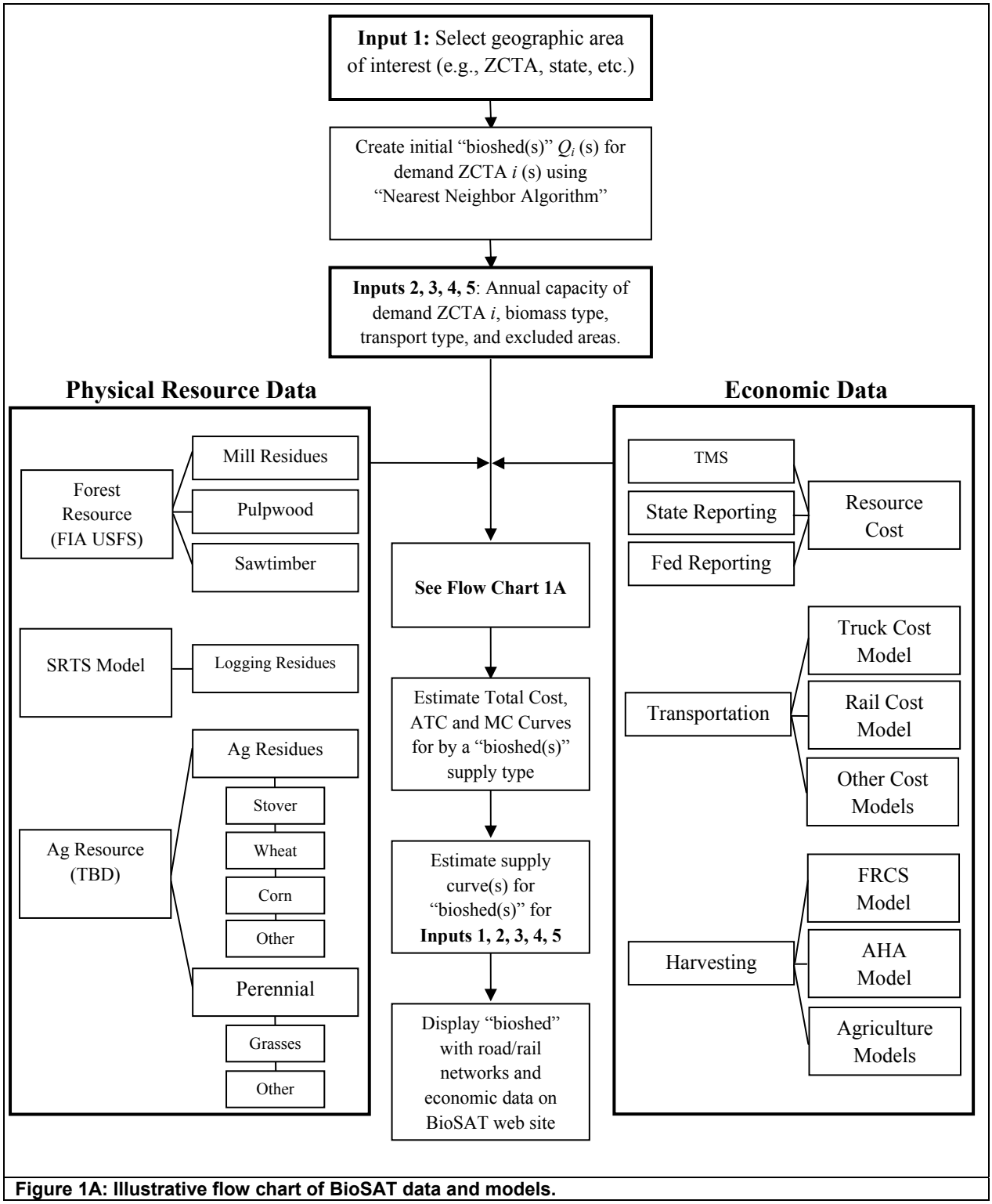


Figure 1A: Illustrative flow chart of BioSAT data and models.

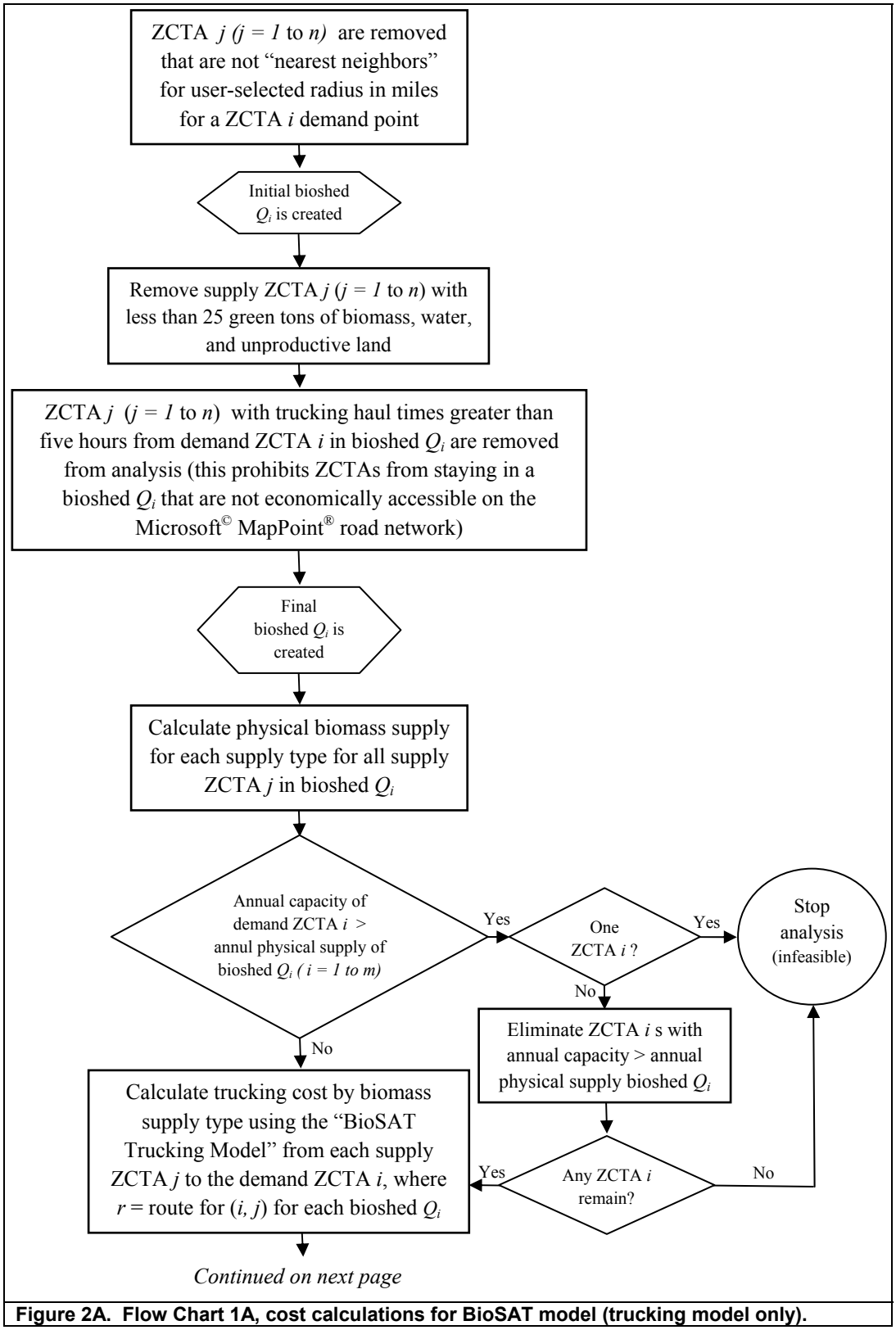


Figure 2A. Flow Chart 1A, cost calculations for BioSAT model (trucking model only).

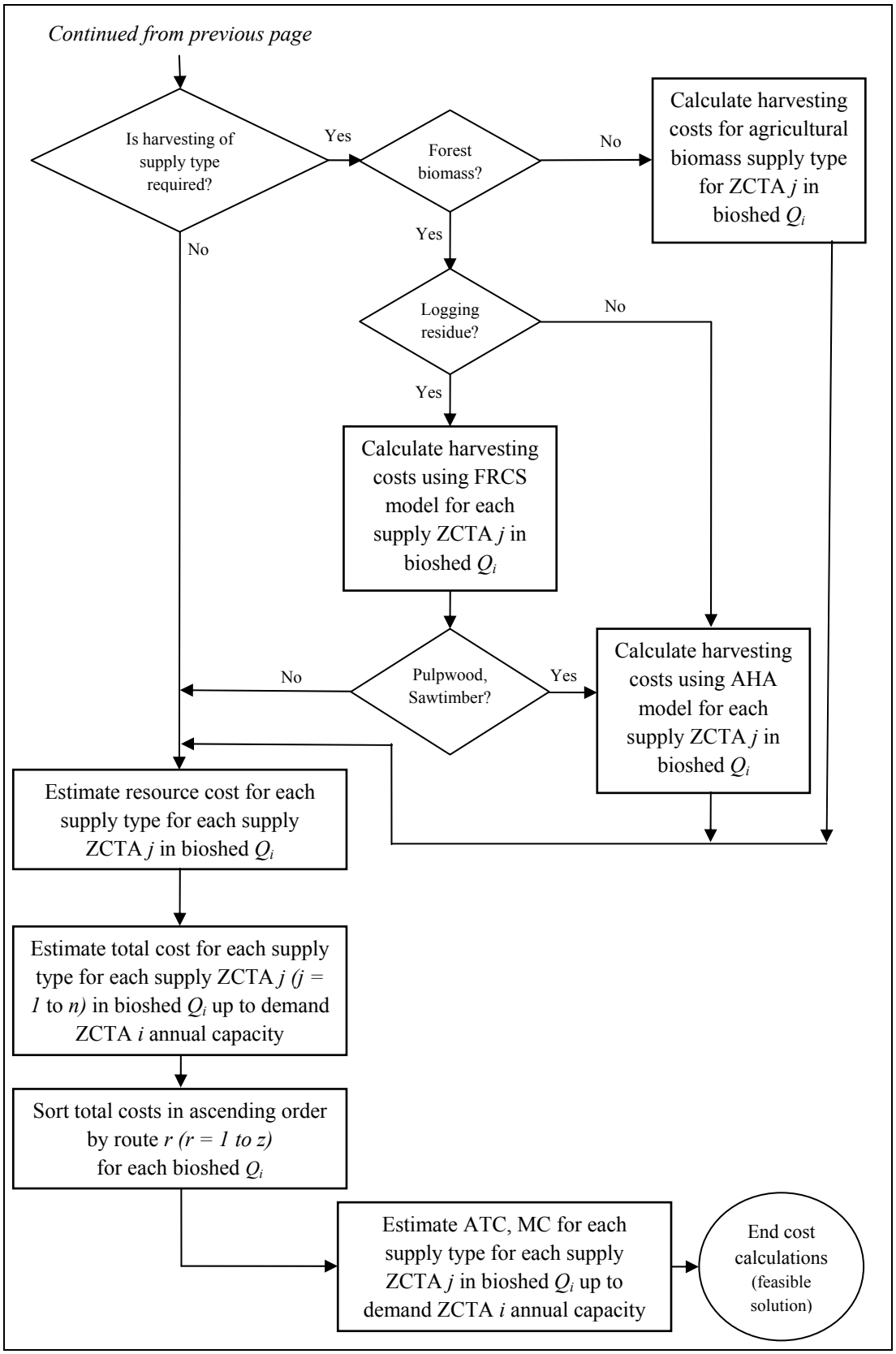

Continued from previous page
Is harvesting of supply type required?
Yes
Forest biomass?
No
Calculate harvesting costs for agricultural biomass supply type for ZCTA j in bioshed Qi
Yes
No
Logging residue?
No
Yes
Calculate harvesting costs using FRCS model for each supply ZCTA j in bioshed Qi
Pulpwood, Sawtimber?
No
Yes
Calculate harvesting costs using AHA model for each supply ZCTA j in bioshed Qi
Estimate resource cost for each supply type for each supply ZCTA j in bioshed Qi
Estimate total cost for each supply type for each supply ZCTA j (j = 1 to n) in bioshed Qi up to demand ZCTA i annual capacity
Sort total costs in ascending order by route r (r = 1 to z) for each bioshed Qi
Estimate ATC, MC for each supply type for each supply ZCTA j in bioshed Qi up to demand ZCTA i annual capacity
End cost calculations (feasible solution)

Figure 2A (cont). Flow Chart 1A, cost calculations for BioSAT model (trucking model only).

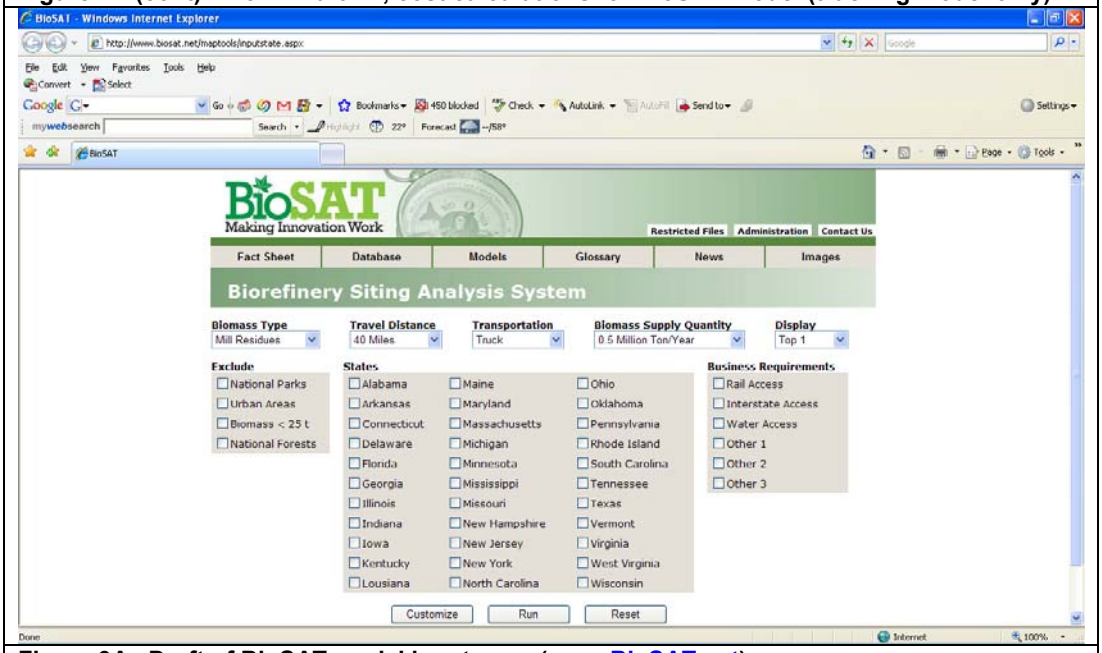


Figure 3A. Draft of BioSAT model input page (www.BioSAT.net)

ACKNOWLEDGEMENTS

The authors would like to acknowledge the USDA Forest Service Southern Research Station, the Southeastern Sun Grant Center and The University of Tennessee Agricultural Experiment station for financial support of this study. A special acknowledgement is given to Ms. Sachiko Hurst, programmer, Office of Bioenergy Programs at The University of Tennessee. The authors also acknowledge Dr. Samuel Jackson, Office of Bioenergy Programs at The University of Tennessee, and Ms. Kerri Norris, Research Technician, of The University of Tennessee Forest Products Center, for their work in support of the BioSAT project. The authors also thank Ms. Yingjin Wang (former Graduate Research Assistant, graduated 09/08) and Ms. Xu (Nancy) Liu (Graduate Research Assistant) for their work on the BioSAT model.

Mixed Multi-Scalar Methods to Assess Wood Biomass Availability on Family Forests in Virginia's Southside

M.D. Brinckman and J.F Munsell

ABSTRACT: *Interest in wood-based bio-energy production systems is increasing. Multi-scalar, mixed-method approaches focusing on both biophysical and social aspects of procurable feedstock are needed. Family forests will likely play an important role in supplying forest-based biomass. However, access depends in large part on the management trends among family forest owners. This paper outlines proposed research to estimate feedstock availability in Virginia's Southside using United States Forest Service Forest Inventory Analysis data, cluster modeling, silviculture guidelines, property-level samples, and landowner management intentions. Aims are to contrast property-level volumes as mediated by owner intentions with values derived using a landscape level analysis.*

KEYWORDS: Bio-energy, Forest Inventory Analysis, Theory of Planned Behavior, wood biomass, ethanol, family forests, disproportionality, silviculture.

Introduction

Rising energy costs, climate change, geopolitical instability, and widespread uncertainty about the future of fossil fuel are profoundly influencing energy strategies in the United States (US) (Duffield and Collins 2006). Expanding production of domestic renewable energy is gaining traction as a potential approach (Dincer 2000). As a result, interest in using wood to generate large-scale bio-energy has resurfaced. Forest-based biomass from private, family forests constitutes an increasingly popular component of wood-based energy solutions in the US (Munsell and Germain 2007).

Forest-based bio-energy is advantageous for several reasons. Using food crops, such as corn, to produce bio-energy is highly controversial (Pimentel 2003). Wood has a higher energy return than other popularized feedstock and conversion technologies are improving (Amidon 2006; Keolian and Volk 2005; Shapouri et al. 2002). Biomass markets are needed to facilitate management on much of the nation's family forests (Munsell et al. 2008). All told, procurement of wood biomass from family forests seems quite appealing. Yet accessing large feedstock volume on family forests may not be entirely straightforward.

Family forest owners are diverse (Butler and Leatherberry 2004). Objectives vary and are increasingly amenity- rather than commodity-based (Kendra and

Matthew D. Brinckman (mdbrinck@vt.edu) is M.S. candidate, Virginia Polytechnic Institute and State University, Forestry, 304 Cheatham Hall (0324), Blacksburg, VA 24061. John F. Munsell (jfmunsel@vt.edu) is assistant professor and extension specialist, Virginia Polytechnic Institute and State University, Forestry, 304C Cheatham Hall (0324), Blacksburg, VA 24061.

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Hull 2005). Moreover, exurbanization is shaping a new period defined by smaller parcels and more owners (Sampson and DeCoster 1997; Butler and Leatherberry 2004). If forest-based bio-energy production systems are to succeed in regions dominated by family ownership, multi-scalar, mixed-method approaches focusing on both biophysical and social aspects of procurable feedstock are needed. The basic premise is that, regardless of scale, available volume does not supersede owner intentions to supply it and vice versa.

To estimate feedstock availability in Virginia's Southside, we plan to use multi-scalar, mixed-method evaluations within the context of an impending bio-energy plant in Mecklenburg County. Our research will estimate supply within a 30-mile road distance woodshed around a proposed ethanol plant. We will use US Forest Service Forest Inventory Analysis (FIA) data, cluster modeling, and silvicultural guidelines to estimate above-ground sustainable yield. We will also survey landowners about their intentions to commercially harvest and inventory their forests to assess above-ground wood biomass potential. We will then contrast this potential with owner intentions to estimate a margin of error for potential accessibility. Our ultimate goal is to contrast the insight gained from landowner surveys and property assessments with FIA estimates of wood biomass availability.

Rationale

Erroneous estimates of wood biomass availability can be detrimental. Financial investors, potential processors, energy consumers, forest owners, and resource managers stand to lose. For example, in 2007 the Nevada Department of Corrections invested 8.3 million dollars in forest-based bio-energy. Adjacent federal forests were assessed and deemed to house sufficient volume. One prison and two local communities expected to receive resulting energy. In the end, however, the actual supply fell dramatically short.

> *Officials from the Nevada Department of Corrections had predicted that the plant would replace energy from electricity and natural gas at Northern Nevada Correctional Center and the neighboring Stewart Conservation Camp. The plant, however, has not run more than three days straight since it opened six months ago. "Wood continues to be an issue for us," said Lori Bagwell, department director of support services. "We do not have an adequate and appropriate supply." -Nevada Appeal, March 2, 2008.*

Much was blamed on public input policies for federal forest management. Adequate stock did not beget adequate supply due to social processes unaccounted for when deriving initial estimates. And though the Nevada example pertains to public forests, its general lesson resonates in terms of family forests. Volume means little if owner inclinations preclude procurement. This is particularly relevant in the eastern US, where most forests are family owned and management objectives and behavior vary widely (Butler and Leatherberry 2004). Thus, forest-based bio-energy systems in the eastern US should be careful to

calibrate volume-based feedstock estimates using management trends among family forest owners.

Context

In Virginia, close to 65% of the forests are family owned (Virginia Department of Forestry 2008). A sizeable percentage of these forests are located in its southern region, or "Southside". Southside Virginia is loosely defined as the area of the state that lies east of the Blue Ridge Mountains and south of the James River. Tobacco production has historically been the dominant economic enterprise. However, tobacco quotas and production have steeply declined over the past decade, leaving local economies in question and forcing landowners to look for profitable alternatives.

In 1999, the Virginia General Assembly created the Virginia Tobacco Indemnification and Community Revitalization Commission (VTICRC) to help address the challenges Southside and Southwest communities face when transitioning out of tobacco production. The VTICRC's regional classifications are displayed in Figure 1. Among many diverse awards, VTICRC, along with Virginia Block Grant, recently allocated $650,000 to install a 12-inch water main to the site of a future 100 million gallon per year Bluestone Bio Energy LLC ethanol plant in Southside's Mecklenburg County.

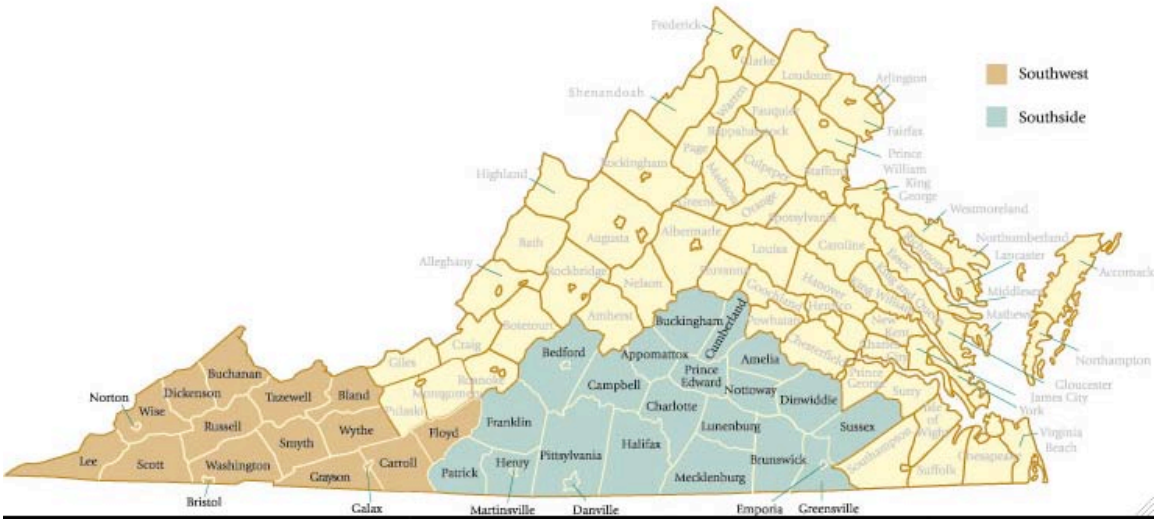


FIGURE 1: VTICRC's classification of Virginia's Southside and Southwest regions. Available at: http://www.tic.virginia.gov/tobmapupdated.shtml

Bluestone Bio Energy LLC was born from Osage Bio Energy, which primarily manages barley to ethanol production. Located in Glen Allen, Virginia, Osage Bio Energy is an auxiliary of Osage Incorporated, which is the largest motor-grade ethanol distributor in the eastern US. Bluestone primarily plans to

procure and process hull-less barley biomass, though substantial interest in wood biomass also exists.

Given Bluestone's attention to wood, forest-based biomass seems like a potentially promising management option in Virginia's Southside. Many former tobacco farmers and other landowners in the area possess ample woodlands and could supply considerable volumes of feedstock. What is more, wood biomass is presently on hand in great quantities, whereas other feedstock, such as the hull-less barley, may take some time before sufficient biomass volumes are obtainable.

Before forest-based bio-energy systems can take root in the Southside, immediate and long-term availability must be understood. To do this, we propose using biophysical and social metrics at different scales to capture both the human and biological constraints on supply. We believe the combination will help construct realistic procurement strategies that are sustainable in light of the region's potential to provide wood biomass. Using this underlying principle within the Southside case described above, our primary research question is to determine the extent to which landscape-level estimates of procurable, above-ground wood biomass correspond to property-level estimates derived using biophysical and social measurements.

Objectives

We will pursue 5 objectives to answer the primary research question. They are: 1) use selected FIA attribute data, cluster modeling, and silvicultural guidelines to estimate immediately available wood biomass across the Bluestone case study woodshed; 2) survey 40 randomly selected family forest owners with 20 acres or more in the case study woodshed about their intention to commercially harvest wood within the next 5 years using sociopsychological scales; 3) measure corresponding FIA attributes within the 40 forests held by case study owners and use aggregated data to estimate immediately available wood biomass across property-level cases; 4) Combine owner intention measurements and property-level wood biomass estimates to explore disproportionate relationships between biophysical and social measurements and adjust aggregated estimates to reflect differences; and 5) compare the adjusted property-level estimate with the original FIA-based, landscape-level estimate to determine potential error range.

Theory

FIA

FIA is managed by the United States Department of Agriculture (USDA) Forest Service. It uses a three-phased sampling protocol to measure and monitor the nation's forests in a standardized manner. In phase one, a uniform pattern of plots is overlaid across the US. Remote sensing data is then used to stratify plots

into groups based on similar condition or cover. On-the-ground measurements of attributes are taken at each plot during phase two. Phase three estimates are then calculated using phase two data and the total area represented by each plot (Forest Inventory and Analysis Program 2008). Using FIA data, it is now possible to make large, landscape estimates from national to county levels.

Silviculture

Silviculture is a system wherein healthy communities of trees and other vegetation are established and maintained for the benefit of people (SAF 2001). It consists of three interdependent component treatments – regeneration, tending, and harvesting. The appropriate and timely use of these treatments should ultimately improve tree and vegetation stability and quality, optimize benefits to a landowner, shorten investment periods, contain costs, and sustain ecological health and productivity (Nyland 2002). Forestry practices based on silviculture can balance a wide variety of commodity and non-commodity values over time. Sawtimber, pulp fiber, small-scale fuelwood, soil stability, ecological habitat, and aesthetics are achievable ends. We will use silviculture's sustainable tenets to guide our landscape- and property-level evaluations.

Theory of Planned Behavior

Several studies characterize management intentions among family forest owners as a way to estimate timber availability (Birch 1994; Karppinen 2005; Kuuluvainen et al. 1996; Young and Reichenbach 1987). We will build upon these sociopsychological investigations by modeling the intentions of case study family forest owners to commercially harvest wood in the next 5 years. Model results will highlight significant social and psychological drivers of harvesting aims on case study forests and, in doing so, help calibrate wood biomass estimates as well as allude to potential political and marketing strategies that may effectively encourage increases in commercial harvests to sustainably supply forest-based bio-energy. We will use Azjen's (2005) Theory of Planned Behavior to structure our model (Figure 2).

The Theory of Planned Behavior states that an individual's attitudes, normative beliefs, and control perceptions about a behavior predict their reported intention to perform it. Intention can then be used as a proxy for predicting actual behavior. In other words, a family forest owner's future harvesting behavior can be predicted by measuring their intention to harvest and their behavioral beliefs and attitude towards doing so (i.e., cost benefit analysis), normative beliefs and how they feel important peers will perceive the behavior, and whether and to what extent they perceive external controls (what barriers they feel may prevent them from performing the behavior). Lastly, the relationship between behavioral intention and actual behavior is influenced by exogenous behavioral controls. Controls, such as a law or the availability of a needed service, can alter an individual's ability to act on their behavioral intention.

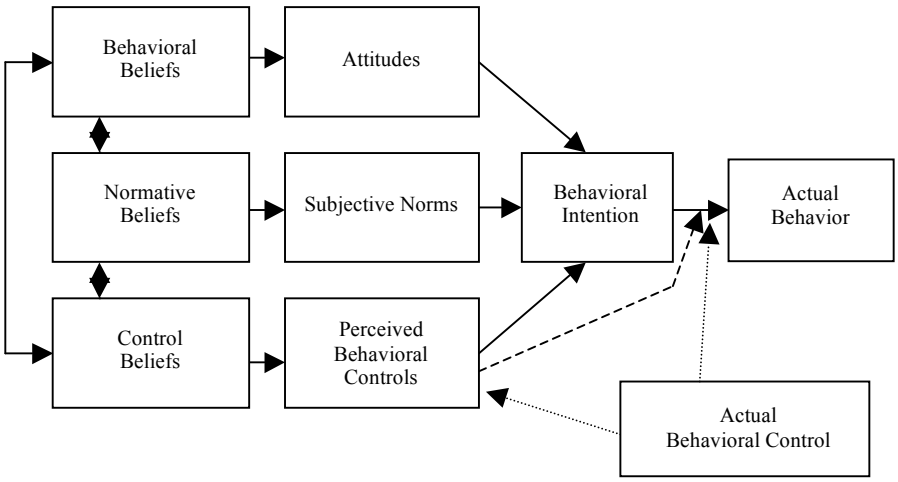


FIGURE 2: The Theory of Planned Behavior (Ajzen 2005).

Disproportionality

Disproportionality in social science is typically associated with between-group variability. For instance, environmental justice researchers often compare racial groups to identify if one group is disproportionately located in areas with environmental problems. Nowak et al. (2006) argues that it is helpful within natural resources sustainability to examine disproportionality within groups because it is possible to more closely examine how group members influence outcomes. That is, a few outliers within a group may help explain a great deal about aggregated group characteristics.

We plan to use within-group variability to explain disproportionality between the social and biophysical aspects of case study wood biomass supply at the property-level. Said differently, we plan to test the extent to which an owner's intention to commercially harvest wood proportionally relates to forest size and procurable wood biomass. Figure 3 displays a hypothesized relationship between mixed, multi-scalar variables. In this hypothesis, a landowner's intention to harvest is proportionally related to parcel size and volume per acre.

The hypothesis stems from trends indicating that owners of larger parcels are more likely to engage in management (Butler and Leatherberry 2004). Yet family forest parcel size continues to decrease and management objectives are changing (Sampson and DeCoster 2000; Kendra and Hull 2005). For example, many case study forests may be ready for harvesting according to silviculture, yet only a small amount of these owners intend to harvest. What is more, those intending to harvest may have disproportionate acreages, which would further affect aggregated wood biomass availability. Such a result would suggest levels of future supply may differ from estimates gleaned using only generalized biophysical capacity or gauging family forest accessibility by solely relying on models of owner harvesting intentions.

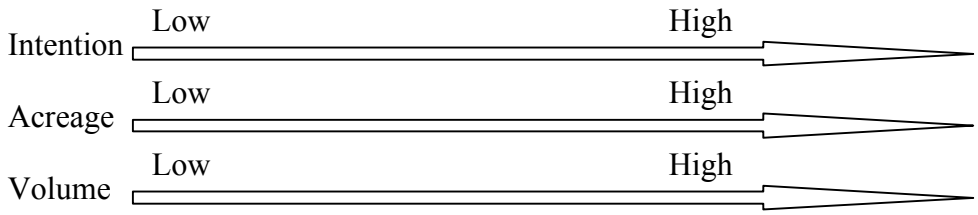


FIGURE 3: Hypothesized relationship between variables effecting wood biomass availability.

Methods

A Geographic Information System (GIS) was used to create the case study's 30-mile road distance woodshed polygon. Bluestone's future location constitutes the polygon's center point. The 30-mile distance was chosen based on Bluestone's procurement interests. A woodshed area polygon was created in ArcMap® using network analyst and ESRI Street Map USA (Figure 4). The polygon totals 1,022,397 acres and includes portions of Mecklenburg, Lunenburg, Charlotte, Halifax, and Prince Edward Counties in Virginia and Warren, Granville, and Vance in North Carolina.

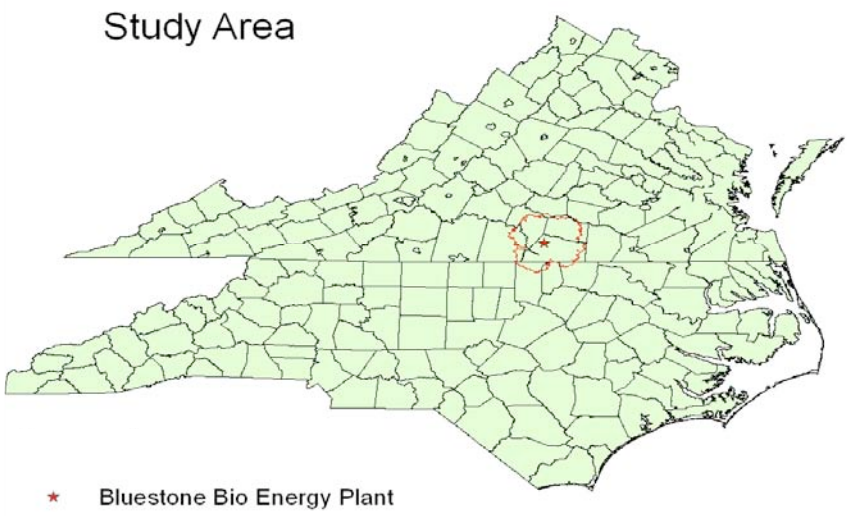


FIGURE 4: Case study polygon, overlaid on a map of Virginia and North Carolina. Plant location is symbolized by a star and is located in Mecklenburg County, Virginia.

To select FIA plots, point locations surveyed between 2000 and 2007 were projected and then chosen using an error threshold of no more than one mile beyond the woodshed polygon boundary. The error was allowed because all FIA plot coordinates have been altered (or *fuzzed*) up to 0.5 miles. For this reason, an uncertainty analysis using 0.5 mile buffers around all plots was conducted to determine upper and lower estimate bounds. For the base estimate, all plot centers

located within the polygon were used. Plots centers located in the polygon with 0.5 mile buffers intersecting the polygon boundary were removed for the lower bound estimate. Plots located outside the polygon, but with buffers intersecting were added and used in the upper bound estimate. The number of plots selected in the uncertainty analysis are annotated in Table 1.

TABLE 1: Number of FIA plots included in the uncertainty analysis.

Selection	# of plots selected
Base Estimate	242
Lower Bound	221
Upper Bound	267

Estimates of wood biomass per acre will be reported for each of the three FIA plot groups listed in Table 1. Each FIA plot is adjusted based on the forested acres it represents. In terms of wood biomass estimations, pre-calculated, easily obtainable biomass values also exist for each FIA tree. At the landscape level, key case study FIA attributes will be used in the cluster analysis to assemble plots into like groups for silvicultural evaluation. Attributes include forest type, site index, growing stock or stocking, and average diameter at breast height (DBH). We will use these data and silviculture stocking guides following Roach and Gingrich (1968) to estimate immediately sustainable wood biomass yield for even-aged conditions. In uneven-aged circumstances, we will estimate wood biomass yield using the Arbogast (1957) curve to distribute sustainable removals across tree diameters.

To identify property-level cases, we will draw a random sample of approximately 1000 case study family forests from the supply woodshed. Real property data from each county and a spatial query using GIS will be used to identify the names and mailing addresses of owners with qualified forested properties. A letter will be sent to each sampled owner to inquire about willingness to participate.

Sampled owners will be able to reply to the letter using a self-addressed stamped postcard. On the card, respondents will indicate their willingness to participate and confirm acreage, forest type, and property location. We will send a reminder letter two weeks following the first mailing and assess non-response bias after six weeks. Property selection will be purposive and based on available information about regional forest characteristics and sampled properties. We will strive to include cover types and property sizes at rates that reflect Southside trends.

For each forest, we will measure corresponding FIA attributes at the property level following Munsell and Germain's (2007) field survey method. The field survey consists of a randomly placed grid system oriented to cover the forested area. Measurements will be taken in $1/10^{th}$ acre fixed-area circular plots randomly selected within the grid. Sampling will cease if the margin of basal area error per plot is $\leq$20% around the mean basal area at the 95^{th} percentile. If not, plots will be added until reaching this threshold. We will use data and silvicultural guidelines to estimate wood biomass yield from sampled family forests.

We will survey owners in person or over the phone with a Likert-type instrument designed using the Theory of Planned Behavior. The instrument will use summated rating scales to measure and operationalize constructs within a theoretical model that will explain an owner's intention to commercially harvest timber in the next 5 years. Summated rating scales allow for a more accurate differentiation among respondents (Spector 1992). We will assume that intention to harvest commercially acts as a proxy for potential to supply wood biomass. Behavioral, normative, control, and intention variables will be measured using bi-polar, multiple response statements on a scale from 1 to 5.

We will use verbal protocol focus groups to test statement validity. The protocol allows researchers to test if questions or statements invoke desired thought processes among respondents using an interpretative-qualitative method (Sudman et al. 1996). We will also use expert review. Cronbach's Alpha will be used to test summated scale reliability. Reliability (which is also referred to as the internal consistency of the instrument) measures the survey instrument's ability to convey the same meaning to the entire sample population.

We will also measure owner characteristics, such as demographics, land acquisition, length of ownership, management plan use, and previous management activities to name a few. We will use these variables when modeling latent aspects of the Theory of Planned Behavior to account for external factors. To measure internal validity, each survey will end with questions involving hypothetical management tradeoffs. We will present owners with a range of tradeoffs to gauge which objectives more strongly affect harvest intentions. Our aim is to test whether the theoretical measurements remain constant when faced with realistic management decisions.

A K-means cluster analysis will be used to group case study FIA plots into types based on selected attributes. K-means cluster analysis clusters samples into a pre-specified number of groups using Euclidean distances between sample and cluster means and multiple iterations to achieve sufficient group convergence. Stepwise multiple regression will be used to model owner intentions. One component model will be tested to identify statistical significance and assess overall change in R^2. Final outputs will include beta coefficients, t-values, F values, R^2, and model significance. Forest size, sustainable wood biomass estimates, and owner intentions will be cross-tabulated to characterize disproportionality.

Implications

Case study results will test the extent to which attitudes, beliefs, objectives, and general landowner characteristics predict an owner's willingness to commercially harvest. Insights will help explain drivers of disproportionate relationships between social and biophysical aspects of potential wood biomass supply. Benefits include gaining a better sense of potential supply relative to a future ethanol plant and demonstrating a multi-scalar, mixed-method approach for assessing wood biomass availability more generally.

Literature Cited

Ajzen, I. 2005. Attitudes, Personality, and Behavior. 2nd Ed. Open University Press / McGraw-Hill, Milton-Keynes, England. 175 p.

Amidon, T.E. 2006. Biorefining: The biorefinery in New York: Woody biomass into commercial ethanol. Pulp & Paper Canada. 107(6): 47.

Arbogast, C. 1957. Marking guides for northern hardwoods under the selection system. Station Paper LS-56. St. Paul, Minn.: U.S. Deparement of Agriculture, Forest Service, Lake States Forest Experiment Station. 20 p.

Birch, T.W. 1994. Private forest-land owners of the United States. USDA Forest Service Northeast Station, Radnor, PA. 183p.

Butler, B.J. and Leatherberry, E.C. 2004. America's Family Forest Owners. Journal of Forestry. 102(7) 4-14.

Dincer, I. 2000. Renewable energy and sustainable development: a crucial review. Renewable & Sustainable Energy Reviews. 4(2000) 157-175

Duffield, J.A. and Collins, K. 2006. Choices: The Magazine of Food, Farm & Resource Issues. 21(1): 9-14.

Forest Inventory and Analysis Program. 2008. The forest inventory and analysis database: database description and users manual version 3.0 for Phase 2, revision1. U.S. Department of Agriculture, Forest Service, Washington Office. Available: http://www.fia.fs.fed.us/library/database-documentation/FIADB_user_manual_v3-0.pdf.

Karppinen, H. 2005. Forest owners' choice of reforestation method: an application of the theory of planned behavior. Forest Policy and Economics. 7(2005): 393-409.

Kendra, A. and Hull, R.B. 2005. Motivations and Behaviors of New Forest Owners in Virginia. Forest Science. 51(2).

Keoleian, G.A. and Volk, T.A. 2005. Renewable Energy from Willow Biomass Crops: Life Cycle Energy, Environmental and Economic Performance. Critical Reviews in Plant Sciences. 24(5-6) 385.

Kuuluvainen, J., Karppinen, H., and Ovaskainen, V. 1996. Landowner objectives and non-industrial private timber supply. Forest Science. 42(3): 300-309.

Munsell, J.F. and Germain, R.H. 2007. Woody Biomass Energy: An Opportunity for Silviculture on Nonindustrial Private Forestlands in New York. Journal of Forestry. 105(8):398-402.

Munsell, J.F., Germain, R.H., and Munn, I.A. 2008. A Tale of Two Forests: Case Study Comparisons of Sustained Yield Management on Mississippi and New York Nonindustrial Private Forestland. Journal of Forestry. 106(8): 431-439.

Nowak, P.J., Bowen, S., and Cabot, P.E. 2006. Disproportionality as a framework for linking social and biophysical systems. Society and Natural Resources. 19: 153-173.

Nyland, R.D. 2002. Silviculture: Concepts and Applications, 2nd ed, McGraw-Hill, New York, 633 pp.

Pimentel, D. 2003. Ethanol fuels: energy balance, economics, and environmental impacts are negative. Natural Resources Research. 12(2): 127-134.

Roach, B.A. and Gingrich, S.F. 1968. Even-aged silviculture for upland central hardwoods. Agric. Handb. 355. Washington, DC.: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 39 p.

Sampson, N. and DeCoster, L. 2000. Forest Fragmentation: Implications for Sustainable Private Forests. Journal of Forestry. 98(3): 4.

Shapouri, H. Duffield, J.A., and Wang, M. 2002. The energy balance of corn ethanol: An update. Agricultural Economic Rep. 814, USDA, Economic Resources Service Office of the Chief Economist, Office of Energy Policy and New Uses, Washington, DC.

Society of American Foresters (SAF). 2001. Dictionary of Forestry. Helms, J.A. (ed.). SAF, Bethesda, MD. 210 p.

Spector, P.E. 1992. Summated Rating Scale Construction: An Introduction. Quantitative applications in the social sciences series, No. 82. SAGE Publications, Newbury Park, CA. 80 p.

Sudman, S., Bradburn, N.M., and Schwarz, N. 1996. Thinking about Answers: the Application of Cognitive Processes to Survey Methodology. Jossey-Bass Publishers: San Francisco. 304 p.

Virginia Department of Forestry. 2008. 2008 State of The Forest: Annual Report on Virginia's Forests. Available at: http://www.dof.virginia.gov/info/index-forms-docs.shtml.

Young, R.A. and Reichenbach, M.R. 1987. Factors influencing the timber harvest intentions of non-industrial private forest owners. Forest Science. 33(2): 381-393.

A Strategic Assessment of Biofuels Development in the Western States

Kenneth E. Skog[1], Robert Rummer[2], Bryan Jenkins[3], Nathan Parker[3], Peter Tittmann[3], Quinn Hart[3], Richard Nelson[4], Ed Gray[5], Anneliese Schmidt[5], Marcia Patton-Mallory[6], and Gayle Gordon[7]

Abstract: *The Western Governors' Association assessment of biofuels potential in western states estimated the location and capacity of biofuels plants that could potentially be built for selected gasoline prices in 2015 using a mixed integer programming model. The model included information on forest biomass supply curves by county (developed using Forest Service FIA data), agricultural biomass supply curves, transportation networks, and capital and operating costs of selected conversion technologies. Results indicate biofuels could potentially provide between 5 and 10 percent of projected transportation fuel demand in the region with fuel price between $2.40 and $3.00 per gasoline gallon equivalence (gge) excluding local distribution costs and taxes. At a target price of $2.40/gge, forest biomass could supply an estimated 11 million oven dry tons per year, or about 9 percent of total feedstock supplied.*

Keywords: biofuels, agriculture residues, wood residues, thinnings, grease, herbaceous energy crops, biomass supply estimates, network analysis

Introduction

The technical feasibility of producing biofuels in the western United States is assessed in a set of reports by the Western Governors' Association (2008a–d) using spatially explicit biomass resource supply curves, a detailed transportation network model for the region, and costs for converting biomass to refined biofuels. The study addresses the widespread concern over environmental, geopolitical, and economic effects of U.S. dependence on petroleum. The study is responding to state and federal legislative bodies who are setting goals for reducing consumption of fossil fuels in the transportation sector using targets for the infusion of so-called low-carbon biofuels into the transportation fuel market. The use of biomass from municipal waste streams, forest thinnings, and herbaceous agricultural residues or energy crops for biofuels production can significantly reduce the net life cycle emissions of greenhouse gases in comparison with crude oil; the benefits from grain and other crops are less certain. This report and the accompanying models represent a significant step forward in

[1] U.S. Forest Service, Forest Products Laboratory, Madison, WI; [2] U.S. Forest Service, Southern Research Station, Auburn, AL; [3] University of California, Davis; [4] Kansas State University, Manhattan, KS; [5] Antares Group, Landover, MD; [6] U.S. Forest Service, Ft. Collins, CO; [7] Western Governors' Association, Denver, CO

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

understanding the potential for meeting policy goals based on near-term technological and infrastructure parameters. The report presents biofuel supply curves estimating potential future supplies of liquid fuels from biomass in the western United States as a function of market price. The combined GIS network analysis and biorefinery optimization model was developed to

- spatially resolve biomass resource quantities and distributions throughout the Western Governors' Association region for major feedstock types,
- map supporting transportation and biofuel-handling infrastructure to estimate biorefinery gate feedstock costs and biofuel distribution costs, and
- optimize biorefinery types, sizes, and locations for competing conversion technologies based on the objective of maximizing producer profit under a market price constraint.

Overall, the analysis estimates biofuel supply curves for the year 2015, and biofuels production capacity is estimated at the regional and state levels.

This paper focuses on describing the methods used to estimate forest biomass supply curves and describing selected overall results of the analysis, including information on all forest and agricultural supply sources and maps indicating the estimated location of biofuels plants using cellulosic feedstocks that would include forest biomass feedstocks. Complete results on biomass supply sources, conversion technologies, spatial analysis with construction of biofuels supply curves, and evaluation of alternative scenarios may be found in Western Governors' Association reports (2008a–d).

Methods and Results for Forest Biomass Supply

Sustainability

Estimates of forest biomass supply were developed for several sources by first identifying sustainability principles to guide their use. Specific guidelines are noted for each source discussed. In general terms, sustainability means today's management actions will not degrade the ecological functioning of a natural system (Helms 1998). In the context of biomass removal from forests, the question of sustainability requires consideration of a wide range of issues, including nutrient cycling and soil productivity, maintenance of biodiversity, water quality, and wildlife habitat. These factors, and resulting constraints on forest operations to address concerns, are generally very site-specific. Soil productivity in certain soil types, for example, may be more sensitive to micro-nutrient levels and thus require retention of some level of woody residue. Wildlife habitat requirements may stipulate retention of snags or maintenance of coarse woody debris. Again, ecological factors including wildlife and endangered species need careful site-specific evaluations in determining biomass availability.

Sustainability is explicitly addressed in this analysis through several assumptions. On Federal lands, vegetation management projects are implemented within the framework of environmental analyses and regulations that ensure

consideration of ecological effects and sustainability. Although less restricted, treatments on private lands are also constrained through various environmental laws and regulations (Ellefson et al. 1997). The potential forest biomass supply that is modeled here is a secondary output of other management objectives. We consider biomass that would be available from forest health treatments, fire hazard reduction work, or treatment of activity fuels after logging where questions of sustainability are addressed in the larger management plan.

The present assessment also assumes ecological considerations and practical limitations would have the effect of reducing the amount of biomass available for removal and utilization. The process used models silvicultural treatments and estimates total available biomass. The total available biomass is then further reduced to reflect material left on site to meet ecological constraints or that is otherwise impractical to remove. The reduced amount is the net biomass available for removal. For example, a previous study (USFS 2003) with limited environmental screens estimated 345 million oven dry tons (odt) of biomass may be available from fire hazard reduction thinnings, whereas with our additional screens—for our Base Case—we estimate 114 million odt tons are currently available. For each estimate, it is assumed these amounts would be harvested over a period of years.

As a final gross check on sustainability, the net annual growth in western forest types was calculated from Forest Inventory and Analysis (FIA) plot data and compared to the estimated biomass removal volumes. While growth, mortality, and removal are not holistic measures of ecological integrity, they provide a benchmark of management intensity and impact. For 2002, the total net annual growth of growing stock on timberland in western states was about 97 million odt per year, and of this 43 million odt was removed (Smith et al. 2003). Growing stock growth does not include growth in tops and branches or in nongrowing stock trees. Our Base Case would use about 13 million odt of biomass per year, which is an amount less than 25% of currently unremoved net growth of growing stock (13/(97 – 43) = 0.24). The estimated fraction would be less if we included, in the denominator, the growth of tops of growing stock trees and growth of nongrowing stock trees.

Biomass Sources

In general terms the forest biomass sources include the following:

- Thinning of timberland with high fire hazard
- Logging residue left behind after anticipated logging operations for conventional products
- Treatment of pinyon juniper woodland
- General thinning of private timberland
- Precommercial thinning on National Forest land in western Oregon and Washington
- Unused mill residue

Our analysis includes supply of biomass from federal lands. But this supply from federal land may not be a viable, because the Energy Independence and Security Act of 2007 would not allow biofuels made using biomass from most federal lands to count toward the biofuels RFS (renewable fuels standard). Supply would be allowed from tribal lands held in trust by the federal government and from all lands in "the immediate vicinity of buildings and other areas regularly occupied by people, or of public infrastructure, at risk from wildfire." The RFS requires that 21 billion gallons of "advanced biofuels" need to be supplied by 2022 and only certain biomass sources may be used in meeting this standard. The only one of our sources that would not be notably reduced by this restriction would be the estimated 2.7 to 4.3 million od tons of biomass per year from general thinning on private land.

Biomass supply estimates were made for each county in selected western states. We make a Base Case supply estimate for each source and for some sources we make a High Case estimate to cover a range of uncertainty about supply from the source. Supply estimates include amounts available at roadside in each county for each of several successively higher costs.

Base Case and High Case estimates of total potential annual supply by source are shown in Table 1. Base Case by state and roadside cost are shown in Table 2 and in Figure 1.

Thinning on Timberland with High Fire Hazard

Thinning of timberland with high fire hazard contributes to forest sustainability by reducing the risk of uncharacteristically severe fire. By conducting a thinning, the intent is to move toward a natural fire regime pattern with natural recurrence of less severe fire. Supply was estimated by simulating thinnings on federal and non-federal land using the FTE 3.0 model (Miles 2005, Skog et al. 2006) and Forest Service FIA plot data (USFS 2008a). It is assumed that timberland with current high fire hazard will be thinned over a period of years with either (1) an uneven-aged thinning (where some trees of all size classes may be taken) or (2) an even-aged thinning (where small-diameter trees are taken first followed by successively larger trees until the hazard reduction target is met). A series of screens were applied to identify about 23 million federal and non federal acres that would receive simulated treatment. One screen excluded from treatment is those forest types where stand replacement fire is the norm (lodgepole pine and spruce-fir). An additional screen excluded treatment of wet climate counties in western Oregon and Washington (see separate source below). These areas were excluded because such treatments would not be consistent with our ecological objectives.

For federal lands, it is assumed that even-aged and uneven-aged treatments are used equally, but for non-federal land, it is assumed only uneven-aged treatments

Table 1: Potential annual wood biomass supply from selected western states (million oven-dry tons).

	Source	Base Case	High Case	WGA CDEAC[a]	BTSR[a]
A	Fire hazard thinning on timberland	5.2	7.5	7.2	
B	Logging residue	4.7	4.1	5.3	5.3
C	Treatment of pinyon-juniper woodland	7.6	11.5		
D	General thin on private timberland	2.7	4.3		
E	Pre-commercial thin on National Forest in western counties of Oregon and Washington	0.3	0.3		
F	Mill residue	0.2	0.2	0.3	0.3
	TOTAL	20.7	27.9		
	Thinning to reduce fire hazard on timberland				10.8
	Thinning on other forest land			9.2	9.2
	TOTAL			22	25.6

[a] BTSR, Perlack et al. (2005); WGA CDEAC, WGA (2006).

Table 2: Base Case cumulative forest biomass supply by state and roadside cost

	Biomass supply (oven dry tons/year) at various roadside costs (in $/oven dry ton)						
State	**$10**	**$20**	**$30**	**$40**	**$50**	**$75**	**$100**
Arizona	53,313	154,025	222,599	225,198	228,874	2,092,106	2,094,275
California	1,271,547	3,366,681	3,966,745	4,046,998	4,104,845	4,263,956	4,268,243
Colorado	82,812	193,561	279,369	324,313	341,516	1,542,596	1,552,011
Idaho	778,692	1,005,643	1,478,387	1,592,434	1,669,077	1,803,476	1,824,399
Kansas	8,720	8,720	8,720	8,720	8,720	8,720	8,720
Montana	628,548	1,053,812	1,554,616	1,694,996	1,768,144	1,850,486	1,882,451
Nebraska	4,971	4,971	4,971	4,971	4,971	4,971	4,971
Nevada	4,799	7,043	7,122	7,195	7,195	1,370,524	1,370,524
New Mexico	68,897	135,084	299,745	326,263	352,722	1,675,499	1,680,423
North Dakota	265	265	265	265	265	265	265
Oregon	924,418	1,628,936	1,712,498	1,764,367	1,824,752	1,850,106	1,851,089
South Dakota	95,407	98,503	112,224	112,224	112,224	112,224	112,224
Texas	3,022	3,022	3,022	3,022	3,022	3,022	3,022
Utah	32,670	48,437	101,966	118,102	128,534	1,776,062	1,787,916
Washington	916,029	1,437,920	1,657,948	1,757,994	1,803,262	1,820,173	1,826,722
Wyoming	81,784	123,925	185,505	204,620	211,075	298,320	301,136
Total	4,955,893	9,270,549	11,595,702	12,191,683	12,569,199	20,472,506	20,568,392

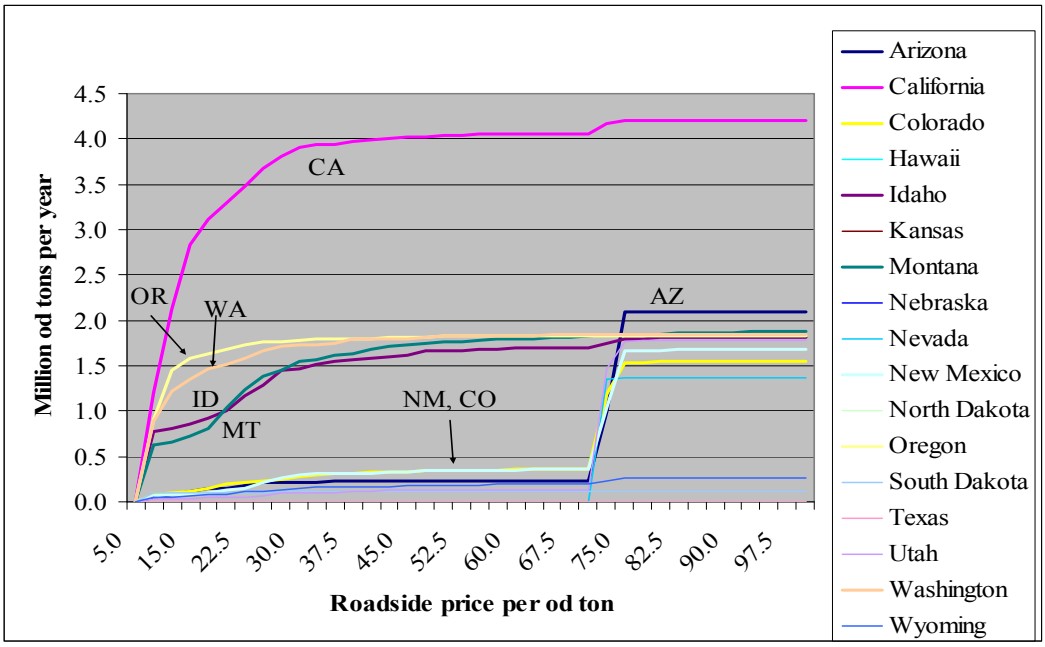


Figure 1: Base Case forest biomass supply by state.

are used. For this source and sources C, D, and E in Table 1, it was assumed biomass volumes identified would be harvested over a period of years. Over this period of harvest, tree growth and mortality will continue and—depending on these growth and mortality rates—additional material would be available for harvest beyond the estimated harvest period. For the Base Case, for sources A and E, we chose a harvest period of 22 years. This time period was previously chosen so fire hazard reduction treatments (source A) would be done on about 500,000 acres per year. For sources C and D, we chose a harvest period of 30 years to match the harvest period used in the DOE/USDA "Billion Ton Supply" report (Perlack et al. 2005) for thinning treatments.

For the source A Base Case, it is assumed that tops and branches of all trees and main stem of trees up to 7 inches diameter at breast height (dbh) are supplied for biofuels, and for the High Case, trees removed up to 9 inches are also supplied for biofuels. Main stems of larger trees not used for biofuels are assumed to be used to make lumber or other higher value products. The cost to remove tops and branches to roadside was assumed to be covered by the cost of removing the whole tree. At roadside there is an assumed $8/dry ton chipping cost. The cost for removing the main stem of trees supplied for biofuels was estimated using the FRCS model (Biesecker and Fight 2006) for wood removals from each FIA forest plot. It was assumed stumpage cost would be $2/odt on private land and $0 on public land. Using these data, wood biomass supply curves were estimated for each county in 12 Western states—Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, South Dakota, Utah, Washington, and Wyoming.

Logging Residue Left behind after Anticipated Logging Operations for Conventional Products

Wood harvested and left on the ground at harvesting sites (or land clearing sites) may be taken to a certain degree subject to limits including (but not limited to) the need to maintain nutrients on site and to retain habitat. For the Base Case supply estimate, we use the allowable removal fractions from the DOE/USDA billion-ton-supply report—65% for logging residue is available for biofuels from harvest sites and 50% from land clearing sites. The High Case is the same as the Base Case for this source because only a Base Case exists for this source. Data on logging residue and land clearing are from the Forest Service 2002 RPA Timber Product Output data base (USFS 2008b). To estimate the roadside cost we assume that whole tree removal will be used (where not already used) to bring out tops and branches to roadside. The cost for removing tops and branches to roadside will be covered by the cost of removing the main stem material. That is, the only cost to provide the wood at roadside will be to chip for $8/odt. It is assumed stumpage cost would be $2/odt on private land and $0 on public land. It is recognized logging residues come from current logging operations that provide sawlogs, pulpwood, posts, and poles. It is assumed if thinning to reduce fire hazard expands and general thinning on private land expands (including biomass for fuels), then the extent of traditional operations will decrease along with associated logging residue. Given the uncertainty about the degree of displacement, we decrease logging residue use for fuels by one-quarter unit for each unit increase in biomass for fuels coming from new thinnings.

Treatment of Pinyon-Juniper Woodland

Pinyon-juniper is a category of woodland forest that produces less than 20 ft^3 per acre per year. Pinyon-juniper forest type has expanded extensively beyond its historic range, and our ecological objective in treating this area over time is to bring the extent of this forest type closer to its historic range. For the Base Case supply estimate, we use allowable removal fractions from the DOE/USDA billion-ton-supply report (table A-6)—45.9% of wood on these public pinyon-juniper lands is available for biofuels and 61.2% of wood on private pinyon-juniper lands is available. This study excludes wood supply from other woodland categories in the West because we could not cite an ecological reason for such treatment.

For the Base Case, we estimate 1/30 of the total volume would be supplied each year (as assumed in the billion-ton-supply report). We made a general estimate that the average cost of harvest would be $60/odt and roadside chipping would cost $12.60/odt, for a total of $72.60/odt. The chipping cost for pinyon-juniper trees is estimated to be higher than for tops and branches of other trees based on case studies that indicate chipper throughput is lower for pinyon-juniper. This is thought to be due in part to the irregular form of pinyon-juniper trees. It was assumed stumpage cost would be $2/odt on private land and $0 on public

land. For the High Case, we assume that the treatments would occur over 20 years and costs would be subsidized at $20/odt based on proposed legislation.

Note that Figure 1 shows that large quantities of biomass from pinyon-juniper land become available in several states when price reaches $72.60. This is because we have a single price estimate for removing this biomass. In reality, the supply would increase more gradually over a range of prices we estimate would be centered on a price of $72.60.

General Thinning on Private Land

It is presumed that as demand and prices for biomass for fuels increase, there will be an increase in operations to harvest both woody biomass and sawlogs/pulpwood in combined operations on private land. Some private land is excluded from this source because it is already treated under the fire hazard reduction thinnings noted above. This source estimates supply from private land acres that have sufficient stocking to warrant thinning but have lower fire hazard. For the Base Case supply estimate, we simulated an uneven-aged thinning on private land FIA timberland plots that were not treated by a fire hazard thinning procedure (source A). The estimation procedure is the same used to estimate biomass from thinning U.S. timberland for the billion-ton-supply report (stands with density greater than 30% of maximum stand density index are thinned back to 30%). Because the thinnings may be heavier than appropriate for lodgepole pine and spruce-fir forest types—they are subject to wind throw if thinned too heavily—we did not treat those forest types.

The Base Case supply is assumed to be provided in equal annual amounts over 30 years. The supply costs were estimated in the same way as for the fire hazard reduction thinnings (source A). For the High Case, trees removed up to 9 inches are also supplied for biofuels and the annual supply is assumed to be provided in equal amounts over 20 years. It is assumed stumpage cost would be $2/odt.

Precommercial Thinning on National Forest Land in Western Counties in Oregon and Washington

We did not simulate fire hazard reduction thinnings on National Forest timberland in counties west of the Cascade Mountains in Oregon and Washington where the thinning objective would not be focused on reducing fire hazard but on maintaining appropriate stocking and habitat conditions. Instead, for source E, we simulated a precommercial thinning of FIA plots to remove trees five to 9 inches dbh in stands up to 40 years old. For the Base Case, it is assumed that 1/22 of this volume could be harvested each year (the same as for source A). The cost to harvest and move wood to roadside was estimated for each treated FIA plot using the FRCS model. Harvest costs for individual plots ranged from a low of $22/odt to about $70/odt for many plots, with some plots costing over $500/odt. It is

assumed stumpage cost on National Forest land is $0/odt. The High Case supply is the same as the Base Case.

Unused mill residue

Forest Service surveys of wood products mills (e.g., lumber, plywood, pulp) periodically estimate amounts of coarse and fine wood and bark residue generated by county and how much goes for various uses (e.g., fuel, fiber input for pulp or panels.) Source F is the estimate of mill residue that goes unused. We assume this entire unused amount is available to make biofuels. The amount supplied is the same for the Base Case and High Case. It is assumed the cost at the mill is $0/odt.

Additional Results and Discussion

The WGA Biofuels Assessment reports used the forest biomass supply curves, supply curves for additional lignocellulosic biomass sources, supply curves for grease and tallow, information on transportation networks, and costs for conversion technologies in a mixed integer programming model to identify the location for biofuels plants given different offered prices for biofuels at fuel terminals.

Additional sources of lignocellulosic feedstocks include corn stover, straw, herbaceous energy crops (e.g., switchgrass), orchard and vineyard waste, and several types of municipal solid waste. Supply curves for lignocellulosic biomass sources are shown in Figure 2. Table 3 shows amounts of all feedstocks supplied when the offered price for biofuels is $2.40/gge at terminals. The types of biofuels produced include ethanol and biodiesel. At $2.40/gge, forest biomass contributes 11.4 million odt of feedstock per year, or about 9% out of a total supply of 130 million odt.

At $2.40/gge, the 130 million odt of feedstocks would produce 7.6 billion gge/year (Figure 3). If each ton of feedstock produces about the same amount of biofuel, then forest biomass would produce 0.6 to 0.7 billion gge of biofuels. The conversion efficiency varies by feedstock, so this is a rough estimate.

The locations of the biofuels plants using lignocellulosic feedstocks are shown in Figure 4. Biomass supply curves in Figure 1 indicate that forest biomass would be supplying plants that are located in California, Oregon, Washington, Idaho, and Montana.

The modeling framework developed for this assessment constitutes a comprehensive framework for spatially explicit integrated analysis of the entire biofuel supply chain. As with any model, its foremost limitation is in the quality of the input data. The results of this modeling effort indicate that there is significant potential to expand biofuels production in the West. Exclusive of resource competition from other energy and product markets, there is the potential

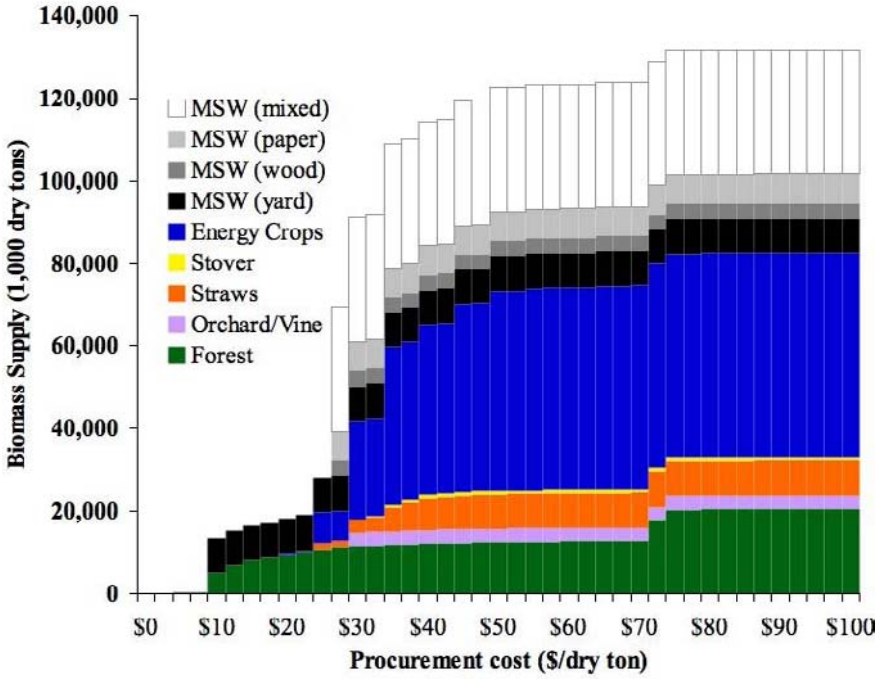


Figure 2. Lignocellulosic biomass feedstock supply by roadside or farm gate cost.

Table 3: Feedstock consumed (million odt) at a biofuels target price of $2.40/gge

Feedstock	Feedstock consumption	Percentage of total
Corn	43.9	33.8
Herbaceous energy crops	43.1	33.2
Municipal solid waste	19	14.6
Forest biomass	11.2	8.6
Straw (wheat, barley, rye, oats)	7.9	6.1
Orchard and vineyard waste	2.9	2.2
Tallow	0.9	0.7
Corn stover	0.8	0.6
Waste grease	0.2	0.2
Total	130	

for the West to supply substantial fractions of renewable fuels under the new federal Renewable Fuel Standard. The WGA Assessment report suggests several key conclusions concerning land use, and transportation infrastructure that will affect biofuels potential:

- Land use policies will have a significant impact on the availability of feedstock.
- Land use policies will affect expansion of herbaceous energy crop production on marginal lands that will be influenced by sustainability standards or research findings.

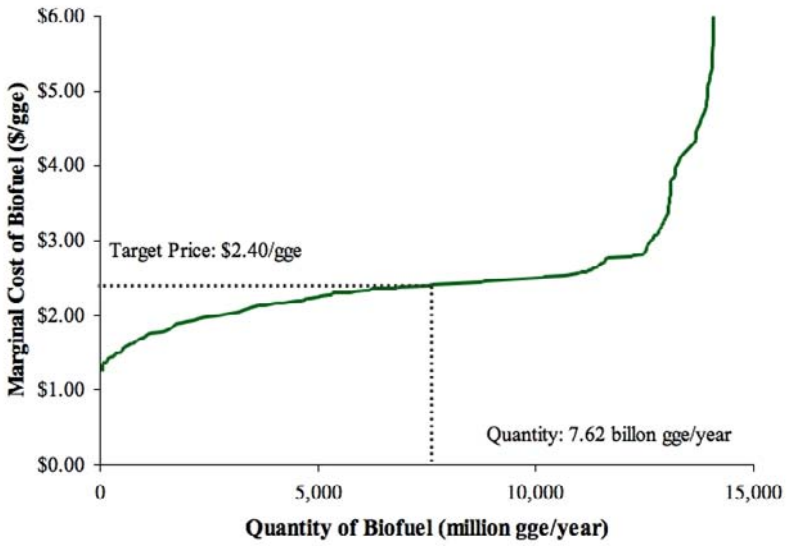


Figure 3: Westwide biofuels supply by delivered cost at fuel terminals.

- Land use policy formulation should carefully explore the possibility of meeting GHG reduction targets under the federal RFS through more sustainable energy crop substitution on lands currently producing corn and other high input crops at low relative yields.
- A more detailed analysis is needed on the capacity of existing transportation infrastructure to meet demands of the biofuel supply chain.
- A spatially explicit analysis should be conducted of the potential for new transportation infrastructure to improve supply chain economics for biofuels production.

References

Biesecker, R.L.; Fight, R.D. 2006. My fuel treatment planner: a user guide. Gen. Tech. Rep. PNW-GTR-663. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 31 p. http://www.fs.fed.us/pnw/data/myftp/myftp_home.htm

Ellefson, P.V., Chen, A.S., Moulton, R.T. 1997. State forest practice regulatory programs: an approach to implementing ecosystem management on private forest lands in the United States. Environmental Forestry 21(3):421-432.

Helms, J.A., ed. 1998. The Dictionary of Forestry. Society of American Foresters, Bethesda, MD. 210 p.

Miles, Patrick D. 2005. Fuel Treatment Evaluator web-application version 3.0. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. http://www.ncrs2.fs.fed.us/4801/fiadb/FTE_Version3/WC_FTE_version3.asp

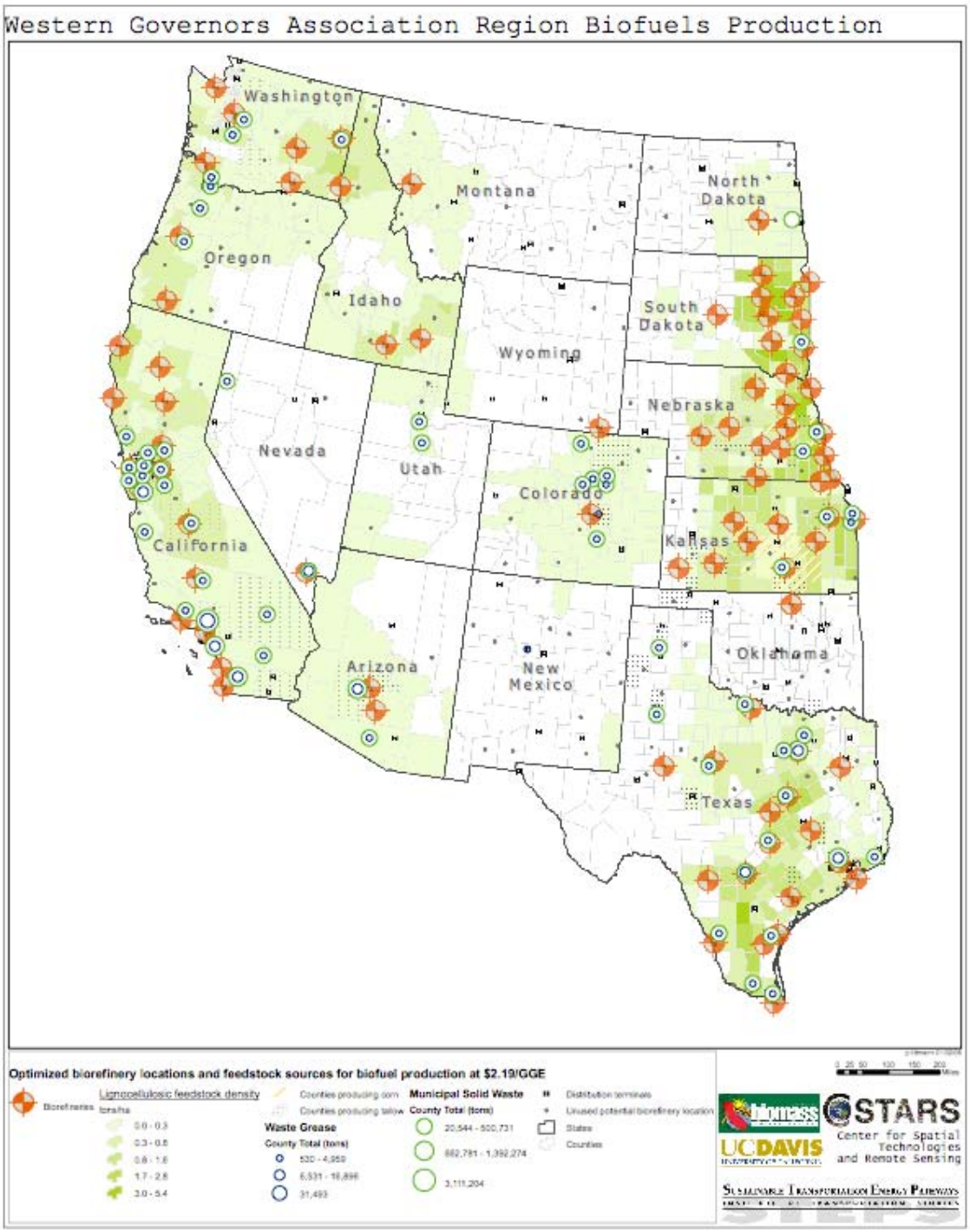


Figure 4: Potential location of lignocellulosic biofuels plants in the West (orange circle markers).

Perlack, R.D. et al. 2005 Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion ton supply. Oak Ridge National Laboratory, Oak Ridge, TN 60 p. http://feedstockreview.ornl.gov/pdf/billion_ton_vision.pdf

Skog, K.E.; Barbour, R.J; Abt, K.L.; Bilek, E.M. (Ted); Burch, F.; Fight, R.D.; Hugget, R.J.; Miles, P.D.; Reinhardt, E.D.; Shepperd, W.D. 2006. Evaluation of silvicultural treatments and biomass use for reducing fire hazard in western states Research Paper FPL-RP-634. Madison, WI: USDA, Forest Service, Forest Products Laboratory. 29 pages.

Smith, W. Brad; Miles, Patrick D.; Vissage, John S.; Pugh, Scott A. 2003. Forest Resources of the United States, 2002. Gen. Tech. Rep. NC-241. St. Paul, MN: USDA Forest Service, North Central Research Station. 137 p.

USFS. 2003. A strategic assessment of forest biomass and fuel reduction treatments in western states. http://www.fs.fed.us/research/infocenter.html

USFS. 2008a. Forest Inventory Data Online. See http://fia.fs.fed.us/tools-data/

USFS. 2008b. Timber Products Output Database. See http://ncrs2.fs.fed.us/4801/fiadb/rpa_tpo/wc_rpa_tpo.ASP

Western Governors Association. 2006. Forest fuel treatment & thinning biomass – Timberland. In: 2006 Biomass Taskforce Report: Clean and diversified energy initiative – Biomass Task Force Report – Supply Addendum. Denver, CO. p 11-12ff. http://www.westgov.org/wga/initiatives/cdeac/Biomass-supply.pdf

Western Governors Association. 2008a. Strategic assessment of bioenergy development in the West: Biomass resource assessment and supply analysis for the WGA region. September 1, 2008. Denver, CO http://www.westgov.org/wga/initiatives/transfuels/Task%201.pdf

Western Governors Association. 2008b. Strategic assessment of bioenergy development in the West: Bioenergy conversion technology characteristics. September 1, 2008. Denver, CO http://www.westgov.org/wga/initiatives/transfuels/Task%202.pdf

Western Governors Association. 2008c. Strategic assessment of bioenergy development in the West: spatial analysis and supply curve development. September 1, 2008. Denver, CO http://www.westgov.org/wga/initiatives/transfuels/Task%203.pdf

Western Governors Association. 2008d. Strategic assessment of bioenergy development in the West: analyses of deployment scenarios and policy interactions. September 1, 2008. Denver, CO http://www.westgov.org/wga/initiatives/transfuels/Task%204.pdf

Utilizing Climate Information



Predicting forest attributes from climate data using a recursive partitioning and regression tree algorithm

Greg C. Liknes[1], Christopher W. Woodall[2], and Charles H. Perry[2]

ABSTRACT: *Climate information frequently is included in geospatial modeling efforts to improve the predictive capability of other data sources. The selection of an appropriate climate data source requires consideration given the number of choices available. With regard to climate data, there are a variety of parameters (e.g., temperature, humidity, precipitation), time intervals (e.g., 30-year normal, seasonal average), and summary statistics (e.g., mean, minimum) which can be selected. In this study, we propose a technique for evaluating the combination of climate parameters that are most closely related to ground observations of forest attributes. Using data from the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service as response variables, recursive partitioning and regression tree analysis was applied using a suite of climate variables from the Daymet database as predictor data. Although model improvement scores for climate variables were modest, the technique provides opportunities for deciding among a wide array of possible climate predictors.*

KEYWORDS: Daymet, climate, forest inventory, data mining

Introduction

Forest composition and structure are, in part, a function of local climatic conditions (Bailey 1995, Whittaker 1975). Geospatial modeling predictions of forest attributes may use climate information to augment topographic and remote sensing data. However, the full range of climate data possibilities are rarely considered because of the daunting number of possible combinations of time interval, descriptive statistics, and spatial resolution. For example, what is the appropriate length of time over which to assess typical rainfall in a given area? Should we consider mean, maximum, or minimum values of climate inputs (e.g., precipitation, solar radiation) and state variables (e.g., temperature, humidity)?

[1]Research Physical Scientist, USDA Forest Service, Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108; Corresponding author email: gliknes@fs.fed.us
[2]USDA Forest Service, Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Ohmann and Gregory (2002) predicted tree species composition and structure in coastal Oregon using a variety of data. They found Landsat Thematic Mapper satellite imagery and climate information to have the first and second most explanatory power, respectively, followed by location, topography, ownership, and geology. The authors transformed temperature and precipitation data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) climate dataset to create eight predictor variables. Their climate predictors captured seasonality, variability, growing season conditions, and continentality, but due to the limitations of the PRISM dataset, other factors such as radiation were not included.

Previous work by Liknes and Woodall (2007) began to assess the climate factors that have the most predictive power relative to forest attributes by examining correlations between forest inventory data and a variety of parameters in the Daymet[3] climate database. In this study, we aimed to build on our previous work by re-examining the Daymet dataset with an improved technique.

Data

Data were analyzed for the states of Michigan, Minnesota, and Wisconsin. These states cover 49 million hectares, of which 21 million hectares are sub-boreal and temperate forests.

Forest Inventory Data

Field data were collected between 2001 and 2006 on nearly 20,000 forested or partially-forested plots by the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service. The 0.4-ha plots in the study area are re-visited every five years, allowing calculation of growth and mortality for a subset of plots visited in both 2001 and 2006.

Climate Data

Climate data used in this study were taken from the Daymet climate database (Thornton et al. 1997). The Daymet raster datasets provide full coverage of the conterminous United States at 1-km resolution for a suite of climate parameters including temperature, precipitation, humidity, and radiation. Additionally, measures of climate variability (interannual standard deviation and day-to-day variability) are available for each parameter, as well as selections of a single-year

[3] Daily Surface Weather and Climatological Summaries (http://www.daymet.org)

average, 18-year annual average, or an 18-year average for a specific month of the year. The 18-year average datasets are for the period between 1980 and 1997.

We considered a subset of the Daymet database, with all selected variables averaged over the 18-year period. See the Appendix for a list of the climate variables used in this study.

Methods

Forest inventory data for individual trees were aggregated to calculate biomass, basal area, and growth for each plot. Climate values at each plot were extracted using a Geographic Information System operation, which assigned each plot to a climate pixel based on proximity to the nearest pixel center.

Recursive Partitioning and Regression Tree Analysis

Although various data mining techniques are available, we chose to examine recursive partitioning and regression tree (*rpart*) analysis (Therneau 1983). The *rpart* algorithm shares many similarities with other data mining techniques, particularly classification and regression tree analysis. For more information on classification and regression tree analysis, refer to Breiman et al. (1983). The following advantages for analyzing multiple predictor variables are provided by *rpart*: it is compatible with the prediction of continuous variables, the resulting models can be presented as intuitive binary trees, and it requires relatively few input choices by the user.

We implemented the *rpart* algorithm using R statistical software[4] and the *rpart* package[5]. More detail on the parameters used with *rpart* appears in the Appendix. Six separate model runs were conducted using the suite of climate parameters as input predictors; biomass, basal area, and growth were response variables. Two predictor sets were used. The first set contained a suite of climate variables representing minimum, maximum, or mean daily values or mean annual total values. The second set included all of the first predictor set and additional climate data related to variability (see Appendix). Model improvement for node splits was used as a basis for determining which climate variables had the strongest relationship to forest attributes where improvement is defined as

$$1 - (SS_{right} + SS_{left}) / SS_{parent} \quad (1)$$

and *SS* is the sum-of-squares, *right* and *left* refer to the sides of the split, and *parent* refers to the node that is split.

[4] The version of R used was 2.8.0. The software is available at http://www.r-project.org.
[5] The version of the *rpart* package used was 3.1-42.

Results

For both basal area and biomass, the top node splits were related to minimum daily air temperature, while for growth the top node splits were related to precipitation frequency (Table 1). Figure 1 shows the top three node splits that resulted when an *rpart* model was created using the second predictor dataset (climate and variability, see Appendix) to predict forest biomass. The primary split for biomass occurred on minimum daily December temperature at a value of -15.375 degrees Celsius. Including interannual and day-to-day variability (second predictor set) did not change the results for biomass, but it did change the results for basal area and growth; interannual variability appeared among the top node splits in both cases.

Table 1: Results from six model runs predicting forest attributes using selected climate variables and recursive partioning and regression tree analysis in Minnesota, Wisconsin, and Michigan.

Response variable	Predictor set	Top three node splits	Model improvement
Basal area	1	January minimum daily air temperature	0.02145327
		Mean daily total shortwave radiation	0.01690332
		Mean precipitation event size	0.006725006
Basal area	2	Interannual variability in December mean daily minimum temperature	0.02569515
		Mean daily total shortwave radiation	0.01612992
		Mean precipitation event size	0.003692035
Biomass	1	December minimum daily air temperature	0.03854474
		Mean daily total shortwave radiation	0.03828900
		Minimum daily air temperature	0.01672963
Biomass	2	December minimum daily air temperature	0.03854474
		Mean daily total shortwave radiation	0.03828900
		Minimum daily air temperature	0.01672963
Growth	1	Precipitation frequency (<0.255)	0.009561471
		Precipitation frequency (<0.275)	0.0035711810
		Cooling degree days	0.005642840
Growth	2	Precipitation frequency	0.009561471
		Interannual variability in December mean daily minimum temperature	0.005102520
		Interannual variability in total precipitation	0.006068496

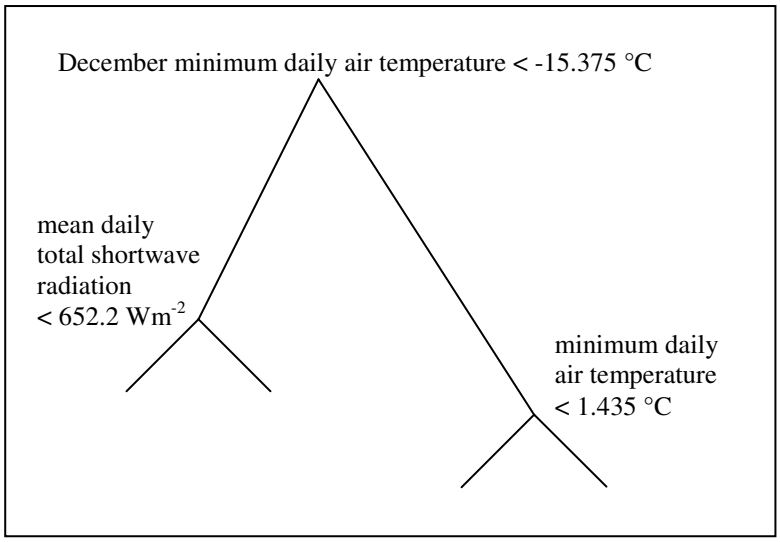


Figure 1: A regression tree resulting from a recursive partitioning and regression tree model of forest biomass in Minnesota, Wisconsin, and Michigan using Daymet climate variables as predictors.

Model improvement values are generally useful as a relative indication of how predictors perform. Improvement of the biomass model due to the predictors at top node splits was approximately three times that of the growth model (0.03 vs. 0.009).

Discussion

Analysis of climate variables with the *rpart* algorithm resulted in modest model improvements. More importantly, the technique allowed for selection of the most influential climate variables from a large set of potential factors. The technique is straightforward to implement and highlights the possible importance of less conventional climate variables (such as mean daily minimum temperatures). The inclusion of some measure of interannual variability of temperature and precipitation may also be warranted for some forest attributes. Furthermore, the predictors selected for model inclusion should depend on the attribute to be modeled. The results presented should not be considered as definitive guidance for the selection of climate variables used in the prediction of forest attributes. For example, future work could include an analysis using *rpart* and the entire Daymet database over a larger geographic area, as well as comparisons of results using other data mining or exploratory data analysis techniques.

Many issues require further exploration to improve the utility of this technique with FIA plot data. For example, there is a mismatch in the areal extent of a

Daymet raster cell (100 ha) and the 0.4-ha FIA plot. The mismatch may have a sizeable impact on the relationship between our response and predictor variables. Additionally, if climate variables are used in conjunction with remotely-sensed data in a model, there may be interactions that are not accounted for in this approach.

A temporal mismatch also may have affected our results. Forest attribute data were obtained between 2001 and 2006, but climatic data were averaged over the period between 1980 and 1997. Contemporaneous climate data may improve model performance, and these data would be available by aggregating monthly PRISM data[6]. However, PRISM has a smaller set of parameters relative to Daymet (e.g., total shortwave radiation is excluded). Although questions surrounding spatial and temporal resolution need to be addressed, *rpart* analysis using climate variables holds promise for improving variable selection and model prediction.

References

Bailey, R.G. 1995. Description of the ecoregions of the United States, 2nd ed. Misc. Publ. 1391. Washington, DC: U.S. Department of Agriculture, Forest Service.

Breiman, L.; Friedman, J. H.; Olshen, R. A.; and Stone, C. J. 1983. Classification and Regression Trees. Belmont, CA: Wadsworth.

Liknes, G.C; Woodall, C.W. 2007. An evaluation of climate parameters in relation to forest inventory data: towards guidance for satellite-based forest attribute modelling. Proceedings of ForestSat 2007 - Forests and Remote Sensing: Methods and Operational Tools; 2007 Nov 5-7; Montpellier, France. [CD-ROM].

Ohmann, J.L.; Gregory, M.J. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, U.S.A. Canadian Journal of Forest Research. 32: 725-741.

Therneau, T.M. 1983. A short introduction to recursive partitioning. Orion Technical Report. Palo Alto, CA: Stanford University, Department of Statistics.

Thornton, P.E.; Running, S.W.; White, M.A. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. Journal of Hydrology, 190: 214-251.

Whittaker, R.H. 1975. Communities and ecosystems, 2nd ed. New York and London: Macmillan.

[6] PRISM Group (http://www.prism.oregonstate.edu/)

Appendix

rpart parameters used:
minisplit = 100
minibucket = 100
cp = 0.0001
method = anova

Sample *rpart* implementation in R:

```
fit <- rpart(Climate$BasalArea ~ ta_a_pa + tx_a_pa + tn_a_pa + td_a_pa + tc_a_pa +
        tf_a_pa + tn_a_pm12 + tn_a_pm01 + tn_a_pm02 + pf_a_pa + pe_a_pa +
        hv_a_pa + rt_a_pa, control=list(minisplit=100, minibucket=100, cp=0.0001),
        method="anova")
```

Appendix Table: Climate variables used as predictors of basal area, biomass, and growth using the *rpart* algorithm.

Climate variable description	Daymet abbreviation	Predictor Set
Mean daily air temperature	ta_a_pa	1,2
Minimum daily air temperature	tn_a_pa	1,2
Maximum daily air temperature	tx_a_pa	1,2
Growing degree days	td_a_pa	1,2
Cooling degree days	tc_a_pa	1,2
Freezing degree days	tf_a_pa	1,2
January minimum daily air temperature	tn_a_pm01	1,2
February minimum daily air temperature	tn_a_pm02	1,2
December minimum daily air temperature	tn_a_pm12	1,2
Mean precipitation event size	pe_a_pa	1,2
Precipitaion frequency	pf_a_pa	1,2
Mean daily total shortwave radiation	rt_a_pa	1,2
Mean daily water vapor pressure	hv_a_pa	1,2
Day-to-day variability in mean daily temperature	tva_a_pa	2
Day-to-day variability in mean daily minimum temperature	tvn_a_pa	2
Day-to-day variability in mean daily maximum temperature	tvx_a_pa	2
Day-to-day variability in mean daily water vapor pressure	hvv_a_pa	2
Day-to-day variability in total shortwave radiation	rvt_a_pa	2
Interrannual variability in mean daily minimum temperature	tn_s_pa	2
Interannual variability in January mean daily minimum temperature	tn_s_pm01	2
Interannual variability in February mean daily minimum temperature	tn_s_pm02	2
Interannual variability in December mean daily minimum temperature	tn_s_pm12	2
Interannual variability in mean daily maximum temperature	tx_s_pa	2
Interannual variability in total precipitation	pt_s_pa	2
Interannual variability in mean precipitation event size	pe_s_pa	2
Interannual variability in precipitation frequency	pf_s_pa	2

Actual Evapotranspiration (AET) and Tree Species Richness in the Eastern U.S.A.

Weihong Fan[1] and Richard H. Waring[2]

ABSTRACT: *Although many studies confirm that competition and disturbance play important roles in determining tree diversity locally, climatic constraints become increasingly important at broader geographic scales. We evaluate the extent that annual actual evapotranspiration (AET) might account for observed variation in tree diversity across the entire eastern U.S. and within 24 Level III ecoregions designated by the Environmental Protection Agency. To estimate tree diversity, we extracted data from a total of 87,137 Forest Inventory and Analysis (FIA) survey plots. For each 1000 km^2 cell with ≥15 plots, logarithmic functions were derived to predict the number of species encountered per hectare (equivalent to 17 fixed-radius plots). Across the region, tree diversity exhibited a humped-shaped pattern with AET (r^2=0.49, P < 0.0001). Stratifying ecoregions by AET, which increases from north to south, winter temperature emerged as a significant variable where AET averaged <800 mm.yr^{-1}, whereas summer temperature accounted for much of the variation where AET was between 800 to 950 mm.yr^{-1}. In ecoregions with AET>1100 mm.yr^{-1}, significant variation in tree diversity was associated with seasonal differences in precipitation. We conclude that across the eastern U.S., AET provides a reasonable prediction of regional variation in tree diversity, especially where AET is < 800 mm.yr^{-1}.*

KEY WORDS: AET, AET stratification, tree species richness; tree diversity, level III ecoregions, FIA database, environmental variables of species diversity; the eastern US.

Introduction

It appears more and more likely that species diversity will be significantly reduced in response to changing climatic conditions over the coming decades. This possibility has lead to increased effort to correlate current biodiversity with climate. At the continental scale, the variable most frequently correlated with biodiversity is actual evapotranspiration (AET), which expresses the annual balance between precipitation and latent heat exchange. AET is also considered a surrogate for net primary production because of a close relationship between production and climatic factors (Lieth 1975). In reference to trees, diversity

[1] Natural Sciences & Mathematics, The Richard Stockton College of New Jersey, Pomona, New Jersey, 08240, USA; e-mail: fanw@stockton.edu

[2] College of Forestry, Oregon State University, Corvallis, Oregon 97331, USA; e-mail: Richard.waring@oregonstate.edu

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

generally increases from cold, dry conditions with low AET to warm, moist tropical environments with high AET (Currie and Paquin 1987, Latham and Ricklafs 1993, Davies et al. 2004). Other environmental variables often associated with variation in biodiversity include potential evapotranspiration and rainfall (Hawkins et al. 2003).

Within regions, physiographic variation and historical factors explain much of the residual variation not account for by AET (Whittaker and Field 2000, Caley and Schluter 1997, Currie et al. 1999, Huston 1999, Whittaker et al. 2001, Sarr et al. 2005). In the United States, the portion east of the Mississippi River contains the largest spatial variation in tree diversity, whether analyzed from species range maps (Currie and Paquin 1987), ecoregions (Waring et al. 2006) or counties (Iverson and Prasad 2001). Studies of tree diversity that rely on range maps have the disadvantage of not being responsive to future climatic limitations. Range maps also overlay areas where the environment may limit a species' presence. As a result, environmental correlations are likely to differ from those based on extensive field surveys (Hurlbert and White 2005).

In the United States, we are fortunate that the federal government has supported and continues to support intensive ground surveys of forested areas throughout the eastern U.S. as part of the Federal Inventory and Analysis (FIA) program (http://fia.fs.fed.us/tools-data/default.asp). Sampling in the western United States is less intensive than in the eastern portion of the country, and to date, most data have been acquired only within variable-radius plots instead of the more robust fixed-radius sampling design. The availability of both fixed- and variable-area plots in the eastern U.S. permits a correlation to be established between the two types of sample plots to provide estimates of tree species present per hectare (described below).

Because the eastern U.S. is a region where rainfall is generally sufficient and evenly distributed throughout the year, reasonable estimates of AET may be made without requiring detailed information on available soil moisture or canopy leaf area index, as would be required to calculate AET in the more arid portions of the western U.S. (Churkina et al. 1999). Based on the literature, we hypothesize that different patterns will emerge between AET and tree diversity depending on the scale of analysis, and on how a region is stratified in reference to AET. Specifically, we expect a humped-shaped relation with AET and tree diversity across the entire region, as resulted from a similar analysis in the northwestern portion of the country (Swenson and Waring 2006). If that is the case, we would expect positive, neutral, or negative relationships between tree diversity and AET as AET increase from north to south. Where a clear relationship between tree diversity and AET is lacking, other environmental factors may emerge as significant, such as winter or summer temperature and precipitation, soil water storage, and topographic relief. If a common set of variables can explain variation in tree diversity across a set of climatically comparable ecoregions, we

should achieve a better understanding of the impact of climatic change on tree species diversity.

Methods

The study covers all forested areas in 31 states in the eastern U. S. For across region analysis, we evaluate the relationship between nine environmental variables and tree richness recorded on all 1000 km^2 (31.6 km x 31.6 km) cells with 15 or more FIA survey plots. A total of 24 ecoregions delineated by the Environmental Protection Agencies' Level III classification (<http://www.epa.gov/wed/pages/ecoregions/ecoregions.htm>) qualified for within ecoregion analysis with at least 20 such cells within each ecoregion (Table 1). Analyses were not conducted at finer scale because the geographic locations of sampling plots are imprecise to maintain privacy of information acquired on private lands (McRoberts et al. 2005 http://ncrs2.fs.fed.us/4801/fiadb/fiadb_documentation/Perturbing-Swapping.pdf).

TABLE 1: The 24 ecoregions recognized by the EPA level III classification and number of qualified 1000 km^2 cells (N)

Code	Name	N
35	South Central Plains	84
39	Ozark Highlands	90
45	Piedmont	156
49	Northern Minnesota Wetlands	22
50	Northern Lakes and Forests	198
51	North Central Hardwood Forests	64
52	Driftless Area	36
56	Southern Michigan/Northern Indiana Drift Plains	23
58	Northeastern Highlands	109
59	Northeastern Coastal Zone	31
60	Northern Appalachian Plateau and Uplands	27
62	North Central Appalachians	28
63	Middle Atlantic Coastal Plain	70
65	Southeastern Plains	310
66	Blue Ridge	46
67	Ridge and Valley	93
68	Southwestern Appalachians	37
69	Central Appalachians	64
70	Western Allegheny Plateau	71
71	Interior Plateau	52
72	Interior River Valleys and Hills	46
75	Southern Coastal Plain	107
82	Laurentian Plain and Hills	42
83	Eastern Great Lakes and Hudson Lowlands	28

Species diversity

One lesson learned from previous studies is that there are advantages, especially when the sample size is small, if comparable areas are sampled (Whittaker et al. 2001, Sarr et al. 2005). In the study, we estimate species

diversity (richness) per hectare. This was a challenge because the largest data sets acquired by the FIA surveys used variable radius plots, where the sampling area varies as a function of prism angle size and the diameter and distance of trees from a central point (Alerich et al. 2004). Moreover, the choice of prism size varied, even within states. Although the sampled area can be estimated from mean tree diameter, large variations around the mean result in questionable estimates of species richness per unit area if made directly from variable radius plot data.

Fortunately, over the last seven years, the FIA surveys have introduced a standard fixed-radius plot layout, where four 7.3 m (24 ft.) radius subplots cover a total of 0.0675 ha (1/6 of an acre) (USDA, Forest Service 2004). However, fixed-radius plots are only available for a limited number of states thus far and the sampling density of fixed-radius plots is sparse compared to variable-radius plots. To utilize the more extensive variable-radius plot data, we compared the two independent estimates of species richness where both sets of samples were available.

FIA plot data were downloaded from the FIADB website, http://fiatools.fs.fed.us/fiadb-downloads/fiadb3.html, of USDA Forest Service. In this data set, plots are identified by their locations as provided by the database, and were then grouped as they occurred in 1000 km^2 cells. All cells containing ≥15 FIA plots were included because an asymptote in species numbers is approached with that sample size, and by setting a higher minimum, the number of qualified cells would be reduced significantly. For each cell with both sets of plots (fixed- and variable-radius), logarithmic functions of species richness with increasing number of plots were calculated and compared. To smooth the curves, we first listed plots of each set in a cell by number of species in ascending order, then in descending order, before randomly reordering the sequence twice. The average species richness values from the four summarizations were used for each 1000 km^2 cell. We standardized the estimates for 17 fixed-radius plots, equivalent to one hectare, based on the cell-specific slope and intercept of the logarithmic function of species number vs. plots number. We generally attained an excellent fit between species richness and increasing plot areas for each 1000 km^2 cell with an $R^2 > 0.95$. The results from variable radius plots were compared with that of fixed radius plots and the relationship statistically assessed.

Environmental variables

Nine environmental variables were selected to correlate with species richness patterns. These included two climatic indices: AET and mean annual precipitation (PPT); two soil properties: soil organic matter content (OM) and soil water holding capacity (SW); and five topographic descriptors: average % slope (SL), range in % slopes (RSL), average elevation (EL), range in elevation (REL), and the length of major roads (LR).

The USGS (2002) provides a GIS Database (http://webgis.wr.usgs.gov/globalgis/), where annual AET, representing the difference between precipitation and latent heat loss, is calculated for the U.S. based on equations developed by Prentice et al. (1993). The equations are applied at monthly time steps and the values then summed for the year. This formulation differs from the more data demanding Penman-Monteith equation by not requiring information on canopy leaf area, and stomatal response to evaporative demand and limitations of soil water. AET values were acquired using ArcGIS/ArcInfo software for all qualified 1000 km^2 cells by averaging values extracted from the 1 x 1 km resolution data. The values of PPT were acquired similarly. Data of SL, RSL, SW, and OM were calculated from the USDA STATSGO database (1994) at a scale of 1:250000. Before intersecting with the soil data layers, each 1000 km^2 cell was further divided into 49 equally spaced cells to capture the spatial variations of each variable. Within each 1000km^2 cell, the 49 smaller cells were then either summed, averaged or the range calculated, depending on the variable. Elevation data at a scale of 1:250000 were downloaded form USGS GeoData website (http://edc.usgs.gov/products/elevation/dem.html) and the values of average elevation and elevation ranges for each 1000 km^2 cell were summarized as with the soil data. Total length of major roads was calculated for each 1000 km^2 cell in GIS using ArcGIS/ArcInfo by overlaying the cells with the road map provided by the USGS Global GIS Database.

Statistical Analysis

Across the entire region, a stepwise regression analysis with all nine abiotic variables was first performed using all qualified 1000 km^2 cells. A variable was included if significant at $P < 0.001$. We then repeated the process separately for each of the 24 defined ecoregions. Finally, we stratified ecoregions by AET and assessed tree richness correlations in reference to differences in seasonal temperature and precipitation.

Results

Species richness

We found that the number of species from the two FIA inventories, fixed and variable radius plots, was significantly and positively related ($R^2 = 0.74$, $P < 0.0001$) when standardized to 17 plots (Figure 1). The slope of the linear function was 0.97 when the intercept was set to 0. With the relationship established between fixed- and variable-radius plots illustrated in Figure 1, we were able to convert estimates of species richness derived from variable-radius plots to richness per hectare (equivalent to 17 fixed-radius plots).

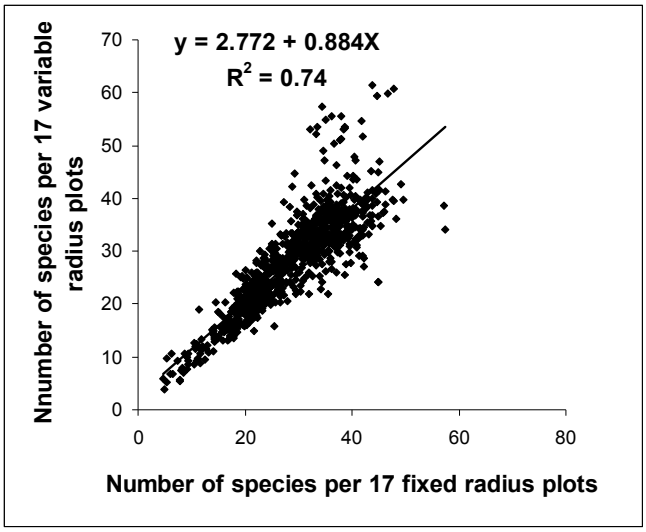


FIGURE 1: The relationship between number of species per 17 variable radius plots and number of species per 17 fixed radius plots ($R^2 = 0.74$). When the intercept was set to 0, the slope of the linear function is 0.97, which demonstrates that estimates of species richness from the two independent FIA inventories are equivalent.

Spatial variation of tree species richness in the eastern U.S. indicates that areas with the highest richness are located in central portion with heterogeneous topography and a favorable climate. Species richness decreases significantly in all directions (Figure 2).

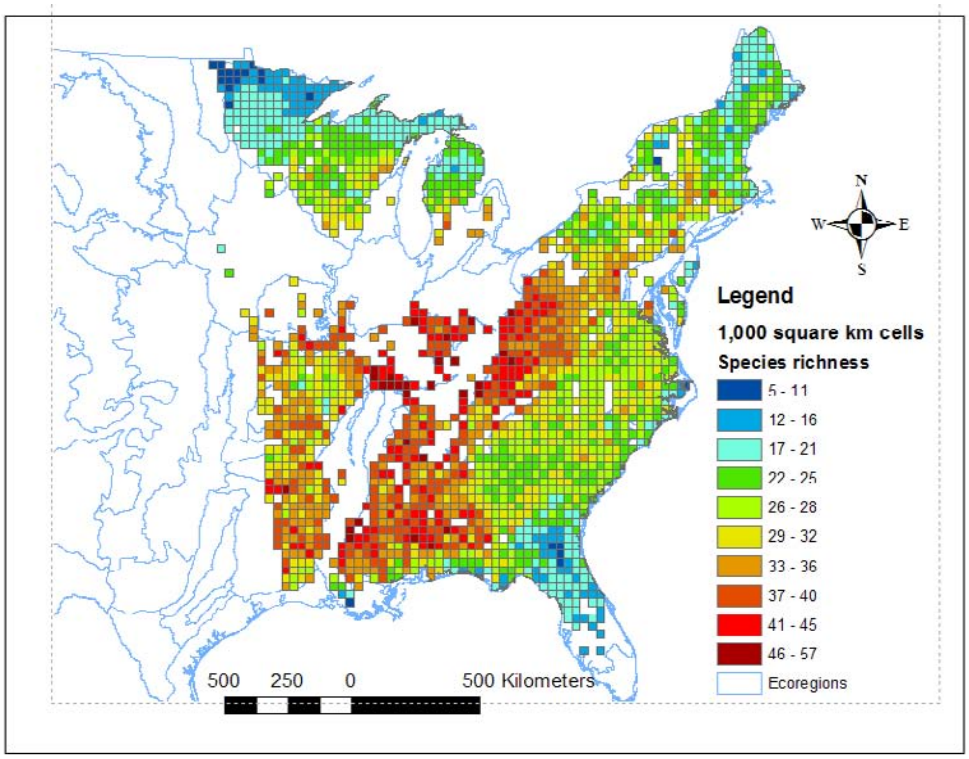


FIGURE 2: Spatial variations of tree species richness estimated for qualified 1000 km cells with ≥15 FIA survey plots. The highest number of species occurs in the central portion of the region where the climate is moderate and the topography heterogeneous.

Across ecoregions

Across all ecoregions, AET was the sole variable accounting for significant variation in species richness ($R^2 = 0.41$, $P < 0.0001$). We found that a 3rd order polynomial equation gave a better fit to the data, raising the R^2 to 49% (Figure. 3). Species richness increases significantly as AET increases to about 950 mm. yr^{-1}, flattens out between 950 - 1100 mm. yr^{-1} before decreasing. Separation of AET values by ecoregions demonstrates that the ecoregions with intermediate AET (950 – 1100), such as ecoregions 65 (Southeastern Plains), 69 (Central Appalachians), 71 (Interior Plateau), and 72 (Interior River Valleys and Hills), have the highest species richness, while ecoregion 75 (Southern Coastal Plain) with very high AET values (1214 mm. yr^{-1}) exhibits much lower species richness, equivalent to values recorded in ecoregion 58 (Northeastern Highlands) with very low AET (700 mm. yr^{-1}).

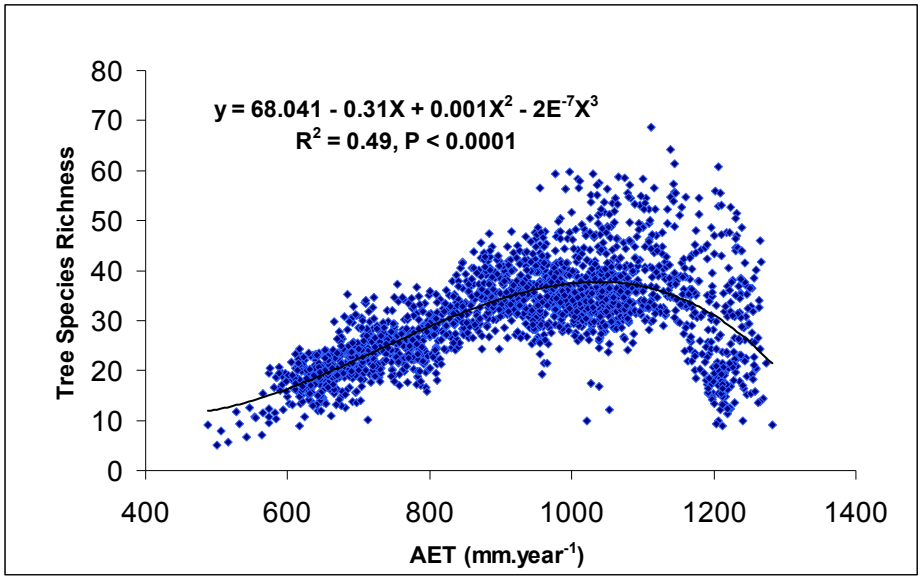


FIGURE 3: The polynomial relationship between AET and species richness across ecoregions in the eastern US. Species richness increases as AET increases to about 950 mm. yr^{-1}, plateaus from 950 mm. yr^{-1} to about 1100 mm. yr^{-1} before decreasing as AET continue to increase above 1100 mm. yr^{-1} .

Within ecoregions

The detailed analyses show that where AET is low in the northern tier of ecoreigons (i.e., < 800 mm.yr^{-1}), winter temperature is important with a few exceptions (Table 2). In ecoregions with AET between 800 – 950 mm.yr^{-1}, summer temperature is generally important. In ecoregions with AET > 1100 mm. yr^{-1}, 26% to 42% of variations in species richness were accounted for by seasonal precipitation. Only between 950 – 1100 mm.yr^{-1}, did neither AET, seasonal temperature nor precipitation account for a significant amount of variation in species richness (Table 2).

TABLE 2: Results of analysis of individual ecoregions where the most important climatic variable is identified. The variables include winter temperature (WT), summer temperature (ST), winter precipitation (WP), and summer precipitation (SP). N is the number of qualified cells within an ecoregion. Ecoregions are listed in ascending order of average annual AET ($mm.yr^{-1}$) to reflect the shift in importance of other variables as AET increases. N/A indicates that no climatic variable was found significant in an ecoregion.

Ecoregion Code	N	Average annual AET	Standard deviation of AET	R^2 with AET function	Best climatic variables	R^2 of the climatic function
49	22	603	41.9	0.36	WT	0.34
82	42	659	41.0	N/A	N/A	N/A
50	198	666	37.2	0.24	ST	0.34
58	109	700	61.2	0.31	WT	0.29
51	64	702	54.4	0.21	WT	0.37
56	23	705	37.4	0.66	ST	0.77
83	28	734	35.0	0.32	N/A	N/A
60	27	778	19.1	0.18	WT	0.42
59	31	782	16.0	N/A	WP	0.40*
52	36	791	15.9	0.22	WT	0.30
62	28	806	14.6	0.24	ST	0.37
70	71	883	33.2	0.22	ST	0.49
67	93	909	83.3	N/A	N/A	N/A
72	46	916	27.0	0.22	SP	0.46
69	64	917	44.6	0.63	ST	0.41
71	52	951	39.0	N/A	N/A	N/A
39	90	960	25.2	N/A	N/A	N/A
66	46	969	60.7	N/A	N/A	N/A
68	37	1016	27.9	N/A	N/A	N/A
45	156	1027	67.0	N/A	N/A	N/A
63	70	1065	77.9	N/A	N/A	N/A
35	84	1079	55.7	N/A	N/A	N/A
65	310	1121	85.7	N/A	SP	0.42*
75	107	1214	26.4	N/A	WP	0.26

* This is an R^2 of a negative relationship.

Figure 4 presents the most important variables selected by stepwise regression for each ecoregion. None of the nine environmental variables accounted for a significant amount of variation in species richness in those ecoregions designated in white. The relative importance of variables shifts with location. Soil water holding capacity is significant in the southern and southwestern portion of the eastern U.S. while AET is dominant in the north. Precipitation is important in the northeastern coastal regions (negative relationships) and in the northwest (positive relationships). Topographic heterogeneity accounts for a significant amount of variation in the central part of the region. In 18 out of the 24 ecoregions, environmental factors account for 17% to 66% of variation in species richness. These factors are generally weak predictors in the south (R^2s are between 0.17-0.4) but become stronger in the more centrally located ecoregions (R^2s are between 0.58 - 0.66) where topography is most varied and species richness highest.

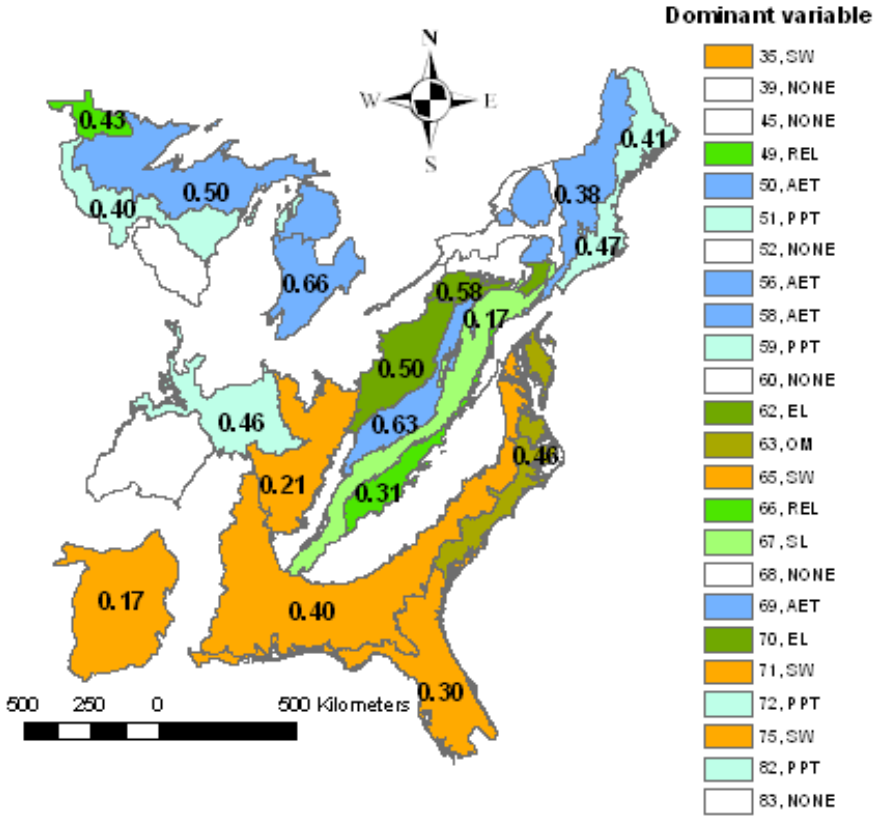


FIGURE 4: Distribution of dominant environmental variables influencing species richness within ecoregions in the eastern US. Ecoregions are color coded to reflect similar variables and labeled with R^2 values of the relationship between the dominant variable and species richness. In general, soil water holding capacity has significant predicting power in the south and southwest, AET is dominant in the north, and topographic heterogeneity account for a significant amount of variation in species richness in the central part of the region.

Discussion

Our hypothesis was not rejected. Different patterns emerge between AET and tree species richness depending on how a region is stratified in reference to AET. Our findings on the relationship between species richness and AET across ecoregions generally support those of other studies conducted at a similar scale (Charles et al. 2001, O'Brien 1993). It is generally acknowledged that AET is closely related to productivity (Lieth 1975, Law et al. 2002). The results of this study are consistent therefore with analyses of FIA survey data in the western U.S. where high species diversity was associated with intermediate levels of productivity and decreased at higher or lower values (Waring et al 2002, Swenson and Waring 2006).

A positive response of species richness to AET is generally expected at lower values of AET, as also observed in reference to Net Primary Production (NPP). At intermediate values of AET or NPP, the relationship is insensitive but in this range tree richness is generally the highest. As values of AET increases further the relation becomes negative. In our study, as in others, neither AET nor

probably productivity accounted for all the observed variation in tree richness; different variables become important depending on the location. For example, we found that winter temperature accounted for a significant amount of variations (29 - 42%) in species richness in the northern tier of ecoregions. Ecoregions 50, 56, and 83 were exceptions, possibly because of deep snow accumulates in the vicinity of large unfrozen lakes during the winter, which may protect more sensitive tree species. In the ecoregions with AET between 800 - 950 mm. yr^{-1}, species richness increased as summer temperature increased. As expected, no significant relationship between AET and specie richness was observed in ecoregions with AET between 950 - 1100 mm. yr^{-1}, where species richness peaked. We interpret the decrease in species richness for the areas with AET > 1100 mm. yr^{-1} as a likely result of a few fast-growing species dominating and restricting light available for other species (Safford et al. 2001, Waring et al. 2002, Swenson and Waring 2006, Hooper and Dukes 2004, Grime 2001).

Within ecoregions, we found that AET was a significant factor in 12 of the 15 ecoregions with AET less than 950 cm. yr^{-1} (Table 2). The variations in species richness and its environmental variables in ecoregion 59, 67 and 82 were considered as special cases. Based on the EPA's description of ecoregion 67, topographic variation is extreme, where high mountains and deep valleys occur over relatively short distances (ftp://ftp.epa.gov/wed/ecoregions/us/useco_desc.doc). As a result, the standard deviation of AET was much higher (84 mm. yr^{-1}) than in other ecoregions (Table 2). With a minimum cell size of 1000 km^2, we expect the full variation in climatic conditions were not captured. Ecoregions 59 and 82 were special cases because, although both are highly forested regions, they had many large lakes that may ameliorate water stress but, with increasing precipitation, foster cloudiness, reducing the amount of light absorbed by the forest canopy, while increasing competition on species adapted to shaded conditions. In ecoregions with AET > 1100 mm. yr^{-1} (i.e. ecoregions 65 and 75) species richness was positively associated to soil water holding capacity (Figure 4). This positive relationship may be a result of large areas of coarse textured soils dominated by pine-oak forests. The explanatory power of soil water holding capacity was between 30 - 40%. The detailed analysis indicated that winter precipitation accounted for 26% of the variation in species richness in ecoregion 75 whereas summer precipitation was negatively related to species richness in ecoregion 65 ($R^2 = 0.42$) (Table 2). A possible explanation of these inverse relationships is that winter precipitation might promote species richness in ecoregion 75 if soil moisture were limiting in the spring whereas summer precipitation might increase productivity and, competition leading to reduced species richness in ecoregion 65.

For this detailed environmental analysis within and across ecoregions we took advantage of both variable- and fixed-radius FIA sampling. As a result of cross-comparison between the two data sets, we have confidence in our ability to utilize variable-area data sets in the western U.S. where fixed-radius sampling is as yet limited. The detailed analysis also indicates that in the search for general responses between environmental variables (or productivity) that the EPA

ecoregion classification level III may be too refined and that a broader stratification may be warranted. Future studies will test this assumption.

Acknowledgement

This study was supported by the National Aeronautics and Space Administration (NASA Grant NNG04GK26G) as part of the Biodiversity and Ecological Forecasting program. We thank Drs. Nicholas Coops and Joanne Nightingale for their helpful suggestions and editing.

Reference

Alerich, C. Klevgard, L. Liff, C. Miles, P. D. and Knight, B. 2004. Online. The Forest Inventory and Analysis Database Description and Users Guide Version 2.0. Available: http://ncrs2.fs.fed.us/4801/fiadb/fiadb_documentation/FIADB_v20_060105_draft.pdf

Caley, M. J. and Schluter, D. 1997. The relationship between local and regional diversity. - -- Ecology 78:70-80.

Charles, H. Godfray, J. and Lawton, J. H. 2001. Scale and species numbers. - TRENDS in Ecology & Evolution 16:400-404.

Churkina, G. Running, S. W. and Schloss, A.L. 1999. Comparing global models of terrestrial net primary productivity (NPP): the importance of water availability. - Global Change Biology 5:46-55.

Currie, D.J. and Paquin, V. 1987. Large-scale biogeographical pattern of species richness of trees. - Nature 329:326-327

Currie, D. J. Francis, A. P. and Kerr, J. T. 1999. Some general propositions about the study of spatial patterns of species richness. - Ecosciences 6(3):392-399

Davies, T. J. Barraclough, J. G. Savolajnen, V. Chase, M. W. 2004, Environmental causes for plant biodiversity gradients. Philosophical Transactions of the Royal Society of London. Series B, - Biological Sciences 359:1645-1656

Grime, J. P. 2001, Plant Strategies, Vegetation Process, and Ecosystem Properties. Second Edition. - John Wiley & Sons, New York

Hawkins, B. A, Field, R. Cornell, H. V. Currie, D. J. Guegan, J. Kaufman, D. M. Kerr, J. T. Mittelbach, G. G. Oberdorff, T. O'Brien, E. M. Porter, E. E. and Turner, J. R. G. 2003. Energy, water, and broad-scale geographic patterns of species richness. - Ecology 84:3105-3117.

Hooper, D. U. and Dukes, J. S.. 2004. Overyielding among plant functional groups in a long-term experiment. - Ecology Letters 7: 95-105

Hurlbert, A. H. and White, E. P. 2005. Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. - Ecology Letters 8:319-327.

Huston, M. A. 1999. Local processes and regional patterns: appropriate scales for understanding variation in the diversity of plants and animals. - Oikos 86:939-401

Iverson, L.R. and Prasad, A. M.. 2001. Potential changes in tree species richness and forest community types following climate change. - Ecosystems 4:186-199.

Latham, R.E. and Ricklafs, R.E. 1993. Global patterns of tree species richness in moist forest: energy-diversity theory does not account for variation in tree species richness. - Oikos 67:325-333.

Law, B. E., Falge, E., Gu, L., Baldocchi, D. D., Bakwin, P., Berbigier, P., Kavis, K., Dolman, A. J., Falk, M., Fuentes, J. D., Golstein, A., Granier, A., Grelle, A., Hollinger, D., Janssens, I. A., Jarvis, P. G., Jensen, N. O., Katul, G. ,Mahli, Y., Matteucci, G., Meyers, T., Monson, R., Munger, W., Oechel, W., Olson, R., Pilegaard, K., Paw, K. T., Thorgeirsson, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S. 2002. Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation. Agricultural and Forest Meteorology 113:97-120.

Lieth, H. 1975. Modelling the primary productivity of the world. pp237-263. In H. Lieth and R. H Whittaker (eds.), Primary Productivity of the Biosphere. - Springer-Verlag, New York.

McRoberts, R. E., Holden, G. R. Nelson, M. D. Moser, W. K. Lister, A. J. King, S. L. LaPoint, E. B. Coulston, J. W. Smith, W. B. and Reams, G. A. 2005. Online. Estimating and circumventing the effects of perturbing and swapping inventory plot locations. Available: http://ncrs2.fs.fed.us/4801/fiadb/fiadb_documentation/Perturbing-Swapping.pdf

O'Brian, E. M. 1993. Climate gradients in woody plant species richness: Towards an explanation based on an analysis of southern Africa's woody flora. - Journal of Biogeography 20: 181-198

Prentice, I. C., Sykes, M. and Cramer, W. 1993. A simulation model for the transient effects of climate change on forest landscapes. - Ecological Modelling 65:51-70

Sarr, D. A., Hibbs, D. E. and Huston, M. A. 2005. A hierarchical perspective of plant diversity. - The Quarterly Reviews of Biology 80:187-212.

Swenson, J. J., and Waring, R. H. 2006. Modelled photosynthesis predicts woody plant richness at three geographic scales across the north-western United States. - Geog. Ecol. And Biodiversity (*in press*).

U.S. Forest Service, 2004. Forest Inventory and Analysis National Core Field Guide Volume 1: Field Data Collection Procedures for Phase 2 Plots (Version 2.0)

USDA National Soil Survey Center, 1994. State Soil Geographic (STATSGO) Data Base: Data use information, Miscellaneous Publication Number 1492

USGS 2002. [CD-ROM]. Global GIS Database Verson 6.2. Available: http://webgis.wr.usgs.gov/globalgis/

Waring, R. H. Coops, N. C. Ohmann, J. L. and Sarr, D. A. 2002. Interpreting woody plant richness from seasonal ratios of photosynthesis. - Ecology 83(11):2964-2970

Waring, R. H. Coops, N. C. Fan, W. and Nightingale, J. N. 2006. MODIS enhanced vegetation index predicts tree species richness across forested ecoregions in the contiguous U.S.A. - Remote Sensing of Environment (in press).

Whittaker, R. J. and Field, R. 2000. Tree species richness modeling: an approach of global applicability. - Oikos 89: 399-402.

Whittaker, R. J. Willis, K. J. and Field, R. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. – Journal of Biogeography 28:453-470.

Landscape-Level Analyses II



Summary and Findings of the 2006 BLM Forest Lands Report

Tim Bottomley[1] and Jim Menlove[2]

***Abstract**: In 2006, the Bureau of Land Management (BLM) contracted with the Forest Service Forest Inventory and Analysis Program (FIA) to assist in the preparation of a report specific to all forest lands under the administration of the BLM. The BLM requested that the FIA provide information on the extent and general conditions of BLM-managed forests and woodlands, within certain statistical reliability parameters.*

The report focused on the 33 million acres of forest land managed by the BLM outside of Alaska by using data gathered from more than 2,500 FIA plots. The analysis included information on the extent of forest land (total acres of occurrence and general location), trees per acre, stand size, stand age, volume and biomass, basal area, stand density index (SDI), and number of snags (standing dead trees). Specific findings include:

- *An increase in BLM forest acres from 55 million, as determined in a 2005 report, to 69 million today.*
- *The high tree densities in many forest types*
- *The expansion of pinyon and juniper woodlands in historical nonforest areas*
- *The unbalanced age class distributions in the aspen and lodgepole pine forest types*
- *The differing sampling intensities in the States*

An upcoming BLM publication will highlight some specific data findings of interest and provide some interpretations of the information to BLM field offices for land use planning.

Keywords: Forest, inventory, FIA, BLM.

[1] United States Department of the Interior–Bureau of Land Management; National Operations Center; Bldg. 50; Denver Federal Center; Denver, CO 80225-0047 USA; e-mail: tim_bottomley@blm.gov

[2] USDA Forest Service; Rocky Mountain Research Station; Interior West FIA; 507 25th Street; Ogden, UT 84401 USA; e-mail: jmenlove@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

Background

The Bureau of Land Management (BLM), an agency within the U.S. Department of the Interior (DOI), administers more than 253 million surface acres[3] of public land in the western United States, including Alaska. About 69 million acres, or 27%, are classified as forested.

The BLM manages these forest lands according to the principles of multiple use and sustained yield as required by the Federal Land Policy and Management Act of 1976 and the Oregon and California Railroad Act of 1937, which covers forest lands in western Oregon. National priorities for these forests include maintaining and restoring forest health, salvaging dead and dying timber, providing high-quality wildlife and fish habitat, and providing economic opportunities in rural communities by making timber and other forest products, including biomass, available from vegetation management treatments.

The last nationwide inventory taken of BLM forest lands occurred in 2001 and only addressed acres of forest land in the predominant forest types; it did not address any condition of these forest lands. In spring 2006, the BLM contracted with the U.S. Department of Agriculture's Forest Service (FS) Forest Inventory and Analysis Program (FIA) to prepare the *BLM Forest Lands Report –2006* (hereinafter referred to as the "*Report*"; Bottomley and Menlove 2006), specific to forest lands under the administration of the BLM. According to the BLM, the FIA program would provide data that are the best combination of availability, consistency, accuracy, and comprehensiveness for BLM forest lands.

The BLM desired information related to the extent and general conditions, with statistical reliability measures, of the forest lands under its jurisdiction. The *Report* did not separate BLM forest lands by resource management objective or legal status (e.g., congressionally designated wilderness areas). While some of the terminology used in the *Report* reflected the early history of the FIA program focusing on wood supply (e.g., timberlands), no suitability or availability of BLM forest lands for commercial wood products were implied in the *Report*.

Figure 1 illustrates the broad distribution of BLM forest lands within 13 conterminous western States and Alaska. Figures 2a and 2b illustrate two extremes in BLM land status patterns, large forested landscapes and small isolated forest tracts. These land status situations, particularly broad distribution of BLM forest lands and the small isolated forest tracts, present challenges with data consolidation and analysis efforts.

[3] The BLM Forest Lands Report–2006 reported a total of 261 million acres according to the most recent data at that time. According to the 2008 BLM Public Land Statistics report, the number of surface acres managed by the BLM was reduced to 253 million because of land tenure changes.

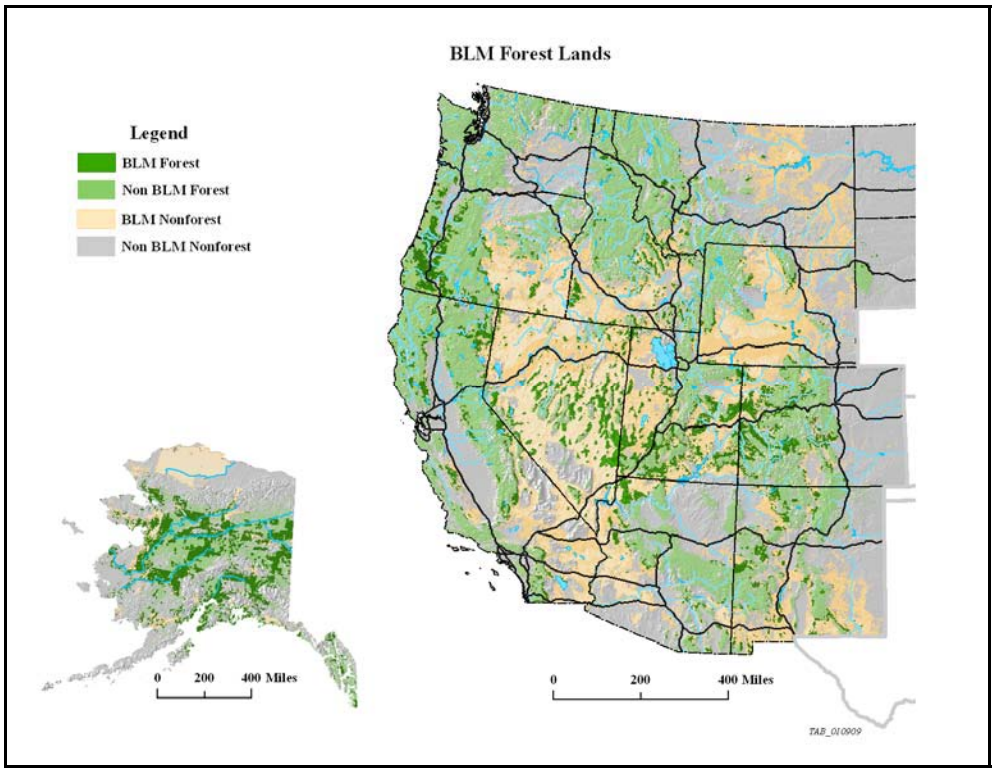


Figure 1: Map of BLM forest lands.

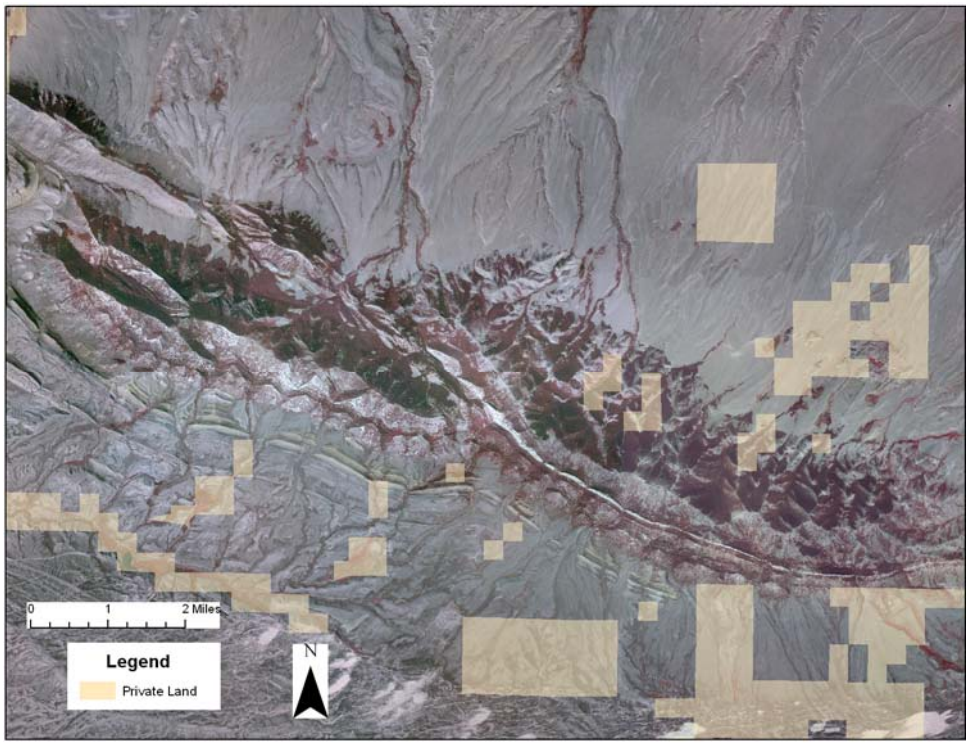


Figure 2a: Aerial photo of an example large landscape of BLM forest lands (Ferris Mountain Range, Wyoming).

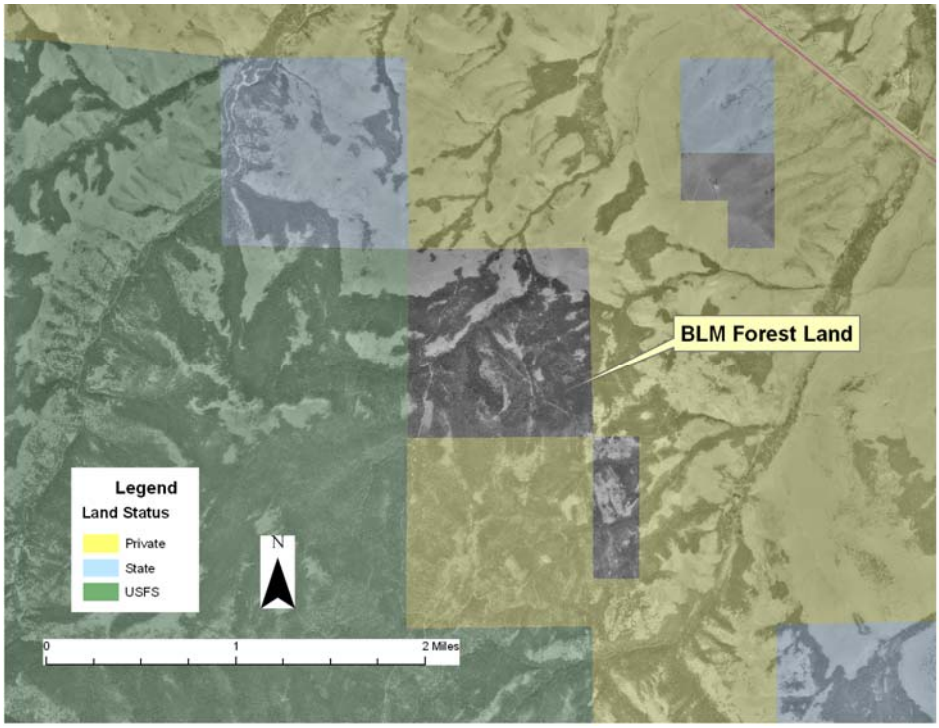


Figure 2b: Map of an example small forested BLM land tract (Fringe of Medicine Bow National Forest, Wyoming).

Data Consolidation Techniques

Because it was necessary to acquire data from several FIA units, two major methods were used. First, data from States within the Interior West FIA unit (IWFIA) were queried from the local IW-FIA FIADB version 1.0 database, by using the most recent inventories through 2005. Second, data from States in the Pacific Northwest and North Central FIA units were downloaded from the public DataMart FIADB version 2.1 data download site (since replaced by FIADB version 3.0 DataMart) by using the 2004 inventory year. When calculating standard errors, the local IW-FIA version 2.1 database was used for Interior West States. Queries from the different database structures were carefully constructed so that the results were compatible, and could be compiled in Microsoft Access databases and Microsoft Excel pivot tables.

In addition to the standard FIA variables that were used, several additional variables were calculated. For forest type, an alternate was used where, if the forest type was "nonstocked," the field forest type was assigned. In this way, the "nonstocked" stand-size class could be assigned to different forest types. Also, a stand density index (SDI) was calculated for each plot condition using the summation method and compared using the summation SDI maximum values for forest types as described in Shaw (2000). This also allowed the calculation of

condition-based trees-per-acre values, which were then used to construct classes by trees-per-acre.

Report Format

BLM forest lands were classified by forest type. The ten most common forest types were then analyzed for the following: extent (total acres of occurrence and general location), trees per acre, stand size, stand age, volume and biomass, basal area, SDI, and number of snags (standing dead trees) given. The *Report* was organized to discuss each of these above items in the above order. An additional section dealt with individual States and discussed the following items: forest land by type, years of data collection, volume estimates, gross growth and mortality rates (IWFIA only) and causes of mortality (IWFIA only). Standard errors were calculated for acres of forest and net live tree volume.

Alaska

BLM forest lands in Alaska were not included in the analysis of plot data because FIA plot data have only been collected in coastal Alaska, whereas the BLM manages large areas of forest land in the Alaskan interior. However, an assessment of BLM forest lands in the Alaskan interior was conducted for the *Report* through remote sensing by using a map developed by the USDA Forest Service–Forest Inventory and Analysis Program & Remote Sensing Applications Center (RSAC; Ruefenacht et al. 2008). With the RSAC map, the IWFIA was able to estimate that there were about 36 million forested acres in Alaska on BLM lands in 14 different forest cover types. However, the lack of plot data prevented including the BLM forest lands in Alaska in the majority of the following analysis.

Findings

On the basis of 2,521 FIA plots, 48 different forest types were reported on BLM forest land in the contiguous United States. The 10 most common types discussed in the *Report* account for 93% of the forest land acres across all States and the majority of forest land in each of the States. Table 1 shows the estimated acres of these 10 forest types.

Table 1: Area estimates for the 10 most common and all other forest types on BLM lands in the contiguous United States.

Forest type		Acreage
Pinyon-Juniper Woodland		16,034,527
Juniper Woodland		6,152,507
Douglas-fir (all)		2,653,015
Coastal Douglas-fir	(1,496,756)	
Noncoastal Douglas-fir	(1,156,259)	
Western Juniper		1,947,558
Ponderosa Pine		1,072,133
Deciduous Oak Woodland		964,765
Cercocarpus Woodland		642,001
Aspen		367,837
Lodgepole Pine		359,306
All other forest types		2,431,203
Totals		32,624,853

Major findings from the *Report* include: (1) a large increase in BLM forest acres from estimates in previous reports, (2) high tree densities in many forest areas, (3) expansion of pinyon and juniper species in historical grass and shrub lands, (4) unbalanced age class distribution in the aspen and lodgepole pine forest types, and (5) the differing sampling intensities by State. Each of these findings is discussed in the following sections.

Large Increase in Acreage

In 2005, the BLM had estimated the total forest land under its jurisdiction at 55.1 million acres, according to 2001 data. The *Report* provided a new estimate of 69.1 million acres, a 25% increase. Table 2 provides a comparison, by State, of these two estimates.

Table 2: BLM Forest Lands. Comparison of 2005 and 2006 estimates. (Note that all acreages are in thousands.)

State	2005 Acreage	2006 Acreage
Alaska	28,279	36,447
Arizona	1,074	1,893
California	2,208	1,449
Colorado	4,110	5,076
Eastern States	0	0
Idaho	892	945
Montana	810	1,325
Nevada	6,274	7,831
New Mexico	985	1,121
Oregon	3,341	3,789*
Utah	6,073	7,825
Washington	50	79
Wyoming	1,004	1,290
Totals	**55,100**	**69,070**

Three key factors account for the majority of this 14 million acre increase.

- Attaining better remote sensing data, particularly in Alaska, resulted in an increase of total forest land from 28.3 million acres to 36.4 million acres, an increase of more than 8 million acres.

- The 2005 estimate relied heavily on FIA estimates for the 2002 RPA, which used a sampling protocol of 10% crown cover for woodlands instead of the 5% crown cover protocol used in the data analysis of the *Report*. Reducing the percent crown cover requirement to 5% resulted in an estimated increase of about 3.3 million acres land classified as forest.

- Changing methodologies in the FIA program (e.g., minimum criteria standards for heights and diameters and appropriate species) over the years have resulted in including acreages in several forest types that were not in the above-mentioned RPA, but are included in the *Report*. These include a large portion of the deciduous oak woodland (1.0 million acres), *Cercocarpus* woodland (0.6 million acres), and mesquite (0.3 million acres). These methodology changes resulted in an estimated increase of approximately 2 million acres.

Future BLM forest land calculations are anticipated to result in a reduction from the 69 million acres, primarily due to the ongoing land tenure adjustments in Alaska. Additionally, changes in the FIA program sampling and measurement protocols (particularly those related to species and minimum crown cover criteria) will also cause either increases or decreases in the calculation of BLM forested acres. For example, elimination of lands with 5–9% forest crown cover would reduce the total forest land base by about 10%.

Although there is a large increase in the estimated amount of BLM forest land, little or no effect on the BLM's Forest and Woodland Management program is anticipated. Whether BLM lands are classified as forest or nonforest has no significance regarding annual appropriations from Congress. Nor does this classification automatically allow or restrict any particular use activity (such as timber harvesting). Decisions on appropriate management activities that may occur on any of these lands are made through individual land use plans and are not influenced by classification as forest or nonforest by the FIA.

High SDIs For Many Tree Species

The *Report* dealt with each of the 10 most common forest types on BLM land and determined the range of and average of trees per acre and basal area per acre. However, given the variety of tree species involved, as well as the variety of stand ages, Stand Density Index (SDI) was chosen as the metric to evaluate tree density. Therefore, the range and average SDI for each of these forest types was determined. Table 3 summarizes SDI information in the *Report* by forest types.

Table 3: Stand Density Index information on BLM lands for major forest types.

Forest type	Total acres (thousands)	Percentage of acres				Acres of Concern (thousands)
		<25% SDI_{max}	25–35% SDI_{max}	>35% SDI_{max}	>60% SDI_{max}	
Pinyon-Juniper	16,035	32	15	52	21	8,338
Combined Juniper	6,153	58	11	31	11	1,907
Coastal DF	1,497	10	6	85	40	1,272
Noncoastal DF	1,156	33	25	42	14	486
Western Juniper	1,948	79	12	9	2	175
Ponderosa Pine	1,072	50	17	33	16	354
Deciduous Oak	965	71	12	17	6	164
Cercocarpus	642	42	17	41	16	263
Aspen	368	51	20	30	13	110
Lodgepole Pine	359	28	4	69	14	248
Totals (includes rounding errors)	**30,195**					**13,318**

One major finding from the data was that more than 40% (13.3 of 32.6 million acres) of the BLM forested lands in the contiguous 48 States are at, or are exceeding, full site occupancy (greater than 35% of SDI_{max}). This high tree density can lead to massive die-offs in forested landscapes when droughts create a shortage of water and too many trees are competing for that resource. The intense competition for water can also predispose these trees to insect and disease predation as their ability to fight off these damaging agents is reduced because of water stress.

Expansion of Pinyon and Juniper Forest Types

Another issue on many BLM lands is the expansion of coniferous trees into what are considered historical grass and shrub lands. For a number of reasons, such as lack of fire, livestock grazing, or cyclic wet periods, there has been an evident trend of trees establishing and growing in areas that, as determined by earlier historical photography, were grass or shrub lands. While the FIA does not directly measure expansion by trees into nonforested areas, the intent of the *Report* was to provide information that may be helpful in understanding the extent of this vegetation change on BLM lands.

The pinyon–juniper, combined juniper, and western juniper forest types seldom exist in even-aged stands because these tree species tend to accumulate gradually on a site. As a result, stands in these forest types often have individual trees that are much older, as well as much younger, than the age determined for a specific stand. However, it may be reasonable to assume that stands that are recently established in otherwise shrub and grassland ecosystems should contain no trees significantly older than the trees present. Therefore, the percentages in each of these forest types in the younger age classes that have no recorded live-tree age of more than 150 years should give an approximation of the degree to which the pinyon and juniper stands could be considered expansion since the beginning of European settlement in the West. Conversely, those stands with trees present that are 150 years old and older give an approximation of the extent of historical woodlands.

From the FIA data analyzed, of the 24 million acres of BLM forest land in either a pinyon or juniper forest type, approximately 12.1 million acres (50%) have both a stand age of 150 years or younger and contain no trees older than 150 years. This indicates that a substantial portion of pinyon and juniper forests may be expanding into historical shrub or grasslands, although unknown portions are regeneration of disturbed stands.

The remaining 50% of the pinyon and juniper forest type has at least some trees that existed on these sites before the interruption of disturbance processes that keep many arid areas in shrub and grass species, including the exclusion of fire. These stands could therefore potentially be considered historical woodlands; however, in many instances they may have been less dense in the past.

Unbalanced Age Class Distribution Concerns with Aspen and Lodgepole Pine Forest Types

The FIA inventory methodology was not considered adequate to provide accurate measurements of recent outbreaks in stand mortality from insects and disease on BLM forest lands, particularly since at least two States (New Mexico and Wyoming) have had no FIA data since 2001. These States were not added to the annual inventory system (see next section for more information) at the time of

this report. Additionally, two States (Idaho and Nevada) had only two annual inventory cycles. However, collected data, even if a little dated, can provide very useful information related to the risk of stands to pending mortality, particularly if the forest stands are reaching the end of their usual life expectancy. For example, two of the major forest types evaluated—aspen and lodgepole pine—have relativity short life spans of 150 (Howard 1996) and 150–200 years (Anderson 2003), respectively. Additionally, both forest types have a usually natural even-aged stand development habit, which places added focus on stand age, because most of the trees in these even-aged stands are approximately the same age.

The extent of aspen decline and die-off on BLM lands was not evaluated in the *Report*. However, a predisposing condition (stand age) was evaluated. Figure 3 shows the present age class distribution of BLM aspen stands. From these data, it is apparent that a large percentage (50%) of the aspen forests are at or approaching life expectancy. These data indicate that the majority of BLM aspen stands are at or approaching a high risk from either aspen decline or aspen die-off.

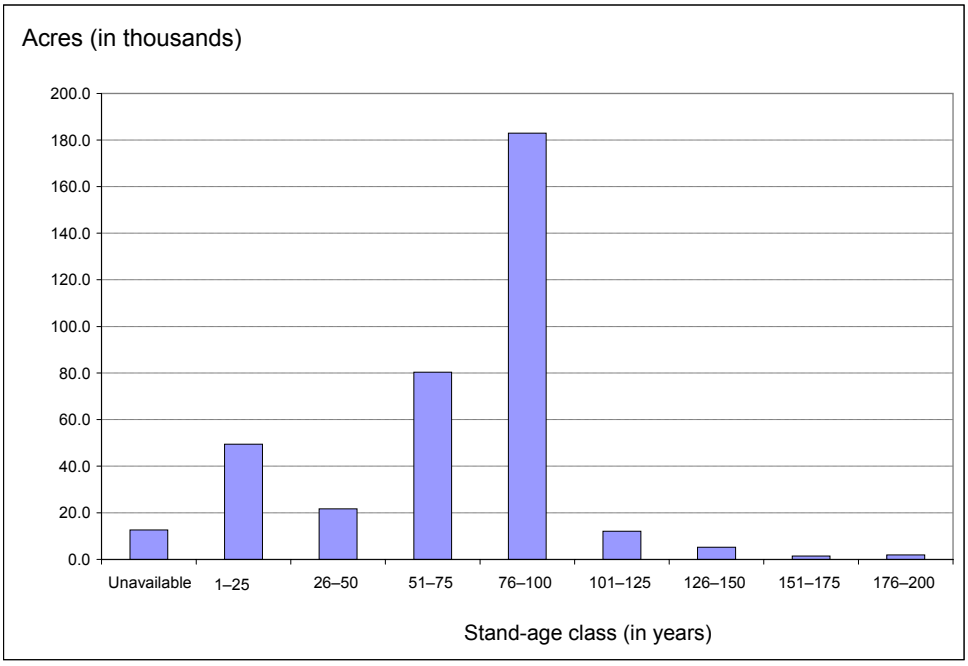


Figure 3: Area of aspen forest by stand-age class, BLM land.

The lodgepole pine forest cover type has a similar unbalanced age class distribution, as shown in Figure 4. This condition could also predispose these stands to large-scale mountain pine beetle mortality.

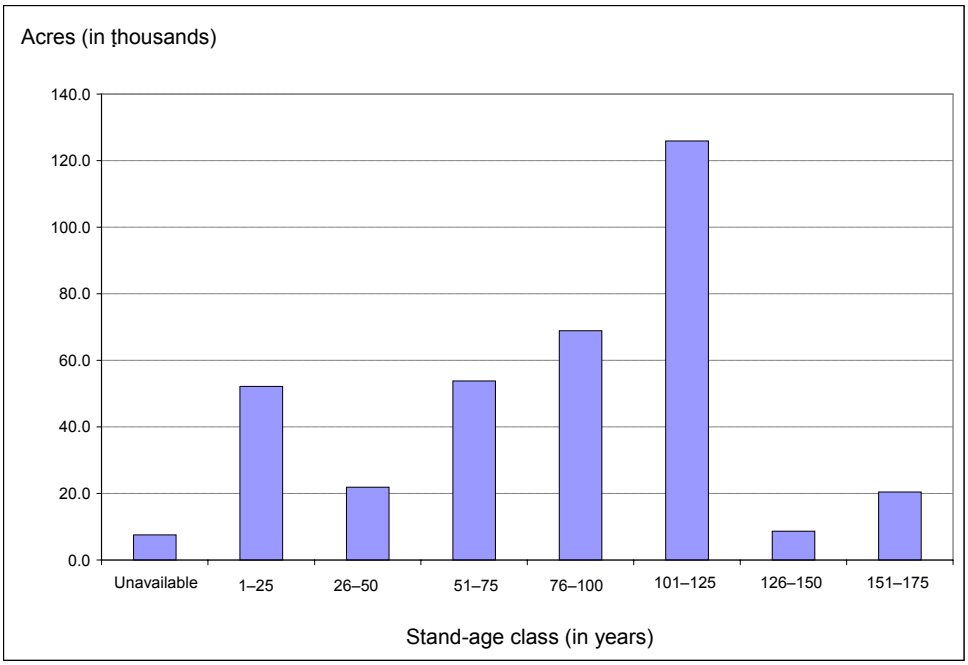


Figure 4: Area of lodgepole pine forest by stand-age class, BLM land.

Differing Sampling Intensities by State

Historically, the FIA has collected inventory data on a State-by-State basis, completing regionally based statewide inventories covering the entire sampling grid for a given State every 7 to 20 or more years. These historical inventories were known as periodic inventories. Beginning in the mid-1990s, the FIA program began making a transition from periodic to annual inventories (Gillespie 1999). The annual inventory samples an evenly distributed 10% to 20% of the sample grid (an annual panel) in each State every year, so that every State's grid is completed every 5 to 10 years. The annual system is better able to detect changes and trends, and efforts are continuing to establish nationally consistent standards for data collection, compilation, and reporting.

As the annual inventory methods were adopted, States were gradually added to the annual system, often while the most recent periodic surveys were being completed in other States. The result is that until all States have at least one full cycle of annual data, land managers with forest lands in more than one State (such as the BLM) will have data where plot intensities and time spans differ from one State to another.

Table 4 provides information on the number of annual cycles each State has had through 2006 and the number of forested acres each FIA plot represents. Percent standard error is provided for the acres of forest land in each State. The percent standard error primarily reflects the acres per plot, with generally lower

values for acres per plot also having low percent standard errors. The major exceptions to that trend are States with few forested plots on BLM land (e.g., North Dakota, South Dakota, and Washington), where the few total plots in the respective States (3, 3, and 5) result in very high percent standard errors (74.63%, 55.72%, and 47.47%). As the number of annual panels increases in each State, percent standard errors will decrease.

Table 4: FIA plots and forest conditions on BLM land at the time of the Report.

State	Total plots	Plots w/ forest conditions	Forested Acres	Acres per plot	Annual Panels	Percent standard error of acres
Arizona	1,192	161	1,893,000	11,758	5	6.98
California	998	104	1,449,000	13,933	4	8.98
Colorado	578	338	5,075,000	15,015	4	4.46
Idaho	412	36	945,000	26,250	2	17.05
Montana	398	75	1,290,000	17,200	3	11.94
Nevada	1,639	293	7,831,000	26,727	2	4.61
New Mexico	2,085	173	1,121,000	6,480	Periodic	6.15
North Dakota	*	3	10,369	3,456	4	74.63
Oregon	899	267	3,789,000	14,191	4	4.38
South Dakota	*	3	25,394	8,465	4	55.72
Utah	2,311	833	7,825,000	9,394	6	2.82
Washington	9	5	79,000	15,800	3	47.47
Wyoming	2,879	230	1,290,000	5,609	Periodic	6.04

* NC-FIA does not assign ownership data to nonforest conditions; therefore, the total number of plots (and the amount of nonforest land) on BLM land is unknown.

In addition, although most of the available data are consistent between FIA units, some variables and summaries of interest were not available in all States or were not completely consistent between States. These different methodologies have direct effects on the *Report*, but predominately on a State basis, and are discussed in more detail in the *Report*.

Next Steps for BLM

The intent of the *Report* was strategic—to give the BLM a broad perspective about the condition of forest lands under its jurisdiction. The *Report* was not intended to be used as the sole source for advocating any particular vegetation treatment or change of condition. Decisions on vegetation management objectives are made at the local level through the BLM's land use planning processes (e.g., resource management plans). The BLM is preparing a publication (Bottomley and Menlove 2009) that will discuss the findings and provide assistance to agency field offices for looking at these potential issues (e.g., high tree densities, pinyon and juniper expansion and densification, and age class distribution issues for aspen and lodgepole pine) raised in the *Report* and determine if these issues are applicable in their specific areas of interest.

Conclusion

FIA data provided an excellent source of information to the BLM for making general statements concerning the condition of the forest lands under its jurisdiction. These general statements, while not site-specific enough to determine specific management activities, are useful for guiding strategic level decisions for the agency.

Literature Cited

Anderson, Michelle D. 2003. Pinus contorta var. latifolia. *In*: Fire Effects Information System, [Online]. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (Producer). Available: http://www.fs.fed.us/database/feis/ [December 8, 2008].

Bottomley, T., and J. Menlove. 2006. BLM Forest Lands Report—2006: Status and Condition. Bureau of Land Management, Denver, Colorado. BLM/ST/ST-07/0001+5000. 111 p.

Bottomley, T. and J. Menlove. 2009. Technical Note 425: Summary and Findings of the 2006 BLM Forest Lands Report. Bureau of Land Management, Denver, Colorado. In press.

Gillespie, A.J.R. 1999. Rationale for a national annual forest inventory program. Journal of Forestry. 97(12): 16–20.

Howard, Janet L. 1996. Populus tremuloides. *In*: Fire Effects Information System, [Online]. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (Producer). Available: http://www.fs.fed.us/database/feis/ [December 8, 2008].

Ruefenacht, B., M.V. Finco, M.D. Nelson, R. Czaplewski, E.H. Helmer, J.A. Blackard, G.R. Holden, A.J. Lister, D. Salajanu, D. Weyermann, K. Winterberger. 2008. Conterminous U.S. and Alaska Forest Type Mapping Using Forest Inventory and Analysis Data. Photogrammetric Engineering & Remote Sensing. Vol. 74(11): November 2008. pp. 1379–1388.

Shaw, J.D. 2000. Application of stand density index to irregularly structured stands. Western Journal of Applied Forestry 15:40–42.

Tree Migration Detection Through Comparisons of Historic and Current Forest Inventories

Christopher W. Woodall[1], Christopher M. Oswalt[2], James A. Westfall[3], Charles H. Perry[4], Mark D. Nelson[5]

***Abstract**: Changes in tree species distributions are a potential impact of climate change on forest ecosystems. The examination of tree species shifts in forests of the eastern United States largely has been limited to modeling activities with little empirical analysis of long-term forest inventory datasets. The goal of this study was to compare historic and current spatial distributions of tree species for sets of northern and southern tree species in the eastern United States using region-wide forest inventories. Based on the results of this study, no conclusions could be drawn about tree migration in the eastern United States. The technique of comparing outer ranges of tree species based on periodic forest inventories may be confounded by inconsistent forest inventory methods across time and space along with tree species identification measurement error. It is suggested that novel tree migration detection methods be developed based on contemporary forest inventories that are consistent across space and time.*

Keywords: Climate change, tree migration, forest inventory

Tree Species Migration

The world's climate is forecasted to change significantly over the next century due to a doubling of pre-industrial atmospheric carbon dioxide concentrations resulting in an increase in mean surface temperatures of 2 to 4.5 degrees C, more episodic precipitation events, and longer growing seasons (IPCC 2007). Climate is an important driver of forest ecosystem functions (Stenseth et al. 2002). Thus, changes in climate should change forest ecosystem attributes and functions.

[1]Research Forester, U.S. Department of Agriculture, Forest Service, Northern Research Station, St. Paul, MN 55108; fax: 651-649-5140, phone: 651 649-5141, e-mail: cwoodall@fs.fed.us.
[2]Research Forester, U.S. Department of Agriculture, Forest Service, Southern Research Station, Knoxville, TN
[3]Research Forester, U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA
[4,5] Research Forester, U.S. Department of Agriculture, Forest Service, Northern Research Station, St. Paul, MN

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Fitness of trees is expected to be impacted by changes in absolute temperatures and the timing and/or magnitude of precipitation events (Saxe et al. 1998, Nabuurs et al. 2002, Sacks et al. 2007), along with a higher probability of catastrophic wildfires in regions of the United States (Westerling et al. 2006). These effects on individual tree fitness are forecasted to subsequently affect tree response to stress agents such as insects and disease (Volney and Fleming 2000; Logan et al. 2003). The culmination of climate change effects on forest ecosystems ultimately may be the migration of tree species (Opdam and Wascher 2004, Walther et al. 2002). There is evidence of past forest migration rates exceeding 50 km per century during episodes of climate change (Schwartz 1992, Noss 2001, Parmesan and Yohe 2003). Currently, forests may need to migrate one order of magnitude faster than in past migrations to adequately respond to current rates of warming (Schwartz 1992). However, modern day fragmentation of forest ecosystems may slow the movement of tree species, potentially reducing tree migration capacity by one order of magnitude (Schwartz et al. 2001, Davis and Shaw 2001, Walther et al. 2002, Opdam and Wascher 2004). Given the substantial implications of climate change impacts on tree species distributions within a relatively short period of time, the detection of tree species migration is critical.

Examination of tree species migration has mainly focused on investigating historic ranges during the past millennia (e.g., Davis and Shaw 2001, Malcolm et al. 2002, McLachlan et al. 2005, Pearson 2006) and simulating future tree species shifts (e.g., Schwartz et al. 2001, Iverson and Prasad 1998, Iverson et al. 1999, Malcolm et al. 2002, McCarty 2001, Iverson et al. 2008). These studies have been invaluable not only for raising awareness about climate change impacts on forest ecosystems, but also for highlighting knowledge gaps. However, holistic assessment of these climate change effect models continues to call for refinement of modeling techniques with little or no empirical validation of these models with current data (e.g., Botkin et al. 2007). Therefore, techniques need to be developed for validating extensive simulations of potential tree species shifts, which are based on poorly understood tree migration dynamics (Malcolm et al. 2002). Remote-sensing products and field-based forest inventories provide data for monitoring forest attributes across large regions. Unfortunately, remote-sensing products are not well suited for identifying individual tree species across large geographic extents, especially in the understory. The alternative is to use forest inventories to track geographic ranges of tree species over a period of decades.

Given the need to monitor the possible migration of tree species across the United States, the goal of this study was to compare tree locations for selected study species using both the oldest and most current digital forest inventories (stored in digital format) across states in the eastern United States. The study had three objectives: 1) for a selected list of northern tree species, compare the minimum latitude by species in year one (oldest inventory) and year two (most current inventory); 2) for a selected list of southern tree species, compare the maximum latitude by species in year one and year two; and 3) discuss the hurdles in using historic and annual forest inventories to monitor tree species migration.

Historic and Current Forest Inventories

During the 1930s, the U.S. Forest Service Forest Inventory and Analysis (FIA) program was charged by Congress to "make and keep current a comprehensive inventory and analysis of the present and prospective conditions of and requirements for the renewable resources of the forest and rangelands of the United States" (McSweeney-McNary Act of 1928) (Gillespie 1999, Frayer and Furnival 1999, Bechtold and Patterson 2005). During most of the 20th century, FIA was the primary source for information about the extent, condition, status, and trends of forest resources across all ownerships in the United States (Smith 2002). However, the national inventory of forest land was conducted only periodically, using sample designs and data management systems that varied by state and inventory period (Gillespie 1999). A variety of plot-level sample designs were used: fixed-radius to variable-radius, clusters of 4 to more than 10 sub-plots, and differing measurement protocols by individual variables (e.g., tree height or length). Additionally, remeasurement periods ranged from between 7 years to more than 20 years. The strategic-scale paradigm of varying sample designs, methods, and dates often confounded regional, cross-state forest resource analyses and digital data management -- two important aspects germane to 21st century analyses.

An annual forest inventory was initiated in 1999 by the FIA program. FIA now applies a nationally consistent sampling protocol using a quasi-systematic design covering all ownerships in the entire Nation (Bechtold and Patterson 2005). A three-phase inventory is now implemented, based on an array of hexagons assigned to separate interpenetrating, non-overlapping annual sampling panels (Bechtold and Patterson 2005). In phase 1, land area is stratified using aerial photography or classified satellite imagery to increase the precision of estimates using stratified estimation. Remotely sensed data may also be used to determine if plot locations have accessible forest land cover (Bechtold and Patterson 2005). In phase 2, permanent fixed-area plots are installed in each hexagon when field crews visit plot locations that have accessible forest land. Field crews collect data on more than 100 variables, including land ownership, forest type, tree species, tree size, tree condition, and other site attributes (e.g., slope, aspect, disturbance, land use) (Smith 2002, USDA Forest Service 2008). Plot intensity for phase 2 measurements is approximately one plot for every 2,428 ha of land (125,000 plots nationally). Briefly, the plot design for FIA inventory plots consists of four 7.2-m fixed-radius subplots spaced 36.6 m apart in a triangular arrangement with one subplot in the center. All trees with a diameter at breast height of at least 12.7 cm are inventoried on forested subplots. Within each subplot, a 2.07-m microplot offset 3.66 m from subplot center is established where all trees with a d.b.h. between 2.54 and 12.7 cm are inventoried.

Data and Methods

Two sets of 18 predominantly northern and southern tree species were selected based on species range maps developed by Little (1971); these sets

generally coincided with past species migration studies (Iverson and Prasad 1998) (Table 1).

Table 1: Common and Latin names of northern and southern tree species in the eastern United States used in this study

Northern Species		Southern Species	
Common Name	**Latin**	**Common Name**	**Latin**
Balsam Fir	*Abies balsamea*	Shortleaf Pine	*Pinus echinata*
Tamarack	*Larix laricina*	Slash Pine	*Pinus elliottii*
White Spruce	*Picea glauca*	Loblolly Pine	*Pinus taeda*
Black Spruce	*Picea mariana*	Baldcypress	*Taxodium distichum*
Red Pine	*Pinus resinosa*	Pignut Hickory	*Carya glabra*
Eastern White Pine	*Pinus strobus*	Flowering Dogwood	*Cornus florida*
Northern White - Cedar	*Thuja occidentalis*	American Holly	*Ilex opaca*
Eastern Hemlock	*Tsuga canadensis*	Sweetgum	*Liquidambar styraciflua*
Sugar Maple	*Acer saccharum*	Yellow-Poplar	*Liriodendron tulipifera*
Yellow Birch	*Betula alleghaniensis*	Southern Magnolia	*Magnolia grandiflora*
Paper Birch	*Betula papyrifera*	Sweetbay	*Magnolia virginiana*
Gray Birch	*Betula populifolia*	Red Mulberry	*Morus rubra*
Black Ash	*Fraxinus nigra*	American Sycamore	*Platanus occidentalis*
Balsam Poplar	*Populus balsamifera*	Southern Red Oak	*Quercus falcata*
Bigtooth Aspen	*Populus grandidentata*	Laurel Oak	*Quercus laurifolia*
Quaking Aspen	*Populus tremuloides*	Blackjack Oak	*Quercus marilandica*
Northern Pin Oak	*Quercus ellipsoidalis*	Water Oak	*Quercus nigra*
Northern Red Oak	*Quercus rubra*	Post Oak	*Quercus stellata*

The oldest forest inventory (year one) that was available in digital format was selected for each eastern state in the Nation (Table 2). The oldest inventories ranged in date from 1977 to 1995. For year two, the most recent annual inventory was selected because all eastern states currently have an FIA annual forest inventory (for more information, see USDA 2007). The most current forest inventories ranged in date from 2002 to 2005. All forest inventory data were taken entirely from FIA's national public database (FIADB 3.0) in 35 eastern states, so all plot latitudes are "fuzzed" as required by law. Because plot locations are perturbed in an unbiased direction not exceeding 1.67 km (typically within a 0.8-km radius of the actual plot location), estimates of maximum species latitudes should not be biased. Annual inventories for each state were first initiated between 1998 and 2003 and continued through 2006, so sample intensities may vary by state. Because FIA inventory is quasi-systematic with sample plots distributed across the geographic extent of each state, varying sample intensities will not bias assessment of tree species locations, but will only affect the precision of the estimates.

Table 2: Measurement years for FIA periodic forest inventories (Year 1) and annual forest inventories (Year 2) by eastern state used in this study

State	Year 1	Year 2	State	Year 1	Year 2
Alabama	1990	2004	Nebraska	1983	2005
Arkansas	1995	2005	New Hampshire	1983	2005
Connecticut	1985	2005	New Jersey	1987	2004
Delaware	1986	2004	New York	1993	2004
Florida	1987	2005	North Carolina	1984	2002
Georgia	1989	2004	North Dakota	1980	2005
Illinois	1985	2005	Ohio	1991	2004
Indiana	1986	2005	Pennsylvania	1989	2004
Iowa	1990	2005	Rhode Island	1985	2005
Kansas	1981	2005	South Carolina	1986	2005
Kentucky	1988	2004	South Dakota	1980	2005
Louisiana	1991	2005	Tennessee	1989	2004
Maine	1995	2003	Texas	1992	2005
Maryland	1986	2004	Vermont	1983	2005
Massachusetts	1985	2005	Virginia	1984	2004
Michigan	1980	2004	West Virginia	1989	2004
Minnesota	1977	2005	Wisconsin	1983	2004
Missouri	1989	2004			

Inventory Comparisons

Differences across minimum and maximum latitudes for northern and southern species between time one and time two indicated no obvious trends (Table 3). For northern species, 9 of 18 study species had higher minimum latitudes in time 2 than in time 1. The average degree difference between inventories for all northern species was 0.18 degrees farther south. Northern pin oak, black ash, and gray birch had some of the greatest shifts northward with their minimum latitude shifts at 2.5 degrees or more. For southern species, only 5 of the 18 study species had maximum latitudes farther north, but the average maximum northward latitude shift across all southern species was 0.02 degrees. Southern magnolia and laurel oak had the largest shifts northward in maximum latitude, with shifts of more than 2.4 degrees northward.

No species migration conclusions can be made given such inconsistent changes in maximum and minimum latitudes between forest inventories. Given that little ecological information can be gleaned from this exercise, perhaps methodologies should be critiqued. This study's technique of comparing maximum and minimum outliers of species ranges based partially on periodic forest inventories has revealed many confounding factors. The two overwhelming factors that complicate species migration detection using periodic data are inconsistent methods/measurement periods and reliance on outliers that may be measurement errors (e.g., species identification). First, only trees with a d.b.h. greater than 2.54 cm can be examined because seedlings were inventoried sporadically using inconsistent sampling methodologies in periodic inventories. Second, the latitudinal shift of mature and/or established trees may be a lagging indicator of climate change effects. Attempting to compare a periodic forest inventory from 1982 to an annual inventory conducted in 2000 may not provide a

sufficient period of time (only 18 years) to indicate the movement of trees with a d.b.h. greater than 2.54 cm.

Table 3: Maximum or minimum latitudes in times 1 and 2 by species in the eastern United States

Northern				Southern			
Common Name	**Time 1 Min. Latitude (deg.)**	**Time 2 Min. Latitude (deg.)**	**Degree Diff.[a]**	**Common Name**	**Time 1 Max. Latitude (deg.)**	**Time 2 Max. Latitude (deg.)**	**Degree Diff.[a]**
Balsam Fir	43.12	42.06	1.06	Shortleaf Pine	40.53	40.06	0.47
Tamarack	39.94	40.66	-0.72	Slash Pine	34.06	34.55	-0.49
White Spruce	40.06	39.89	0.17	Loblolly Pine	39.8	39.09	0.71
Black Spruce	40.03	42.46	-2.43	Baldcypress	39.01	39.01	0.00
Red Pine	38.45	39.17	-0.72	Pignut Hickory	43.19	43.09	0.10
Eastern White Pine	35.11	31.49	3.62	Flowering Dogwood	45.96	43.99	1.97
Northern White-Cedar	42.79	41.35	1.44	American Holly	39.81	41.97	-2.16
Eastern Hemlock	34.19	34.34	-0.15	Sweetgum	41.49	40.14	1.35
Sugar Maple	31.47	29.11	2.36	Yellow-Poplar	42.95	42.39	0.56
Yellow Birch	35.17	34.53	0.64	Southern Magnolia	33.83	36.25	-2.42
Paper Birch	39.75	38.48	1.27	Sweetbay	39.17	39.51	-0.34
Gray Birch	38.46	40.96	-2.50	Red Mulberry	44.99	43.38	1.61
Black Ash	35.58	38.36	-2.78	American Sycamore	43.31	42.91	0.40
Balsam Poplar	39.37	35.54	3.83	Southern Red Oak	40.36	39.37	0.99
Bigtooth Aspen	37.28	36.44	0.84	Laurel Oak	34.22	39.14	-4.92
Quaking Aspen	38.78	40.51	-1.73	Blackjack Oak	41.01	40.36	0.65
Northern Pin Oak	38.54	41.1	-2.56	Water Oak	39.8	38.71	1.09
Northern Red Oak	31.25	31.69	-0.44	Post Oak	40.54	40.31	0.23

[a]Degree Difference = Time 1 latitude – time 2 latitude

Furthermore, the oldest forest inventories (pre-1970s) are currently not digitized, disallowing comparisons to current inventories. Third, to examine tree species shifts across large geographic extents, multiple state inventories need to be used. Because inventories were periodic before 1999, comparing periodic to annual inventories would mean, for example, comparing a 1978 inventory to 2001 for one state while comparing a 1986 inventory to 1999 in an adjoining state. Fourth, FIA field crews attain measurement repeatability of only 95-98 percent for tree species identification (Pollard et al. 2006). Developing range maps based on maximum spatial distributions from periodic forest inventories places too much reliance on single observations that may be prone to measurement error. Fifth, even if some portions of periodic inventory plots were remeasured during the most recent annual inventory, sample protocols have changed. Hence, examining tree species migration over long periods of time in the eastern United States is almost completely confounded by changes in plot locations (sample intensities), plot sampling configurations/protocols, and non-matching periodic inventory dates from state to state. Overall, there appears to be many obstacles to using periodic forest inventories to monitor tree migration. Robust tree migration monitoring requires development of novel techniques, especially given the potentially profound impacts that such migration could have on the total environment and society as a whole.

Future Directions

Over thirty years ago, the overarching purpose of conducting forest inventories was state-level forest resource assessment (e.g., growing-stock volumes and forest-type distributions). The use of such data across state boundaries to monitor tree migration was not foreseen at that time and now is nearly precluded by a host of confounding factors. Although there may be some ways to avoid some of the confounding factors stemming from periodic inventories (e.g., conducting state-level monitoring of individual refugia), there is no statistically robust way forward to use periodic inventories to monitor species migrations along the eastern United States. Continued research is strongly suggested in this area. It is suggested that new indicators of tree species migration be developed using only annual forest inventory data (since 1999). These data are systematically balanced across the entire eastern United States using consistent sample protocols, digital database management, and thorough documentation. Perhaps comparing seedling to mature tree distributions by species may be a new indicator to consider. Nonetheless, high quality and consistent forest inventory across large-scales provides the best opportunity to monitor tree species migration for the foreseeable future.

Literature Cited

Bechtold, W.A.; Patterson, P.L., eds. 2005. Forest Inventory and Analysis National Sample Design and Estimation Procedures. Gen. Tech. Report. SRS-GTR-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Botkin, D.B.; Saxe, H.; Araujo, M.B.; Betts, R.; et al. 2007. Forecasting the effects of global warming on biodiversity. Bioscience. 57: 227-236.

Davis, M.B.; Shaw, R.G. 2001. Range shifts and adaptive responses to quaternary climate change. Science. 292: 673-679.

Frayer, W.E.; Furnival, G.M. 1999. Forest survey sampling designs, a history. Journal of Forestry. 97: 4-10.

Gillespie, A.J.R. 1999. Rationale for a national annual forest inventory program. Journal of Forestry. 97: 16-20.

Grace, J.; Berninger, F.; Nagy, L. 2002. Impacts of climate change on the tree line. Annals of Botany. 90: 537-544.

IPCC. 2007. Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Solomon, S.; Qin, D.; Manning, M.; Chen, Z.; Marquis, M.; Avery, B.; Tignor, M.; and Miller, H.L. (eds.) Cambridge, UK, and New York, NY: Cambridge University Press. 996 p.

Iverson, L.R.; Prasad, A. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. Ecological Monographs. 68: 465-485.

Iverson, L.R.; Prasad, A.; Schwartz, M.W. 1999. Modeling potential future individual tree-species distributions in the eastern United States under a climate change scenario: a case study with Pinus virgininia. Ecological Modelling. 115: 77-93.

Iverson, L.R.; Prasad, A.M.; Matthews, S.N.; Peters, M. 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. Forest Ecology and Management. 254: 390-406.

Little, E.L. 1971. Atlas of United States trees. Volume I. Conifers and important hardwoods. Misc. Pub. 1146. Washington, D.C.: U.S. Department of Agriculture Forest Service.

Logan, J.A.; Regniere, J.; Powell, J.A. 2003. Assessing the impacts of global warming on forest pest dynamics. Frontiers in Ecology and Environment. 1: 130-137.

Malcolm, J.R.; Markhma, A.; Neilson, R.P.; Garaci, M. 2002. Estimated migration rates under scenarios of global climate change. Journal of Biogeography. 29: 835-849.

McCarty, J.P. 2001. Ecological consequences of recent climate change. Conservation Biology. 15: 320-331.

Nabuurs, G.J.; Pussinen, A.; Karjalainen, T.; Erhard, M.; Kramer, K. 2002. Stemwood volume increment changes in European forests due to climate change – a simulation study with the EFISCEN model. Global Change Biology. 8: 304-316.

Noss, R.F. 2001. Beyond Kyoto: forest management in a time of rapid climate change. Conservation Biology. 15: 578-590.

Opdam, P.; Wascher, D. 2004. Climate change meets habitat fragmentation: linking landscapes and biogeographical scale levels in research and conservation. Biological Conservation. 117: 285-297.

Parmesan, C.; Yohe, G. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature. 421: 37-42.

Pollard, J.E.; Westfall, J.A.; Patterson, P.L.; Gartner, D.L.; Hansen, M.; Kuegler, O. 2006. Forest inventory and analysis national data quality assessment report for 2000 to 2003. Gen .Tech. Rep. RMRS-GTR-181. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 43 p.

Sacks, W.J.; Schimel, D.S.; Monson, R.K. 2007. Coupling between carbon cycling and climate in a high-elevation subalpine forest: a model-data fusion analysis. Oecologia. 151: 54-68.

Saxe, H.; Ellsworth, D.S.; Heath, J. 1998. Tree and forest functioning in an enriched CO2 atmosphere. New Phytologist. 139: 395-436.

Schwartz, M.W. 1992. Modeling effects of habitat fragmentation on the ability of trees to respond to climatic warming. Biodiversity and Conservation. 2: 51-61.

Schwartz, M.W.; Iverson, L.R.; Prasad, A.M. 2001. Predicting the potential future distribution of four tree species in Ohio using current habitat availability and climatic forcing. Ecosystems. 4: 568-581.

Smith, W. B. 2002. Forest inventory and analysis: a national inventory and monitoring program. Environmental Pollution. 116: S233-S242.

Stenseth, N.C.; Mysterud, A.; Ottersen, G.; Hurrell, J.W.; Chan, K.S.; Lima, M. 2002. Ecological effects of climate fluctuations. Science. 297: 1292-1296.

U.S. Department of Agriculture, Forest Service. 2008. Forest Inventory and Analysis National Core Field Guide, Version 4.0. 2008. Available at www.fia.fs.fed.us/library/field-guides-methods-proc (Feb. 20 2008), USDA Forest Service, Forest Inventory and Analysis National Office.

U.S. Department of Agriculture. 2007. Forest inventory and analysis, fiscal year 2006 business report. U.S. Department of Agriculture, Forest Service. FS-884.

Volney, W.J.A.; Fleming, R.A. 2000. Climate change and impacts on boreal forest insects. Agriculture, Ecosystems, and Environment. 82: 283-294.

Walther, G.R.; Post, E.; Convey, P.; Menzel, A.; et al. 2002. Ecological responses to recent climate change. Nature. 416: 389-395.

Westerling, A.L.; Hidalgo, H.G.; Cayan, D.R.; Swetnam, T.W. 2006. Warming and earlier spring increase western U.S. forest wildfire activity. Science. 313: 940-943.

From Genes to Ecosystems: Measuring Evolutionary Diversity and Community Structure with Forest Inventory and Analysis (FIA) Data

Kevin M. Potter[1]

***Abstract**: Forest genetic sustainability is an important component of forest health because genetic diversity and evolutionary processes allow for the adaptation of species and for the maintenance of ecosystem functionality and resilience. Phylogenetic community analyses, a set of new statistical methods for describing the evolutionary relationships among species, offer an innovative approach for assessing the health of forest communities from an evolutionary perspective. Forest Inventory and Analysis data are ideal for conducting phylogenetic community analyses for forest tree species at broad scales. FIA data from 100,000 plots across the conterminous United States were used to investigate the evolutionary characteristics of forest tree communities. This required generating a phylogenetic "evolutionary tree" of the 311 forest tree species inventoried by FIA, based on recent gene sequencing studies of several plant groups. Phylogenetic diversity was quantified for each plot; this statistic sums the evolutionary age of the species present, using the phylogenetic tree. This is a more meaningful way to quantify biodiversity than species richness because, rather than weighting all species on a plot equally regardless of relatedness, it measures their cumulative evolutionary age. General patterns at the ecoregion section scale were similar between mean plot-level species richness and phylogenetic diversity, but important differences also existed. Additionally, the analyses described each plot's phylogenetic community structure, or whether the species were more clustered or dispersed across the 311-species reference evolutionary tree than expected by chance. The most phylogenetically overdispersed sections were located in the Interior West, while the most phylogenetically clustered included several in the Upper Midwest, New England, California and the Southeastern Coastal Plain. Communities that are phylogenetically clustered consist of more closely related species and therefore may be more susceptible to threats such as pests and climate change. Phylogenetically dispersed communities may be more resilient to these pressures, because greater evolutionary diversity is expected to translate into a greater likelihood that more species will be unaffected by, or will be adaptable to, environmental changes.*

Keywords: Biodiversity, evolutionary biology, landscape genetics, forest health monitoring.

---

[1] North Carolina State University; Department of Forestry and Environmental Resources; Forest Health Monitoring Program; 3041 Cornwallis Road, Research Triangle Park, NC 27709; kevinpotter@fs.fed.us.

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

The integration of three diverse scientific disciplines – landscape ecology, conservation biology, and evolutionary genetics – can offer innovative and valuable insights about the health of forested ecosystems, particularly at large scales (figure 1). Approaching conservation from a landscape ecology perspective is vitally important because large-scale processes affect biodiversity and because management decisions are often made at regional levels. At the same time, it is imperative to consider genetic processes in a landscape ecology context because ecological dynamics at a variety of scales can affect the evolution of species. Understanding conservation biology in the light of evolutionary processes, meanwhile, is necessary to measure and manage forest biodiversity. Forest health monitoring work that integrates all three disciplines of landscape ecology, conservation, and genetics is now possible because of the availability of recent advancements in computing, statistical analysis, and molecular biology. This work is particularly relevant to assessing forest genetic sustainability across large scales using U.S. Forest Service Forest Inventory and Analysis (FIA) Phase 2 inventory data.

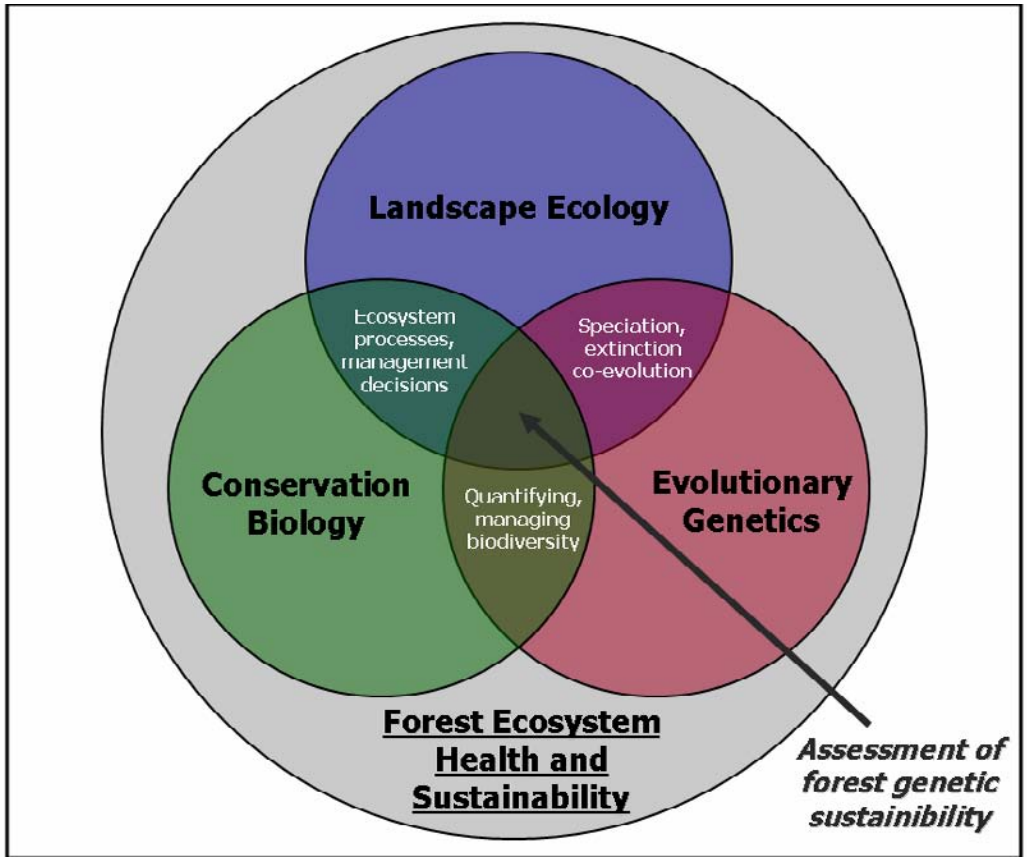


Figure 1: Relationships among landscape ecology, conservation biology and evolutionary genetics within the context of monitoring forest health and sustainability.

Forest genetic sustainability is an important component of forest health because genetic diversity and evolutionary processes allow for the adaptation of species and for the maintenance of ecosystem functionality and resilience. Genetic diversity within species, for example, reflects the recent and existing integrity of evolutionary and ecological processes (Brown et al. 2000). At the same time, existing evolutionary lineages will generate future biodiversity, and are therefore a cornerstone of environmental health (Erwin 1991). The importance of forest genetic diversity is reflected in the Criteria and Indicators for the Conservation and Sustainable Management of Temperate and Boreal Forests, which includes three indicators of genetic diversity under Criterion 1 (conservation of biological diversity):

(1) number and geographic distribution of forest associated species at risk of losing genetic variation and locally adapted genotypes,

(2) population levels of selected representative forest associated species to describe genetic diversity, and

(3) status of on-site and off-site efforts focused on conservation of genetic diversity (Montréal Process Working Group 2008).

The first two of these indicators attempt to quantify the number and population levels of individual species at risk of losing genetic variability, while the third indicator seeks to evaluate the success gene conservation efforts for forest species, presumably those at risk. All three are valuable measurements, but each is limited to a handful of indicator species and therefore does not address the evolutionary diversity present across entire forest communities. A new approach, however, can now combine nationwide FIA forest inventory data, an ecoregion-scale perspective, and recent advancements in gene sequencing and molecular systematics to quantify evolutionary diversity for communities of species across large spatial scales.

This approach, community phylogenetic analyses (Webb and others 2002), offers the ability to synthesize evolutionary biology and landscape ecology in the context of assessing forest community health, both in terms of biodiversity and community resilience to stress. Community phylogenetic analyses are now possible across broad taxonomic species groupings (forest trees, for example) and across large spatial scales because of recent advances in the ability to create robust phylogenies describing the relationships among species. These phylogenies are hypothesized "family trees" of species (figure 2) that are developed by surveying molecular systematic studies, which compare differences and similarities in gene sequences to determine the evolutionary relationships among the species and to estimate how long ago species diverged from each other. Fossil evidence from palaeobotanical studies can further calibrate the dates at which species or species groups diverged. The branches within these phylogenetic trees often are measured in millions of years. Constructing such

phylogenetic trees does not always require new gene sequencing work, because molecular systematic studies now have been published for a wide variety of plant groups.

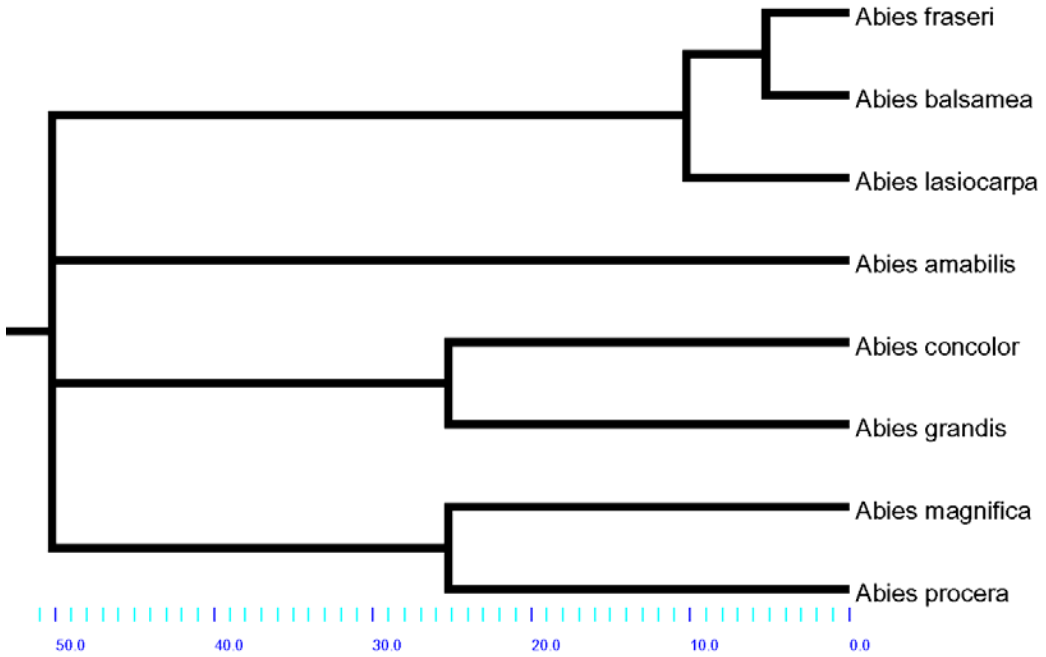


Figure 2: A phylogenetic tree of the hypothesized evolutionary relationships among true fir (*Abies*) species of North America inventoried by FIA. The scale bar at the bottom of the figure depicts time in millions of years. Such phylogenetic trees are hypotheses of the relationships among species, and can be improved over time with additional gene sequencing studies and with additional information unearthed in the fossil record.

Such phylogenetic trees are necessary for two types of phylogenetic community analyses that are useful in describing evolutionary relationships within forest tree communities at broad spatial scales: the generation of phylogenetic diversity statistics and the quantification of phylogenetic community structure.

Phylogenetic diversity statistics (Faith 1992, Webb and others 2006) are particularly meaningful as a measure of biodiversity because they quantify the cumulative evolutionary age and evolutionary potential of all the species in the community of interest. This is typically done by first generating a phylogenetic tree encompassing the species present in a community, based on existing molecular systematics research. Next, the lengths of the branches of that phylogenetic tree, usually measured in millions of years of evolutionary time, are measured and summed. This approach may be more meaningful than traditional biodiversity metrics such as species richness and abundance, which weight the value of all species equally regardless of their relatedness (figure 3). Maintaining the evolutionary potential of groups of species measured in this way has become an increasingly important conservation goal (Rodrigues and Gaston 2002, Sechrest and others 2002, Soltis and Gitzendanner 1999).

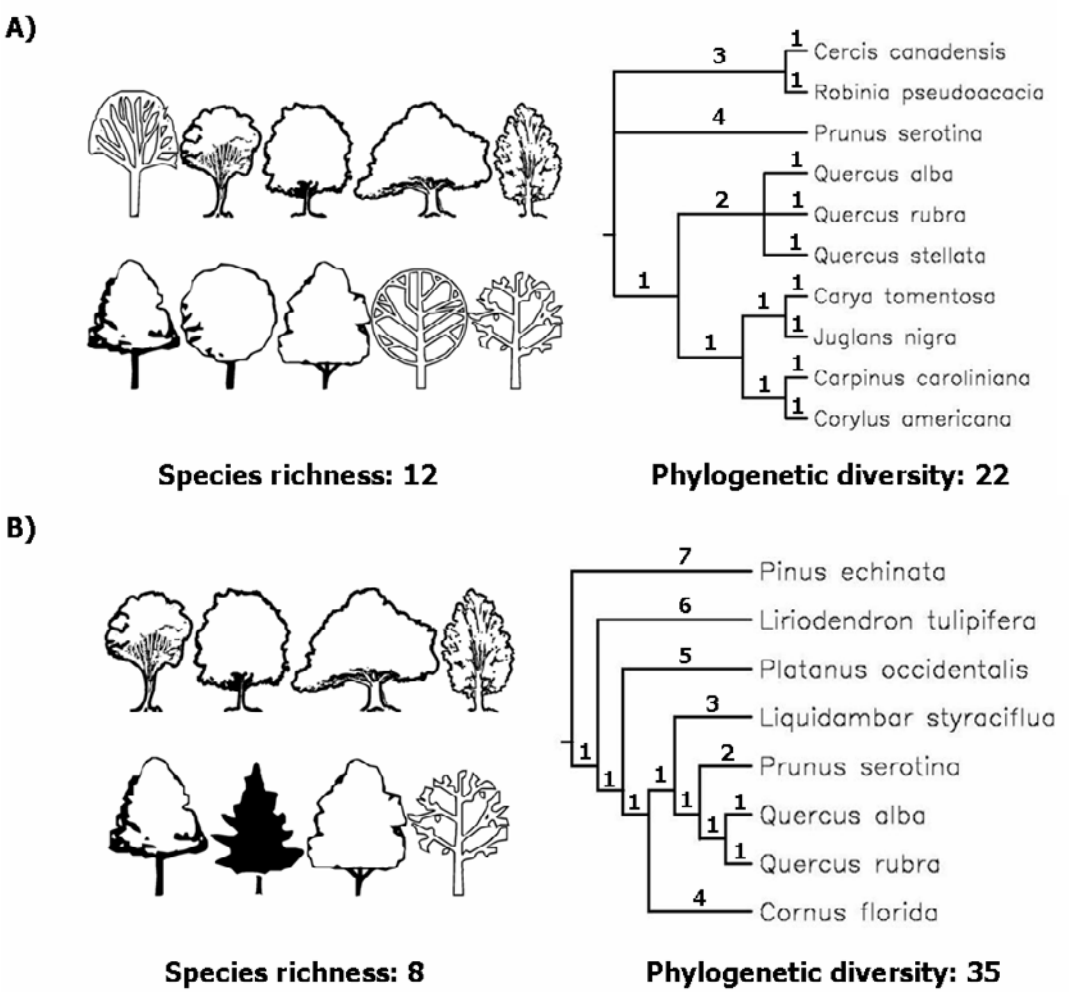


Figure 3: A comparison between two measurements of biodiversity, species richness and phylogenetic diversity. Species richness is the number of species present in the community, while phylogenetic diversity is the summed evolutionary age of all the species in the community, as measured by the phylogenetic tree encompassing the species present. Community (A) has greater species richness than community (B), but community (B) has greater phylogenetic diversity because the species in the community contain greater evolutionary distance. The branch lengths shown on the phylogenetic trees are for demonstration purposes, and do not depict actual evolutionary relationships among the species.

A second set of phylogenetic analyses focuses on determining whether the species within a specific community are more phylogenetically clustered or dispersed on the evolutionary tree of life than expected by chance (Webb 2000). Such an analysis would, for example, compare the phylogenetic tree of the species within a single FIA plot to the phylogenetic tree that includes all the tree species in North America. If the species on the FIA plot are more closely related than expected by random chance, when compared to the larger reference phylogenetic tree, then the community is phylogenetically clustered (figure 4a). If the species on the plot are less closely related than expected by chance, then the community is phylogenetically overdispersed (figure 4b). This type of analysis is useful from

a forest health perspective because it offers a way to quantify a community's potential evolutionary resilience in the face of pests, pathogens, climate change and other stresses. That is because communities made up of species that are overdispersed on the phylogenetic tree possess greater-than-expected evolutionary diversity, and may therefore encompass a higher proportion of species unaffected by a given stressor or able to adapt to it. Phylogenetically clustered communities, meanwhile, contain less evolutionary diversity and may be at greater cumulative risk from stressors that might affect several closely related species, such as sudden oak death (*Phytophthora ramorum*).

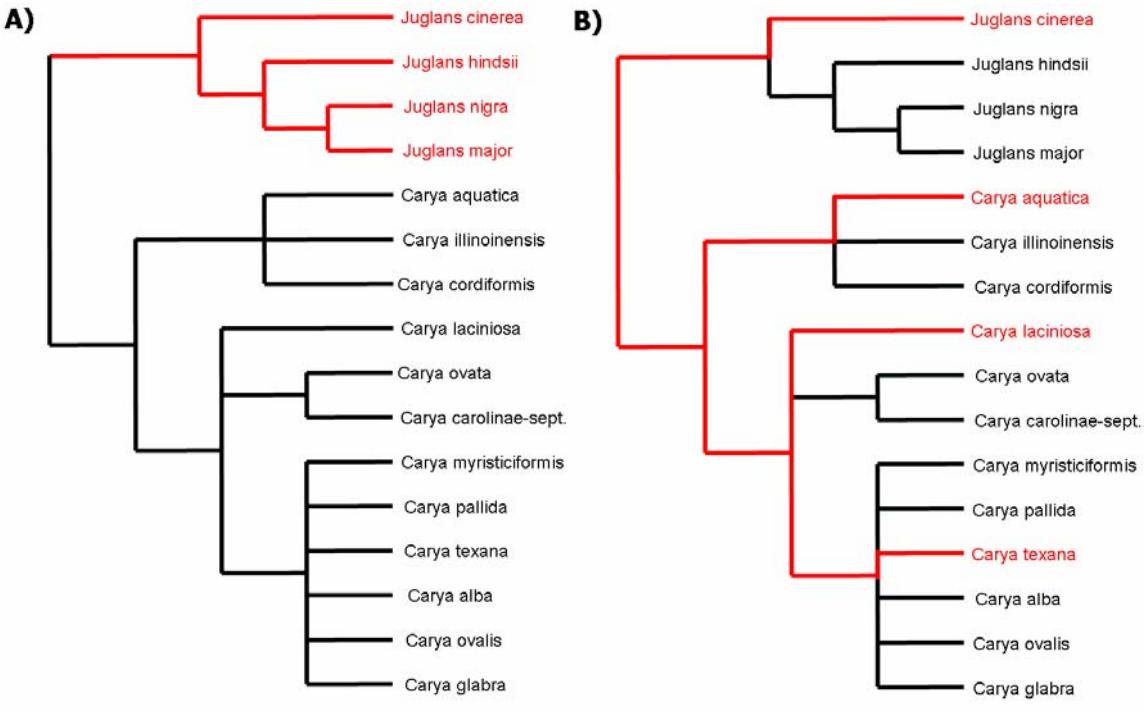


Figure 4: Phylogenetic trees of hypothetical tree communities representing the two types of phylogenetic community structure, (A) phylogenetic clustering and (B) phylogenetic overdispersion. The species in red are those present in the community; the species in black and red together encompass the phylogenetic reference tree of species that could exist in the community.

The work described in this paper had three main objectives:

(1) Assess the usefulness of FIA tree inventory datasets for phylogenetic diversity and community structure analyses across large ecoregion scales.

(2) Compare mean FIA plot-level tree species richness to mean plot phylogenetic diversity at the ecoregion section scale.

(3) Test for correlations between mean plot measures of phylogenetic community structure and environmental variables at the ecoregion section scale.

Methods

The software package Phylocom 3.41 (Webb and others 2007) was used to calculate phylogenetic diversity and phylogenetic community structure statistics for the forest trees present on each of 102,304 one-sixth-acre Forest Inventory and Analysis (FIA) plots across the 48 conterminous United States. These plots represented the latest available FIA Phase 2 tree and sapling inventory data (trees ≥ 1 inch dbh) as of November 2007 (Forest Inventory and Analysis Program 2007). A phylogenetic reference supertree (figure 5), encompassing the 311 forest tree species inventoried by FIA across the conterminous United States, was constructed for the estimation of phylogenetic distance among species in units of millions of years (Potter in review).

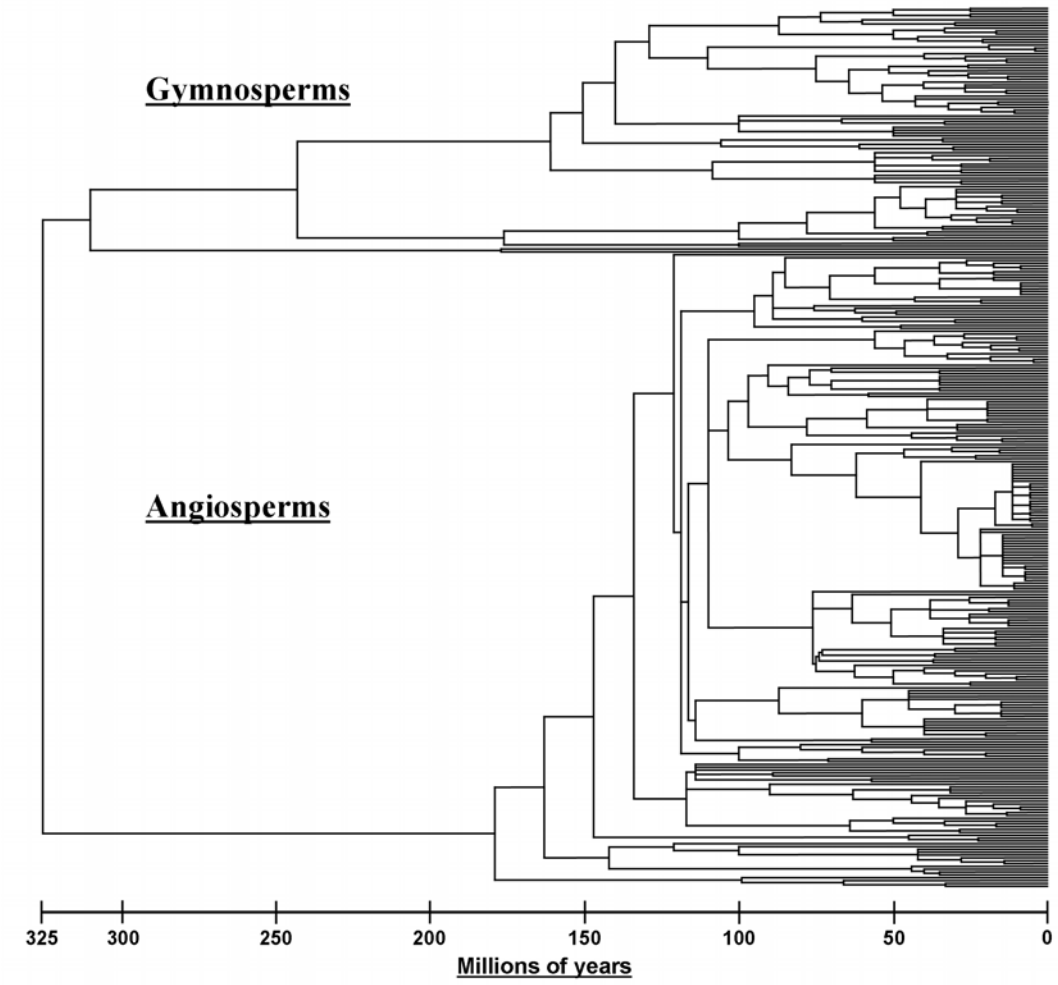


Figure 5: The phylogenetic supertree used in this analysis, incorporating 311 North American forest tree species inventoried by FIA. Evolutionary relationships and branch lengths were based on a survey of approximately 70 recent molecular systematic and paleobotanical studies, with the exception of basal angiosperm relationships, which taken from Wikstrom *et al.* (2001). Branch lengths are measured in millions of years.

Based on its approximate geographic coordinates, each FIA plot was assigned with ArcMap 9.2 (ESRI 2006) to the appropriate ecoregion section (Bailey 1995), using the most recent geographical information system (GIS) mapping of Bailey's hierarchical system of ecoregion domains, divisions, provinces and sections (Cleland and others 2005). Ecoregion sections containing fewer than 25 plots were excluded from the analyses to ensure an adequate sampling.

Species richness was calculated for each FIA plot, as was Faith's (1992) index of phylogenetic diversity (PD). This quantifies the total evolutionary history represented by the species in a community by determining the total phylogenetic branch length distance of the species on the plot, divided by the total branch length distance of the reference phylogenetic supertree encompassing all the North American tree species inventoried by FIA. Mean PD and species richness values were calculated for ecoregion sections.

Additionally, two measures of phylogenetic community structure were calculated for each FIA plot. Both statistics measure whether the configuration of the phylogenetic tree, encompassing the species in a given community, is more or less clustered than expected by chance, as compared to a regional species pool. In this analysis, the regional species pool for each plot consisted of all the species inventoried on all the FIA plots within the same ecoregion section as the plot in question. (See Potter (in review) for details about the calculation of these statistics.)

(1) The Nearest Taxon Index (NTI) is a standardized measure of the *branch-tip phylogenetic clustering or overdispersion* of the species on the FIA plot, regardless of the arrangement of the higher level groups in the phylogenetic tree (Webb and others 2006). In other words, NTI measures whether evolutionary diversity among the species on the plot is higher (overdispersed) or lower (clustered) than expected by chance, as compared to the ecoregion section species pool.

(2) The Net Relatedness Index (NRI) is a standardized measurement of *basal or tree-wide phylogenetic clustering or overdispersion* of the species on the FIA plot (Webb and others 2006). In other words, NRI measures whether evolutionary diversity among the deeper phylogenetic ranks (families, orders, classes and divisions) present on the plot is higher (overdispersed) or lower (clustered) than expected by chance, again as compared to the ecoregion section species pool.

NTI and NRI measure different evolutionary characteristics of communities, so the two statistics could show different results for a plot, for example indicating clustering by one metric and overdispersion by the other. NTI and NRI values are negative when species on a plot do not occur together with closely related species (overdispersed), and positive when species occur with other closely related species (clustered) (Kembel and Hubbell 2006). The means for plot-level NRI

and NTI values were calculated across ecoregion sections, with a *t* test determining whether the section-level mean values were significantly different from 0.

The possibility of large-scale correlations of these two phylogenetic community structure measures with ecological variables was tested for several climate (PRISM Group 2008), soil (Miller and White 1998) and topographic (United States Geological Survey 1996) variables at the ecoregion section scale. The ecoregion section means of each of these was determined across the forested area of the section, using a forest cover map (1 km^2 resolution) derived from Moderate Resolution Imaging Spectroradiometer (MODIS) imagery by the U.S. Forest Service Remote Sensing Applications Center (United States Department of Agriculture Forest Service 2008).

Results and Discussion

Values of both mean plot species richness and phylogenetic diversity tended to be higher in ecoregion sections of the eastern United States than in the western part of the country (figure 6). The sections with the lowest means for both statistics were located in the interior West. While the general pattern was similar between these two measures of biodiversity, important differences were apparent. Specifically, phylogenetic diversity was higher relative to species richness in several ecoregion sections along the Pacific coast, from central California to Washington, as well as in sections in the Northeast and in the Great Lakes region.

Not surprisingly, the 10 ecoregion sections with the highest mean species richness values are all located in the southeastern United States. However, only six of the 10 sections with the highest mean phylogenetic diversity are located in the Southeast, with the other four in the Northeast. Ecoregion sections with high phylogenetic diversity relative to species richness appear to have a combination of both high angiosperm (flowering plant) and gymnosperm (cone-bearing) species richness (Potter in review).

Meanwhile, the analysis of phylogenetic community structure detected regional patterns in phylogenetic overdispersion and clustering (figure 7). Both the branch-tip Nearest Taxon Index (NTI) and tree-wide Net Relatedness Index (NRI) metrics indicated the existence of phylogenetic clustering in some of the most northerly ecoregion sections, including those in the Northeast, the Great Lakes region, the northern Great Plains, and the Pacific Northwest, as well as along the Appalachian and Sierra Nevada mountain chains. Slight to moderate overdispersion for both metrics was present for sections in the Interior West, particularly in the Southwest.

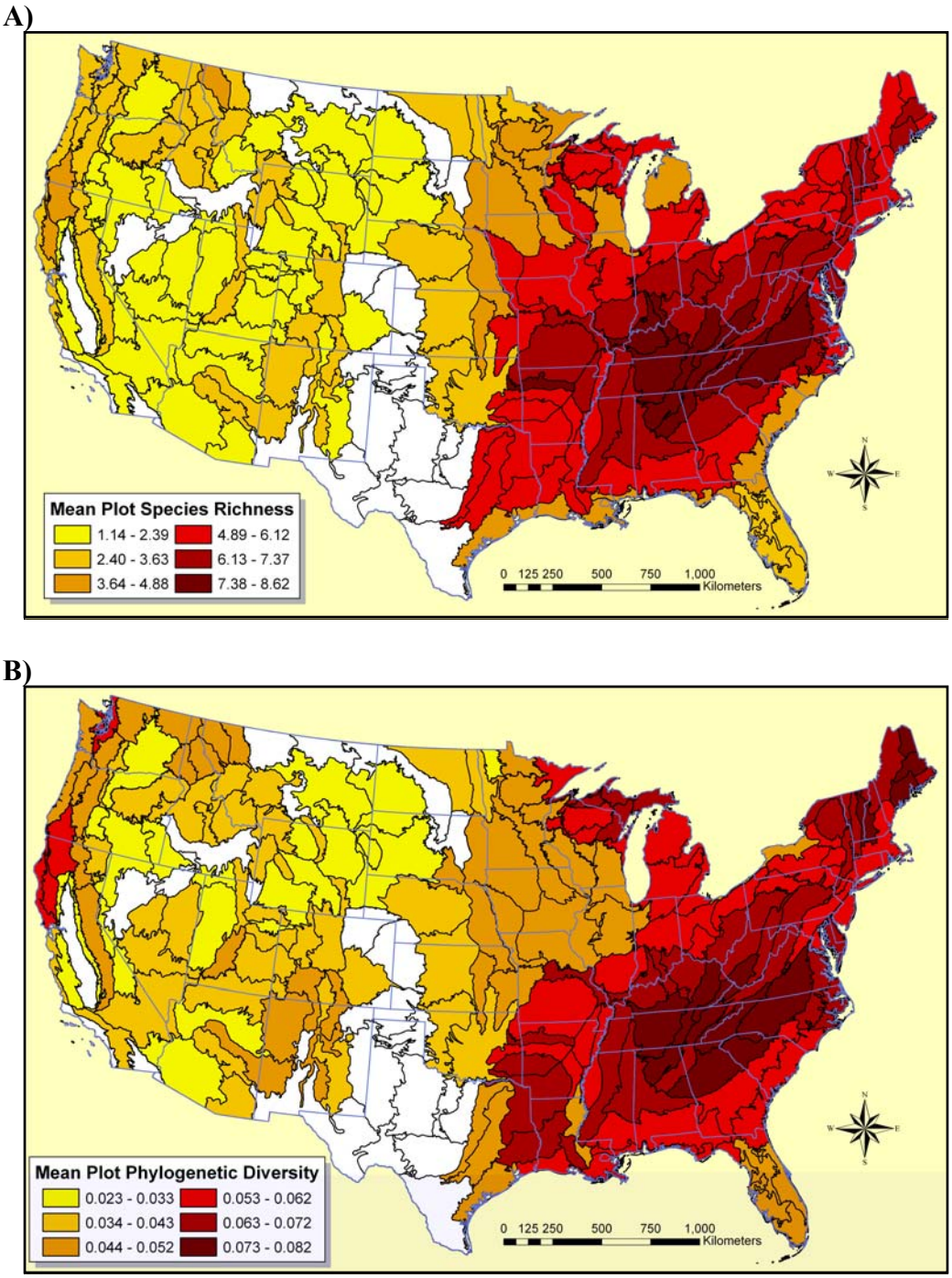


Figure 6: Mean FIA plot-level forest tree species richness (A) and Faith's index of phylogenetic diversity (B) across ecoregion sections. The two statistics were divided into six equal interval classes for comparison purposes.

At the same time, some differences were apparent between the two measures. Perhaps the most notable difference is in the Southeast, where six sections were overdispersed using NTI, which quantifies clustering at the tips of the phylogenetic tree, but where none were overdispersed using the NRI, which

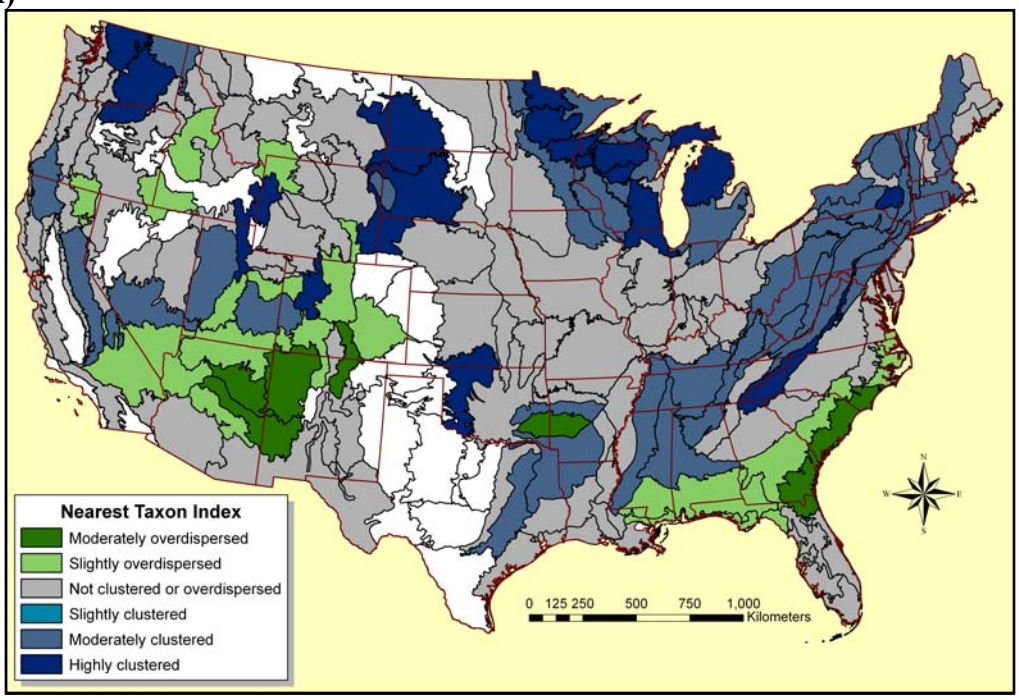
measures tree-wide clustering. In fact, four of these six sections were *clustered* using the NRI. This is most likely the situation because gymnosperm diversity is low in these sections, resulting in tree-wide phylogenetic clustering, while their angiosperm diversity is moderate, with angiosperm species evenly spaced across the phylogenetic tree on many plots (Potter in review).

Such differences may have important forest health implications. A community consisting of species that are overdispersed across the phylogenetic tree but clustered toward the branch tips may contain fewer species susceptible to a given threat. That community, however, may be more likely to lose important ecological functions provided by any species that are eliminated. Meanwhile, a community of species clustered across the phylogenetic tree but overdispersed at the tips may better retain its ecological functionality in such a situation, but might encompass more species at risk of elimination.

It is worth noting that it is possible for a community to possess high phylogenetic diversity but to still have a clustered phylogenetic community structure, and, as a result, to be more susceptible to certain threats. This is because of important differences in the two metrics. Phylogenetic diversity measures the total evolutionary diversity present in a community, while NRI and NTI are indices that quantify the configuration of the evolutionary relatedness among the species. Ecoregions section with high mean plot-level phylogenetic diversity and a high degree of clustering include the Eastern Upper Peninsula (212R), the Blue Ridge Mountains (M221D), the Catskill Mountains (211I), the Maine-New Brunswick Foothills and Lowlands (211B), and the Klamath Mountains (M261A).

The patterns of phylogenetic community structure detected in this study may be driven at least in part by ecological interactions among species within the community. Phylogenetic clustering, for example, may be caused by environmental filtering, the process that occurs when closely related species tend to co-occur because they share similar tolerances to the abiotic environment (Cavender-Bares and others 2004, Tofts and Silvertown 2000). Because the species present in such communities share much evolutionary history and an affinity for similar environmental conditions, they may be particularly susceptible collectively to a certain threats, such as insects and diseases targeting specific families or genera of tree species. Phylogenetic overdispersion in a community, meanwhile, may indicate the existence of competitive exclusion. This occurs when closely related species compete for similar environmental niches and exclude each other from a community when they share limiting resources (Cavender-Bares and others 2004, Tofts and Silvertown 2000). The ecological integrity of such communities may be less at risk from changing conditions because they encompass a wider variety of evolutionary adaptations to respond to those changes.

A)



B)

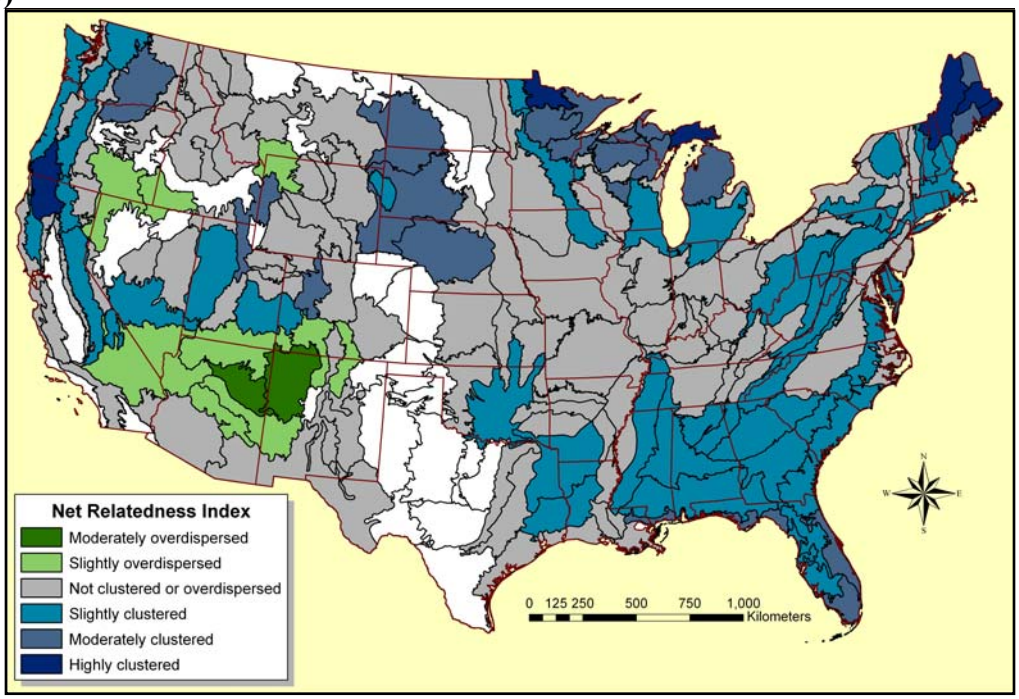


Figure 7: Mean FIA plot-level measures of phylogenetic clustering across ecoregion sections, using (A) the Nearest Taxon Index, a measure of clustering at the tips of the phylogenetic branches, and (B) the Net Relatedness Index, a measure of clustering throughout the phylogenetic tree. A *t*-test was used to determine whether the mean index values were significantly different from 0, with negative index values overdispersed and positive values clustered.

Several correlations were detected between mean phylogenetic community structure values and environmental characteristics at the ecoregion section scale (table 1). These correlations suggest that competitive exclusion for resources among related species (phylogenetic overdispersion) is an important process in ecoregion sections with less "hospitable" environments. Meanwhile, environmental filtering, which occurs when more closely related species tend to occupy the same kinds of environments (phylogenetic clustering), may be more common when environmental conditions are more hospitable, such as at sites with moister and less compacted soils, at lower elevations, and with greater amounts of precipitation.

For example, soil available water capacity was the environmental variable with the strongest positive correlation phylogenetic clustering across ecoregion sections, indicating communities with wetter soils were generally more phylogenetically clustered. Greater soil acidity (lower pH) also was positively associated with phylogenetic clustering, as was depth to bedrock. Among the climate variables, annual precipitation had a weak but significant correlation with NRI, indicating that forest tree communities in ecoregion sections with more precipitation were more phylogenetically clustered using the phylogenetic tree-wide metric. Standard deviation of monthly maximum and minimum temperatures was negatively correlated with both NRI and NTI, suggesting that variability in temperatures was associated with greater phylogenetic overdispersion of communities in the ecoregion. Finally, three different topographical measures were important in predicting phylogenetic structure, with ecoregion section elevation mean, elevation variation (standard deviation) and elevation range all correlated with greater phylogenetic overdispersion.

Table 1: Correlations between mean plot-level phylogenetic clustering statistics across ecoregion sections and environmental variables at the section scale.

Environmental variable	NRI	NTI
Soils		
Available water capacity (AWC)	0.475***	0.463***
Bulk density	-0.519***	-0.37***
pH	-0.355***	-0.249**
Depth to bedrock	0.34***	0.321***
Climate		
Mean annual precipitation (1971-2000)	0.226**	ns
Monthly maximum temperature (1971-2000) (mean)	ns	ns
Monthly maximum temperature (1971-2000) (standard deviation)	-0.377***	-0.383***
Monthly minimum temperature (1971-2000) (mean)	ns	ns
Monthly minimum temperature (1971-2000) (standard deviation)	-0.345***	-0.347***
Topography		
Elevation (mean)	-0.413***	-0.41***
Elevation (standard deviation)	-0.274***	-0.281***
Elevation (range)	-0.315***	-0.339***

*** p<0.001; ** p<0.01; ns, not significant

Conclusions and future work

The work outlined in this paper demonstrates the applicability of phylogenetic community analysis to data sets at regional and even national scales. Generating these phylogenetic diversity and community structure statistics at large scales is one approach for investigating the role evolutionary biology plays in shaping processes and patterns in the natural world. The results of this study indicate that differences exist between phylogenetic diversity and species richness within ecoregion sections, and suggest that phylogenetic diversity may be a more meaningful measure of biodiversity because it accounts for evolutionary relationships among the species in a community. Further, the analyses of phylogenetic community structure detected intriguing regional patterns of phylogenetic clustering and overdispersion. Additional research is necessary to better understand the causes of these patterns, and to further explore the correlations, at regional scales, between phylogenetic clustering and several environmental variables.

The types of analyses described in this paper have potential for investigating the evolutionary aspects of forest health. For example, these analyses could assess whether regions with more phylogenetically clustered forest tree communities are more susceptible to major environmental changes, because of the possibility that a higher proportion of species will be affected by a given stressor. Another important area of research using these methods could quantify the evolutionary impacts of the loss of species, in terms of loss of ecological functionality and evolutionary potential.

Finally, this study establishes that Forest Inventory and Analysis data are suitable for addressing evolutionary biology questions at regional scales. In fact, few if any other available data sets have the extent and resolution necessary to conduct phylogenetic community analyses for forest communities at broad scales. FIA inventory data have considerable potential for use in many such future studies, including several with forest health applications. Future research, for example, could incorporate FIA data to:

(1) assess whether nonnative invasive forest tree species are more phylogenetically related than expected by chance (Strauss and others 2006), which could allow for the identification of groups of closely related nonnative species that may be likely to become invasive.

(2) determine the phylogenetic community effects of the loss of species or groups of species as a result of forest tree insects or pathogens, such as chestnut blight, hemlock woolly adelgid, and white pine blister rust.

(3) test for phylogenetic signal in forest tree susceptibility to insect and disease infestation, particularly from those with multiple hosts, and to broad environmental changes such as climate change.

(4) quantify the phylogenetic diversity and community structure of entire forest plant communities, and of strata within those communities, using FIA Phase 3 data.

Acknowledgments

I wish to thank Frank Koch, Mark Ambrose and Barb Conkling from the Forest Health Monitoring Group at North Carolina State University for their insights about this research and for their technical and administrative support. I further thank Kurt Riitters from the U.S. Forest Service and Cam Webb of the Arnold Arboretum at Harvard University for their technical support. This research was supported in part through Research Joint Venture Agreement 06-JV-11330146-124 between the USDA Forest Service and North Carolina State University.

Literature Cited

Bailey, R.G. 1995. Description of the ecoregions of the United States. 2nd ed., rev. Misc. Publ. No. 1391. Washington, D.C.: U.S. Department of Agriculture. 108 p. + map.

Cavender-Bares, J.; Ackerly, D.D.; Baum, D.A.; Bazzaz, F.A. 2004. Phylogenetic overdispersion in Floridian oak communities. American Naturalist. 163(6):823-843.

Cleland, D.T.; Freeouf, J.A.; Keys, J.E. [and others], 2005. Ecological Subregions: Sections and Subsections for the conterminous United States.[CD-ROM]. Washington, D.C.: United States Department of Agriculture, Forest Service.

Erwin, T.L. 1991. An evolutionary basis for conservation strategies. Science. 253(5021):750-752.

ESRI. 2006. ArcMap 9.2. Redlands, California: Environmental Systems Research Institute Inc.

Faith, D.P. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation. 61:1-10.

Forest Inventory and Analysis Program. 2007. The Forest Inventory and Analysis Database: Database Description and Users Guide Version 3.0. Arlington, Virginia: United States Department of Agriculture, Forest Service. 230 p.

Kembel, S.W.; Hubbell, S.P. 2006. The phylogenetic structure of a neotropical forest tree community. Ecology. 87(7):S86-S99.

Miller, D.A.; White, R.A. 1998. A Conterminous United States Multi-Layer Soil Characteristics Data Set for Regional Climate and Hydrology Modeling. [Online]. http://www.soilinfo.psu.edu/index.cgi?soil_data&conus. [Date accessed: July 28, 2008].

Montréal Process Working Group. 2008. The Montréal Process, third edition. [Online]. http://www.rinya.maff.go.jp/mpci/rep-pub/1995/santiago_e.html. [Date accessed: May 22, 2008].

Potter, K.M. in review. Evolutionary diversity and phylogenetic community structure of forest trees across the conterminous United States. In: Potter, K. M.; Conkling, B. L., ed. Forest Health Monitoring 2008 National Technical Report. Asheville, North Carolina: United States Department of Agriculture, Forest Service, Southern Research Station.

PRISM Group. 2008. 30-arcsec (800 m) gridded 30-year normal climate data (1971-2000). [Online]. http://www.prism.oregonstate.edu/products/matrix.phtml. [Date accessed: July 30, 2008].

Rodrigues, A.S.L.; Gaston, K.J. 2002. Maximising phylogenetic diversity in the selection of networks of conservation areas. Biological Conservation. 105(1):103-111.

Sechrest, W.; Brooks, T.M.; da Fonseca, G.A.B. [and others]. 2002. Hotspots and the conservation of evolutionary history. Proceedings of the National Academy of Sciences of the United States of America. 99(4):2067-2071.

Soltis, P.S.; Gitzendanner, M.A. 1999. Molecular systematics and the conservation of rare species. Conservation Biology. 13(3):471-483.

Strauss, S.Y.; Webb, C.O.; Salamin, N. 2006. Exotic taxa less related to native species are more invasive. Proceedings of the National Academy of Sciences of the United States of America. 103(15):5841-5845.

Tofts, R.; Silvertown, J. 2000. A phylogenetic approach to community assembly from a local species pool. Proceedings of the Royal Society of London Series B-Biological Sciences. 267(1441):363-369.

United States Department of Agriculture Forest Service. 2008. National forest type data development. [Online]. http://svinetfc4.fs.fed.us/rastergateway/forest_type/. [Date accessed: May 13, 2008].

United States Geological Survey. 1996. GTOPO30 - Global Topographic Data. [Online]. http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html. [Date accessed: July 30, 2008].

Webb, C.O. 2000. Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. American Naturalist. 156(2):145-155.

Webb, C.O.; Ackerly, D.D.; Kembel, S.W. 2007. Phylocom: Software for the Analysis of Community Phylogenetic Structure and Character Evolution. Version: 3.41.

Webb, C.O.; Ackerly, D.D.; McPeek, M.A.; Donoghue, M.J. 2002. Phylogenies and community ecology. Annual Review of Ecology and Systematics. 33:475-505.

Webb, C.O.; Gilbert, G.S.; Donoghue, M.J. 2006. Phylodiversity-dependent seedling mortality, size structure, and disease in a Bornean rain forest. Ecology. 87(7):S123-S131.

Wikström, N.; Savolainen, V.; Chase, M.W. 2001. Evolution of the angiosperms: Calibrating the family tree. Proceedings of the Royal Society of London Series B-Biological Sciences. 268(1482):2211-2220.

Fine Scale Mapping Applications



Combining FIA Plot Data with Topographic Variables: Are Precise Locations Needed?

Stephen P. Prisley[1], Huei-Jin Wang[2], Philip J. Radtke[3], and John Coulston[4]

***Abstract**: Plot data from the USFS FIA program could be combined with terrain variables to attempt to explain how terrain characteristics influence forest growth, species composition, productivity, fire behavior, wildlife habitat, and other phenomena. While some types of analyses using FIA data have been shown to be insensitive to precision of plot locations, it has been suggested that terrain-based models may require the use of precise plot coordinates. This study compares results obtained from a variety of terrain-based analyses conducted in the Blue Ridge of North Carolina using both precise and perturbed (fuzzed and swapped) FIA plot locations, and documents differences between field-estimated slope and aspect and GIS-derived slope and aspect. Digital elevation model (DEM) data were used to derive simple topographic parameters such as elevation, slope percent, azimuth of aspect, terrain curvature, flow accumulation, slope position, and compound topographic index. These values were then compared in a pairwise fashion for plots using precise and perturbed coordinates. Correlations between precise and perturbed plot locations ranged from r = -0.006 to r = 0.383, except for precise versus perturbed plot elevations where r = 0.929. Second, a simple, terrain-based forest site quality index (FSQI) was calculated for the each plot. This index defines site quality classes for forest productivity based on azimuth of aspect, slope percent, and slope position. FSQI classifications were compared for precise and perturbed plot coordinates; at best only 40% of plots resulted in the same productivity class (out of 5). Finally, field-obtained estimates of slope and aspect were compared with GIS-derived estimates from precisely-located plots to assess their level of agreement. Correlations between field-measured and GIS-derived values were r = 0.6 for slope and r = 0.4 for aspect. Results of these experiments indicate that perturbed plot locations may not be suitable for such fine-scale applications.*

Keywords: terrain, slope, aspect, elevation, site quality, productivity

Introduction

Forest scientists have known for decades that strong linkages exist between forest productivity and site conditions such as topography. Many efforts have attempted to quantify these linkages so as to be better able to model and predict

[1] Associate Professor, Dept. of Forestry, Virginia Tech, Blacksburg, VA 24060; prisley@vt.edu
[2] Graduate Research Assistant, Dept. of Forestry, Virginia Tech, Blacksburg, VA 24060; wanghj@vt.edu
[3] Associate Professor, Dept. of Forestry, Virginia Tech, Blacksburg, VA 24060; pradtke@vt.edu
[4] Head, Methods and Techniques Research, USDA Forest Service, Southern Research Station FIA, Knoxville, TN 37919. jcoulston@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

forest site productivity for assessment and management (Davis and Goetz 1990, Bolstad et al. 1998, Franklin 1995). The availability of advanced spatial analysis software and consistent, reliable, national coverage of digital elevation model (DEM) data has enhanced our ability to characterize and quantify topographic conditions at locations where productivity estimates are available. It is natural, then, to look to FIA data as source for consistent estimates of forest productivity across a large geographic area.

A current research effort at Virginia Tech is attempting to evaluate and augment the southern variant of the Forest Vegetation Simulator (FVS) for southeastern mixed forests. As part of this project, productivity data from FIA plots are being related to topographic conditions which are thought to be drivers of forest growth. However, comparison of productivity measures from FIA plots and topographic conditions at the plot locations must be conducted at a relatively fine spatial scale, on the order of 10 to 30 meters, the resolution of the most widely-used DEM data from US Geological Survey.

For a variety of compelling reasons, publicly available FIA data do not report actual plot coordinates. A mechanism called "fuzzing" adds a random error (up to about 1.6 km) to the plot location, and a subset of plots are "swapped" with other plot locations (LaPoint 2005; Guldin et al. 2006). This process of fuzzing and swapping are referred to here collectively as perturbing (McRoberts et al. 2005). Several authors have investigated the reliability of results obtained from perturbed plot locations relative to actual (hereinafter referred to as "precise") plot locations.

McRoberts and others (2005) discussed the impacts of plot location perturbation on model-based and design-based estimation procedures. They note the effects of perturbed plot locations decrease as the size of the sampling unit increases and the spatial autocorrelation increases. The authors propose a variety of ways in which the FIA program may help avoid modeling problems with perturbed plot locations, such as providing a variety of model-based maps of estimates of forest attributes that users could access via the Internet.

Coulston and others (2006a) evaluated biomass estimates derived from kriging and residual kriging at FIA plots in Minnesota. They noted no difference between kriged estimates of biomass from the perturbed and precise plot locations. It should be noted, however, that the only variable obtained from a disparate spatial dataset was leaf area index (LAI), which came from 1-km MODIS imagery. Therefore, it is unlikely that the independent LAI variable showed much difference between perturbed and unperturbed locations.

In another sample application, Guldin and others (2006) computed inventory parameters such as forest area, numbers of live and growing stock trees, and volume from circular woodsheds of varying radii, using perturbed and precise plot locations. The only variation in the analysis, therefore, would be which plots fell into or out of the compact circular woodsheds because of location perturbation.

The authors reported that differences were trivial. A biomass prediction case study involving additional spatial layers indicated that the model developed from perturbed coordinates was no worse than the low performance from the model from precise coordinates ($R^2 = 0.43$). The authors concluded that perturbed FIA data can be used with confidence for similar applications.

In a subsequent article, Coulston and others (2006b) used simulation to develop spatial layers of different resolutions and levels of spatial autocorrelation. These layers were then used in kriging and linear regression models in which the dependent variables came from FIA plots, and models were compared between precise and perturbed locations. For kriging, no differences were noted. For regression, the authors noted that perturbed locations affected model R^2, and that the affects were most evident at finer resolutions and in datasets with lower levels of spatial autocorrelation. Furthermore, differences were most pronounced in models that had higher initial R^2. The authors suggest that regression modeling is only appropriate with very coarse resolution datasets (1-2 km).

The research underway at Virginia Tech differs substantially from most of the applications reported above. In our efforts, forest productivity estimates from FIA data are being linked to terrain characteristics which may change dramatically over distances that are very small relative to the scale at which FIA plots are perturbed. Our situation is most similar to the fine-scale, low autocorrelation linear regression scenarios reported by Coulston and others (2006b) in which perturbed plot locations resulted in substantially poorer model performance. Therefore, we hypothesize that the terrain characteristics extracted from GIS layers at perturbed plot locations will differ substantially from the conditions present at precise plot locations. Such differences would likely prevent adequate prediction of productivity from terrain conditions.

In order to conduct this research, the Virginia Tech Department of Forestry entered into a Memorandum of Understanding with the US Forest Service as part of the Privacy Policy Study Group in 2005. This agreement provided limited access to precise plot coordinates in a closely regulated setting. As part of the agreement, Virginia Tech is required to report differences that would be obtained from perturbed versus precise plot locations. This paper reports our evaluation procedure and documents the results.

Methods

Study Area

This study was conducted using data from the mountain FIA unit in western North Carolina (Figure 1). This area comprises 21 counties covering approximately 17,870 square kilometers. Most of the study area is in the Blue Ridge province of the southern Appalachian Mountains. Elevations in the study area range from 266 to 2033 m.

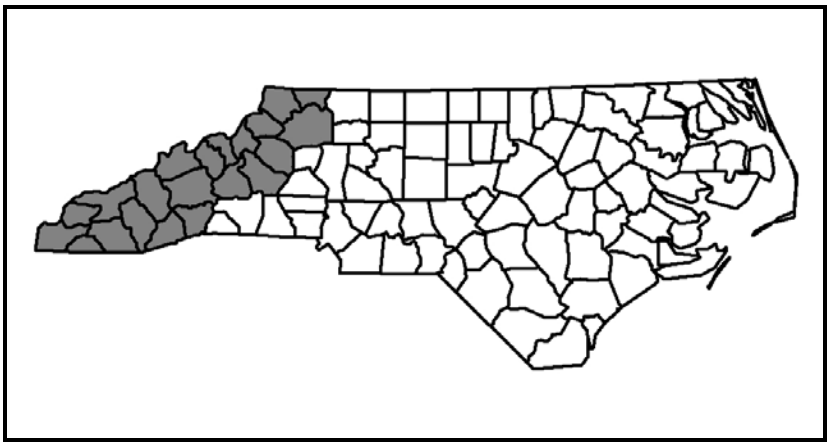


Figure 1: Map of the study area, the mountain FIA unit in western North Carolina.

Terrain Data

Digital elevation data were obtained from the US Geological Survey "seamless" web site (http://seamless.usgs.gov) and consisted of 10m grid cells containing elevation in meters. From this dataset, additional terrain datasets were developed using functions in ArcGIS 9.2 software. These additional layers included slope percent, azimuth of aspect, terrain shape index, terrain curvature, and flow accumulation. Slope percent was computed with the ArcGIS slope tool, which applies the Horn (1981) algorithm using elevations at eight adjacent cells. Azimuth of aspect and terrain curvature were calculated using the default ArcGIS algorithms. Because of the difficulties in dealing with circular data, azimuth of aspect was also transformed using a cosine transformation. Terrain shape index (TSI), an indicator of local landform convexity or concavity (McNab 1989), was computed as the difference between the elevation at a cell and the average elevation of a 35m circular neighborhood. High positive values represent areas of convexity such as ridges, and more negative values represent concave landforms such as coves or drainages. Flow accumulation is a hydrological-based indicator of landscape position. The flow accumulation algorithm computes the number of grid cells whose runoff would flow eventually through a given cell. Low numbers represent areas near ridge crests and high numbers represent valleys, bottoms, and drainageways.

Two additional terrain descriptors were also derived: slope position and forest site quality index (FSQI). To quantify the relative positions of points on the landscape, an indicator of slope position was required. Slope position was defined as the percentage of the flow path distance from a stream to a ridge. To obtain this metric, the flow accumulation layer was used to define streams, which were then masked from the analysis. The flow length tool in ArcGIS was then used to determine both the uphill and downhill flow length for each cell. Uphill flow length is the number of cells along the flow path uphill to a ridge. Downhill flow length is the number of cells along the flow path downhill to a stream. Slope position for a grid cell is then calculated as the downhill flow length at that cell

divided by the sum of uphill and downhill flow lengths. The resulting value for slope position (0 -1.0) is then reclassified into codes for six landscape positions: summit, shoulder, backslope, footslope, terrace, and floodplain (Cotton et al. 2008).

Forest site quality index is an ordinal value integrating slope, aspect and the slope position class defined above (Meiners et al. 1984). The FSQI values have been shown to be correlated with upland oak site index, and are being applied in ongoing forest productivity research. FSQI scores were developed using the slope, aspect, and slope position class layers developed for western North Carolina. Scores were categorized into five classes corresponding to site index ranges (Cotton et al. 2008).

FIA Data

FIA plot data from cycle 7 (ca. 2002) were obtained from the Southern Research Station for North Carolina. Both published (perturbed) and actual (precise) coordinates for each plot were obtained. A total of 1,022 plots were used in this analysis.

Analysis Approach

A spatial dataset of perturbed and precise plots was created from the FIA tables containing coordinates of plot locations. All data were projected to the UTM Zone 17, NAD 83 coordinate system and plots were overlaid with the raster terrain datasets, extracting cell values for each plot location from each layer. Scatterplots and correlation estimates were produced to compare terrain values for perturbed versus precise locations. Categorical FSQI site index classes for perturbed and precise plot locations were compared in a contingency table.

Results

Perturbed versus Precise Coordinates

The influence of perturbed locations on terrain-based models was examined by comparing topographic variables derived from precise versus perturbed coordinates. Despite the trivial contribution to bias in topographic GIS derivatives, perturbed locations did influence the precision of derivatives (Figure 2 and Table 1). Dispersion of differences between the two plot locations had a wide range and its corresponding standard errors were far from zero. Examples of observed discrepancies between precise and perturbed plots illustrate these differences: (1) aspect obtained from a precise coordinate was toward north with azimuth 7°, but that obtained from a perturbed coordinate was toward south with azimuth 175°; (2) a precise location has a steep slope of 99%, but its perturbed location has a relatively flat slope of 2%. Maps of locations with similar terrain discrepancies are depicted in Figure 3 (without actual plot locations). The

strength of topographic correlation was weak or none (r = 0 – 0.383) between perturbed and exact locations, except the elevation variable (r = 0.929).

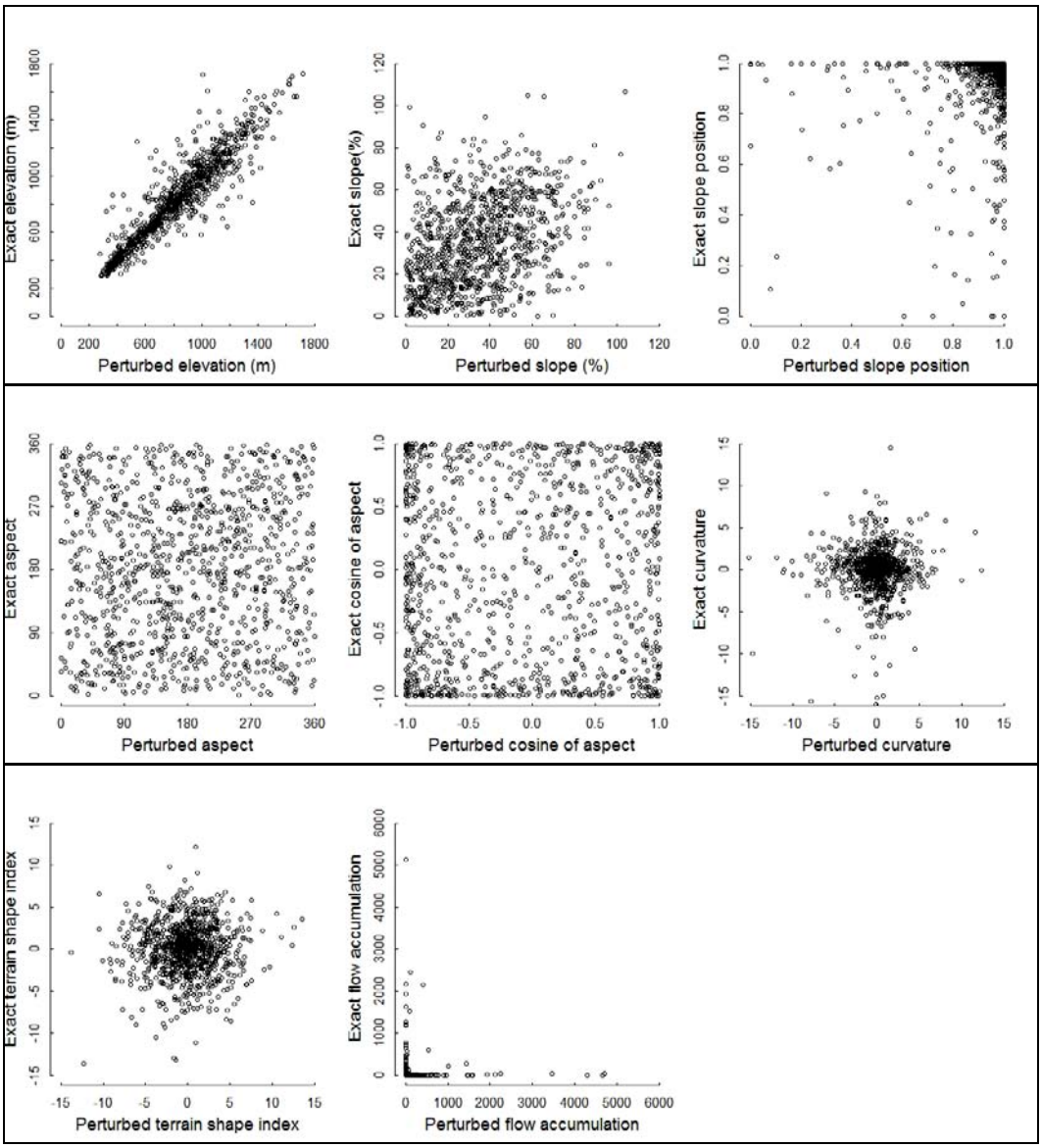


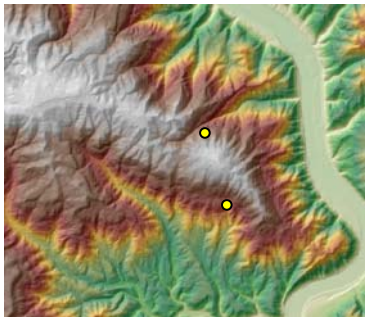
Figure 2: Scatterplots of terrain variables for perturbed and precise coordinates; incuding elevation, slope, slope position, aspect, cosine of aspect, curvature, terrain shape index, and flow accumulation (sample size = 1022).

Forest site quality index (FSQI) consists of scores from 3 to 16. Only 15.4% of plots had the same score for perturbed and precise locations, resulting in a correlation of 16.7%. When aggregated to five classes, FSQI exhibited only a 40% overall agreement (Table 2). In this agreement analysis, no evidence existed to support the assertion that perturbed coordinates could provide sufficient information in evaluating site quality (Table 2). After chance agreement was excluded (using the kappa statistic), the agreement rate was 5%.

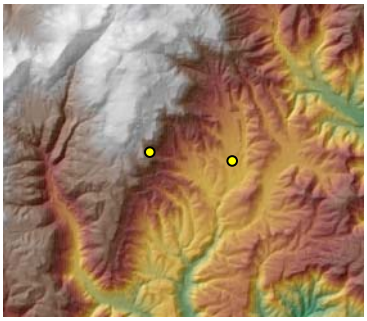
Table 1: The performance of perturbed coordinates is evaluated by comparing to GIS-derived terrain values from precise coordinates.

	r	*Bias*	*SE* [a]	$\sqrt{MSE}$ [b]
Elevation (m)	0.929	6.06	112.9	113.06
Slope (%)	0.383	-0.18	23.14	23.14
Slope Position (%)	0.201	~ 0	0.18	0.18
Aspect (°)	0.044	3.25	139.63	139.67
Cos(aspect)	0.093	0.03	0.95	0.95
Curvature	0.038	0.09	3.53	3.53
Terrain Shape	0.036	0.13	4.37	4.37
Flow Accumulation	-0.006	-49.07	2811.69	2812.12

[a] *SE*: Standard Error
[b] *MSE*: Mean Square Error



Aspect difference



Slope difference

Figure 3: Examples of results for aspect and slope based on perturbed and precise coordinates. Perturbed and precise plots differed in aspect by approximately 170 degrees; perturbed and precise plots differed in slope by 97%. These maps depict similar terrain conditions but do not indicate real plot locations.

Table 2: The contingency table of forest site quality index (FSQI) for perturbed versus exact coordinates: FSQI is computed based on GIS-derived aspect, slope, and slope position for each plot. 1022 plots are used to evaluate the FSQI agreement between perturbed and actual coordinates. Overall agreement of FSQI classes between perturbed and precise coordinates is 40%. The proportion of FSQI agreement is 0.05 (kappa statistic), after chance agreement is excluded.

		Precise coordinates				
	FSQI	3.0-4.5	4.5-7.5	7.5-10.5	10.5-13.5	13.5-16.0
Perturbed coordinates	3.0-4.5	1	3	1	4	1
	4.5-7.5	0	53	78	31	3
	7.5-10.5	3	84	254	166	12
	10.5-13.5	0	42	152	98	10
	13.5-16.0	0	4	10	6	2

Field-Measured versus GIS-Derived Variables

Field-obtained slope and aspect values were available for 767 FIA plots for comparing the performance between field and GIS-derived slope and aspect using precise coordinates. The linear association between field and GIS-derived variables was stronger than the correlation between perturbed and precise plot

locations. For slope, $r = 0.611$ and for aspect $r = 0.405 \sim 0.546$ (Figure 4). Only 53% of plots had GIS-derived slope within 10% of the field measurement (±10% is the MQO for subplot slope). For aspect, only 21% of plots had a GIS-derived value within 10 degrees of the field measurement (±10° is the MQO for aspect).

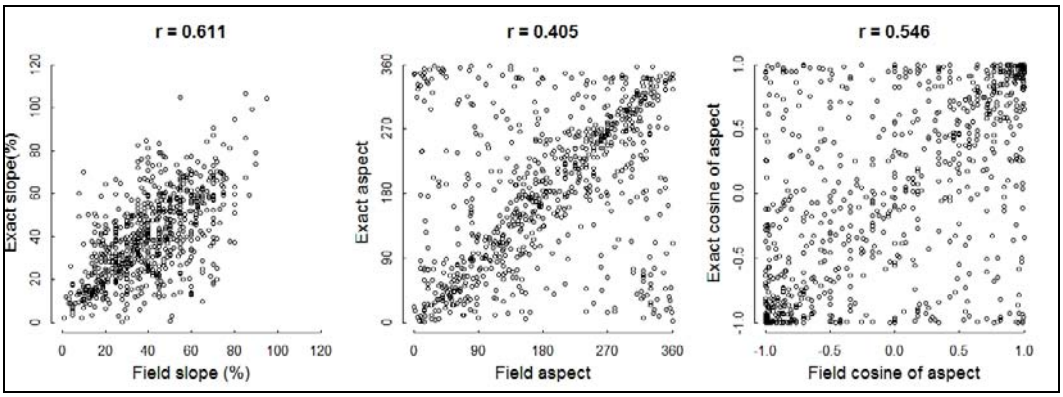


Figure 4: Linear correlations between field-derived and GIS-derived values for slope, aspect, and cosine of aspect (sample size = 767).

Discussion and Conclusions

We know that microscale variation in landform and terrain affects the type and quality of the vegetation in mountainous regions, largely due to temperature and soil moisture limiting to plant growth. Terrain information can be easily derived through GIS procedures because of widely available digital elevation models, and terrain derivatives can improve predictive models (Davis and Goetz 1990). The perturbed coordinate system, however, cannot provide useful information derived from fine-scale digital elevation data. In addition, it is unlikely that FIA can provide all possible terrain variables from field or even GIS measurements: the variety of terrain-related, GIS-derived variables currently used in ecological and hydrological applications is large and growing. Thus, it appears that FIA data in publicly-available form (perturbed locations) are not useful in conjunction with fine-scale spatial applications.

It also appears that the terrain variables that are currently measured in the field (slope and aspect) do not correlate with GIS-derived variables as strongly as we might hope. The algorithms broadly used in GIS-based terrain analysis do not measure the same things as are measured in FIA field procedures. For example, slope is measured in the field using a percent-scale clinometer, observing from the uphill to downhill edges of a subplot. The default GIS approach computes a weighted average gradient in N-S and E-W directions. Therefore, it is not appropriate to interchange field slope/aspect with GIS-derived slope/aspect in either developing or applying predictive models.

Clearly, modeling forest productivity or other phenomena that vary at spatial scales on the order of currently available digital elevation models cannot be adequately performed with the publicly-available FIA plot locations.

Researchers pursuing such endeavors must find other avenues to accomplish this research.

Opportunities for spatial modeling with FIA data with ancillary geospatial data have increased proportional to the availability of geospatial data (e.g. DEMs). These opportunities will continue to increase as more geospatial information at higher resolution becomes available. However, the usefulness of the ancillary data is related to the accuracy of the precise plot locations. Although errors in precise locations were not considered in this analysis, these errors create additional uncertainty for the types of analyses presented here. For example, the FIA program typically uses recreation-grade GPS units to collect coordinates of each inventory plot. These types of GPS units have under-canopy locational accuracies of approximately 7.5 m (Bolstad et al. 2005). We recommend that the accuracy of the precise location should be available to clients who are using these precise coordinates. This information is particularly important when conducting research to develop models based on fine resolution (e.g. 10 m) geospatial data and FIA plot attributes. To keep pace with the increased resolution of ancillary geospatial data, we also recommend that the FIA program adopt new, more accurate, GPS technology as it becomes available and affordable.

Acknowledgements

This research was supported by the USDA Forest Service, Southern Research Station under Agreement No. SRS 05-CA-11330134-251. The authors gratefully acknowledge the support of the Southern Research Station, specifically Cathryn Greenberg and David Loftis. We also acknowledge valuable assistance from Chad Keyser of the Forest Management Service Center, Bill Burkman and Sam Lambert of the SRS FIA unit, and Claudia Cotton and Spencer Riddile of the Virginia Tech Department of Forestry.

Literature Cited

Bolstad, P., Jenks, A., Berkin, J., Horne, K. 2005. A comparison of autonomous, WAAS, real-time, and post-processed global positioning system (GPS) accuracies in northern forests. Northern Journal of Applied Forestry 22(1). 5-11.

Bolstad, P.V., Swank, W., and J. Vose. 1998. Predicting southern Appalachian overstory vegetation with digital terrain data. Landscape Ecology 13:271-283.

Cotton, C.; Prisley, S.; Fox, T. 2008. Mapping upland hardwood site quality and productivity with GIS and FIA in the Blue Ridge of North Carolina. In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2008. Forest Inventory and Analysis (FIA) Symposium 2008; October 21-23, 2008; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Coulston, J.W.; Reams, G.A.; McRoberts, R.E.; Smith, W.B. 2006a. Practical considerations when using perturbed forest inventory plot locations to develop spatial models: a case study. In: McRoberts, R.E.; Reams, G.A.; VanDeusen, P.C.; McWilliams, W.H., eds. 2006. Proceedings of the sixth annual forest inventory and

analysis symposium; 2004 September 21-24; Denver, CO. Gen. Tech. Rep. WO-70. Washington, D.C.: U.S. Department of Agriculture, Forest Service. 126 p.

Coulston, J.W.; Ritters, K.H.; McRoberts, R.E.; Reams, G.A.; Smith, W.B. 2006b. True versus perturbed forest inventory plot locations for modeling: a simulation study. Canadian Journal of Forest Research 36: 801-807.

Davis, F.W., Goetz, S., 1990. Modeling vegetation pattern using digital terrain data. Landscape Ecology 4, 69-80.

Franklin, J. 1995. Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography 19:474-499.

Guldin, R.W.; King, S.L.; Scott, C.T. 2006. Vision for the future of FIA: paean to progress, possibilities, and partners. In: McRoberts, R.E.; Reams, G.A.; VanDeusen, P.C.; McWilliams, W.H., eds. 2006. Proceedings of the sixth annual forest inventory and analysis symposium; 2004 September 21-24; Denver, CO. Gen. Tech. Rep. WO-70. Washington, D.C.: U.S. Department of Agriculture, Forest Service. 126 p.

Horn, B.K.P. 1981. Hill shading and the reflectance map. Proc. IEEE 69(1):14-47.

LaPoint, E. 2005. Access and use of FIA data through FIA spatial data services. In: McRoberts, R.E.; Reams, G.A.; VanDeusen, 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.

McNab, W. 1989. Terrain shape index: quantifying the effect of minor landforms on tree height. Forest Science 35:91-104.

McRoberts, R.E.; Holden, G.R.; Nelson, M.D.; Liknes, G.C.; Moser, W.K.; Lister, A.J.; King, S.L.; LaPoint, E.B.; Coulston, J.W.; Smith, W.B.; Reams, G.A. 2005. Estimating and circumventing the effects of perturbing and swapping inventory plot locations. Journal of Forestry 103(6): 275-279.

Meiners, T.M., D.W. Smith, T.L. Sharik, and D.E. Beck. 1984. Soil and plant water stress in an Appalachian oak forest in relation to topography and stand age. Plant and Soil 80:171-179.

Using interpreted large scale aerial photo data to enhance satellite-based mapping and explore forest land definitions

Tracey S. Frescino[1], Gretchen G. Moisen[2]

Abstract: *The Interior-West, Forest Inventory and Analysis (FIA), Nevada Photo-Based Inventory Pilot (NPIP), launched in 2004, involved acquisition, processing, and interpretation of large scale aerial photographs on a subset of FIA plots (both forest and nonforest) throughout the state of Nevada. Two objectives of the pilot were to use the interpreted photo data to enhance satellite-based mapping capabilities by providing training information at a moderate scale, and to refine definitions of forest land by facilitating explorations of forest land definition changes. This study examines the usefulness of NPIP data for fulfilling these objectives by exploring relationships between photo-interpreted information and FIA ground data with Moderate Resolution Imaging Spectroradiometer (MODIS) data using machine learning, Random Forests models. Photo-interpreted data were determined to be valuable training data for models using MODIS imagery, even in areas of low tree cover, and were found to be slightly more effective than using field data.*

Keywords: Nevada, photo interpretation, large scale photography (LSP), MODIS

INTRODUCTION

With increasing pressure for natural resource information and concurrent decreases in available funds, there is a need for exploring alternatives to the time-consuming, high cost ground data traditionally collected by inventories, such as the USDA Forest Inventory and Analysis (FIA) program. In 2004, the Interior-West FIA program launched the Nevada Photo-based Inventory Pilot (NPIP) to

[1] United States Forest Service; Rocky Mountain Research Station; Forest Inventory and Analysis Program; 507 25th Street; Ogden, UT 84404 USA; tfrescino@fs.fed.us

[2] United States Forest Service; Rocky Mountain Research Station; Forest Inventory and Analysis Program; 507 25th Street; Ogden, UT 84404 USA; gmoisen@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

explore the use of GPS-controlled, large-scale aerial photography (LSP) for enhancing our existing inventory by complementing our field sample, augmenting our traditional FIA estimation process, providing strategic level information on nonforest lands not traditionally inventoried by FIA, and increasing efficiencies, particularly in the extensive, slow-growing woodland forests in the West.

NPIP involved acquisition and processing of LSP on a subset of FIA plots (both forest and non-forest) throughout the state of Nevada during two consecutive field seasons. Nevada was a particularly good study area for several reasons: it was not yet funded for annual inventory (Gillespie 1999); it has the most incomplete and outdated periodic data in the Interior West; it is predominantly nonforest federal lands; and of the forested lands, it is mostly woodland forest types. The specific objectives of the pilot were as follows:

- ***Exceed information requirements*** by providing strategic level information on all cover type groups in a consistent fashion. LSP provides an opportunity to characterize vegetated and non-vegetated cover types typically not sampled by FIA, such as shrublands and grasslands.

- ***Speed up inventory timeline*** by producing meaningful estimates after 2 years of data collection, at a lower cost. The data from NPIP can supplement the sparse annual data collected in the state to produce more precise and timely estimates of field-based attributes, such as area by forest type, area by owner, and area by land class, as well as basal area by tree species group.

- ***Reduce inventory costs*** in prefield and potentially in future field efforts particularly on marginal forest lands. IW-FIA's ground sample is based on *a priori* interpretation of aerial digital imagery (e.g. DOQ). LSP provides higher resolution imagery and improved temporal control for making better decisions and thus, eliminating unnecessary field visits and reducing costs.

- ***Enhance mapping capabilities*** by providing training information at a moderate scale. In the past, FIA map products have relied on the acre-size, FIA plot data for training moderate-scale, and satellite-based map products of forest attributes. NPIP photos offer the opportunity to collect training data at a scale that is comparable to the satellite imagery which should improve accuracy of these map products.

- ***Provide data to evaluate*** forest land definition changes. FIA needs to define and account for all forest land in the U.S. in a consistent and efficient manner. NPIP photos can provide data to explore the impact of definitional changes of forest land on forest estimates, using criteria such as percentage of tree crown cover.

The focus of this study is based on the last 2 objectives of the pilot: enhancing mapping capabilities and evaluate definitions of forest land. The study explores

relationships between photo-interpreted information and Moderate Resolution Imaging Spectroradiometer (MODIS) data, examining photo data as compared to FIA field data and analyzing different thresholds of tree cover interpreted from the photo plots. This paper first describes the data available, photo, field, and MODIS imagery, and then how these data were used for addressing these 2 objectives.

DATA

Photo Data

GPS-controlled LSP was collected in 2004 and 2005 in the state of Nevada, consisting of 28,629,728.6 hectares (70,745,600 acres) (http://dcnr.nv.gov/nrp01/land01.htm), including water but excluding 3,514,290 hectares (8,684,000 acres) of restricted air space. Nevada is less than 14 percent forested and of these forested lands, the dominant forest type is pinyon-juniper woodland type, covering approximately 7 million acres (Born et al. 1992). Other woodland types found in Nevada include *Cercocarpus* species and Gambel oak (*Quercus gambelii*). Timberlands cover less than 1 percent of the total forest land and include white fir, aspen, and limber pine with bristlecone pine, subalpine fir and Engelmann spruce (*Picea engelmannii*) at higher elevations and black cottonwood (*Populus trichocarpa*) in the riparian areas (Born et al. 1992).

The sample survey design for NPIP follows the systematic sampling design of the national FIA program (Reams et al. 2005). The FIA program conducts a comprehensive inventory of forest lands across all ownerships in the United States. Permanently-established ground plots are measured annually based on a systematic sample of regularly spaced hexagons, each representing approximately 6000 acres. The plots are delineated into 5 panels, each panel represents 20 percent of the data and is measured on an annual cycle. In the western region, panels are divided again into subpanels; each subpanel representing half a panel or 10 percent of the data. A subpanel is measured every year over a ten year cycle.

The state was pre-stratified into 3 initial strata using a pixel-based, 250-meter resolution map of predicted timberland forest, woodland forest, and nonforest areas (Figure 1; Blackard et al. 2004). All FIA locations (i.e., all 10 subpanels) within the timberland and woodland strata were photo-sampled, totaling 1,455 plots. In addition, one-tenth of the FIA locations (i.e., one subpanel) within the nonforest stratum were photo-sampled, totaling 877 plots, resulting in an overall total of 2,332 photo plots.

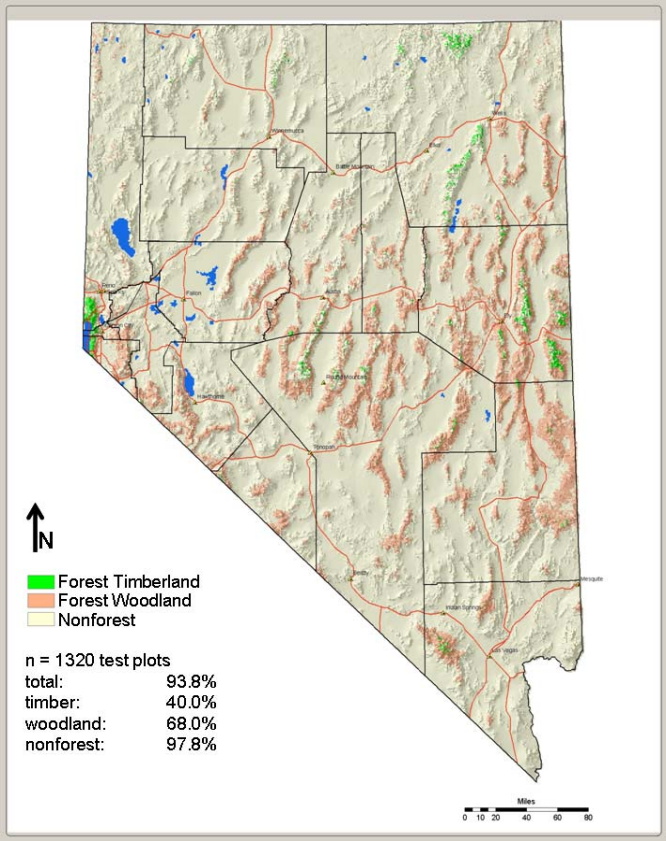


Figure 1: Stratification map.

In 2004, 395 locations were acquired by contract with the Remote Sensing Application Center (RSAC) using a direct-to-digital camera, DCS645C, with a 55-mm lens. Flights were flown 3,000 feet above ground with a 2,002-ft swath width, resulting in a 0.49-ft ground sample distance (GSD). In 2005, a contract through the USDA Aerial Photography Field Office (APFO) to Aerial Services, Inc. (ASI) provided photos for 1,937 locations. The ASI photography used natural color 9x9-inch film and a 6-inch lens. These plots were flown at a scale of 1:5,000 feet with a 3,750-ft swath width. Stereo triplicate photographs were acquired for all plots, converted to a digital Tagged Image File Format (TIFF), and geolocated to each FIA X-Y plot center location.

The NPIP photo-interpreted plot sample design consisted of a dot grid sample within a 250 meter radius circle covering approximately 20 hectares (48 acres) of land. There were a total of 49 points per plot each representing about an acre size (Figure 2) with the center point of each plot straddling the FIA field plot center. Point generation and photo interpretation were accomplished using the Digital Mylar Image Sampler tool developed by the Remote Sensing Applications Center (RSAC) (Clark et al. 2004). Each point was assigned a value of condition class, defined as an area of homogenous vegetation having similar characteristics (USDA 2006), and an object class identifying the object in the photograph the point fell on, such as a tree or shrub (USDA 2006).

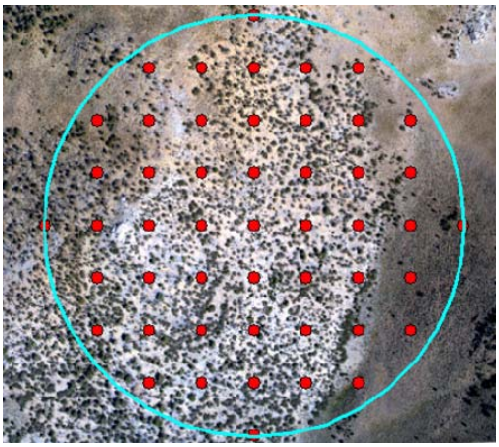


Figure 2: NPIP plot sample design.

The individual point data were summarized to plot level and condition level attributes. One attribute analyzed in this study was percent cover of live trees, calculated by summing the points that fell on live trees and dividing by the number of points per plot or condition, respectively. More details on photo sampling procedures are documented in Frescino et al., In Press.

FIA Field Data

As mentioned previously, the FIA sampling design is based on a nationally consistent and uniform spatial distribution of field plots divided into annual panels across the United States. In the Interior-West, where and when funding is available, a subpanel of data is measured each year, representing 10 percent of the total plots. In Nevada, 2 subpanels (20% of the entire sampling grid) of data were collected in years 2004 and 2005, totalling 1216 plots; of these, 381 plots sampled forest land. Live tree crown cover was measured using four 25-ft transects and included all live trees 1.0 inch and greater (USDA 2006). These data were collected at the condition level and summed to plot level.

MODIS Data

We used 16-day, cloud-free, composites of MODIS imagery for spring, summer, and fall of 2005. These included visible-red (RED) and near-infrared (NIR) bands and 2 vegetation indices, normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI). NDVI measures energy absorption of chlorophyll and cell wall reflectance in vegetation cover and is defined as (NIR – RED) / (NIR + RED) (Tucker 1979). NDVI is powerful in its reduction of multiplicative noise among bands but is sensitive to canopy background changes

(Huete et al. 2002). EVI incorporates the blue visible spectral band, helping to reduce this sensitivity in background noise while improving sensitivity in high biomass regions. EVI is defined as G*(NIR – RED) / (NIR + C_1*RED – C_2*BLUE+ L), where BLUE is visible-blue band, C_1 and C_2 are aerosol resistance coefficients, G is a gain factor, and L is the canopy background adjustment factor (Huete et al 2002).

METHODS

Random Forests

In meeting both our objectives of enhancing mapping capabilities and evaluating definitions of forest land, we used a machine learning algorithm, Random Forests (Breiman, 2001), to model relationships between response variables collected on photos or field plots, and the MODIS data. Random Forests is a classifier that generates multiple classification predictive models referred to as trees, and outputs the class that occurs most frequently. For each tree a random subset of the training data (approximately two thirds) is used to construct the tree, with the remaining data points used to construct out-of-bag (OOB) error estimates. For categorical responses, the OOB error is the average number of times the class was not correctly classified, yielding OOB estimates of percent correctly classified (PCC) and Kappa, an index that compares agreement against that which might be expected by chance. For a continuous response, the output includes the total percent variance explained, root mean square error (RMSE), and Pearson's correlation (Cor). At each node of each tree, a random selection of predictors is chosen to determine the split. Random Forests will not overfit data, therefore the only penalty of increasing the number of trees is computation time. For this study, we used 500 trees to build the models and relied on the OOB error estimates to compare different models.

Enhance Mapping Capabilities

The first objective addressed in this paper was enhancing mapping capabilities by providing training information at a moderate scale, compatible with the MODIS data. Recent mapping efforts, regionally (Blackard et al. 2004; Ruefenacht et al. 2004) and nationally (Blackard, et al. 2008; Ruefenacht et al. 2008), have coupled FIA data with MODIS imagery with the assumption that the acre-size (4047 sq meter) FIA sample plot is representative of the 15-acre (250-sq

meter) pixel size of MODIS. This mismatched scale has not previously been analyzed, thus the impact on the models and map products are unknown. The NPIP sample plot is 500 meters in diameter and therefore provides data for a sampled area that is thought to better match the area of the MODIS pixel. To understand this impact of scale better, we explored relationships between MODIS data and the compiled photo-interpreted data at both the plot and condition levels using Random Forests and compared the output from these models to output from models generated using FIA ground plot data.

We focused on 2 different questions:

1. How does using photo plots as training data differ from using field plots as training data for generating models with MODIS imagery?
2. How does the cost of photos and field training data affect models with MODIS imagery?

For the first question, we considered only those plots that met the following two conditions: 1) both field and photo data were available (hereafter referred to as collocated), and 2) both field and photo plots had only one condition. For the latter we assumed the homogeneity within the plots would minimize noise or model error and therefore the error we were seeing would be related to differences in scale. We modeled 4 different responses, 3 binary and 1 continuous: presence of forestland (F/NF); presence of pinyon-juniper forest type (PJ Type); presence of pinyon-juniper species (PJ Spp); and percent live tree cover. Forestland was defined as lands having 10 percent or greater cover of live trees; pinyon-juniper forest type included plots with a condition classified as pinyon-juniper; pinyon-juniper species included plots having at least one pinyon or juniper tree; and percent live tree cover ranged from 0 to 100 percent of live trees on the plot.

The second question examined how the number of field plots and photo plots affected model error, where the number of plots was based on fixed cost scenarios. Here, we looked at percent live tree cover at plot- and condition- levels from both field and photo plots and modeled this data as a function of MODIS spectral information using Random Forests. The cost, per plot, of the NPIP inventory was approximately $300 (Table 1) compared to an estimated $1500 per FIA field plot. We analyzed 4 different cost scenarios based on these estimates, starting as high as $432,000 and halving the cost each time down to $54,000 (Figure 3). From the total number of available plots in each cost scenario, we examined plot level information of those plots having only one condition and condition-level information of conditions that covered 40 percent or greater area of a plot. These data were randomly selected with replacement for each cost scenario. We generated Random Forests models of live tree percent cover using plot and condition data for each scenario and compared the results of each model.

Table 1: Cost and time constraints for the photo sampling components of NPIP. The costs of per plot estimates.

Description	Cost	Time constraints
Photo acquisition	$120 per triplicate (direct to digital)	Photo acquisition should be initiated before the desired flight month.
Photo-to-digital	$20 (scanning)	
Hard copy printing (8x8)	$5 (per image)	Local, professional photo shop.
Geolocation	$5	
Interpretation	$80 (includes field training) (average > one plot per hour)	Difficult to interpret more than 6 hours per day.
Quality control	$30 (includes data compilation and editing and minimal field validation)	Sharing tasks led to timing and feedback issues.
General cost		
1 Tb of computer storage space	$2,000 (per each; two were purchased, one for back-up)	
Total	$260 per plot + $4,000 storage space	

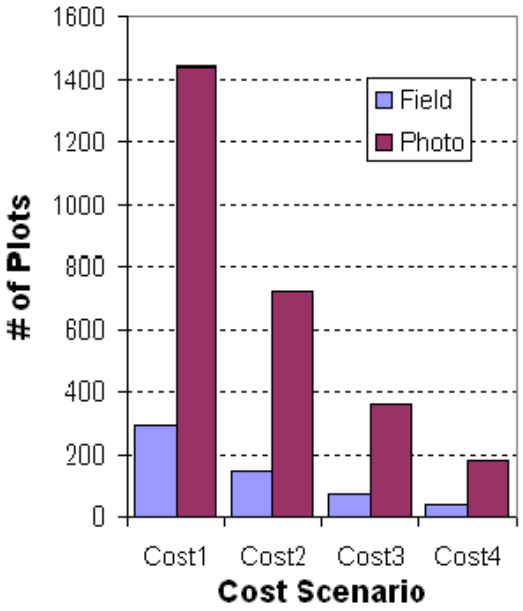


Figure 3: Number of plots used for each cost scenario. Cost1: $432,000; Cost2: $216,000; Cost3: $108,000, Cost4: $54,000

Evaluate Definitions of Forest Land

The second objective addressed in this paper was evaluating definitions of forest land by facilitating explorations of forest land definition changes. Frequent

discussions within FIA review the definition of forest land, including whether to base the definition on stocking or crown cover and what percent crown cover is appropriate for all users. For this objective, we explored what the potential would be to capture different percentages of crown cover using moderate scale imagery. Specifically, we looked at classification accuracies of percent live tree cover interpreted from LSP and MODIS spectral data at different thresholds of cover, 5%, 10%, 15%, 20%, 25%, and 30%. We again used the Random Forests to generate models of each threshold as a binary response and compared the percent correctly classified (PCC) and Kappa values calculated from the OOB error.

RESULTS/DISCUSSION

Enhancing Mapping Capabilities

There were a total of 682 photo and field collocated plots having only one condition. For the photo plots, 118 of these were forested with 115 plots classified as pinyon-juniper forest type and the remaining 564 plots non-forested. Of the field plots, 114 were forested with 112 classified as pinyon-juniper forest type and the remaining 568 plots non-forested. The density, or probability distributions of percent live tree cover for the field and photo collocated, forested plots were very similar as shown in figure 4, with the photo tree cover generally a little higher than the field tree cover, overall. As expected, paired t-tests revealed similar results.

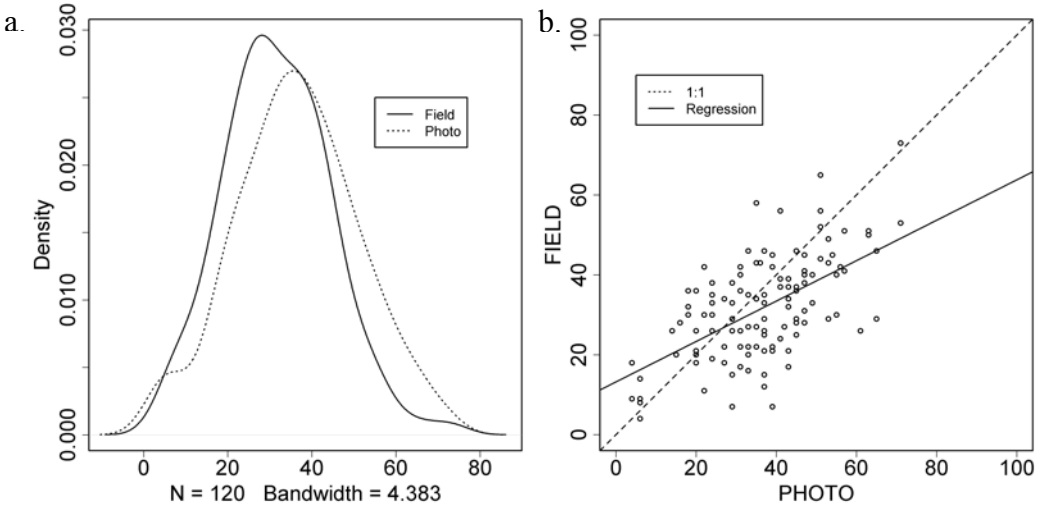


Figure 4: Percent live tree cover for the field and photo collocated, forested plots. a. Density, or probability distributions. b. Scatterplot of photo versus field with 1:1 and regression lines.

The OOB error rates for the field and photo predictions with binary responses are displayed in figure 5. In all cases, the models using photo data had a slightly lower error rate than the models using field data. The continuous response, percent live tree cover, showed similar results for the photo models with greater percent variance explained, a lower RMSE, and a higher correlation. The percent variance explained, RMSE, and correlation for the field model was 77.14, 6.30, and 0.88, relatively, whereas the photo model was 84.67, 5.99, and 0.92, relatively. Figure 6 shows the relationship between the predicted and observed values resulting from the field and photo models, where the observations were greater than zero. Although observed values equal to zero were included in model development, they were excluded from the graphs to get a better picture of the spread for the non-zero values. These results suggest an advantage using photo information at a similar scale as the MODIS pixel over using the acre-size field plot.

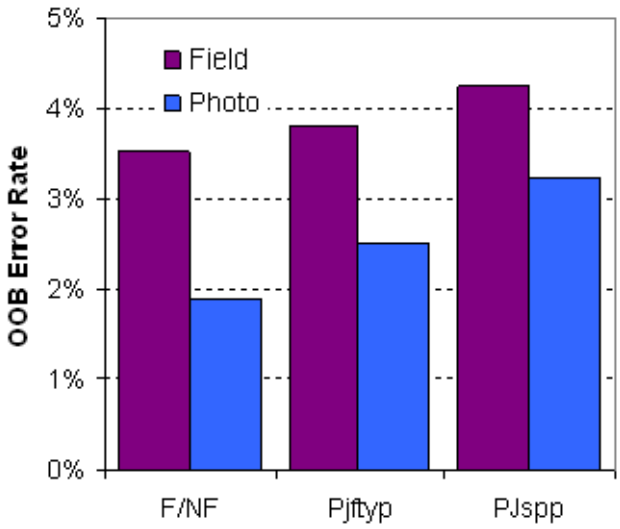


Figure 5: The OOB error rate for Random Forests binary models for field and photo collocated single condition plots.

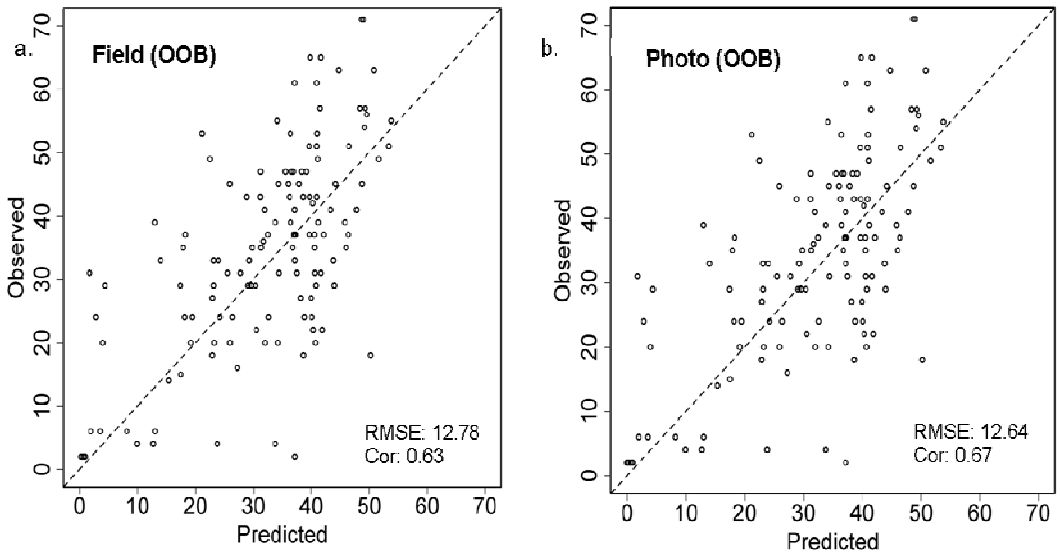


Figure 6: The OOB error for Random Forests model of percent live tree cover for field and photo collocated single condition plots, where observed values are greater than zero.

Table 2 shows the total number of plots used in each cost scenario, the number of single condition plots, and the number of conditions that covered greater that 40 percent of the plot area. Here, it is interesting to note the much higher number of training data available when using condition-level information, especially with the photo plots. The larger size of the photo plots is conducive to sampling more conditions.

The results of the Random Forests models for percent live tree cover under all cost scenarios are displayed in figure 7. This figure shows the OOB variance explained from the photo plot-level (PHOTO-c1) and condition-level (PHOTO-cnd) models along with the field plot-level (FIELD-c1) and condition-level (FIELD-cnd) models.

Table 2: Number of plots used for comparing different cost scenarios.

		Field			Photo		
Cost Scenario	Cost	Total	Single Condition	Conditions >=40%	Total	Single Condition	Conditions >= 40%
1	$432,000	288	248	285	1440	613	1520
2	$216,000	144	123	142	720	306	757
3	$108,000	72	62	71	360	154	375
4	$ 54,000	36	34	36	180	70	190

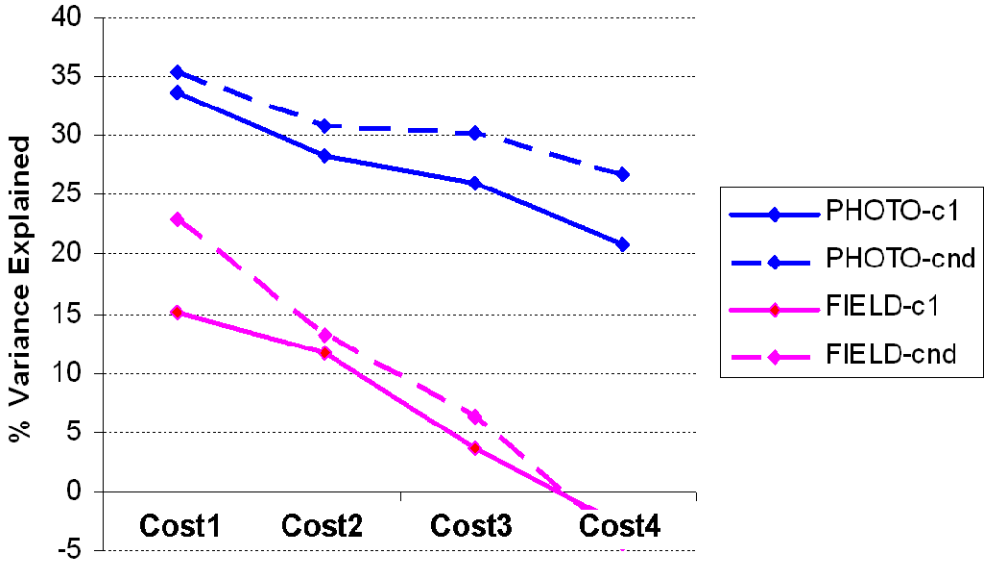


Figure 7: OOB variance explained for percent live tree cover for four cost scenarios, and four plot selection criteria

Overall, the models using photo data explained more variance than the models using field data at the both plot and condition levels. Cost scenario 1, having the highest cost ($432,000) had the highest percent variance explained in all cases, as would be expected. As the scenarios decreased in cost, or in the number of plots used, the variance explained in the models decreased as well. Using the condition-level information seemed to be more effective in each cost scenario with slightly higher percent variance explained in each case. Again, we see the photo data adding valuable information to our models, and at a much lower cost. We also see the positive affect of using condition-level data for training.

Refine Definitions of Forest Land

We used a total of 2328 photo plots from NPIP. Figure 8 shows the division of plots for each forest definition threshold. The OOB PCC and Kappa values resulting from fitting each of these five datasets in a binary forest/nonforest Random Forests model are shown in figure 9. The PCC and Kappa values are fairly high for cover thresholds at 5 percent up to 15 percent but then begin to decline at 20 percent. This result shows that there is potential to successfully model forestland at cover less than 20 percent. Further testing is needed to determine if this is a true effect or an artifact of the training data.

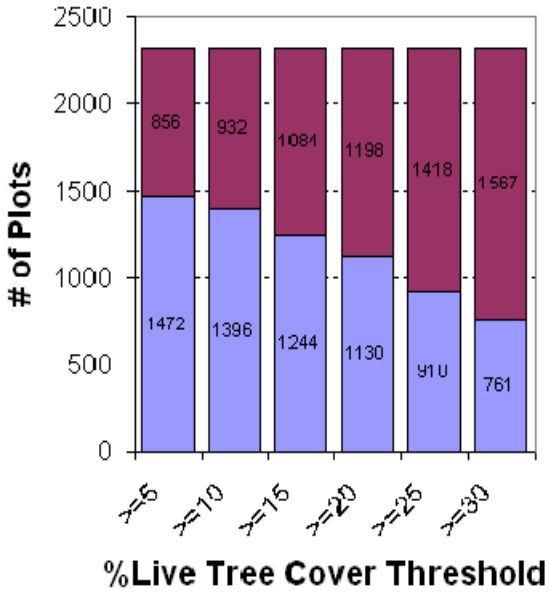


Figure 8: The number of forest (blue) and nonforest (purple) plots by threshold used in the binary Random Forests models.

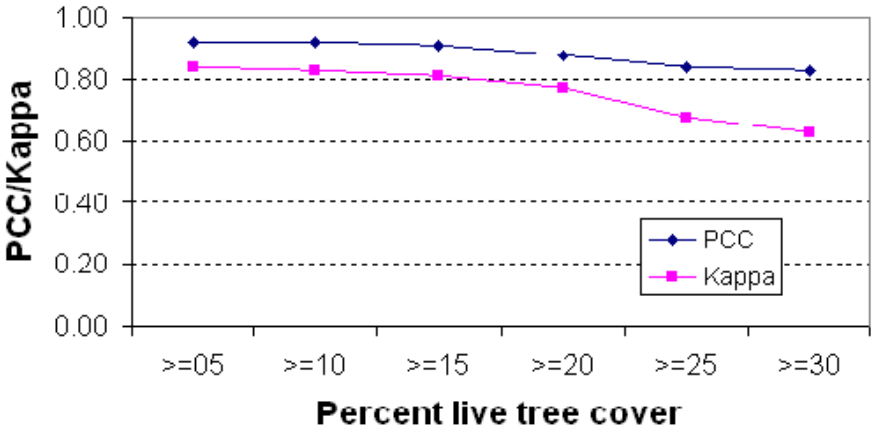


Figure 9: OOB PCC and Kappa values from models of percent live tree cover at different thresholds.

CONCLUSION

Large scale aerial photography can provide valuable information as training data for mid-scale modeling and mapping efforts as a link between MODIS imagery and ground data and as auxiliary information for tree cover estimations, even at the lower extremes. Using the photo data to establish relationships with MODIS imagery was found to be slightly more effective than using field data

according to the OOB error estimates from Random Forests models. This was true for the binary responses of forest presence, pinyon-juniper presence, and pinyon-juniper presence, as well as percent live tree cover.

The cost scenario analysis showed the affect of using photo data versus field data for modeling percent tree cover at different costs. The photo interpretation data look promising as an alternative or ancillary data source for providing information at lower costs. It was also determined that the addition of condition-level information was valuable for increasing the variance explained in the models.

LSP is also a valuable data source for generating spatial products of not only forest, but rangeland conditions including presence of shrub type and, more specifically, sage (Figure 10).

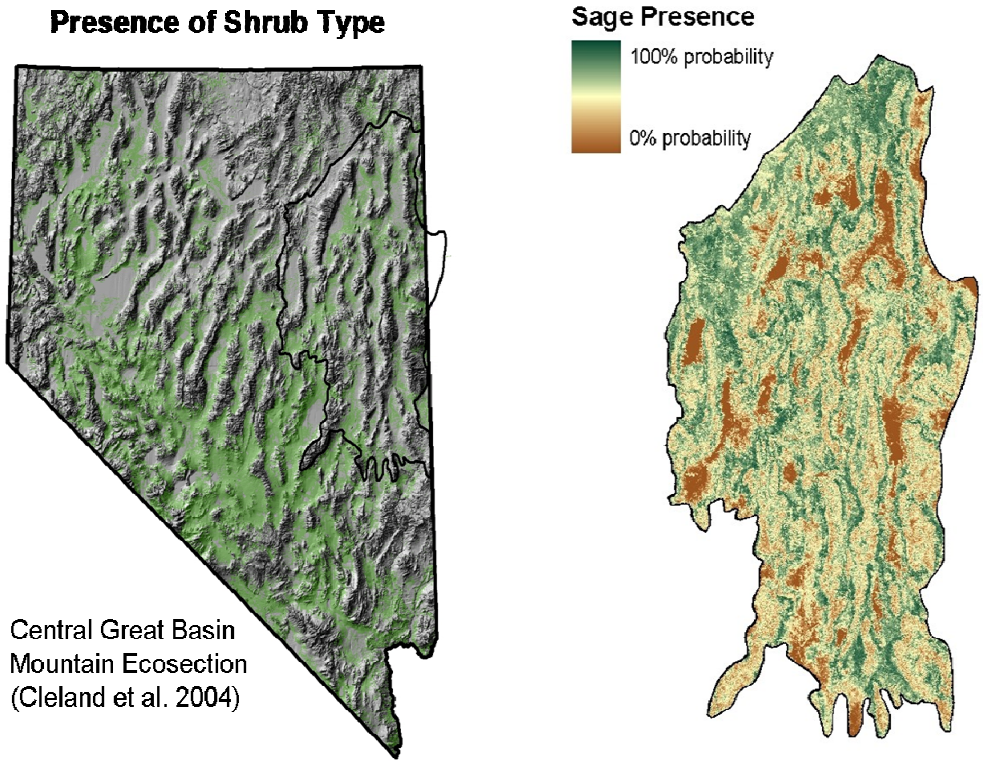


Figure 10: Spatially-explicit maps generated from Random Forests models. a. Presence of shrub type. b. Presence of sage species.

This paper was a start at analyzing the valuable data available from NPIP and the relationships with MODIS imagery. Further research is needed to investigate this relationship with MODIS (and other predictors) in more detail, as well as to continue exploring the use of condition-level information as training data. Also,

looking at cost scenarios including both photo and field data could provide more insight into cost savings and efficiencies of large scale inventories.

REFERENCES

Blackard, J.A.; Finco, M.V.; Helmer, E.H.; Holden, G.R.; Hoppus, M.L.; Jacobs, D.M.; Lister, A.J. Lister; Moisen, G.G.; Nelson, M.D.; Riemann, R.; Ruefenacht, B.; Salajanu, D.; Weyermann, D.L.; Winterberger, K.C.; Brandeis, T.J.; Czaplewski, R.L.; McRoberts, R.E., Patterson, P.L.; Tymcio, R.P. 2008. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. Remote Sensing of Environment 112(4):1658-1677. Biomass maps for Conus and Alaska are available on-line at http://fsgeodata.fs.fed.us/rastergateway/biomass.

Blackard, J.A.; Moisen, G.G.; Tymcio, R.P. 2004. Modeling investigations into mapping of forest attributes with FIA data for the state of Wyoming. Proceedings of the tenth Forest Service remote sensing applications conference, April 5–9, 2004, Salt Lake City, UT. Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing. Unpaginated CD-ROM.

Born, J.D.; Tymcio, R.P.; Casey, O.E. 1992. Nevada Forest Resources. U.S. Department of Agriculture, Forest Service, Intermountain Research Station, Resource Bulletin INT-76.

Breiman, Leo. 2001. Random Forests, Machine Learning 45:5-32.

Clark, J.; Finco, M.; Warbington, R.; Schwind, B. 2004. Digital Mylar: a tool to attribute vegetation polygon features over high-resolution imagery. RSAC-4107-RPT1. Salt Lake City, UT: U.S. Department of Agriculture, Forest Service, Remote Sensing Applications Center. 8 p.

Frescino, T.S.; Moisen, G.G.; Megown, K.A.; Nelson, V.J.; Freeman, E.A.; Patterson, P.L.; Finco, M.V.; Brewer, K. In Press. Nevada photo-based inventory pilot (NPIP) photo sampling procedures. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station Gen. Tech. Rep. XXX.

Gillespie, A.J.R. 1999. Rationale for a national annual forest inventory program. Journal of Forestry 97(12):16-20.

Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83:195-213.

Reams, G.A.; Smith, W.D.; Hansen, M.H.; Bechtold, W.A.; Roesch, F.A.; Moisen, G.G. 2005. The forest inventory and analysis sampling frame. *In*: Bechtold, W.A.; Patterson, P.L. eds. The enhanced forest inventory and analysis program-national

sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 15 p.

Ruefenacht, B.; Moisen, G.G.; Blackard, J.A. 2004. Forest type mapping of the Interior West. Proceedings of the tenth Forest Service remote sensing applications conference, April 5–9, 2004, Salt Lake City, UT. Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing. Unpaginated CD-ROM.

Ruefenacht, B.; Finco, M.V.; Nelson, M.D.; Czaplewski, R.; Helmer, E.H.; Blackard, J.A.; Holden, G.R.; Lister, A.J.; Salajanu, D.; Weyermann, D.; Winterberger, K. 2008. Conterminous U.S. and Alaska Forest Type Mapping Using Forest Inventory and Analysis Data. Photogrammetric Engineering and Remote Sensing.

Swets, D.L.; Reed, B.C.; Rowland, J.D.; Marko, S.E. 1999. A weighted least-squares approach to temporal NDVI smoothing. *In*: 1999 ASPRS Annual Conference: From Image to Information, Portland, Oregon, May 17-21. Proceedings: Bethesda, Maryland, American Society for Photogrammetry and Remote Sensing.

Tucker, C. J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Evironment, 8:127-150.

US Department of Agriculture. 2006. Interior West Forest Inventory and Analysis Field Procedures V3.0. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, Utah.

System Improvements



The Role of Pre-Field Operations at Four Forest Inventory Units: We can see the trees, not just the forest

Sara A Goeking[1] and Greg C. Liknes[2]

Abstract:

The Forest Inventory and Analysis (FIA) program attempts to inventory all forested lands throughout the United States. Each of the four FIA units has developed a process to minimize inventory costs by refraining from visiting those plots in the national inventory grid that are undoubtedly nonforest. We refer to this process as pre-field operations. Until recently, the pre-field process of differentiating forest from nonforest plots received very little attention or scrutiny. No national pre-field manual currently exists, nor does any other form of formal documentation of the pre-field process. Pre-field specialists from all four FIA units gathered in December 2007 to discuss the role of pre-field operations, to identify national commonalities, to share techniques and technologies, and to explore opportunities to increase efficiency and expand functionality. The meeting revealed that each of the four FIA pre-field teams serves the same two fundamental roles: 1) identifying plot locations that are potentially forested and therefore require visitation by a field crew, and 2) subsequently preparing field materials. Each pre-field team currently practices some form of quality assurance, yet no common procedures exist regarding pre-field quality assessment and quality control. Participants agreed that additional information could be derived from the photo-interpretive aspect of this process via the collection of land use and crown cover data from all FIA plots, including those that are not visited by field crews. Such information might facilitate estimates of nonforest conditions with trees, e.g., urban forest, and land-use changes, e.g., the development of formerly forested lands. This paper documents the current status of pre-field operations, suggests areas for future process improvements and collaboration among the four FIA units, and evaluates the potential of the pre-field process to provide additional information to the inventory.

Keywords: pre-field, forest canopy, canopy cover, crown cover, nonforest land use, aerial photography, National Agriculture Imagery Program, image segmentation

[1] USDA Forest Service; Rocky Mountain Research Station; 507 25th St.; Ogden, UT 84401. Email: sgoeking@fs.fed.us

[2] USDA Forest Service; Northern Research Station; 1992 Folwell Ave; St. Paul, MN 55108. Email: gliknes@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

The Forest Inventory and Analysis (FIA) program attempts to inventory all forested lands throughout the United States. The national FIA sampling grid includes a nontrivial proportion of nonforest plots, so a major means of increasing the program's efficiency has been to minimize unnecessary field visits to nonforest plots. Each of the four FIA units has developed a process to reduce inventory costs by refraining from visiting plots that are undoubtedly nonforest (Reams et al. 2005). We refer to these processes as pre-field operations.

Pre-field specialists distinguish potentially forested plots from plots that are undoubtedly nonforest by examining aerial photographs, maps, and previous inventory data, if any exist. They observe individual trees, patches of trees, and entire forest areas as they examine each plot, and their interpretation focuses on whether each plot meets the FIA definition of forest. Therefore, the pre-field process involves evaluating the land use and tree cover of every plot in the national FIA grid. Plots that are deemed to be nonforest by pre-field specialists are then omitted from the field inventory.

Two aspects of this process are particularly remarkable. First, pre-field operations have the ability to affect estimates of total forest area by constraining the sample of potentially forested plots. For this reason, the pre-field process should be examined for consistency, repeatability, and efficiency. Currently no national pre-field manual exists, nor does any other formal documentation or consistent quality assurance of pre-field protocols. Second, pre-field specialists are in a unique position to quantify and document variables pertaining to land use and tree cover at all FIA plots, including nonforest plots that are not visited by field crews.

Thus, pre-field operations not only have a major impact on the overall forest inventory, but as Reams et al. (2005) suggested, they may also be currently under-utilized. The purpose of this paper is to document the current state of FIA pre-field operations nationwide, and to suggest potential process improvements that increase efficiency while also providing additional information for the inventory.

The Current Status of Pre-field Operations

Pre-field specialists from each of the four FIA units (Interior West, Northern Research Station, Pacific Northwest, and Southern Research Station, hereafter referred to as IW, NRS, PNW, and SRS, respectively) met in Salt Lake City on December 4-5, 2007, to discuss the current state of pre-field operations. The objectives of this meeting were to discuss the role of pre-field operations, identify national commonalities, share techniques and technologies, and explore opportunities to increase efficiency and expand functionality. Participants in this meeting ascertained that pre-field operations are remarkably similar at each of the

four FIA units, and they were thus able to summarize the status of pre-field nationwide.

Pre-field operations at all four FIA units encompass two basic functions. First is identification of plot locations that are potentially forested and therefore require visitation by a field crew. The second function is preparation of field materials such as old and new data sheets, and maps and imagery for documentation and navigational purposes.

The basic decision rule for distinguishing potentially forested plots from nonforest plots is the same at all units. If there is any evidence that a plot might meet the FIA definition of forest, it is scheduled for field visitation. The difficulties of identifying "field visit" plots vary due to some units being several cycles into the inventory, and some being in their first cycle. NRS and SRS have previous inventory data to inform their interpretation of aerial imagery, whereas IW and PNW are trying to determine whether a plot that has never been visited might be forested. The portion of the entire sample grid that is scheduled for sampling in any given year is referred to as that year's panel (Bechtold and Patterson 2005, Reams et al. 2005).

Digital imagery provides the basis for classification of each panel's plots into visit and non-visit plots, where non-visit plots are then assumed to be synonymous with nonforest. After the December 2007 meeting, all units began using the Image Server provided by the Remote Sensing Applications Center, except in states where National Agriculture Imagery Program (NAIP) imagery was recently made available and was not yet updated in the Image Server. Pre-field specialists in Alaska are at a distinct disadvantage due to the spatially and temporally variable availability of high-quality aerial imagery in that state.

Most of the developmental effort for the pre-field process has been directed at the development of technological tools and file structures to support pre-field operations. All units have a user interface that is linked to aerial imagery and previous inventory data (where it exists), and is also used for pre-field data entry. These tools and file structures are parallel in function but different in practice. Each interface was built with a different programming language and uses a different database structure. Therefore, the effort required to develop graphical user interfaces and pre-field database has been replicated at each of the four FIA units.

All units include, or are about to implement, quality assurance programs as part of the pre-field process. Some units assess repeatability of visit/non-visit determinations based on aerial imagery, while others assess the accuracy of equating non-visits with nonforest, i.e., "non-visit" plots are field-visited and checked to make sure they are nonforest. The quality assurance processes at SRS, NRS, and PNW assess the repeatability of visit/non-visit determinations. SRS and NRS also field-verify a portion of both visit and non-visit plots. PNW plans

to begin field verification in the upcoming season, and IW plans to implement a quality assurance process that assesses the repeatability of visit/non-visit determinations and field-verifies a portion of both visit and non-visit plots.

Each pre-field team assembles some type of plot packet (or jacket, folder, etc.) for use by field crews. The packets may contain imagery (stereo photo pairs or printed digital imagery), travel maps, and field reference sheets. Much of this work is done manually by pre-field specialists and field staff.

Recommendations for Future Collaboration and Process Improvement

Participants at the December 2007 meeting developed a list of three major areas for improvement and collaboration: development of pre-field user interface tools, automated preparation of field materials, and standardization of the quality assurance process. The first two areas would involve collaboration among specialized programmers and database managers, or creating a pre-field programming group. The third area would require initiating a focused quality assurance protocol that could evaluate the repeatability of visit/non-visit determinations, evaluate the uncertainty associated with these determinations, and include documentation of the pre-field process.

User interface tools

The development of pre-field user interfaces requires specific programming expertise. Each unit currently uses its own user interface, yet the basic requirements of each pre-field team are very similar. As FIA migrates to new databases and data centers, old tools will likely become obsolete. The development of a template, which could then be adapted to each unit's specific requirements, would make more efficient use of programmers' resources.

The ideal interface for pre-field operations would include the following characteristics: (1) it would permit on-screen inspection of each plot, including links to georeferenced aerial imagery, maps and other ancillary spatial data, and previous inventory data, (2) the link to aerial imagery and spatial data must be flexible enough to integrate new data sources as they become available, (3) the interface would also permit pre-field data entry and connect directly to the relevant data tables, and (4) data entry portions of the interface must also be sufficiently flexible to incorporate each unit's regional pre-field variables, e.g., in IW, distance from a maintained road to the plot center.

Automated preparation of field materials

The pre-field function of preparing field materials could be made much more efficient by automating some of the basic tasks. Field materials include new data forms (referred to as plot sheets, reference sheets, drawsheets, etc., at different

units); previous inventory data forms; printed imagery and/or aerial photographs; and topographic maps. Streamlining this process could be accomplished by developing programs or scripts that generate printable maps including topographic maps and imagery, and also retrieve scanned and archived versions of previous inventory data. The ideal tool would print these materials, along with new field data sheets, as one bundle per plot.

The ideal tool proposed here could eventually lead to the creation of virtual plot packets, which would not only streamline the pre-field process but would also enable remote access by field crews anywhere. This ideal field preparation tool would require that previous inventory data be available digitally, and that old field data sheets exist as scanned, digital copies. Most FIA units have already scanned, or are in the process of scanning, their previous inventory data sheets. However, distribution of virtual plot packets to remote duty stations could be constrained by the availability of hardware and intranet connections.

Since the December 2007 meeting, Doug Shipley of SRS has developed a prototype tool that addresses one function of our hypothetical ideal tool. This tool relies on attributes of a point shape file to designate the plots of interest, and then produces print imagery using either NAIP imagery or the RSAC Image Server (Doug Shipley, pers. comm.). At least two FIA pre-field teams are currently using this tool to prepare imagery for inclusion in field packets.

Quality assurance

Due to the potential impact of visit/non-visit determinations on the overall forest inventory, it is important to know whether those determinations are both accurate and repeatable. An effective quality assurance program would include both quality control to evaluate the repeatability of visit/non-visit determinations, and quality assurance to evaluate the uncertainty associated with these determinations. It would also include documentation of the pre-field process, and why a particular plot was, or was not, sent to the field.

A quality control program would involve setting standards and processes to control the data acquisition process. This would include standards for training of pre-field specialists, as well as hot and/or cold checks to evaluate the need for more training. Hot and cold checks involve making a determination with a priori knowledge of the initial determination. There are two possible types of pre-field hot or cold checks: 1) a field visit to determine whether the pre-field determination was correct, and 2) a comparison of the determination made by a pre-field expert using the same imagery and materials the original specialist used. This type of quality control is similar to the method that is currently used for quality control in the field.

A quality assessment program would assess whether the error in pre-field visit/non-visit determinations is within a range of acceptable errors. The smaller

the acceptable error rate is, the larger the sample of plots must be in order to be confident that the actual error rate is below the acceptable error. This process would be similar to the blind checks performed in the field, where repeatability is assessed using a large number of blind checks that compare two sets of independent observations. In the case of pre-field quality assurance, the blind checks could be based on two remote, image-based determinations, or on a field visit as compared to a pre-field determination.

Another aspect of quality assurance is documentation of the overall process. At the December 2007 meeting, all participants agreed that the pre-field process should include documentation of the data and source of the imagery that was used to exclude nonforest plots from the field inventory. Documentation should include enough information to indicate why each plot was designated as either a visit or non-visit plot.

Region-Specific Quality Issues: It is clear that although pre-field operations are similar at the four FIA units, unique challenges exist at each unit. Three of these issues are briefly summarized here: GPS coordinate quality, ownership data quality, and packet tracking and accountability.

Most plots within NRS and SRS have been visited multiple times due to the shorter five-year inventory cycle in those regions (Reams et al. 2005). As a result, pre-field specialists at those two units exert substantial effort examining multiple sets of GPS coordinates to identify the one that appears to be the most accurate. Their main objective is to provide field crews with the most accurate coordinates possible to enable easier re-location of the plot center. In addition, accurate plot coordinates enable better linkages between plots and other geospatial datasets. This is not yet a problem at IW or PNW because they are still in their first inventory cycle in most states.

The issue of plot ownership is somewhat problematic for IW and PNW, where permission-to-access requests are sent to private landowners and public land management agencies prior to the field season. Thus it is critical to identify plot ownership during preparation of field materials. Map servers are available online for some counties, but where they are not available field crews must visit county courthouses to identify the owners of privately owned plots. ALP (the Forest Service's Automated Lands Program) may eventually help resolve issues on national forest lands, but other lands such as BLM and state may still prove problematic and thus require a new process for accurately identifying land ownership. Specific units are examining different protocols of updating plot ownership data prior to field data collection, but no standard protocol exists.

Within all units, field plot packets tend to travel more widely than nearly any FIA staff member, as the packets are sent from pre-field teams to field supervisors to field crews to quality control crews and back to the office. Therefore, it may be useful to create systems to track plot packet locations and thus maintain

accountability for field materials. Currently, SRS has a fairly extensive tracking and accountability system, and IW has a packet database and check-in/out system for office use.

Expanding the Utility of Pre-field Operations

Pre-field specialists from each unit agreed that additional information could be derived from the photo-interpretive aspect of making forest/nonforest determinations. Specifically, they were confident that it is possible to document crown cover and land use at all FIA plots; they just need a formal protocol to guide the process. The pre-field process involves assessing whether each plot meets the current FIA definition of forest, which is defined as land that is ≥10% stocked by forest trees of any size (or ≥5% cover in IW), or land formerly having such cover, and is not currently developed for a nonforest land use (USDA 2007). Thus, the process involves making judgments about the land use and tree cover of each plot, but documentation of these variables is not standard protocol.

Although nonforest land use is not part of the national protocol, all units currently record a land use code for nonforest plots as part of the pre-field process. Pre-field observers use the national FIA land use categories and interpret the land use at the plot center.

Recording land use and crown cover variables would meet several objectives. First, it would document why a plot was, or was not, scheduled for field visitation. Second, it would facilitate categorization of nonforest plots for adjunct vegetation inventories, such as urban forests or rangelands (Figure 1). Third, recording land use would enable tracking of land use changes over time, such as conversion of forest to agriculture or developed areas. Finally, the variables would allow for a flexible future definition of forest. At least within IW, a historically inconsistent definition of "forest land," as well as procedural changes in forest inventory, have

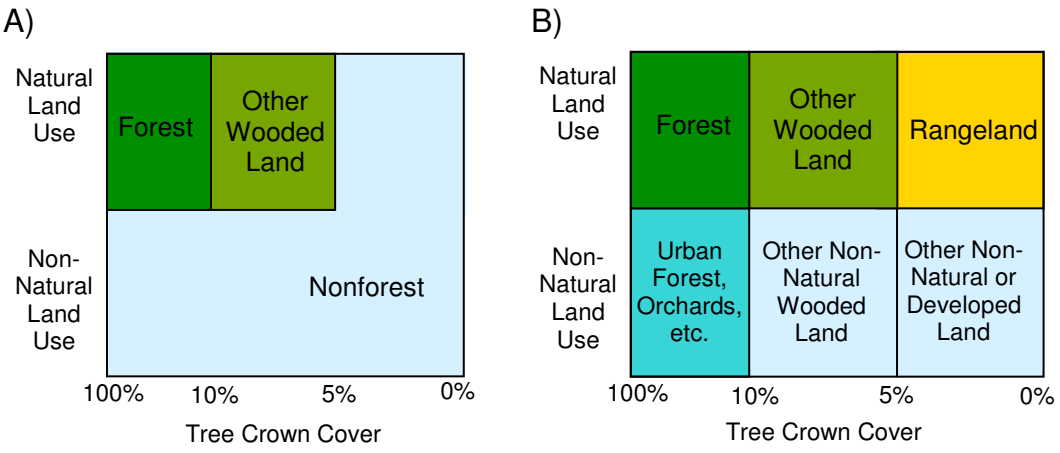


Figure 1: The relationship between land use and crown cover. A) The current ability to categorize FIA plots according to land use and crown cover. B) The ability to categorize FIA plots if land use and crown cover are recorded for all FIA plots.

led to difficulties in comparing forested area over time. Knowing the history of land use and crown cover at each plot would enable analysis of long-term dynamics as inventory definitions and procedures change.

Methods of remotely estimating crown cover

As mentioned above, pre-field specialists at the December 2007 meeting agreed that it is possible to classify the land use and quantify the crown cover of each FIA plot as part of the pre-field process. At this meeting, a few potential methods were discussed, but none was selected as the group agreed that more study was need before selecting a preferred technique.

The Interior West Pilot Study: We recently completed a pilot study to identify a method of remotely estimating crown cover that produces accurate, efficient, and repeatable estimates for any given plot. The analysis for this pilot study is in progress (Goeking et al., in prep.) but preliminary results are presented here.

The Interior West (IW) FIA area was selected for the pilot study because crown cover is a regional field variable in IW. This variable provided us with ground-truth data against which we could compare various methods for remotely estimating crown cover.

We compared three methods of remotely estimating crown cover: ocular estimates, dot-count samples, and image segmentation. These three methods were applied at 100 plots that were sampled from the population of all single-condition plots that were visited in five IW-FIA states (Arizona, Colorado, Idaho, Montana, and Utah) during the 2007 field season. Each method of estimating cover was applied to the area inside a 1-acre circle around the plot location center. All estimates were based on 1-m resolution NAIP imagery.

Ocular estimates were made by three observers aided by a visual crown coverage scale. Dot counts consisted of 50 random points within each plot's 1-acre sample circle, with a minimum distance of 2 m between points (Figure 2). Image segmentation was accomplished using the Feature Analyst tool provided by the Forest Service's Remote Sensing Applications Center (Figure 3).



Figure 2: Examples of three plots showing the distribution of dots used for dot counts.

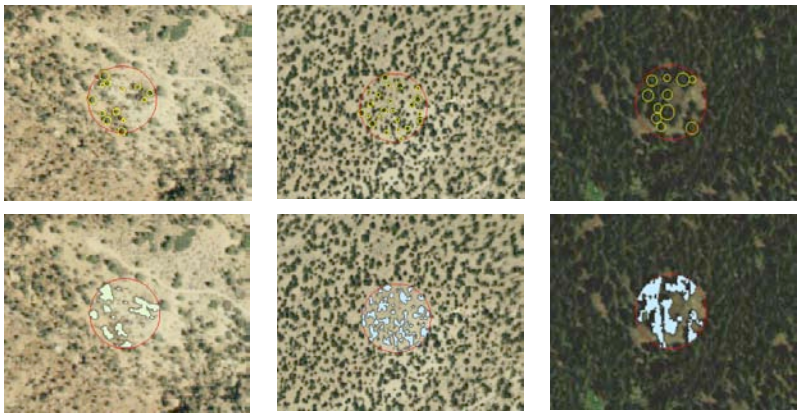


Figure 3: Examples of three plots showing the polygons digitized for use in image segmentation (top) and the output classification of crown cover (bottom).

We assessed the accuracy of the three methods by comparing our estimates against field-recorded crown cover. Korhonen et al. (2006) found line intersect sampling to be an accurate and unbiased method of estimating canopy cover as compared to other ground-based methods, although their sample intersects were denser than those used by IW. Field transects in IW consist of four 25-ft transects, one in each of the cardinal directions, on each subplot (Figure 4). Intercepts were recorded in 1-ft increments. Intercepts are defined as points within the crown perimeter of tally tree species of any size. Each subplot has a total of 100 possible "hits" of tree cover; an entire plot has 400 possible hits (USDA 2007).

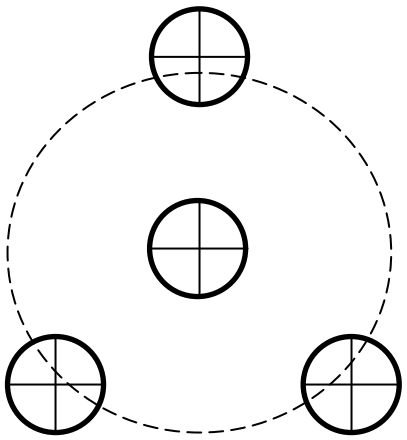


Figure 4: Configuration of transects used for measuring crown cover on IW FIA plots. Dashed circle represents an area one acre in size. Transects are shown by four solid lines per subplot. Each transect includes 25 trancepts that are spaced 1 ft. apart.

Results from the Interior West Pilot Study: Initial results are depicted in Figure 5. Dunnett's nonparametric test for multiple comparisons against a control indicated that only image segmentation produced estimates that were not

significantly different from field transects (α=0.05), although dot-count estimates were only marginally statistically different from the field transects (p=0.06). All methods tended to over-estimate crown cover (Figure 6). Ocular estimates were the most biased, followed by dot-counts and image segmentation, in that order. A more rigorous statistical analysis of these data appears in Goeking et al. (in prep.).

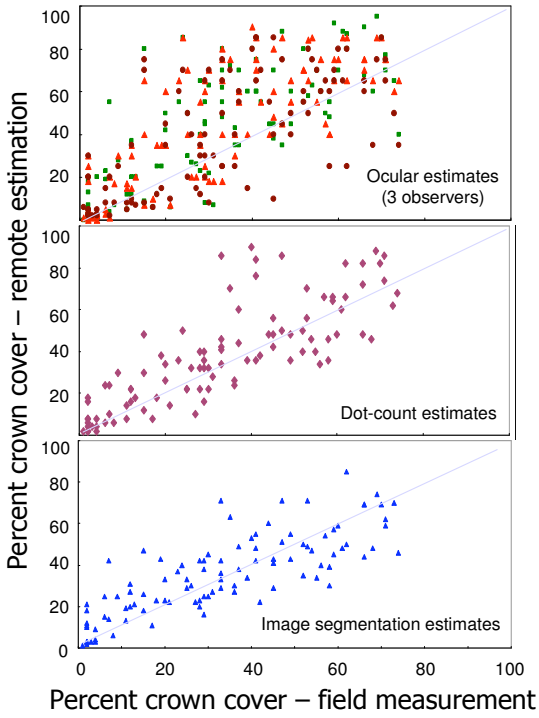


Figure 5: Relationship between field-measured crown cover and remotely estimated crown cover using ocular estimation, dot counts, and image segmentation methods (n=100).

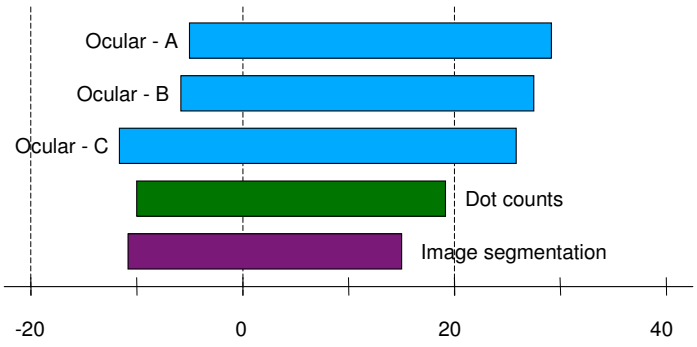


Figure 6: Differences between the crown cover estimates produced ocular estimates (three observers), dot counts, and image segmentation methods compared to field-measured crown cover. The center of each bar represents the mean difference between the estimates and the field-measured values. The width of each bar represents one standard deviation around the mean of these differences. The position of each bar relative to zero indicates the bias of that method of remotely estimating crown cover, relative to field measurements.

Logistical Considerations: Ocular estimates differed widely among the three observers, indicating that they are not very repeatable. The one consistency among the three interpreters was an overall bias toward overestimation. In general, ocular estimation is by far the fastest method of estimating crown cover, but these estimates are also the most biased, least accurate, and least precise.

Dot counts were less efficient than ocular estimates, but slightly more efficient than image segmentation. They also produced the most consistent and precise estimates, being slightly more precise than image segmentation. However, this method tends to overestimate crown cover at all levels, although the bias is less than that shown by ocular estimates.

Image segmentation is the most time-intensive due to the image pre-processing and on-screen digitizing that it requires, but it is also the least biased and most accurate overall. This method has the unique advantage of producing polygon output, which could be useful for estimating landscape-level parameters such as patch size, perimeter to area ratios, and other landscape metrics. However, image segmentation tends to overestimate crown cover at plots with very low crown cover (<20%). In woodland areas of the Mojave and Sonoran deserts, tally and non-tally species appear very similar in NAIP imagery, and thus lead to over-classification of crown cover. Analyses in progress are investigating whether it is possible to reduce the bias of image segmentation at low crown cover by reducing over-classification of shadows and non-tally species.

Summary

The four FIA pre-field teams fulfill the same fundamental functions within their respective units, so their parallel roles present several opportunities to increase efficiency by reducing duplication of effort. The primary recommendation of the pre-field workshop participants was to create a programming group that could assist pre-field teams in developing new tools for plot inspection, photo-interpretation, and data entry; and for automating the preparation of field materials. The pre-field process could also be made more robust by developing a standard quality assurance protocol that would document the pre-field process, provide guidelines for training and quality control, and assess the error rate of visit/non-visit determinations.

As pre-field specialists determine whether each plot should be visited or not, they evaluate the land use and tree cover of every plot in the national FIA grid. They already record a land use variable, even on nonforest plots. Since objective methods exist for remotely quantifying crown cover, the utility of pre-field operations could be expanded by recording both land use and crown cover for all FIA plots. Doing so would not only provide clear documentation of pre-field protocols, but would also enable analysts to estimate the magnitude of other vegetation inventories, track land use changes at all FIA plots, and estimate the impact of changing the definition of forest. There are tradeoffs among the various

potential methods of estimating crown cover, including expediency vs. cost; accuracy, bias, and precision; and potential ability to assess landscape-level parameters.

In summary, pre-field operations may have a major impact on the overall forest inventory by potentially constraining the field sample, and may also be currently under-utilized. Collaboration among pre-field teams may lead to increased efficiency and accuracy, and simultaneously contribute additional information to the inventory.

Acknowledgments

Pre-field staff from each of the four FIA units contributed to this paper through their participation in the national pre-field workshop on December 4, 2007. Participants included Jennifer Bakken and Jason Toombs of IW-FIA; Perry Colclasure, Melissa Patterson, and Gerry Bednarczyk of PNW-FIA; Doug Shipley of SRS-FIA; and Dale Gormanson, Jeff Wazenegger, and Rich McCullough of NRS- FIA. Phyllis Adams of PNW-FIA and Paul Patterson of IW-FIA also provided valuable guidance regarding quality assurance of the pre-field process. Thanks also to Ken Brewer, Kevin Megown, and Mark Finco of the Remote Sensing Applications Center for graciously hosting the workshop.

IW analysts John Shaw and Larry Deblander helped focus the purpose and need for estimating crown cover at all FIA plots. Jennifer Bakken and Jason Toombs assisted with the evaluation of various methods of remotely estimating crown cover. Doug Shipley of SRS, Chris Toney of RMRS, and other reviewers at Rocky Mountain Research Station provided valuable editorial comments.

References

Bechtold, W.A., and P.L. Patterson (eds.). 2005. The enhanced Forest Inventory and Analysis program – national sampling design and estimation procedures. USDA For. Serv. Gen. Tech. Rep. SRS-80.

Korhonen, L., Korhonen, K.T., Rautiainen, M., and Stenberg, P. 2006. Estimation of forest canopy cover: a comparison of field measurement techniques. Silva Fennica 40(4): 577-588.

Reams, Gregory A.; Smith, William D.; Hansen, Mark H.; Bechtold, William A.; Roesch, Francis A.; Moisen, Gretchen G. 2005. The forest inventory and analysis sampling frame. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 21-36

USDA Forest Service. 2007. Interior West Forest Inventory and Analysis field procedures, version 3.01. Available online at http://www.fs.fed.us/rm/ogden/data-collection/field-manuals.shtml; last accessed Jan. 22, 2009.

A Stem-map Model for Predicting Tree Canopy Cover of Forest Inventory and Analysis (FIA) Plots

Chris Toney[1], John D. Shaw[2], and Mark D. Nelson[3]

Abstract: *Tree canopy cover is an important stand characteristic that affects understory light, fuel moisture, decomposition rates, wind speed, and wildlife habitat. Canopy cover also is a component of most definitions of forest land used by US and international agencies. The USDA Forest Service Forest Inventory and Analysis (FIA) Program currently does not provide a national standard measurement of tree canopy cover, and most regional FIA units do not measure canopy cover in the field.*

This paper describes a model for predicting canopy cover of FIA plots by mapping the locations of trees ≥ 5.0 in. diameter within the plot, and statistical modeling of sapling contribution to total cover. The model was developed with an operational focus, including the requirement that it scale efficiently to national applications. Coefficients for species-specific crown width equations have been stored in lookup tables with surrogates assigned to FIA tree species lacking equations. Modeling was supported by field measurements on 12,070 FIA plots distributed across the eight-state Interior West FIA region. Refinements to the model included adjustments to crown width equations for small-diameter trees, stem-mapping of microplot subsamples to support cover estimation of sapling-stage plots, and the use of spatial statistics to derive predictor variables describing the spatial pattern of overstory trees. Model predictions were compared to field measurements of canopy cover by line-intercept method on 1,454 single-condition plots from the Interior West FIA 2006 field season. The mean absolute difference between field-measured and model-predicted values was ± 7.9% canopy cover, with mean bias of -0.7% canopy cover. The relationship between field-measured and predicted values was approximately linear with approximately constant variance and a correlation coefficient r = 0.875.

FIA produces estimates of forest land area based on a definition of forest land that includes a minimum threshold of tree stocking. Proposed changes to the FIA definition from one based on stocking to one based on canopy cover could affect estimates of forest land area, but the amount and variability of this change is not fully understood. We made a preliminary assessment of the effect of a canopy cover-based definition on forest area estimates for a subset of states within the Northern Research Station FIA unit.

Keywords: canopy cover, crown width, FIA, land cover, line intercept, Ripley's K function, spatial pattern, stand height, stem-mapping

---

[1] USDA Forest Service; Rocky Mountain Research Station; Inventory, Monitoring, and Analysis Program and the LANDFIRE Project; Missoula, MT 59808; email christoney@fs.fed.us
[2] USDA Forest Service; Rocky Mountain Research Station; Inventory, Monitoring, and Analysis Program; Ogden, UT84401; email jdshaw@fs.fed.us
[3] USDA Forest Service; Northern Research Station Forest Inventory and Analysis; St. Paul, MN 55108; email mdnelson@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

Canopy cover is defined as the proportion of the forest floor covered by a vertical projection of the tree crowns (Jennings et al. 1999). Canopy cover influences the forest microclimate by affecting understory light, surface temperature, surface moisture levels, and wind speed (Jennings et al. 1999, Christopher and Goodburn 2008). It is a key stand characteristic used in a variety of applications (Shaw 2005). For example, canopy cover is often a component of wildlife habitat suitability models (Gill et al. 2000, Bond et al. 2004, Vospernik et al. 2007), and regional management plans for some species require maintaining certain levels of canopy cover (Fiala et al. 2006). In fire behavior simulation models such as FARSITE (Finney 2004), increasing canopy cover has a moderating effect on wind speed which is a driver of fire spread rate and partially determines situations where a ground fire is likely to transition to a crown fire. By shading the surface, canopy cover also determines dead fuel moisture levels under a given weather scenario. Canopy cover thresholds are a component of most international definitions of forest land, which has implications for reporting and carbon accounting (Lund 2002).

Plot-based data on canopy cover commonly are used to support these and other applications, but the Forest Inventory and Analysis (FIA) program of the USDA Forest Service currently does not have a national standard canopy cover measurement and most FIA regional units do not measure tree canopy cover in the field. However, FIA does use a national standard plot design and tree measurement protocol. An approach for modeling canopy cover, optimized for this plot design, could provide canopy cover data that are estimated consistently across the nation. Our objective was to develop a model for predicting tree canopy cover of FIA plots along with software for efficient data processing.

FIA currently produces estimates of forest land area based on a definition of forest land that includes a minimum threshold of tree stocking. Land that formerly was stocked (e.g., clear-cuts) is still considered forest if not developed for other uses (Bechtold and Patterson 2005: glossary, page 80). A proposed change to the FIA definition of forest land could replace the minimum stocking threshold with a threshold of minimum canopy cover, but presumably would retain some form of the land-use requirement in the definition. This change could affect estimates of forest land area, but the amount, direction, and spatial variability of this potential effect are not fully understood. A second objective was to make a preliminary assessment of the effect of a canopy cover-based definition on forest area estimates for a subset of states within the Northern Research Station (NRS) FIA unit.

Methods

Canopy Cover Data and Modeling Approach

Additional background on the data and modeling approach was presented by Shaw (2005). The current FIA standard plot design was described in detail by Bechtold and Scott (2005) and is summarized in figure 1.

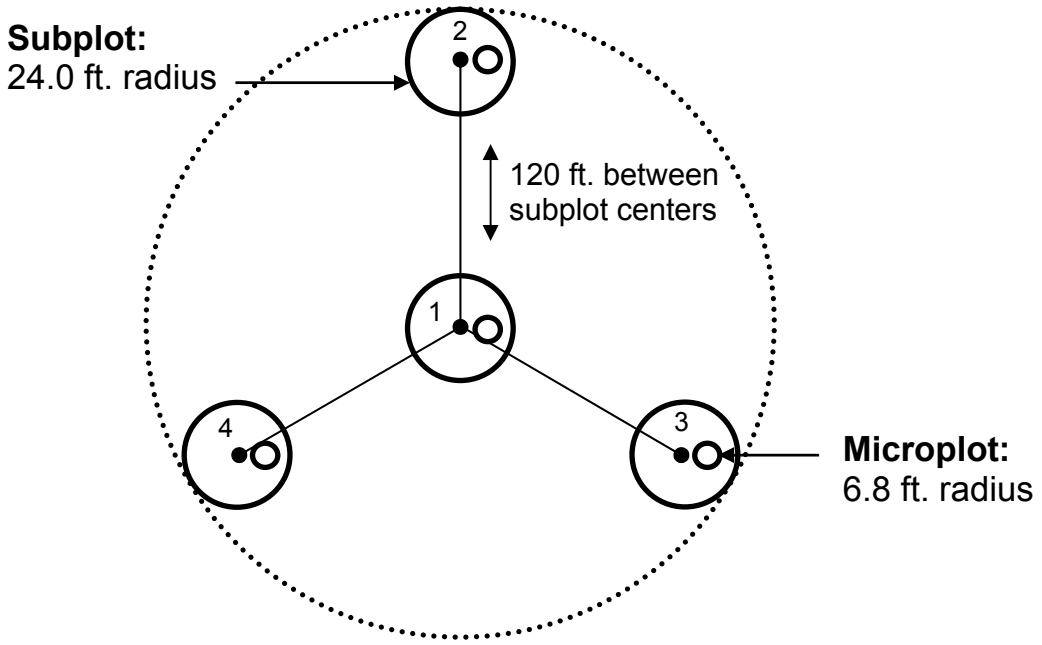


Figure 1: FIA national standard plot layout. Microplot centers are 12 ft. at 90º azimuth from subplot centers. The minimum circle enclosing all four subplots is approximately 1.5 ac. Not shown are optional 58.9-ft. radius macroplots surrounding each subplot.

Details of the tree measurements are described in the FIA field manual (USDA Forest Service 2005). Diameters are generally measured at breast height (DBH), but for some western woodland species (USDA Forest Service 2007) diameters are measured at the root collar (DRC). DBH is measured to the nearest tenth of an inch at a point 4.5 feet above ground level on the uphill side of the tree. DRC is measured to the nearest tenth inch at the ground line or at the stem root collar, whichever is higher. Reference herein to stem diameter means DBH or DRC.

A key feature of the tree measurement protocol is that a coordinate is recorded for each tree ≥ 5.0 in. diameter in the subplots, and for each tree ≥ 1.0 in. diameter but < 5.0 in. diameter in the microplots. Trees ≥ 1.0 in. diameter but < 5.0 in. diameter are denoted as saplings by FIA, and are measured in the microplots. Coordinates are recorded as distance and azimuth from the microplot centers for saplings, and from subplot centers for trees ≥ 5.0-in. diameter.

The Interior West FIA (IW-FIA) unit comprises Idaho, Montana, Wyoming, Colorado, Utah, Nevada, Arizona, and New Mexico. IW-FIA measured canopy cover of trees ≥ 1.0-in. diameter using the line intercept canopy cover (LIC) method during 1995-2006 (USDA Forest Service 2006). Four 25-ft transects were established in each subplot, in the cardinal directions beginning 1 foot from the subplot centers. The length of crown interception of live tally tree species ≥ 1.0-in. diameter was recorded along each transect. Canopy cover was calculated by FIA condition class within the plots, by dividing the total live crown interception length by the total length of transects within each condition (400 feet total transect length for single-condition plots). Only single-condition plots were used to develop the canopy cover model, meaning all four subplots were classified as having the same FIA forest condition status. Approximate locations of the IW-FIA plots used for canopy cover modeling are shown in figure 2.

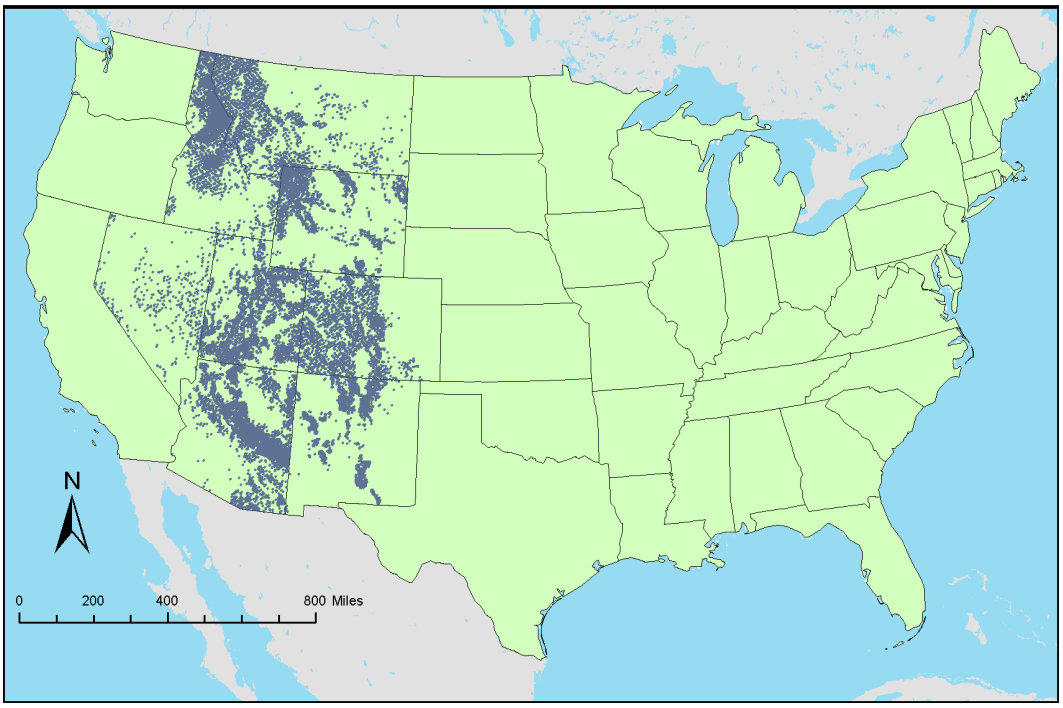


Figure 2: Approximate locations of 12,070 single-condition Interior West FIA plots (blue dots) used to develop a stem-map model for predicting tree canopy cover.

Precision of the LIC measurement was assessed using blind check plots during 2000-2003 (Pollard et al. 2006). A target tolerance of ± 10% canopy cover was specified for the measurement. Blind check data showed that measurements were within tolerance 88% of the time, and were within 2x tolerance 99.1% of the time (n = 101 plots). We assumed that this level of measurement precision was adequate for using LIC as calibration and validation data in our modeling.

Bechtold (2003, 2004) developed regression equations that predict tree crown width from stem diameter and other predictor variables for 140 tree species in the United States. We used the equations from Bechtold (2003, 2004) that predict crown width from stem diameter. Surrogates were assigned to species lacking an equation based on similarities in tree crown shape. Since the model-fitting data used by Bechtold (2003, 2004) did not include trees < 5.0 in. diameter, we estimated crown width adjustment factors for saplings based on data from Bragg (2001).

The crowns of trees ≥ 5.0-in. diameter were modeled as symmetrical, circular polygons with area estimates based on the species-specific crown width equations. The modeled tree crowns were mapped within each subplot by placing the center of each crown at the coordinates recorded by field crews for stem-locations (figure 3). Canopy cover of trees ≥ 5.0-in. diameter was estimated by computing the proportion of the subplot polygons intersected by the mapped tree crowns, and then averaging the four subplot values to get a plot-level estimate (hereafter, referred to as SMC_{subp}, i.e., stem-map canopy cover of trees ≥ 5.0-in. diameter in the subplots). Canopy cover, by definition, was constrained to the 0 to100% range. The procedure was repeated for saplings in the microplots to compute a separate estimate of sapling canopy cover at the plot level, not accounting for overlap by the larger trees (SMC_{micr}). Geometric computations were done with a custom C program using the Geometry Engine Open Source library (http://trac.osgeo.org/geos/).

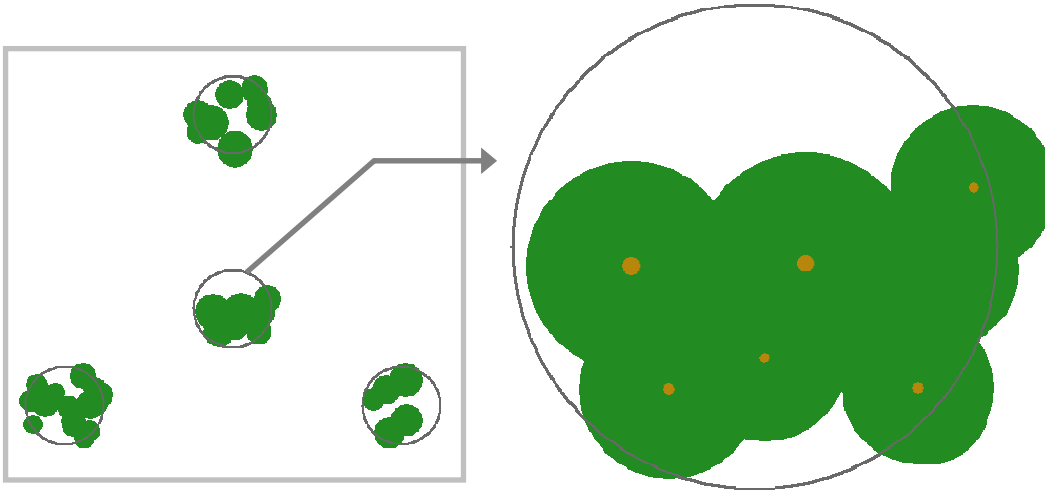


Figure 3: Example of mapping modeled crowns of trees ≥ 5.0-in. diameter in an FIA plot based on stem coordinates recorded by field crews. Measured stem diameters are drawn to scale as brown circles in the expanded view of subplot 1.

Because only trees ≥ 5.0-in. diameter can be stem-mapped across the entire FIA subplot cluster, we expected that SMC_{subp} would under-estimate LIC on average since LIC included all trees ≥ 1.0-in. diameter, including those < 5.0-in. To provide model-estimated canopy cover of trees ≥ 1.0-in. diameter, we developed a linear regression equation to predict the contribution of saplings to

total canopy cover. SMC_{subp} was compared to LIC, and a residual was calculated for each plot as $e = LIC - SMC_{subp}$. We assumed that e was due primarily to the exclusion of saplings from SMC_{subp}, and we expected e would be correlated with SMC_{micr}. However, since saplings may be overtopped by larger trees, a portion of the sapling cover may not contribute to total canopy cover in a given plot. The contribution of saplings to total canopy cover could also depend on other stand characteristics such as the density, height, and spatial pattern of larger trees.

Estimates of Ripley's K function (Ripley 1991, Venables and Ripley 2002) were used to account for the spatial pattern of trees ≥ 5.0-in. diameter. Ripley's K is useful for summarizing certain aspects of a pattern of points within a defined observation window, in this case, tree stem locations within the four subplots of an FIA plot. It can suggest whether an observed point pattern is either more clustered or more regularly spaced than would be expected for a random arrangement of the points. K(r) is a function of a distance r around each point in the pattern. The **spatstat** package version 1.11-7 (Baddeley and Turner 2005) for R (R Development Core Team 2006) was used to estimate K(r) for distances of 6, 8, 10, and 12 feet. The **Kest** function in **spatstat** estimates K(r) from a homogenous point pattern in a window of arbitrary shape. We specified the border method (Ripley 1991) in **spatstat** for edge correction of the point-pattern calculations. For regression modeling we worked with the square-root transformation $L(r) = \sqrt{K(r)/\pi}$, which stabilizes variance (Stoyan and Penttinen 2000).

Potential predictor variables considered for a regression model to predict e were:

SMC_{micr} = stem-mapped canopy cover estimate of saplings in the microplots
numSaplings = total number of saplings measured in the microplots
numTrees = total number of live trees ≥ 5.0 in. diameter measured in the subplots
meanTreeHtBAW = basal area-weighted mean height (ft.) of trees ≥ 5.0 in. diameter
meanL = mean of L(6 ft), L(8 ft), L(10 ft), and L(12 ft)

Single-condition IW-FIA plots measured during 1995 through 2005 were used for model fitting, and the 2006 plots were withheld as a validation set. Crown width models for certain western oak species (e.g., Gambel oak *Quercus gambelii*) had relatively large prediction errors (Bechtold 2004) and a review of plot photos in oak woodland forest types suggested that LIC would have relatively large measurement errors in these forest types. Plots classified as oak woodland forest types (FIA forest type codes 925 and 926) were removed from the model fitting data so they would not influence regression estimates. No plots were removed from the validation set. Linear regression models were fit using the **lm** function in R 2.4.1 (R Development Core Team 2006).

Smooth curves fitted by local polynomial regression (Venables and Ripley 2002) were used in exploratory analysis and as an aid to the interpretation of scatterplots in the Results. Smooth regression curves were fit with the **loess.smooth** function in R using the default span parameter of 2/3 (R Development Core Team 2006).

Assessment of Forest Area Estimates in NRS

Four of the 24 states within the NRS-FIA were selected to represent a range of conditions across the region, e.g., a range in latitude and longitude, temperature, precipitation, and dominant forest types. Selected states included: Michigan (MI), Missouri (MO), Pennsylvania (PA), and South Dakota (SD). Only plots from annual inventory years 2002 to 2006 were included, which was the most current evaluation group as of 5 August 2008. Compared with a base federal sampling intensity of one plot per approximately 6,000 acres, sampling intensity during this period was tripled in MI, doubled in southern MO, and at single (base) intensity in PA and SD. Nearly 375,000 tree records were queried from 10,075 plots where all conditions were forest land (i.e., 100 percent forested plots) and plots had tree records.

Per-state estimates of forest land area and accompanying sampling errors were produced using FIA's EVALIDator web-based estimation tool (http://www.fia.fs.fed.us/tools-data/other/). Estimates were produced for 1) all forest land with stocking of at least 10 percent, including seedlings, 2) all forest land with stocking of at least 10 percent, excluding seedlings, and 3) all forest land with model-estimated canopy cover of at least 10 percent (excludes seedlings). For comparison, estimates also were produced for 1) all forest land area, including all plots of any stocking, regardless of tree presence or condition proportion (equivalent to FIA's reported estimates of forest land area), and 2) forest land area of any stocking, but based on the queried subset of plots having tree records and 100 percent forested conditions. The last two estimates were used to determine 1) the reduction in estimated forest area resulting from constraining plots to a minimum of 10 percent stocking or canopy cover, and 2) the reduction in estimated forest area resulting from constraining plots to those with trees and 100 percent forested conditions.

Results

Canopy Cover Model

As expected, SMC_{subp} tended to underestimate LIC (figure 4). The mean difference (LIC - SMC_{subp}) was 9.0% canopy cover, while the mean absolute difference (|LIC - SMC_{subp}|) was ±11.6% canopy cover (n = 12,070 plots).

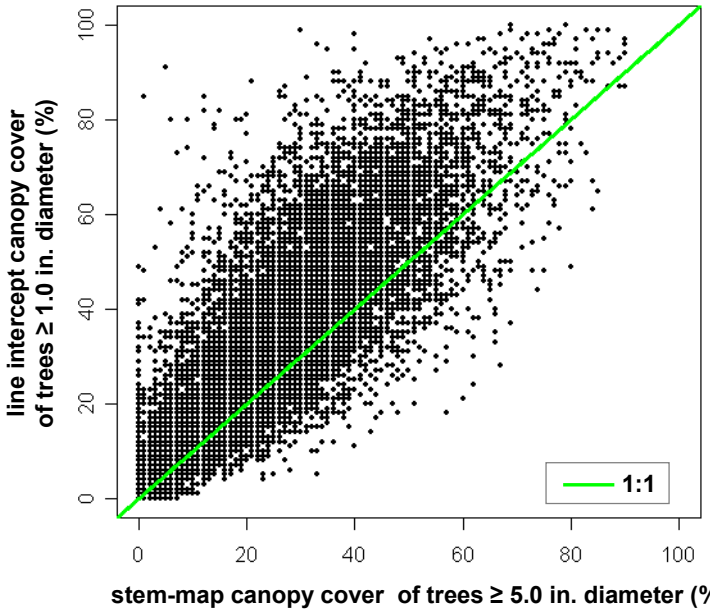


Figure 4: Comparison of the field-measured line intercept canopy cover of trees ≥ 1.0 in. diameter versus canopy cover of trees ≥ 5.0 in. diameter estimated by stem-mapping (n = 12,070 plots).

The data set shown in figure 4 was split into two subsets to examine the effect of sapling presence on the mean difference between stem-map and line-intercept canopy cover. Figure 5a shows the comparison of SMC_{subp} versus LIC for the subset of plots that had at least one sapling recorded in the microplots. SMC_{subp} underestimated LIC by an average of 11.4% canopy cover for these plots, while the mean absolute difference was ±13.7% canopy cover (n = 7,794 plots). Figure 5b shows the comparison of SMC_{subp} versus LIC for the subset of plots that had no saplings detected by the microplot subsample. SMC_{subp} underestimated LIC by an average of 4.7% canopy cover for these plots, while the mean absolute difference was ±7.7% canopy cover (n = 4,276 plots). The plots shown in figure 5a had low to high sapling canopy cover, while the plots shown in figure 5b likely had zero to low sapling canopy cover since the microplots are a $1/12^{th}$ subsample of each subplot. Some of the remaining 4.7% mean difference between SMC_{subp} and LIC is likely due to sapling cover not detected by the microplot sample.

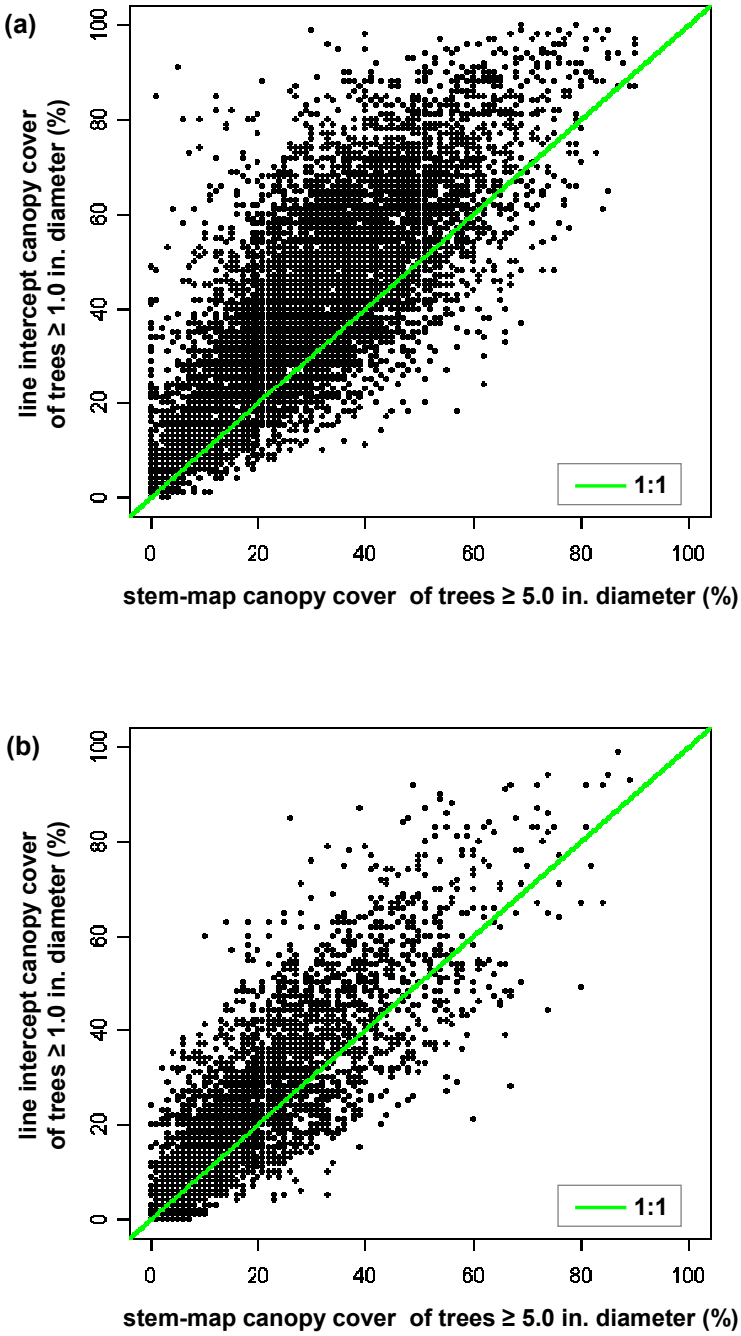


Figure 5: (a) Comparison of the field-measured line intercept canopy cover of trees ≥ 1.0 in. diameter versus canopy cover of trees ≥ 5.0 in. diameter estimated by stem-mapping, for the subset of plots that had at least one sapling recorded in the microplots (n = 7,794 plots), and (b) comparison of the field-measured line intercept canopy cover of trees ≥ 1.0 in. diameter versus canopy cover of trees ≥ 5.0 in. diameter estimated by stem-mapping, for the subset of plots that had no saplings detected by the microplot sample (n = 4,276 plots).

Residual canopy cover from the comparison of SMC_{subp} to LIC was positively correlated with SMC_{micr} (figure 6). In plots having low (<10%) canopy cover of trees ≥ 5.0 in. diameter, we assumed that overtopping of saplings by the crowns of larger trees was negligible, and total canopy cover was estimated by $SMC_{subp} + SMC_{micr}$. In plots having $SMC_{subp} \geq 10\%$, we assumed that saplings could be overtopped to some extent by the larger trees. For these plots, we estimated the sapling component of total canopy cover by multiple linear regression, taking into account the vertical structure and spatial pattern of the larger trees in addition to SMC_{micr}. The residual canopy cover (e = LIC - SMC_{subp}) shown in figure 6 was the dependent variable. Summary statistics for the variables used in regression analysis are shown in table 1, and estimated regression coefficients for predicting e are in table 2.

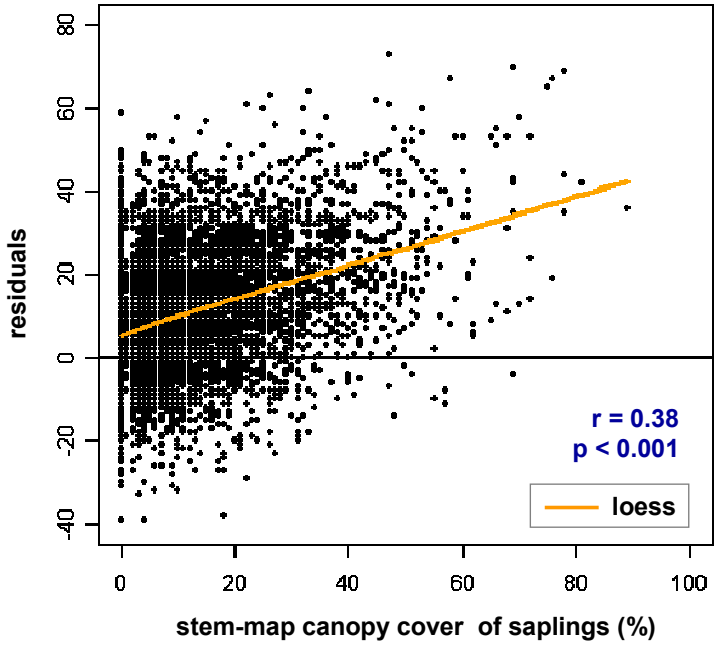


Figure 6: Relationship between canopy cover of saplings estimated by stem-mapping the microplots, and residuals calculated as the field-measured line intercept canopy cover of trees ≥ 1.0 in. diameter minus the canopy cover of trees ≥ 5.0 in. diameter estimated by stem-mapping the subplots. The orange line is a smooth regression curve fitted with the **`loess.smooth`** function in R.

Table 1: Ranges, means, and standard deviations of the variables used in a linear regression model to estimate the sapling component of total canopy cover. Residual canopy cover (e = LIC - SMC_{subp}) was the dependent variable.

Variable	Minimum	Mean	Std. dev.	Maximum
residual canopy cover (%)	-39	10.0	13.2	73
SMC_{micr} (%)	0	10.1	11.9	89
numSaplings	0	3.6	5.2	60
numTrees	1	27.3	17.2	132
meanTreeHtBAW (ft)	4.2	42.3	25.0	137.9
meanL	0	9.4	3.9	35.3

Table 2: Empirical model coefficients for estimating the sapling component of total canopy cover.

Variable	Coefficient	Standard error	p
(Intercept)	-8.036	0.699	< 0.001
SMC_{micr}	0.211	0.020	< 0.001
numSaplings	0.552	0.045	< 0.001
numTrees	-0.131	0.008	< 0.001
ln(meanTreeHtBAW)	4.367	0.201	< 0.001
meanL	0.222	0.034	< 0.001
RMSE	11.75		
d.f.	8,577		
R^2	0.21		

The final model-predicted canopy cover of trees ≥ 1.0 in. diameter was generated by the following steps:

1. Obtain estimates of SMC_{subp} and $SMC_{micr.}$
2. If $SMC_{subp} < 10$, use SMC_{micr} as an estimate the sapling component of total canopy cover. If $SMC_{subp} \geq 10$, estimate the sapling component of total cover using the regression coefficients in table 2, with predicted values restricted to ≥ 0.
3. Predicted canopy cover = SMC_{subp} + the estimated sapling component from step 2, with the restriction of modeled canopy cover ≤ 100.

Model Testing

A validation data set consisted of IW-FIA plot measurements during 2006 that were not included in the model-fitting data that were used to estimate the regression coefficients in table 2. Figure 7 shows the comparison of model-predicted canopy cover of trees ≥ 1.0 in. diameter versus LIC for the validation set. Model-predicted canopy cover was essentially unbiased with respect to LIC (mean difference -0.7), while the mean absolute difference was ±7.9% canopy cover (n = 1,454 plots). The correlation between the model-predicted and field LIC values was r = 0.875.

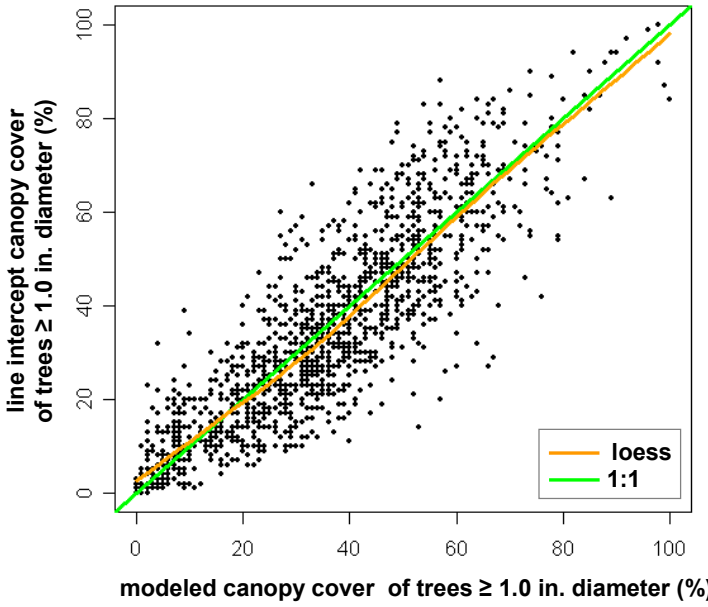


Figure 7: Comparison of field-measured line intercept canopy cover of trees ≥ 1.0 in. diameter versus model-predicted canopy cover of trees ≥ 1.0 in. diameter for the validation plots measured in 2006 (n = 1,454 plots).The orange line is a smooth regression curve fit to the data points with the **`loess.smooth`** function in R.

Forest Area Estimates

Estimates of all forest land area were 19,544,600 acres in MI, 15,078,279 acres in MO, 16,599,569 acres in PA, and 1,734,724 acres in SD (figure 8). Estimates of forest land area decreased when constrained to minimum thresholds of 1) at least 10 percent stocking of trees and seedlings (0.5 – 7.4 percent decrease), 2) at least 10 percent stocking of trees ≥ 1 in. DBH (0.4 – 9.6 percent decrease), or 3) at least 10 percent model-estimated canopy cover of trees ≥ 1 in. DBH (0.3 – 8.0 percent decrease). This equated to excluding nonstocked forest land from FIA estimates, and also excluding seedlings from the latter two criteria. South Dakota, the state with the lowest mean stocking and canopy cover, showed the largest percent decreases. However, none of differences among forest land area estimates based on the varying criteria for stocking and canopy cover within each state were statistically significant. Decreases due to excluding nonstocked plots and seedlings from the minimum canopy cover-based estimates were nearly identical to decreases due to excluding nonstocked plots and seedlings from the minimum stocking-based estimates.

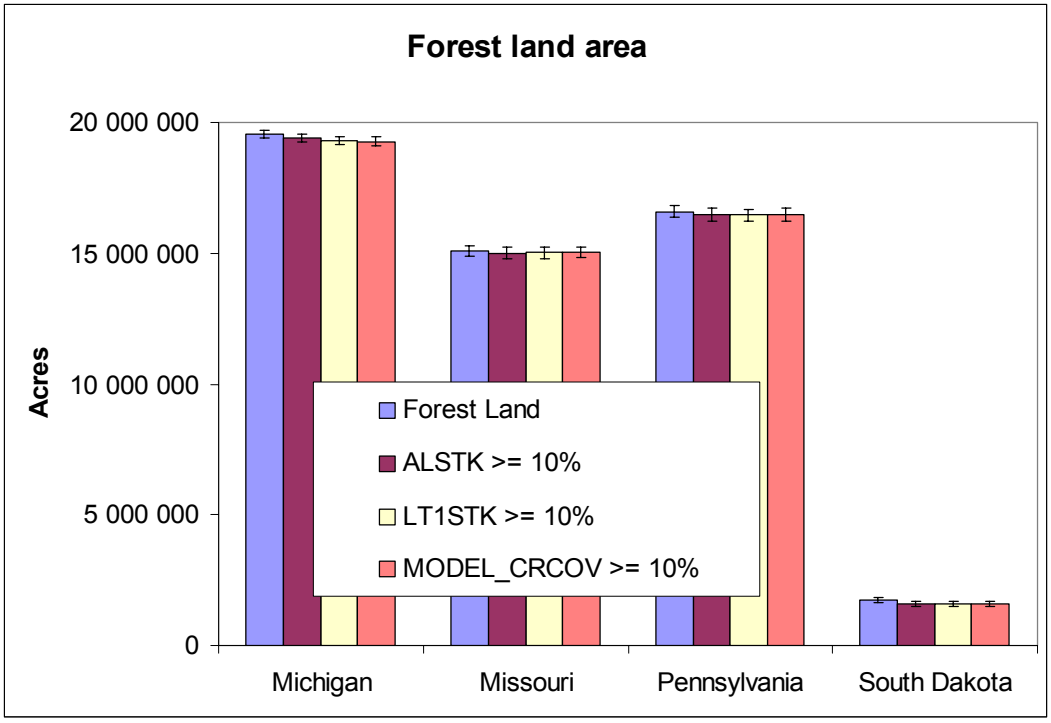


Figure 8: Forest land area estimates for selected states in NRS-FIA based on different criteria for stocking and canopy cover. Forest land = current FIA estimates of forest land area that include a land-use component in the definition of forest land. ALSTK = estimates based only on a minimum threshold for all tree stocking (trees, saplings, seedlings). LT1STK = estimates based only on a minimum threshold for stocking of trees ≥ 1 in. DBH (excludes seedlings). MODEL_CRCOV = estimates based only on a minimum threshold for model-estimated canopy cover of trees ≥ 1 in. DBH (excludes seedlings). Bars indicate 95% confidence intervals.

Discussion

The canopy cover model described here makes use of the tree spatial information available for FIA plots. Aspatial methods for estimating plot-level canopy cover may assume a random tree distribution (e.g., Crookston and Stage 1999). Compared with aspatial methods, a model based on stem-mapping could avoid bias when tree distributions are nonrandom, especially in stands with regular tree spacing such as plantations (Christopher and Goodburn 2008). Explicitly incorporating tree spatial pattern into the model is also expected to improve precision of the estimates compared with an aspatial approach.

Some characteristics of modeled canopy cover may be beneficial for certain applications. Model-estimated canopy cover is fully defined with respect to species and individual tree variables, e.g., “canopy cover of species considered tree life form by FIA (USDA Forest Service 2007) and having minimum stem diameter of 1 in.” In contrast, estimates of plot-level canopy cover derived from aerial photos generally are limited in terms of species and size class

discrimination, and may confuse tree canopy with background vegetation and shadows. Modeling canopy cover from tree data provides the flexibility of estimating various components of the total cover, such as canopy cover by species or canopy cover within different height layers. Field measurement of various canopy cover components, in addition to total tree cover, may be too costly to implement in some inventories since the time required on each plot would increase. A canopy cover model optimized for FIA plots provides a large data set at low additional cost, with prediction errors that may be tolerable in several applications.

FIA's current definition of forest land includes both a land-cover and a land-use component, and also implies potential future stocking. Although a minimum threshold of 10% stocking is specified, land that formerly was stocked (e.g., clear-cuts) is still considered forest if not developed for other uses (Bechtold and Patterson 2005). A proposed shift from a stocking- to a cover-based definition of forest land could retain a land-use requirement and likely would imply a minimum potential future canopy cover. Our analysis suggests that differences in forest area estimates resulting from the proposed change to a cover-based definition could be negligible. These results should be considered as preliminary since they included only four selected states in the NRS-FIA region, and model-estimated canopy cover was subject to the caveats described below.

Caveats and Current Limitations

The canopy cover model described above contains two parts: a geometric model of tree crowns within FIA subplots and an empirical model of the sapling contribution to total canopy cover that accounts for potential overlap from the crowns of larger trees. The empirical model was based on line-intercept field data collected within the IW-FIA region. Performance of the empirical model when extrapolated to other regions is currently unknown. It would be desirable for FIA to measure line-intercept canopy cover on at least a subset of plots in all regions. A $1/16^{th}$ subset like the FIA Phase 3 sample (Reams et al. 2005) should provide sufficient data for regional validation and calibration of canopy cover models, while adding only slightly to the time it takes to complete the plot measurements.

Canopy cover estimates based on mapping modeled tree crowns within the subplots currently do not include edge correction (not to be confused with edge correction applied to estimates of Ripley's K which deal strictly with point patterns derived from the stem coordinates). A possible source of bias is the presence of trees with stems outside the subplot boundary, but having crowns that cover a portion of the subplot. Since the line-intercept reference data used in the present study included trees $\geq$ 1.0 in. diameter, bias due to edge effects is confounded with bias due to omitting the saplings from the subplot stem-maps. Our analysis suggests that the remaining bias due to edge effects is probably small once saplings have been accounted for (cf. Nelson et al. 1997). However,

incorporating an edge correction method (e.g., Williams et al. 2003) would be desirable in a future version of the model.

Model development has focused so far on canopy cover estimation at the whole-plot level for single-condition plots. The model is expected to extend to multi-condition plots in which all condition classes are forested and could be used to estimate canopy cover of the whole plot footprint when a portion of the plot is nonforest. The model currently does not estimate canopy cover by condition, and multi-condition plots have not been used in model development or testing. The BOUNDARY table in FIADB (USDA Forest Service 2007) could be used to construct irregular subplot polygons when multiple conditions have been mapped within a subplot. This could support canopy cover estimation by condition, but the edge effects discussed above would become more problematic in small polygons. Future development of the model should address multi-condition plots.

Seedlings were not included in the IW-FIA line intercept measurements of canopy cover taken during 1995-2006, and we did not attempt to incorporate the seedling component of total cover into the model. However, in 2007 IW-FIA began including seedlings in the line-intercept measurements. It is possible that sufficient data are now available to account for seedlings in model-estimated canopy cover. The national standard seedling count from the microplots is a potential predictor variable. Future model development could address the seedling component of total cover.

The stem-map model also could be extended to produce canopy cover estimates by individual tree species as well as estimates within specified height layers. This functionality could be useful in applications such as wildlife habitat analysis and vegetation classification.

Software for Data Processing

Production of plot-level canopy cover estimates has been automated, requiring only a subset of the FIADB TREE table (USDA Forest Service 2007) as input (figure 9). The software avoids the use of a GIS for performance considerations and is based on open-source libraries. Crown widths are calculated using coefficients stored in lookup tables that have been populated for FIA tree species nationally based on Bechtold (2003, 2004). The software and lookup tables currently accommodate crown width equations of the quadratic form fit by Bechtold (2003, 2004) as well as a three-parameter nonlinear power function (Bragg 2001) and nonlinear power function in logarithmic scale (Shaw 2005). Users can mix these equation types if needed and substitute localized equations if available. Stem-map output variables are listed in table 3.

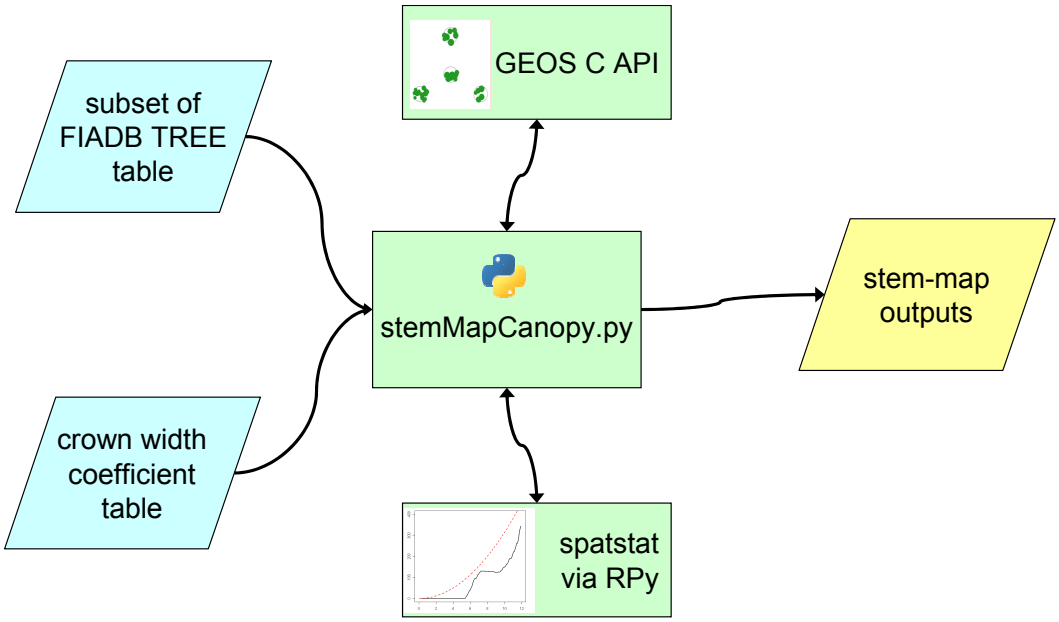


Figure 9: Flow chart describing a data processing system that implements the stem-map canopy cover model version 1.6. Lookup tables containing coefficients for crown width equations are populated for FIA tree species nationally. Batch processing is done with a Python program. Overlay of tree crowns on subplots and microplots is implemented in a C function using the Geometry Engine Open Source (GEOS) library. Estimates of Ripley's K are obtained from the **Kest** function in the **spatstat** package for R, called from Python via the **RPy** interface. The tabular output includes plot ids along with canopy cover estimates at the whole plot and subplot levels, several stand height metrics, and tree spatial pattern statistics.

Table 3: Outputs from FIA stem-map model version 1.6.

Variable	Description
map_crcov_subp	Estimated canopy cover of trees ≥ 5-in. diameter from crown-mapping the subplots (average of four subplots)
map_crcov_subp_s*n*	Estimated canopy cover of trees ≥ 5-in. diameter in subplot *n* from crown-mapping (*n* = 1 – 4)
map_crcov_micr	Estimated canopy cover of saplings from crown-mapping the microplots (average of four microplots)
map_crcov_micr_s*n*	Estimated canopy cover of saplings in the microplot of subplot *n* from crown-mapping (*n* = 1 – 4)
model_crcov	Estimated total canopy cover of trees ≥ 1-in. diameter, derived from crown-mapped subplots plus a regression estimate of the sapling component
meanTreeHt	Mean height of trees ≥ 5.0 in. diameter
meanTreeHtBAW	Basal-area weighted mean height of trees ≥ 5.0 in. diameter
meanTreeHtDom	Mean height of canopy dominant/co-dominant trees ≥ 5.0 in. diameter
meanTreeHtDomBAW	Basal-area weighted mean height of canopy dominant/co-dominant trees ≥ 5.0 in. diameter
maxTreeHt	Height of the tallest tree ≥ 5.0 in. diameter
predomTreeHt	Predominant tree height (mean height of the tallest trees ≥ 5.0 in. diameter comprising 16 trees per acre)
meanSapHt	Mean height of saplings (≥ 1.0 in. diameter and < 5.0 in. diameter)
maxSapHt	Height of the tallest sapling
K_*r*ft	Estimates of Ripley's K function at *r* feet (*r* = 6, 8, 10, and 12)
L_*r*ft	Estimates of Ripley's L function at *r* feet (*r* = 6, 8, 10, and 12)
G_*r*ft	Estimates of the nearest neighbor distribution function at *r* feet (*r* = 6, 8, 10, and 12)

Acknowledgements

Many thanks go to the data collection and compilation staff of the USDA Forest Service FIA Program who made this work possible. Portions of this work were performed under a Memorandum of Understanding between FIA and the inter-agency LANDFIRE Project (www.landfire.gov) in support of vegetation mapping in LANDFIRE. Sara Goeking and Will McWilliams provided helpful comments on the paper.

Literature Cited

Baddeley, A. and R. Turner. 2005. Spatstat: an R package for analyzing spatial point patterns. *J. Stat. Software* 12:(6)1-42.

Bechtold, W.A. 2003. Crown-diameter prediction models for 87 species of stand-grown trees in the eastern United States. *South. J. Appl. For*. 27(4):269-278.

Bechtold, W.A. 2004. Largest-crown-width prediction models for 53 species in the western United States. *West. J. Appl. For*. 19(4):245-251.

Bechtold, W.A., and P.L. Patterson (eds.). 2005. *The enhanced Forest Inventory and Analysis program – national sampling design and estimation procedures*. USDA For. Serv. Gen. Tech. Rep. SRS-80.

Bechtold, W.A. and C.T. Scott. 2005. The forest inventory and analysis plot design. P. 27-42 in *The enhanced Forest Inventory and Analysis program – national sampling design and estimation procedures*, Bechtold, W.A., and P.L. Patterson (eds.) USDA For. Serv. Gen. Tech. Rep. SRS-80.

Bond, M.L, M.E. Seamans, and R.J. Gutiérrez. 2004. Modeling nesting habitat selection of California spotted owls (*Strix occidentalis occidentalis*) in the central Sierra Nevada using standard forest inventory metrics. *For. Sci*. 50(6):773-780.

Bragg, D.C. 2001. A local basal area adjustment for crown width prediction. *North. J. Appl. For*. 18(1):22-28.

Christopher, T.A. and J.M. Goodburn. 2008. The effects of spatial patterns on the accuracy of Forest Vegetation Simulator (FVS) estimates of forest canopy cover. *West. J. Appl. For*. 23(1):5-11.

Crookston, N.L. and A. R. Stage. 1999. *Percent canopy cover and stand structure statistics from the Forest Vegetation Simulator*. USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-24. 11 p.

Fiala, A.C.S, S.L. Garman, and A.N. Gray. 2006. Comparison of five canopy cover estimation techniques in the western Oregon Cascades. *For. Ecol. Manage*. 232:188-197.

Finney, M.A. 2004. FARSITE: Fire Area Simulator-model development and evaluation. Res. Pap. RMRS-RP-4, USDA For. Serv., Ogden, Utah. 47 p.

Gill, S.J., G.S. Biging, and E.C. Murphy. 2000. Modeling conifer tree crown radius and estimating canopy cover. *For. Ecol. Manage*. 126:405-416.

Jennings, S.B., N.B. Brown, and D. Sheil. 1999. Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other methods. *Forestry* 72(1):59-73.

Lund, H.G. 2002 When is a forest not a forest? *J. Forestry* 100(8):21-28.

Nelson, R.F., T.G. Gregoire, and R.G. Oderwald. 1997. The effects of fixed-area plot width on forest canopy height simulation. *For. Sci*. 44(3):438-444.

Pollard, J.E., J.A. Westfall, P.L. Patterson, D.L. Gartner, M. Hansen, and O. Kuegler. 2006. *Forest Inventory and Analysis national data quality assessment report for 2000 to 2003*. USDA For. Serv. Gen. Tech. Rep. RMRS-GTR-181. 43p.

R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reams, G.A., W.D. Smith, M.H. Hansen, W.A. Bechtold, F.A. Roesch, and G.G. Moisen. 2005. The Forest Inventory and Analysis sampling frame. P. 11-26 in *The enhanced Forest Inventory and Analysis program – national sampling design and estimation procedures*, Bechtold, W.A. and P.L. Patterson (eds.) USDA For. Serv. Gen. Tech. Rep. SRS-80.

Ripley, B.D. 1991. *Statistical inference for spatial processes*. Cambridge University Press. 154 p.

Shaw, J.D. 2005. Models for estimation and simulation of crown and canopy cover. P. 183-191 in *Proceedings of the fifth annual forest inventory and analysis symposium*, McRoberts, R.E., G.A. Reams, P.C. Van Deusen, and W.H. McWilliams (eds.) USDA For. Serv. Gen. Tech. Rep. WO-69.

Stoyan, D. and A. Penttinen. 2000. Recent applications of point process methods in forestry statistics. *Stat. Sci*. 15(1):61-78.

USDA Forest Service. 2005. *Forest Inventory and Analysis national core field guide. Volume I: field data collection procedures for phase 2 plots*. Available online at http://fia.fs.fed.us/library/field-guides-methods-proc/; last accessed Nov. 10, 2008.

USDA Forest Service. 2006. *Interior West Forest Inventory and Analysis field procedures, version 3.0*. Available online at http://www.fs.fed.us/rm/ogden/data-collection/field-manuals.shtml; last accessed Jul. 22, 2008.

USDA Forest Service. 2007. *The forest inventory and analysis database: database description and users guide version 3.0*. Available online at http://fiatools.fs.fed.us/fiadb-downloads/fiadb3.html; last accessed Mar. 27, 2008.

Venables, W.N., and B.D. Ripley. 2002. *Modern applied statistics with S, fourth ed*. Springer Science, New York, NY. 495 p.

Vospernik, S., M. Bokalo, F. Reimoser, and H. Sterba. 2007. Evaluation of a vegetation simulator for roe deer habitat predictions. *Ecol. Model*. 202:265-280.

Williams, M.S., P.L. Patterson, and H.T. Mower. 2003. Comparison of ground sampling methods for estimating canopy cover. *For. Sci*. 49(2):235-246.

Correct County Areas with Sidebars for Virginia

Joseph M. McCollum[1]
Dale Gormanson[2]
John Coulston[3]

Abstract: *Historically, Forest Inventory and Analysis (FIA) has processed field inventory data at the county level and county estimates of land area were constrained to equal those reported by the Census Bureau. Currently, the Southern Research Station FIA unit processes field inventory data at the survey unit level (groups of counties with similar ecological characteristics). In short, the estimation unit has changed from an individual county to groups of counties. However, changing the estimation unit to a survey unit does not constrain the total area of each county when expansion factors are used. Here we provide a method to constrain published FIA estimates of total area at the county level to equal those of the Census Bureau. To accomplish this, we derive an expansion factor formula. Virginia serves as a case example of the techniques because of its unique geo-political boundaries.*

Keywords: *FIA, forest inventory, area control, sub-population, expansion factor*

Introduction

Forest Inventory and Analysis (FIA) is a national program of the USDA Forest Service charged with conducting and maintaining comprehensive forest resource inventories on both public and private lands in the United States. FIA inventories are design-based to provide strategic and extensive information on the quantity, status, and use of the forest resource. The sampling frame is based on a tessellation of the United States into approximate 6,000-acre hexagons derived from a larger hexagon grid developed for the Environmental Monitoring and Assessment Program (EMAP) methodology (White *et al.* 1992). Each hexagon contains one plot for the base national program. The hexagonal array is comprised of non-overlapping, interpenetrating panels.

[1]Information Technology Specialist, Southern Research Station, 4700 Old Kingston Pike, Knoxville TN 37919
[2]Supervisory Forester, Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108
[3]Supervisory Research Forester, Southern Research Station, 4700 Old Kingston Pike, Knoxville, TN 37919

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

FIA estimates are based on a three-phase sampling scheme. The primary objective of Phase 1 is to use remote sensing to stratify land areas to increase the precision of forest resource estimates. The Phase 1 stratification assigns each Phase 2 plot to a stratum and identifies, or estimates, the area of the stratum within the estimation unit. The primary objective of Phase 2 is to obtain field observations and measurements of the traditional FIA suite of forest resource variables. Phase 3 is a subsample of the Phase 2 plots (1 in 16) on which additional forest health attributes are measured.

Bechtold and Patterson (2005) provide the compilation procedures for the national FIA program. Estimates of forest attributes are made for each population or estimation unit using these procedures. An estimation unit is typically defined as either a county or an FIA multi-county survey unit of known area (determined by the Census Bureau). The estimation procedures do not require plot level expansion factors. However, for public distribution of the FIA database (FIADB) expansion factors are used so that users can match estimates from National Information Management Systems Compilation System (NIMSCS) at the estimation unit level using the expansion factors. In addition, the expansion factors can be used to make population estimates for arbitrarily shaped areas such as wood procurement regions surrounding mills. However, when expansion factors are developed at the FIA survey unit level (groups of counties), there is no area control at finer spatial resolution than the survey unit. Therefore county-level area totals based on expansion factors will not equal the area of the county reported by the Census Bureau.

As noted above, total area estimates provided by the Census Bureau are used for area control. In Virginia, there are independent cities in no county at all (Figure 1), but reported by the Census Bureau as county-equivalents. While the Census Bureau does provide area estimates for each independent city, about 1/3 of Virginia's untabulated cities have no FIA plot at all, and thus no mechanism for area control (expansion factors).

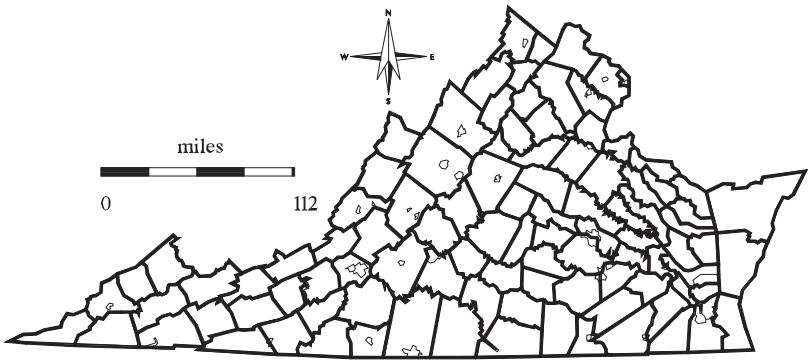


Figure 1: Virginia map showing independent cities

Since 1966, FIA has assigned each of Virginia's cities to one or more counties, as shown in Figure 1. The Tidewater cities of Chesapeake, Hampton, Newport News, Suffolk, and Virginia Beach are reported as counties. Each of those cities is reported as a county-equivalent, because they were once counties. Newport News was Warwick County, Hampton was Elizabeth City County, Virginia Beach was Princess Anne County, Chesapeake was Norfolk County, and Suffolk was Nansemond County.

The objectives of this study are 1) to document rules used by FIA for assigning independent cities to the appropriate county in Virginia and 2) to extend the formula for calculating expansion factors to address situations where the population and estimation unit are defined by FIA survey units but the subpopulation of interest has a known area that a client wants to use for area control.

Methods

Independent Cities

As noted previously, independent cities are separated from counties in Virginia. FIA combined the independent cities with certain counties, but did not outline rules for doing so. What follows are a set of formal rules.

a) A city should be tabulated with its county of origin at the time it became an independent city.
b) If a city is created from multiple counties, the historical county lines should be used.
c) If a city consolidates with a county or another city, the new jurisdiction shall inherit the lineage of the old jurisdiction. Consequently, if a city's county of origin no longer exists, the city may be tabulated under its own name.
d) With all other acquisitions short of consolidation, the new jurisdiction does not inherit lineage.

Rule (a) covers most cases. Rule (b) covers Galax and Petersburg. Rule (c) covers Richmond (tabulated with Henrico and Chesterfield Counties), and the Tidewater cities. Rule (d) covers a number of other cases. These rules capture what FIA has actually done, with two exceptions – Petersburg and Alexandria.

FIA split Petersburg between Dinwiddie and Prince George Counties. McCollum and Jacobs (2002) surmised the historical line dividing Dinwiddie and Prince George Counties to be a line northward from the point of intersection of Dinwiddie, Prince George, and the city of Petersburg**.** The Laws of Virginia**,** Chapters 19 and 31, 1752, (Hening 1823) show that this interpretation was correct south of Lieutenant Run, but otherwise the county line followed that creek, and also that Pocahontas Island was part of Chesterfield County.

The other exception is Alexandria. Alexandria County was ceded to the District of Columbia in 1791 to form the capital city of the United States, a diamond shape that was ten miles square. The District took exclusive jurisdiction by 1801. In 1846, it was retroceded to Virginia. For the Census of 1900, the Census Office tabulated the city of Alexandria separately from Alexandria County. In 1920, Alexandria County renamed itself Arlington County. Since 1915, the city of Alexandria has spread into what was Fairfax County (DenBoer and Sinko 2005; Minnesota Population Center 2004). Since the city of Alexandria was created from Arlington County, it should be tabulated entirely with Arlington County.

Expansion Factors

The techniques used to estimate population totals do not require the use of expansion factors. In fact, Bechtold and Patterson (2005) discourage the use of expansion factors. However, expansion factors are included in the public version of the FIADB. There are two motivating factors for including expansion factors: 1) external clients can easily produce population estimates that match estimates based on stratified estimation when the estimation unit is the same (i.e. survey unit) and 2) external clients can also produce estimates for user defined subpopulations (e.g. wood procurement regions). Based on supplemental material to Bechtold and Patterson (2005) expansion factors for each stratum within an estimation unit are calculated using [1].

$$F_h = \frac{A_T W_h}{n_h} \quad [1]$$

where

F_h is the expansion factor for stratum h
A_T is the total area of the population (survey unit),
W_h is the weight of stratum h, and
n_h is the number of plots in stratum h.

For any subpopulation of known area within the population, the expansion factor can be defined by:

$$F_{hc} = fF_h \quad [2]$$

where

$$f = \frac{A_C}{\sum_{i=1}^{n}\sum_{h=1}^{H}\delta_i(c)F_h} \quad [3]$$

A_C is the area of the subpopulation (e.g. county),

n is the total number of plots in the population (survey unit),
H is the total number of strata,
$\delta_i(c)$ is an indicator function equal to 1 if plot i is in subpopulation C, 0 otherwise, and all other notation is as above.

In general, f will not be equal to 1 for every county in the unit. However, on average, f will be 1. Therefore, if a survey unit consists of C counties, numbered $c = 1$ to C, then:

$$\sum_{c=1}^{C}\sum_{i=1}^{n}\sum_{h=1}^{H}\delta_i(c)\frac{A_T W_h}{n_h} = \sum_{c=1}^{C} A_c = A_T \quad [4]$$

Case Study

We use Virginia as a case study to document how FIA combines independent cities into counties and FIA survey units. We then demonstrate how external clients can recalculate expansion factors to produce estimates for individual counties within the estimation unit (FIA survey unit in this case). The purpose of the exercise is to produce county-level estimates where the total area of the subpopulation is constrained to equal the total area of the county reported by the Census Bureau (including independent cities).

Results

Independent cities were merged into their respective counties using the decision rules presented in this paper. However because the rules for combining independent cities with counties have not been consistent overtime and the boundaries of independent cities have changed over time, the total land area for some individual counties (e.g. Arlington County) has changed (Table 1). In the Knight and McClure report (1966), the city of Alexandria was split between Arlington and Fairfax Counties, apparently along the "D.C. Diamond" line. In subsequent reports, Arlington County meant Arlington County proper. Decade by decade, estimates of land area changed to mirror the Census Bureau's. In 2005, FIA changed its methodology from using the county as area control to using the survey unit as area control.

Table 1: Estimates of land in Arlington County (FIA)

Year	Acres	Source
1966	20500	Knight and McClure 1966
1976	16640	Sheffield 1976
1986	16614	Brown 1986
1992	16614	Johnson 1992
2001	16557	Miles 2001
2005	15218	FIDO[a]
2006	17504	FIDO
2007	16943	FIDO

[a]Forest Inventory Data Online (FIDO) is a modernized delivery system that provides access to the National FIA databases. It allows flexible user specified tables and maps of forest statistics to be generated through a web browser (USDA Forest Service 2008).

By definition the total area of the survey unit calculated from expansion factors using equation [1] equals the total area of the survey unit calculated from expansion factors using equation [2]. County level estimates of total area differ slightly (table 2). Data is taken from the FIA Datamart, available at *http://fiatools.fs.fed.us/fiadb-downloads/datamart.html*. For example, the total area of Accomack County was 902,292 acres based on using expansion factors from equation [1] and was 838,426 acres based on expansion factors from equation [2]. The main point here is that when using the expansion factors from equation [2], the total area estimate matches the total area from the Census Bureau for each county. Forest area estimates will also differ slightly. In absolute terms, the largest difference in forest area estimates was observed in Southampton County and the smallest difference was observed in Newport News.

Table 2: Estimates for accessible forest land and all land for each county in survey unit 1

County	Equation 1		Equation 2	
	Forest	Total Area	Forest	Total Area
	acres			
Accomack	107,858	902,292	95,926	838,426
Brunswick	269,650	366,187	266,808	364,396
Caroline	244,851	334,892	249,181	344,886
Charles City	91,566	136,533	83,158	130,698
Chesterfield	164,719	312,681	163,029	307,407
Dinwiddie	223,947	304,611	241,397	330,595
Essex	79,743	163,477	84,258	182,980
Gloucester	88,706	192,481	90,768	184,318
Greensville	122,214	166,137	142,338	194,413
Hanover	182,262	280,199	197,689	303,411
Henrico	57,868	179,261	54,572	173,913
Isle of Wight	106,279	222,418	99,897	232,166
James City	63,159	124,641	55,687	120,565
King and Queen	136,543	200,117	142,287	208,849
King George	73,695	108,444	82,773	120,186
King William	103,273	176,187	107,823	182,818
Lancaster	44,413	134,452	38,731	148,062
Mathews	24,654	175,815	29,571	161,257
Middlesex	52,581	132,173	44,922	134,882
New Kent	83,349	118,952	97,137	143,027
Northampton	25,673	574,274	20,540	509,017
Northumberland	57,096	192,872	56,894	182,825
Prince George	106,301	191,113	107,838	195,923
Richmond	78,181	149,916	74,917	138,481
Southampton	239,535	359,385	264,726	390,880
Surry	142,175	193,540	148,079	198,595
Sussex	258,098	319,428	253,765	315,411
Westmoreland	87,960	154,375	88,944	161,689
York	46,548	205,838	42,130	188,124
Chesapeake	93,003	310,663	96,909	316,047
Hampton	1,168	88,799	866	87,187
Newport News	5,879	74,395	5,710	76,192
Suffolk	175,330	272,116	179,207	274,603
Virginia Beach	35,856	341,854	31,881	318,287
Total	3,674,133	8,160,518	3,740,358	8,160,516

Discussion

Southern Research Station FIA produces FIA survey unit population estimates and does not produce county-level population estimates. However, several clients are interested in or require county-level estimates. When county-level estimates are required we suggest that clients recalculate the expansion factors so that there is area control at the county-level. Here we provide a method for adjusting expansion factors for this purpose.

Virginia provides a particular challenge for using Census Bureau estimates of total area for area control. The independent city boundaries can change over time and therefore it is important to have a set of decision rules in place for allocating the area of each independent city to the correct county. Admittedly, this is less of an issue when an FIA survey unit is used as the estimation unit. This is because independent cities are more likely to straddle multiple counties than they are to straddle multiple FIA survey units.

Here we present a technique to recalculate expansion factors that results in area control at the county level. This technique adjusts the expansion factor by dividing a known area (A_C in equation 2) by the calculated area based on expansion factors. The same technique can be used for wood procurement regions. These areas are generally used to determine wood volume estimates within some distance of, for example, a plywood mill. The total area of the region is therefore known and equation 2 can be used, provided that the procurement region is contained within one survey unit.

Our objectives were to document rules used by FIA for assigning independent cities to the appropriate county in Virginia and extend the formula for calculating expansion factors to address situations where a client requires subpopulation estimates with area control. We recommend that the formal rules described here be used by FIA for allocating the area of independent cities to the appropriate county. We also recommend that external clients who require sub-population estimates consider adjusting the expansion factors so that area control can be achieved at the sub-population level.

Literature Cited

Bechtold, William A.; Patterson, Paul L., Editors. 2005. The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Brown, Mark J. 1986. Forest statistics for Virginia, 1986 Resour. Bull. SE-087. Asheville, NC: U.S. Department of Agriculture, Southeastern Forest Experiment Station. 74 p..

DenBoer, G., and Sinko, P.T. Virginia Historical Counties. Data Set. Laura Rico-Beck, digital comp. Atlas of Historical County Boundaries, ed. by John H. Long. Chicago: The Newberry Library, 2005. Available online from http://www.newberry.org/ahcbp, last accessed 1 June 2008.

Hening, W.W. 1823. The Statutes At Large; Being A Collection of All the Laws of Virginia from the First Session of the Legislature, in the Year 1619.

Johnson, Tony G. 1992. Forest statistics for Virginia, 1992 Resour. Bull. SE-131. Asheville, NC: U.S. Department of Agriculture, Southeastern Forest Experiment Station. 76 p..

Knight, Herbert A.; McClure, Joe P. 1966. Virginia's timber, 1966 Resour. Bull. SE-8. Asheville, NC: U.S. Department of Agriculture, Southeastern Forest Experiment Station. 64 p..

McCollum, Joseph; Jacobs, Dennis 2005. TIGER 2000 and FIA In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; McWilliams, William H.; Cieszewski, Chris J., eds. Proceedings of the fourth annual forest inventory and analysis symposium; Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 223-229.

Miles, P.D. 2008. Cycle 7, Virginia 2001. Forest Inventory Mapmaker version 2.1. St. Paul, MN: U.S. Department of Agriculture, Forest Service. North Central Research Station. [Available only on internet: www.ncrs2.fs.fed.us/4801/fiadb/index.htm]. Last accessed 1 June 2008.

Minnesota Population Center. 2004. National Historical Geographic Information System: Pre-release Version 0.1. Minneapolis, MN: University of Minnesota 2004. Available online at http://www.nhgis.org. Accessed April 21, 2008.

Sheffield, Raymond M. 1976. Forest statistics for the Northern Piedmont of Virginia 1976 Resour. Bull. SE-39. Asheville, NC: U.S. Department of Agriculture, Southeastern Forest Experiment Station. 40 p..

USDA Forest Service. 2008. *Forest Inventory Data Online* [Online]. Available: http://199.128.173.26/fido/mastf/index.html. [2008, November 12].

White, D., Kimerling, J., and S.W. Overton. 1992. Cartographic and geometric components of a global sampling design for environmental monitoring. Cartographic and Geographic Information Sciences 19: 5-22.

Statistical Considerations



An assessment of the impact of FIA's default assumptions on the estimates of coarse woody debris volume and biomass

Vicente J. Monleon[1]

***Abstract**: Currently, Forest Inventory and Analysis estimation procedures use Smalian's formula to compute coarse woody debris (CWD) volume and assume that logs lie horizontally on the ground. In this paper, the impact of those assumptions on volume and biomass estimates is assessed using 7 years of Oregon's Phase 2 data. Estimates of log volume computed using Smalian's formula are known to be biased, overestimating volume. On the other hand, volumes estimated from the diameter at the point of intersection between the log and the transect are approximately unbiased, regardless of log shape, but may be more variable. In Oregon, Smalian's formula overestimated CWD volume and biomass by 3.6 and 4.2 percent, respectively, compared with the intersection diameter method, or 1.7 and 2.0 times the standard error of the estimates. The impact on the variance of the estimates was negligible. The assumption that logs lie horizontally would result in an underestimation of the total CWD volume. The sensitivity to this assumption was examined under several scenarios, suggesting that the bias may be between 4.1 and 8.1 percent of the volume estimates.*

Keywords: Down woody debris, line intersect sampling, Smalian's formula, intersect diameter, log inclination, Oregon.

Introduction

Coarse woody debris (CWD), defined as large pieces (logs) of down and dead wood in different stages of decay, plays a key role in ecosystem structure and function (Harmon et al. 1986). CWD may also account for a relatively large proportion of the total biomass and carbon pools in many forests ecosystems. In Oregon, for example, CWD volume and biomass are estimated to be 44% and 18% the net volume and biomass in live trees, respectively (Donnegan et al. 2008).

The Forest Inventory and Analysis (FIA) program uses line intersect sampling (LIS) to estimate the volume and biomass in down logs. Details of the measurement and sampling design, and estimation procedures, can be found in Woodall and Monleon (2008). In essence, a log is tallied if its centerline is intersected by the transect. Then, a LIS estimator for the total volume of down logs from the *i*-th plot, $\hat{\tau}_i$, is:

[1] United States Forest Service; Pacific Northwest Research Station; Forest Inventory and Analysis Program; 620 SW Main Street, Suite 400; Portland, OR 97205 USA; vjmonleon@fs.fed.us

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

$$\hat{\tau}_i = K \sum_{j=1}^{n_i} \frac{v_{ij}}{l_{ij} \cos\left(\theta_{ij}\right)} \quad [1]$$

where v_{ij} is the measured volume from the *j*-th log; $l_{ij} \cos\left(\theta_{ij}\right)$ is the length of the horizontal projection of the *j*-th log's centerline (l_{ij} is the length of the centerline and θ_{ij} its inclination); *K* is a constant term that depends on transect length; and n_i is the number of down logs tallied in the plot. To estimate total biomass, the volume v_{ij} is multiplied by the log's bulk density and by a reduction factor to account for loss of mass due to decay (Harmon et al. 2008). Once an estimate for each plot is available, population estimates are obtained following standard FIA procedures (Scott el at. 2005).

Equation 1 requires that the individual log's volume be known, but volume is almost always estimated, not measured. FIA currently uses Smalian's formula to estimate volume for logs attributed as decay class 1, 2 3 or 4 logs (Waddell 2002, Woodall and Monleon 2008):

$$v_{ij} = \frac{1}{2}\left[\pi \left(\frac{DL_{ij}}{2} \right)^2 l_{ij} + \pi \left(\frac{DS_{ij}}{2} \right)^2 l_{ij} \right] \quad [2]$$

where DL_{ij}, DS_{ij}, and l_{ij} are the large- and small-end diameters and length of the log, respectively. This equation is the average of the volume of two cylinders, one with cross-section equal to that of the large end of the log, and another with cross section equal to that of the small-end. It gives the volume of a frustum of a paraboloid and, therefore, if the log's shape is different, bias would result. Husch et al. (1972: 120) indicate that Smalian's formula tends to overestimate the volume of logs, reporting biases as large as 12%. They suggest that "Unless one is willing to accept a rather large error, Smalian's formula should not be used unless it is possible to measure the sections of the tree in 4-foot lengths."

Alternatively, volume may be estimated from the diameter at the point of intersection between the log's centerline and the transect (van Wagner 1968):

$$v_{ij} = \pi \left(\frac{DI_{ij}}{2} \right)^2 l_{ij} \quad [3]$$

where DI_{ij} and l_{ij} are the intersection diameter and length of the log, respectively. This method gives an unbiased estimate of volume, without any assumptions about the shape of the logs. It can be motivated as a crude Monte Carlo approach to volume estimation. The volume of a log of arbitrary shape may be estimated by taking a random sample of points along the log's centerline, measuring the area of the cross-section at each point, multiplying by the log's length, and averaging. Because the point of intersection is a random point, multiplying the area of the cross-section at the intersection by the log length is basically a Monte Carlo

approximation to the volume based on a single sample. It follows that the estimator is unbiased, but it may be highly variable. Therefore, using intersection diameter instead of Smalian's formula may result in greater variance.

Equation 1 also requires knowledge of the inclination of the log, to calculate the length of the horizontal projection of the log's centerline. Currently, FIA does not measure log inclination, implicitly assuming that it is 0 or that it can be ignored. Logs may not lie horizontally because the terrain is not flat, they are supported but other trees or structures, or they are snags leaning more that 45 degrees from vertical (which are logs by FIA definition). Ignoring log inclination underestimates the total volume, with the bias being proportional to the reciprocal of the cosine of the inclination. So, if the inclination is 10%, actual volume is only 0.5% greater than the reported volume. However, this figure rises to 41% when the inclination is 100% (45 degrees).

The objective of this study is to evaluate the impact of the current assumptions regarding individual log volume estimation and log inclination on the estimates of total CWD volume and biomass in Oregon.

Materials and methods

Data

CWD data were collected in all FIA Phase 2 plots in Oregon, between 2001 and 2007 (7 panels). Two, 58.9 foot-long transects were measured in each subplot, for a total of 8 transects and 471.2 ft per plot. If the centerline of a log was intersected by the transect, the length, decay class and diameter at the intersection point were recorded. If the decay class was 1 through 4, the large- and small-end diameters were also recorded. Only logs in forestland were tallied. Details of the field procedures can be found at USDA Forest Service (2007). In total, the dataset includes 7,115 plots (51% forested) and 58,241 logs.

Individual log volume estimation

To assess the effect of the method used to estimate individual log volume, total volume and biomass and their variances were estimated after computing the volume of each individual log using either Smalian's formula (eq. 2) or the intersection diameter formula (eq.3). Only decay class 1-4 logs were included in the analysis because, lacking end diameter measurements, Smalian's formula can not be computed for decay class 5 logs. Log inclination was set to 0 (FIA default assumption). Estimates were obtained as described in Woodall and Monleon (2008).

Log inclination

Log inclination was not recorded in the field, but the slope (inclination) of the subplot was recorded. To estimate the impact of inclination, four scenarios based on the slope of the subplot were considered. We assumed that logs were lying on

the ground, so that their inclination depended only on the slope of the terrain and orientation of the log with respect to the slope. Therefore, the impact on the estimates of logs that did not lie on the ground because they are supported was not considered. The four scenarios were (Fig. 1):

A. Default: logs lie horizontally (θ_{ij} in eq. 1 equals 0). This is equivalent to assuming that all logs are oriented perpendicular to the slope.
B. Log orientation is uniform, independent of the slope. Logs may be facing downslope or across the slope, or in any other orientation, with equal probability
C. Log orientation is predominantly downslope, according to the histogram in figure 1.
D. All logs are oriented downslope (θ_{ij} in eq. 1 equals the subplot slope)

Scenario D would provide an upper bound for the bias, while scenario B, in which the orientation of the pieces are not affected by the slope, would be a reasonable lower bound. Log volume was calculated from the intersect diameter, so that all decay classes could be included. For scenarios B and C, estimates are based on the average of 1000 simulations from the appropriate log orientation distribution.

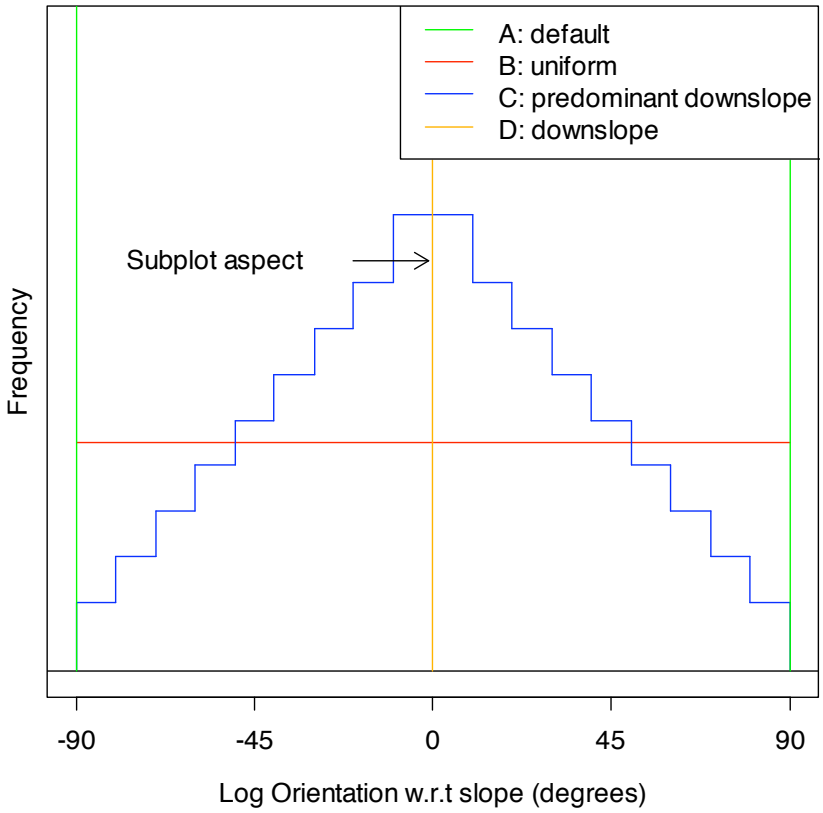


Figure 1: Distribution of the orientation of logs with respect to subplot slope for the four scenarios considered in this study. For scenarios A and D, all pieces are oriented perpendicular to the slope (-90 or 90 degrees) or downslope (0 degrees), represented here by vertical lines at those points.

Results

Individual log volume estimation

Using intersect diameter to estimate log volume, instead of the default Smalian's formula, reduced the estimated total CWD volume on Oregon's forest lands by 1,378 million cubic feet (3.6%) and biomass by 11.2 million tons (4.2%) (Table 1). The practical significance of the difference between the two methods, rather than its statistical significance, is most relevant in this case. Nevertheless, the estimated differences were 10 times greater than their respective standard errors. Therefore, the difference between the parameters estimated by those two methods is highly statistically significant.

Because the estimator based in the intersection diameter is unbiased, a more relevant comparison may be between the bias, estimated by the difference between the two methods, and the sampling error. Estimated bias was 1.7 and 2.0 times the standard error of Smalian's volume and biomass estimates, respectively. Therefore, the bias, not the sampling error (SE), dominates the overall error.

Table 1: Estimated total CWD volume and biomass in decay class 1-4 down logs on Oregon forest land, by individual log volume equation.

Individual log volume equation	Volume		Biomass	
	Total	SE	Total	SE
	-- million cubic feet --		-- million tons --	
Smalian's formula	38,170	824	267.4	5.7
Intersection diameter	36,795	798	256.2	5.4
Difference	-1,378	138	-11.2	1.0

The estimate of log volume using Smalian's formula may be less variable than that from using the diameter at the intersection. The results from this analysis indicate that, at the scale of this study, any effect on the SE of the estimate is negligible. The SE of the intersection diameter method was 798 million ft^3, while that of Smalian's formula was 824 million ft^3. However, the estimated volume and biomass were also greater when using Smalian's formula. Relative to the estimates of the total, the SE was almost identical between both methods: 2.160% (Smalian) and 2.170% (intersection) for volume, and 2.118% (Smalian) and 2.126% (intersection) for biomass.

Log inclination

Accounting for the effect of plot slope increased the estimated CWD volume between 1.72 and 3.38 billion ft^3, or 4.1 to 8.1 %, compared with the current FIA default scenario (Table 2). The magnitude of the difference for scenarios B, C and D was 2.0, 2.7 and 3.9 times the SE of the default scenario, respectively.

Table 2: Estimated total CWD volume in down logs on Oregon forest land, by log inclination scenario. Volume was calculated from the intersection diameter, all decay classes included.

	Volume (billion ft^3)		Difference with scenario A	
Scenario	**Total**	**SE**	**Total**	**%**
	-- billion cubic feet --		-- billion cubic feet --	
A: default	41.92	0.86		
B: uniform	43.64	0.90	1.72	4.1
C: predominantly downslope	44.21	0.92	2.29	5.5
D: downslope	45.30	0.94	3.38	8.1

Discussion

The assumptions evaluated in this study, regarding individual log volume estimation and piece inclination, had a very significant effect in the estimation of total CWD volume and biomass in Oregon. Standard sampling and estimation procedures implicitly assume that the measurement error and bias are negligible compared with the sampling error. In this study, however, the magnitude of the estimated bias was several times that of the estimated sampling error, a result that has important consequences for estimation. For example, confidence intervals constructed with biased totals and their standard errors would be unreliable, because the true coverage is much less than the nominal coverage.

The trade-off between sampling error and measurement bias depends on the sample size. While sampling error decreases as sample size increases, the bias does not. For regional studies that include a large number of plots, such as this study, the sampling error is very small and bias dominates. Nevertheless, the estimated magnitude of the biases (over 4% of the estimated value in most cases) seems significant enough to be of concern even for questions involving much smaller areas and sample sizes. Reducing the bias may involve changes in measurement techniques and protocols, which in some cases may result in higher costs. However, there is a balance between allocating resources to increase the sampling size or reduce the measurement error, and both aspects should be considered.

Regarding the method used to estimate individual piece volume, both theoretical and practical results support the use of diameter at the intersection point. Neither Smalian's nor the intersection point method requires the measurement of the length of individual pieces, because the length (l_{ij}) factors out of eqs. 2 and 3. However, the intersection diameter method only requires the measurement of a single diameter and avoids having to leave the transect path to measure the diameters at both ends of the piece. This method is unbiased and, at the scale of this study, the impact on the variance is negligible. Additional assessments using individual plots did not seem to indicate increased variance when using intersection diameter, compared with Smalian's formula. The diameter at the intersection method attempts to estimate the average cross section of all the pieces. Thus, the rate of convergence is driven by the number of

intersections between transects and logs, not by the number of plots. Therefore, as long as a moderate number of intersections are included, the unbiased intersection diameter method is likely to perform as well as or better than Smalian's formula.

Accounting for log inclination may require additional field measurements. However, ignoring this variable would have a very significant impact on the estimator, at least in mountainous areas such as Oregon. The assumption that the orientation of the logs is uniform with respect to the slope (scenario B) seems to be a realistic lower bound for the bias caused by ignoring log inclination. Even then, the estimated bias was 4.1% of the total volume and 2 times the estimated SE. The actual bias due to ignoring log inclination is likely to be greater than that reported in this study, because the scenarios ignored the effect of leaning snags or supported logs. Supported snags are likely to be more frequent after disturbances such as hurricanes, so ignoring log inclination may substantially underestimate the disturbance impact. This study suggests that measurements directed to reducing this bias may be a reasonable allocation of resources, even at the cost of reduced sample size or transect length.

Biases and differences between methods have been compared with the estimated totals and their standard errors. To put those figures in perspective, the estimated differences can be compared with statewide carbon emissions in the State (Table 3). Depending on the assumption considered, the difference in the estimated carbon pools in CWD ranged between 32 and 65% the total anthropogenic carbon emissions in the state in 2000, highlighting the importance of accurate measurements of pools such as CWD at the regional level.

Table 3: Comparison between current carbon estimates in CWD and 2000 Oregon emissions.

	CO_2 equivalents	**Percent 2000 OR CO_2 emissions[a]**
	- million metric tons -	--- percent ---
Current estimate	505.8	
Intersect diameter	-19.3	32.4
Inclination:		
B: uniform	+19.8	33.2
C: pred. down	+26.3	44.1
D: downslope	+38.8	65.0

[a]Emissions estimated at 59.54 million metric tons of CO_2 equivalents. Source: Oregon Department of Energy, http://oregon.gov/ENERGY/GBLWRM/Oregon_Gross_GhG_Inventory_1990-2005.htm, accessed on October 10, 2008.

Acknowledgements

Many people from the FIA program have been involved in the design of the inventory and data collection and compilation. Their contribution was invaluable. The author would like to thank P. Adams, D. Azuma, C. Scott and C. Woodall for their thoughtful reviews and to J. Fried for coordinating the review process.

References

Donnegan, J.; Campbell, S.; Azuma, D., tech. eds. 2008. Oregon's forest resources, 2001-2005: five-year Forest Inventory and Analysis report. Gen. Tech. Rep. PNW-GTR-765. Portland, OR: U.S. Forest Service, Pacific Northwest Research Station. 186 p.

Harmon, M.E.; Franklin, J.F.; Swanson, F.J.; Sollins, P.; Gregory, S.V.; Lattin, J.D.; Anderson, N.H.; Cline, S.P.; Aumen, N.G.; Sedell, J.R.; Lienkaemper, G.W.; Cromack, K.; Cummins, K.W. 1986. Ecology of coarse woody debris in temperate ecosystems. Advances in Ecological Research 15: 133-302.

Harmon, M.E.; Woodall, C.W.; Fasth, B.; Sexton, J. 2008. Woody detritus density and density reduction factors for tree species in the United States: a synthesis. Gen. Tech. Rep. NRS-29. Newton Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 84 p.

Husch, B.; Miller, C.I.; Beers, T.W. 1972. Forest mensuration, 2nd ed. New York: Ronald. 410 p.

Scott, C.T.; Bechtold, W.A.; Reams, G.A.; Smith, W.D.; Westfall, J.A.; Hansen, M.H.; Moisen, G.G. 2005. Sample-based estimators used by the Forest Inventory and Analysis National Information Management System. In: Bechtold, W.A.; Patterson, P.L., Eds. The enhanced Forest Inventory and Analysis program – national sampling design and estimation procedures. Gen. Tech. Rep. SRS-8-. Asheville, N.C.: US Department of Agriculture, Forest Service, Southern Research Station. pp. 43-67.

US Department of Agriculture, Forest Service (2007). Field instructions for the annual inventory of Washington, Oregon and California, 2007. http://www.fs.fed.us/pnw/fia/local-resources/pdf/field_manuals/2007_field_manual.pdf, Accessed Sept 1, 2008.

Van Wagner, C.E. The line-intersect method in forest fuel sampling. Forest Science 14: 20-26.

Waddell, K.L. 2002. Sampling coarse woody debris for multiple attributes in extensive forest resource inventories. Ecological Indicators 1: 139-153.

Woodall, C.W.; Monleon, V.J. 2008. Sampling protocol, estimation, and analysis procedures for the down woody materials indicator of the FIA Program. Gen. Tech. Rep. NRS-22. Newton Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 68 p.

Field Results for Line Intersect Distance Sampling of Coarse Woody Debris

David L.R. Affleck[1]

***Abstract**: A growing recognition of the importance of downed woody materials in forest ecosystem processes and global carbon budgets has sharpened the need for efficient sampling strategies that target this resource. Often the aggregate volume, biomass, or carbon content of the downed wood is of primary interest, making recently developed probability proportional-to-volume designs such as line intersect distance sampling (LIDS) of interest. This research presents field results from LIDS and conventional line intersect sampling (LIS) inventories of coarse woody debris (CWD) in three stands in northwestern Montana, USA. Field crews quickly adapted to the distance-limited selection protocol of LIDS, though its variable-length transect design increased data collection times relative to LIS in all three stands. Yet aggregate CWD volumes were estimated more precisely with LIDS, more than offsetting the time differential and rendering LIDS more efficient than LIS for volume estimation given a fixed sampling time or budget. Conversely, LIS generally performed more efficiently than LIDS with regards to the estimation of CWD abundance and aggregate length. Its transect design means that LIDS could be readily integrated into Forest Inventory and Analysis based programs to assess and monitor CWD stocks. However, the utility of LIDS, or of any other sampling strategy, must be determined by the relative importances of the estimable CWD attributes of interest.*

Keywords: downed woody materials; size-biased sampling; Lubrecht Experimental Forest.

Introduction

Coarse woody debris (CWD) is an important component of many forest ecosystems. Partially decayed wood on or in the forest floor provides habitat for diverse plant and animal species; regulates soil moisture and nutrient flux; and can modify wildfire behavior and severity (Brown *et al.* 2003; Harmon *et al.* 1986; Jonsson and Kruys 2001). Downed wood also represents a carbon reservoir (*e.g.*, Canadian Council of Forest Ministers 2006) and a potential source of biomass energy. Recognition of these functions has prompted the identification of CWD as an indicator of forest health and sustainability. This, in turn, has sharpened the need for efficient sampling strategies that specifically target the downed wood resource.

Line intersect sampling (LIS; Kaiser 1983) is one of the most commonly used survey methods for downed wood, and is currently implemented in the USDA Forest Inventory and Analysis (FIA) program (see Bechtold and Patterson 2005). Applying LIS, logs or fragments of CWD that are crossed by a fixed-length transect

[1] Department of Forest Management, College of Forestry & Conservation, The University of Montana, 32 Campus Drive, Missoula, MT 59812; email: david.affleck@umontana.edu.

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

are selected into a sample. The search procedure of LIS is straightforward and the likelihood of detection errors small. However, in stand- and forest-level applications, LIS can require an inordinate sampling intensity to achieve estimates of acceptable accuracy, particularly when the aggregate volume or biomass of CWD is of interest (cf. Pickford and Hazard 1978; Woldendorp *et al.* 2004). Line intersect distance sampling (LIDS; Affleck 2008) is a novel CWD survey method that incorporates many elements of the search protocol of LIS, but also utilizes a probability proportional-to-volume selection design. As such, LIDS offers the potential for more rapid and accurate assessments of the aggregate volume of CWD resources, as well as its biomass and carbon content.

This report presents preliminary field research on the relative performance of LIDS and LIS based on downed wood sampling conducted in 2008 in northwestern Montana, USA. The next section provides an overview of the LIDS methodology and estimators and the following section describes the field sampling program. The closing section discusses the field results and comments on the perceived advantages and limitations of LIDS.

Line Intersect Distance Sampling

LIDS is a transect-based sampling method similar to LIS. However, like the perpendicular distance sampling strategy of Williams and Gove (2003), LIDS also incorporates a distance-limited selection criterion that allows for probability proportional-to-volume sampling of CWD.

Sampling is conducted from a point s located uniformly at random inside the tract of interest.[1] A line transect initiated at s is then oriented in the direction θ, also selected uniformly at random. These two design elements supply the basis for inference and obviate the need for any conditions concerning the spatial distribution or orientation of CWD particles on the tract of interest. In practice, the point s will be one of a collection of sample points $\{s_1, s_2, \ldots, s_T\}$ located independently or on a randomly positioned systematic grid. Similarly, radial transects (see Affleck *et al.* 2005) consisting of $M > 1$ segments can be employed in LIDS, in which case the first segment is oriented in the direction $\theta_1 = \theta$ and the orientations of the other segments follow at prescribed intervals (*e.g.*, in a Y-shaped transect $\theta_2 = \theta + 120^\circ$ and $\theta_3 = \theta + 240^\circ$). Where the LIDS strategy departs from LIS is in the fact that the transect as a whole, as well as any individual transect segments, have no fixed length. Instead, a critical distance constant k (m^{-1}) determines the sampling intensity; the role of this constant is analogous to that of Grosenbaugh's (1958) point-sampling constant k.

A particle of CWD crossed by a LIDS transect segment is selected into the sample provided that two conditions are met. First, the central axis of the particle, not simply a branch or lobe, must be intersected by the transect segment. The central axis of a particle is the longest line segment connecting two points on the particle's boundary. Denote the length of the i^{th} particle's central axis by l_i (m). The

[1] See Appendix A for a summary of the notation.

central axis of a freshly fallen log can often be taken as the pith of the main stem and its length will then coincide with the log's length. An odd-shaped particle of CWD can have a central axis that is only partially within the core of wood, however (see *e.g.*, Affleck 2008).

The second condition for particle selection stipulates that the distance from the sample point to the particle is less than a critical distance determined, in part, by k. Specifically, if the m^{th} transect segment crosses the central axis of the i^{th} particle at the point $x_{m,i}$, then the distance (in m) from s to $x_{m,i}$ must not exceed the critical distance ($D_{\text{crit},i}$):

$$D_{\text{crit},i} = \frac{k\,a(x_{m,i})}{M}\,,$$

where $a(x_{m,i})$ is the cross-sectional area (m^2) of particle i contained within the intersection plane. This intersection plane is the vertical plane lying perpendicular to the particle's central axis and containing the point $x_{m,i}$. For the idealized needle-shaped particle lying on flat ground, this cross-sectional area will be a circular face of wood; if this particle is tilted from the horizontal, then $a(x_{m,i})$ measures an elliptical surface. More generally, $a(x_{m,i})$ measures the area of an approximately elliptical section of particle i or the sum of the areas of several such sections if multiple branches of particle i are cut by the intersection plane.

It follows from these conditions that the probability of selecting the i^{th} particle on a given segment of a LIDS transect is directly proportional to its cubic volume (Affleck 2008). This result, in turn, implies that a design-unbiased estimator of the volume of CWD per unit area is given by

$$\widehat{V}_s = F\,n_s\,, \tag{1}$$

where n_s is the total number of particle selections made on the transect located at s and $F = \frac{10000\pi}{2k}$ is a constant volume factor ($\text{m}^3\cdot\text{ha}^{-1}$). Estimating aggregate CWD volume per unit area following LIDS is thus similar to estimating aggregate tree basal area per unit tract area following horizontal point sampling: a count of the number of selected particles (or trees) is simply blown-up by a volume (or basal area) factor.

Other CWD population parameters can be estimated if measurements such as l_i and $a(x_{m,i})$ are taken on the selected particles (see Affleck 2008). For example, a design-unbiased estimator of the number of particles per unit area based on a single transect located at s is

$$\widehat{N}_s = F \sum_{m=1}^{M} \sum_{i \in \mathscr{C}_{s,m}} \frac{1}{l_i \cos(\phi_i)\, a(x_{m,i})}\,, \tag{2}$$

where ϕ_i is the tilt ($^\circ$) of the central axis of particle i with respect to the horizontal plane and the second summation extends over the collection $\mathscr{C}_{s,m}$ of particles selected by the m^{th} segment of the transect located at s. A corresponding design-unbiased estimator of the aggregate length of the central axes of CWD particles per

unit tract area is

$$\widehat{L}_s = F \sum_{m=1}^{M} \sum_{i \in \mathscr{C}_{s,m}} \frac{1}{\cos(\phi_i)\, a(x_{m,i})} \,. \tag{3}$$

For comparison, design-unbiased estimators of the aggregate volume, abundance, and length of CWD following LIS are given in Appendix B.

With multiple LIDS transects located on the tract, transect-specific estimates can be combined and standard errors estimated. For example, transect-based estimates of aggregate CWD volume per hectare can be averaged together:

$$\widehat{V} = \frac{1}{T} \sum_{t=1}^{T} \widehat{V}_{s_t} \,,$$

where $\widehat{V}_{s_t}$ is the estimate obtained from the transect located at sample point s_t ($t = 1, 2, \ldots, T$). Also, if sample points are located independently, then normal-theory confidence intervals can be constructed from the estimated standard error

$$s(\widehat{V}) = \sqrt{\frac{\sum_{t=1}^{T} \left(\widehat{V}_{s_t} - \widehat{V}\right)^2}{T(T-1)}} \,,$$

or from a bootstrapping procedure (Efron and Tibshirani 1994).

Field Data Collection

To assess quantitative and qualitative differences in the practical application of LIS and LIDS, a collection of stands in the Lubrecht Experimental Forest in northwestern Montana were identified for field survey. Stands were selected to cover a range of (overstory) forest types and a variety of downed wood assemblages. All were second-growth, managed stands and ranged in age from approximately 40 years to over 150 years. Field results from three of these stands (Table 1), all sampled in the summer of 2008, are discussed below.

Table 1: Characteristics of selected stands, sampling intensities, LIS transect lengths (L), and LIDS volume factors (F).

Stand	Leading species	Basal area ($m^2 \cdot ha^{-1}$)	Sample points	L (m)	F ($m^3 \cdot ha^{-1}$)
A	*Pseudotsuga menziesii v. glauca*	30.1	60	48	10
B	*Pinus contorta*	22.7	28	9	15
C	*Larix occidentalis*	44.8	30	42	8

Sample points were distributed uniformly and independently at random within each stand. Each sample point served to locate a Y-shaped LIS transect of total length L (m) as well as a Y-shaped LIDS unit with volume factor F. The two sampling units were superimposed and their orientations were selected uniformly

and independently at random. Each segment of a transect was initiated at a 2 m distance from the sampling point. Pilot sampling was conducted in each stand to determine the LIS transect lengths and LIDS volume factors that would result in the selection of 6-9 particles per sample point. As a result, distinct transect lengths and volume factors were used within each stand (Table 1).

Following the definition of Valentine *et al.* (2008), a log or segment of downed wood was considered a particle of CWD if it exceeded 7.5 cm in diameter at any point along its central axis. That is, a connected fragment of downed wood was deemed either a single particle of CWD or an element of fine woody debris – logs were not partitioned into multiple coarse and fine woody segments. Dead wood within live trees, woody material suspended more than 2 m above the forest floor, and downed wood buried beneath the litter layer in the forest floor were not considered. Partially rooted snags or stumps were considered CWD particles only if they leaned more than 45° from the vertical.

The length of the central axis and its tilt from the horizontal were measured on every selected particle. Cross-sectional diameters were obtained for every lobe (*e.g.*, branch or stem fork) within the intersection plane. Generally, two diameters were measured on each lobe, with one diameter being made in the horizontal plane and the other taken at right angles (but still in the vertical intersection plane). The cross-sectional area of the selected particle was then obtained from the sum, over all lobes, of the products of paired diameters. Additionally, the species of the particle was determined, at least to the level of softwood, hardwood, or unknown, and decay class was assessed on the five point scale of Meidinger (1998).

Both LIS and LIDS were implemented at each sampling location, with the order of implementation determined randomly and in advance of sampling. Only the first method applied at a given point was timed. Two person crews were used throughout and separate data collection times were obtained for each of the three segments of the Y-shaped units. The elapsed times included all elements of segment orientation and layout; particle selection and measurement; and travel back to the sample point. All distances were measured using tapes and were corrected for slope. Particle species and decay information was collected primarily to render the collection times consistent with those of standard CWD sampling programs.

The relative efficiency of LIS and LIDS were determined from the average implementation times and from the observed precision of stand parameter estimators. Relative efficiency was calculated as

$$\mathsf{RE} = \frac{\bar{\mathsf{t}}_{\mathsf{LIDS}} \times s^2(\widehat{Y}_{\mathsf{LIDS}})}{\bar{\mathsf{t}}_{\mathsf{LIS}} \times s^2(\widehat{Y}_{\mathsf{LIS}})},$$

where $\bar{\mathsf{t}}_{\mathsf{method}}$ is the average transect sampling time for a particular method and $s(\widehat{Y}_{\mathsf{method}})$ is the corresponding standard error of an attribute estimator. This measure of efficiency is both attribute- and stand-specific. RE can be interpreted as the time needed to achieve a given margin of error using LIDS relative to the time needed to achieve the same margin of error with LIS.

Results & Discussion

Across all three stands, LIS was implemented more rapidly than LIDS (Fig. 1) even though more particles tended to be selected and measured when LIS was applied (Fig. 2). Therefore, the time differential between the two methods can likely be attributed to the variable-length transect design of LIDS. Whereas only a fixed length of line needed to be followed for LIS, in LIDS the crew had to traverse a minimum of 30-50 m along each segment with the total distance traveled depending largely on sighting conditions (*i.e.*, on density of vegetation and variability in micro-topography). Due to the distance-dependent selection criterion of LIDS, determining whether particles were "in" or "out" of the sample was more involved than for LIS. However, with the aid of a simple look-up table (*e.g.*, Table 6 in Appendix B) the crew quickly became familiar with the size-distance relationships implied by a given volume factor. Distances and intersection diameters seldom needed to be measured to determine selection: sufficiently large particles near the sample point and overly small particles far from the sample point were easily identified. Moreover, particles encountered at distances approaching their critical distances were often marked as 'borderline' and evaluated for selection subsequent to the field work. In general, the length (l_i) measurements were the most time-consuming, save on particles having multiple lobes or branches crossing the intersection plane.

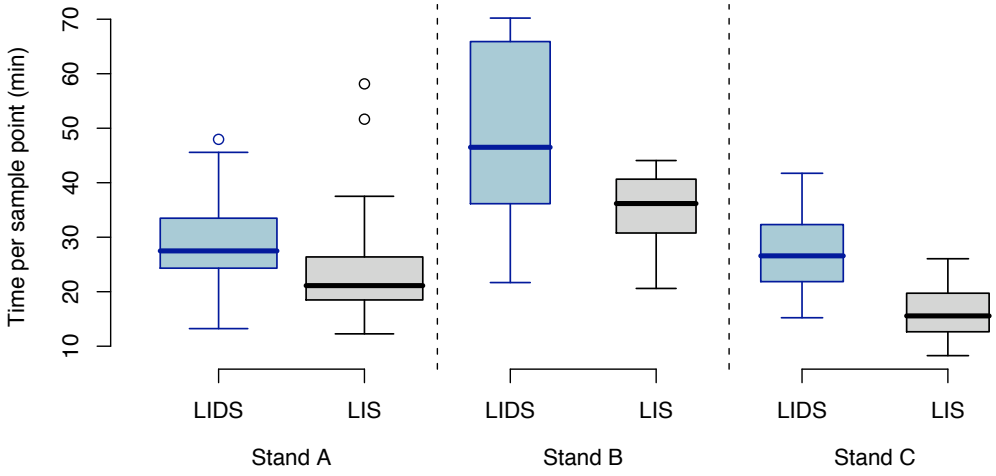


Figure 1: Distributions of transect data collection times by sampling method and stand.

With regards to CWD volume, the two methods produced similar estimates for stands A and C (Table 2). LIDS and LIS estimates differed by approximately 32% in stand B although their confidence intervals from 1000 bootstrap replications overlapped. Notably, in all three stands the time differential favoring LIS (Fig. 1) was more than offset by increased precision on the part of LIDS (Table 2). That is, LIDS permitted the collection of more information on aggregate volume per sample point and per unit time, resulting in favorable (≤ 1) relative efficiencies. In each stand, the difference in precision between the two methods was more pronounced than the differences in estimated tract volumes, markedly so in the case of stand B.

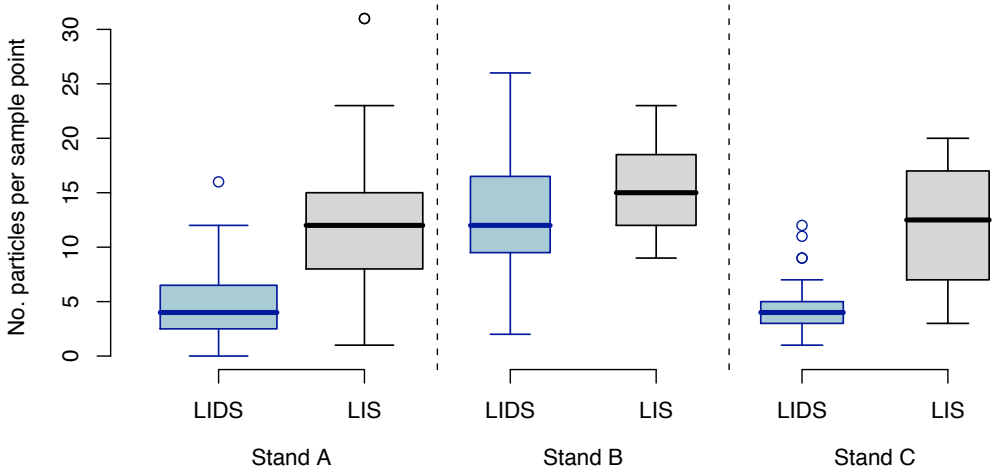


Figure 2: Distributions of transect particle counts by sampling method and stand.

Table 2: CWD volume estimates and corresponding standard errors with lower bounds (LB) and upper bounds (UB) of 90% confidence intervals from 1000 bootstrap replications.

		Volume ($m^3 \cdot ha^{-1}$)			Precision ($m^3 \cdot ha^{-1}$)			
Stand	**Strategy**	**LB**	$\widehat{V}$	**UB**	**LB**	$s(\widehat{V})$	**UB**	**RE**[a]
A	LIDS	41.33	47.50	55.00	3.78	4.38	5.25	0.74
	LIS	37.34	45.41	55.50	4.25	5.62	7.60	
B	LIDS	166.61	189.11	216.43	12.39	15.01	19.72	0.09
	LIS	199.64	260.87	426.90	17.99	57.48	104.94	
C	LIDS	30.67	37.33	43.92	3.16	4.08	5.18	0.66
	LIS	32.22	40.84	54.73	4.91	6.56	8.33	

[a] Relative efficiency of LIDS *vs.* LIS for volume estimation; values below 1 indicate that LIDS is more efficient.

Note also that in one sense the relative efficiencies reported in Table 2 understate the performance of LIDS: if aggregate CWD volume is of central interest, then length measurements and in many cases cross-sectional diameter measurements need not be taken, offering the potential for considerable time savings with LIDS.

Within-stand differences between LIDS and LIS estimates of the abundance or aggregate length of CWD were larger than those for volume (Tables 3 and 4). Excepting stand B, LIS offered improved precision in abundance estimates and thus was more efficient than LIDS. This result is consistent with previously reported simulation research (Affleck 2008) and presumably results from lower variation in particle lengths relative to variation in particle volumes. The probability proportional-to-length LIS design also explains why LIS estimated the aggregate length of CWD more precisely (Table 4) in all three stands. In addition, although aggregate CWD length is rarely of central interest, the relative efficiency estimates of Table 4 understate the performance of LIS in the sense that measurements of length are not required to estimate this attribute using LIS (cf. equation 6 in Appendix B).

The results for stand B are notable for the marked relative efficiency of LIDS in the estimation of aggregate volume (Table 2) as well as the relative inefficiency of

Table 3: CWD abundance estimates and corresponding standard errors with lower bounds (LB) and upper bounds (UB) of 90% confidence intervals from 1000 bootstrap replications.

		Abundance (#·ha^{-1})			Precision (#·ha^{-1})			
Stand	**Strategy**	**LB**	$\widehat{N}$	**UB**	**LB**	$s(\widehat{N})$	**UB**	**RE**[a]
A	LIDS	1319.7	1807.4	2645.3	266.8	366.7	520.6	14.45
	LIS	1210.2	1365.1	1553.6	90.2	106.3	129.1	
B	LIDS	3014.3	3496.5	4252.0	264.0	353.4	514.2	0.88
	LIS	3392.3	3914.4	5092.0	249.4	442.1	745.7	
C	LIDS	997.5	1455.1	2260.6	230.1	393.3	628.9	6.92
	LIS	1074.0	1354.4	1708.3	145.2	195.3	246.3	

[a] Relative efficiency of LIDS *vs.* LIS for abundance estimation; values below 1 indicate that LIDS is more efficient.

Table 4: CWD length estimates and corresponding standard errors with lower bounds (LB) and upper bounds (UB) of 90% confidence intervals from 1000 bootstrap replications.

		Length (m·ha^{-1})			Precision (m·ha^{-1})			
Stand	**Strategy**	**LB**	$\widehat{L}$	**UB**	**LB**	$s(\widehat{L})$	**UB**	**RE**[a]
A	LIDS	4176.3	5107.3	6698.3	536.1	729.3	1003.0	10.95
	LIS	3665.2	4025.2	4467.0	198.9	242.8	307.3	
B	LIDS	21231.6	23932.3	26859.1	1369.7	1704.7	2272.6	2.10
	LIS	24819.3	27052.6	29301.5	1200.5	1383.5	1658.4	
C	LIDS	4501.0	7038.1	11328.6	1019.3	1931.7	3150.1	41.88
	LIS	3690.1	4325.9	4908.5	342.4	389.9	454.8	

[a] Relative efficiency of LIDS *vs.* LIS for length estimation; values below 1 indicate that LIDS is more efficient.

LIS in abundance estimation (Table 3). This even-aged pine stand of approximately 90 years had recently undergone intense self-thinning, generating a considerable quantity of CWD in the process (Fig. 3). However, the downed wood exhibited little variation in decay class, shape, or size. It therefore was anticipated that the statistical advantages of LIDS with regards to volume estimation would be less pronounced in this stand: a probability proportional-to-volume design is less effective, in principle, if the particles' volumes are relatively uniform. Of course, part of the apparent efficiency of LIDS derives from the poor performance of LIS. The latter is presumably the result of using such a short transect ($L = 9$ m). However, there were few areas in this stand with exposed ground and any increase in the transect length would have greatly increased the average sampling times and particle counts.

In sum, these preliminary field results support previous theoretical and simulation research (Affleck 2008) indicating the potential efficiency of LIDS for CWD volume estimation. Moreover, it is clear that LIDS is a practical alternative to LIS in the field: crews can successfully apply the LIDS selection protocol with the aid of simple lookup tables (see Appendix C). Therefore, where the aggregate volume of CWD is of central interest, LIDS appears to be a viable and accurate sampling strategy. Since particle mass and carbon content tend to scale in proportion to particle volume, it also follows that LIDS could offer real advantages over LIS in estimating the aggregate biomass or carbon content of downed wood (cf. Valentine



Figure 3: Photograph taken in stand B; the downed logs in this even-aged *Pinus contorta* stand showed less variation in size relative to the other stands surveyed.

et al. 2008). Ultimately, however, the utility of LIDS or any other strategy, LIS included, will be dictated by the set of descriptive population parameters of interest and the relative importance attached to each. If the aggregate biomass, length, and abundance of CWD ($\#\cdot ha^{-1}$) are to be estimated with comparable precision, then a sampling strategy that provides reasonably accurate estimates of all three parameters must be sought. Strategies such as LIDS, LIS, and fixed-area plot sampling that tend to be more nearly optimal for one of these parameters may be relatively inefficient overall.

Acknowledgments

This research was supported in part by a USDA-CSREES McEntire-Stennis Cooperative Forestry grant. The author is indebted to Ryan Schlect, Motoshi Honda, and Alissa Corrow for invaluable assistance in the field.

Literature Cited

Affleck, D.L.R. (2008) A line intersect distance sampling strategy for downed wood inventory. *Canadian Journal of Forest Research*, **38**, 2262–2273.

Affleck, D.L.R., T.G. Gregoire, and H.T. Valentine (2005) Design unbiased estimation in line intersect sampling using segmented transects. *Environmental and Ecological Statistics*, **12**, 139–154.

Bechtold, W.A. and P.L. Patterson (Eds.) (2005) The Enhanced Forest Inventory and Analysis Program – National Sampling Design and Estimation Procedures, General Technical Report SRS-GTR-80, USDA Forest Service, Southern Research Station.

Brown, J.K., E.D. Reinhardt, and K.A. Kramer (2003) Coarse woody debris: Managing benefits and fire hazard in the recovering forest. General Technical Report RMRS-GTR-105, USDA Forest Service, Rocky Mountain Research Station.

Canadian Council of Forest Ministers (2006) Criteria and indicators of sustainable forest management in Canada: National status 2005. Natural Resources Canada, Canadian Forest Service, Ottawa.

Efron, B. and R.J. Tibshirani (1994) *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Grosenbaugh, L.R. (1958) Point-sampling and line-sampling: Probability theory, geometric implications, synthesis. Occasional Paper 160, USDA Forest Service, Southern Forest Experiment Station.

Harmon, M.E., J.F. Franklin, F.J. Swanson, P. Sollins, S.V. Gregory, J.D. Lattin, N.H. Anderson, S.P. Cline, N.G. Aumen, J.R. Sedell, G.W. Lienkaemper, J. K Cromack, and K.W. Cummins (1986) Ecology of coarse woody debris in temperate ecosystems. *Advances in Ecological Research*, **15**, 133–302.

Jonsson, B.G. and N. Kruys (Eds.) (2001) *Ecology of woody debris in boreal forests*, Ecological Bulletins, **49**.

Kaiser, L. (1983) Unbiased estimation in line-intercept sampling. *Biometrics*, **39**, 965–976.

Meidinger, D. (1998) Coarse woody debris. In *Field Manual for Describing Terrestrial Ecosystems*, Land Management Handbook No. 25. B.C. Ministry of Environment, Lands, and Parks and B.C. Ministry of Forests, Victoria, B.C.

Pickford, S.G. and J.W. Hazard (1978) Simulation studies on line intersect sampling of forest residue. *Forest Science*, **24**, 469–483.

Valentine, H.T., J.H. Gove, M.J. Ducey, T.G. Gregoire, and M.S. Williams (2008) Estimating the carbon in coarse woody debris with perpendicular distance sampling. In *Field Measurements for Forest Carbon Monitoring: A Landscape-Scale Approach*, C.M. Hoover, (Ed.), chap. 6. Springer-Verlag, New York.

Williams, M.S. and J.H. Gove (2003) Perpendicular distance sampling: an alternative method for sampling downed coarse woody debris. *Canadian Journal of Forest Research*, **33**, 1564–1579.

Woldendorp, G., R.J. Keenan, S. Barry, and R.D. Spencer (2004) Analysis of sampling methods for coarse woody debris. *Forest Ecology and Management*, **198**, 133–148.

Appendix A: Symbols Used

Table 5: Symbols and their units.

Symbol	Units	Definition
s_t	(m, m)	coordinate location of sample point t ($t = 1,2,\ldots,T$)
θ_m	$^\circ$	orientation of transect segment m ($m = 1,2,\ldots,M$)
T	–	number of sample points (*i.e.*, number of LIDS or LIS transects)
M	–	number of segments on a radial transect
k	m^{-1}	LIDS critical distance constant
F	$\mathrm{m}^3{\cdot}\mathrm{ha}^{-1}$	LIDS volume factor ($F = \frac{10000\pi}{2k}$)
L	m	total length of a LIS transect
n_{s_t}	–	number of CWD particles selected by a transect at s_t
$\mathscr{C}_{s_t,m}$	–	set of particles selected by the m^{th} segment of the transect at s_t
l_i	m	length of the central axis of particle i
ϕ_i	$^\circ$	tilt of the central axis of particle i from the horizontal
$x_{m,i}$	(m, m)	intersection point of the central axis of particle i and transect segment m
$a(x_{m,i})$	m^2	cross-sectional area of particle i in the vertical plane containing $x_{m,i}$ and oriented perpendicular to the central axis of particle i
$\widehat{V}$	$\mathrm{m}^3{\cdot}\mathrm{ha}^{-1}$	estimated aggregate CWD volume per unit tract area
$\widehat{N}$	ha^{-1}	estimated CWD abundance per unit tract area
$\widehat{L}$	$\mathrm{m}{\cdot}\mathrm{ha}^{-1}$	estimated total length of CWD central axes per unit tract area
$s(\widehat{V})$	$\mathrm{m}^3{\cdot}\mathrm{ha}^{-1}$	estimated standard error of $\widehat{V}$
$s(\widehat{N})$	ha^{-1}	estimated standard error of $\widehat{N}$
$s(\widehat{L})$	$\mathrm{m}{\cdot}\mathrm{ha}^{-1}$	estimated standard error of $\widehat{L}$
$\bar{\mathrm{t}}$	min	average sampling time per transect

Appendix B: Line Intersect Sampling Estimators

Corresponding to the LIDS estimators in equations 1-3 are design-unbiased LIS estimators of, respectively, aggregate volume ($\mathrm{m}^3{\cdot}\mathrm{ha}^{-1}$), abundance ($\#{\cdot}\mathrm{ha}^{-1}$), and length ($\mathrm{m}{\cdot}\mathrm{ha}^{-1}$):

$$\widehat{V}_s = F_{\mathrm{LIS}} \sum_{m=1}^{M} \sum_{i \in \mathscr{C}_{s,m}} a(x_{m,i}) \tag{4}$$

$$\widehat{N}_s = F_{\mathrm{LIS}} \sum_{m=1}^{M} \sum_{i \in \mathscr{C}_{s,m}} \frac{1}{l_i \cos(\phi_i)} \tag{5}$$

$$\widehat{L}_s = F_{\mathrm{LIS}}\, n_s\,, \tag{6}$$

where $F_{\mathrm{LIS}} = \frac{10000\pi}{2L}$ is the LIS length factor ($\mathrm{m}{\cdot}\mathrm{ha}^{-1}$) that is determined by the total transect length L.

Appendix C: LIDS Reference Table

Table 6: Critical distances for fixed minimum round diameter (MRD), and MRDs for fixed critical distances, across varying volume factors (F in $m^3 \cdot ha^{-1}$) for a three-segment LIDS transect.

MRD	D_{crit} **(m)**[a]			D_{crit}	**Minimum round diameter (cm)**[b]		
(cm)	$F = 8$	$F = 10$	$F = 15$	**(m)**	$F = 8$	$F = 10$	$F = 15$
1	0.05	0.04	0.03	1	4.4	4.9	6.0
2	0.21	0.16	0.11	2	6.2	7.0	8.5
3	0.46	0.37	0.25	3	7.6	8.5	10.5
4	0.82	0.66	0.44	4	8.8	9.9	12.1
5	1.29	1.03	0.69	5	9.9	11.0	13.5
6	1.85	1.48	0.99	6	10.8	12.1	14.8
7	2.52	2.02	1.34	7	11.7	13.0	16.0
8	3.29	2.63	1.75	8	12.5	13.9	17.1
10	5.14	4.11	2.74	10	13.9	15.6	19.1
12	7.40	5.92	3.95	12	15.3	17.1	20.9
14	10.08	8.06	5.37	14	16.5	18.5	22.6
16	13.16	10.53	7.02	16	17.6	19.7	24.2
20	20.56	16.45	10.97	20	19.7	22.1	27.0
25	32.13	25.70	17.13	25	22.1	24.7	30.2
30	46.26	37.01	24.67	30	24.2	27.0	33.1

[a] $D_{crit} = \frac{\pi^2 MRD^2}{8MF}$

[b] $MRD = \sqrt{\frac{8MF}{\pi^2 D_{crit}}}$

A comparison of forest height prediction from FIA field measurement and LiDAR data via spatial models

Yuzhen Li[1]

ABSTRACT: *Previous studies have shown a high correspondence between tree height measurements acquired from airborne LiDAR and that those measured using conventional field techniques. Though these results are very promising, most of the studies were conducted over small experimental areas and tree height was measured carefully or using expensive instruments in the field, which is not feasible in a practical forest inventory context. In this study, 105 plots located west of the Kenai Mountains, Kenai Peninsula, Alaska were measured and LiDAR data over the same set of field plots were acquired. Plot tree height, stand height, LiDAR mean height and LiDAR 90th percentile height were computed. Using the Matern covariance model for constant mean Gaussian spatial process, ordinary kriging was implemented and contour maps of predicted plot-level height from field height measurements and from LiDAR data were produced over the entire region along with maps of estimated standard error. Results indicate that at 300m by 300m pixel resolution, the spatial trends of predicted plot-level height are similar between field measurements and LiDAR measurements. The distribution of predicted stand height is very similar to the distribution of predicted LiDAR mean height with mean difference of only 0.28m. The mean of predicted plot tree height is comparable to the mean of predicted LiDAR 90th percentile height, but the distribution of predicted LiDAR 90th percentile height has much heavier tails.*

KEYWORDS: LiDAR, plot-level height, Gaussian process**,** Ordinary kriging

Introduction

Forest height is a crucial inventory attribute for calculating timber volume, forest biomass, site potential, and silvicultural treatment scheduling. Measuring height by current photogrammetric or field survey techniques is time consuming and expensive. As a new emerging remote sensing tool, airborne laser scanning system - Light Detection and Ranging (LiDAR) data have been studied to derive height information. Two different approaches have been used to obtain height measurements from LiDAR data. The first approach is to identify individual trees using a canopy height model and extract their height. The second approach is to regress plot-level or stand-level height on derived LiDAR metrics which describe vertical and horizontal distribution of forest canopy (Anderson et al 2006, Hyyppä et al 2000, Maltamo et al 2004, Næsset 2002, Persson et al 2002). Many studies have reported that accuracy of height estimate from LiDAR data is comparable

[1] College of Forest Resources, University of Washington, Seattle, WA, PO box 352100, USA; email: yzhli@u.washington.edu

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

with field height measurement, while others found LiDAR tends to underestimate individual tree height because of the low probability of a small-footprint laser pulse intercepting the apex of tree top (Anderson et al 2006, Gaveau and Hill 2003, Hyyppä et al 2000, Yu et al 2004). Though these results are promising, most of reported studies were conducted over small areas and field heights were measured carefully or using more expensive and accurate instruments than the hand-held rangefinder that is commonly used in forest inventory practice such as the Forest Inventory and Analysis (FIA) program in the USDA Forest Service. The accuracy of LiDAR-derived height when compared to field height measurement is not clearly understood in an operational forest inventory setting.

Another issue with large-area operational forest inventory is the accuracy of plot locations. Often less accurate, easy-to-carry GPS receivers are used to get a location for field plots. Thus may introduce inaccurate geographical co-registration of field plots with LiDAR data. For example, Rockwell Precision Lightweight GPS Receiver is used in the FIA program and its accuracy is 7 meters under forest canopy (Hoppus and Lister 2006). If field plots are poorly georeferenced, it is likely that the empirical regression relationship between field height and LiDAR metrics will be affected.

Forests are spatially structured and spatially close forests tend to be more similar due to the fact that they may experience similar living conditions, such as nutrient supply. Spatial models describing spatial correlations have been frequently used to determine forest biophysical parameters and characterize forest ecosystem structure (Biging and Dobbertin 1995, Lappi, J. 2001, Stoyan and Penttinen 2000, Stoyan and Stoyan 1998, Zawadzki et al 2005). In this study, instead of linking field plots with LiDAR data directly, a stationary spatial process was assumed for plot-level height, and then spatial models were applied to predict plot-level height at unobserved locations both from field inventory and LiDAR data respectively. The particular objective is to produce maps of predicted plot-level height over a large region, and then compare the distributions of heights predicted from operational field inventory and from LiDAR measurements.

Study area and data description

The study area is located in the west of the Kenai Mountains, Kenai Peninsula, Alaska (151.804W to 149.498W, 59.580N to 61.456N, Figure 1). The area covers approximately 5,000 square miles. A total of 105 forest inventory field plots are located in this study area. Each field plot consists of a cluster of four circular subplots approximately 1/24 acre in size with a radius of 24.0 ft, and each subplot contains a 6.8-foot fixed-radius microplot (Bechtold and Patterson 2005). Within each subplot, the height of trees with diameter at breast height of 5.0 inches or greater were measured; within each microplot, the height of saplings (1.0-4.9 inches DBH) and seedlings were measured. At each subplot center, a polygon type, which is a unique combination of land cover type, forest density, forest stand size and forest stand origin, was determined and the size of the polygons

was collected (field procedures for coastal Alaska inventory 2003). Two aggregated plot-level heights, plot tree height and stand height, were defined and calculated for the purpose of this study. Plot tree height is defined as the average height of individual trees with DBH equal or greater than 5 inches weighted by polygon area. Stand height is defined as the average height of trees with DBH equal or greater than 5 inches, seedlings, and sapling weighted by polygon area.

LiDAR data were collected over each field plot and its surrounding area. A total of 105 LiDAR virtual patches were produced with each covering an area of approximately 9 hectares (300m by 300m, Figure 1). Multiple flights were made for some plots, because the initial flight didn't cover the entire field plot. For each 300m by 300m LiDAR coverage, a digital terrain model (DTM) was generated using returns classified by the data provider as bare-earth points, then all LiDAR returns were spatially registered to the DTM using their coordinates. The relative height of each return was computed as the difference between its vertical Z coordinate and the terrain surface height. Returns with a relative height value less than 2 m were excluded to eliminate ground returns, rocks, stumps and low vegetation. The remaining points were considered to be laser canopy hits. Finally, the laser canopy hits within the boundary of a 144-foot fixed radius plot containing the four subplot plots were extracted, and LiDAR plot mean height and 90th percentile height were calculated. The large plot was used instead of four individual subplots to decrease the effect of inaccurate field plot locations that results from poor GPS positions or azimuth and distance errors when locating the individual subplots.

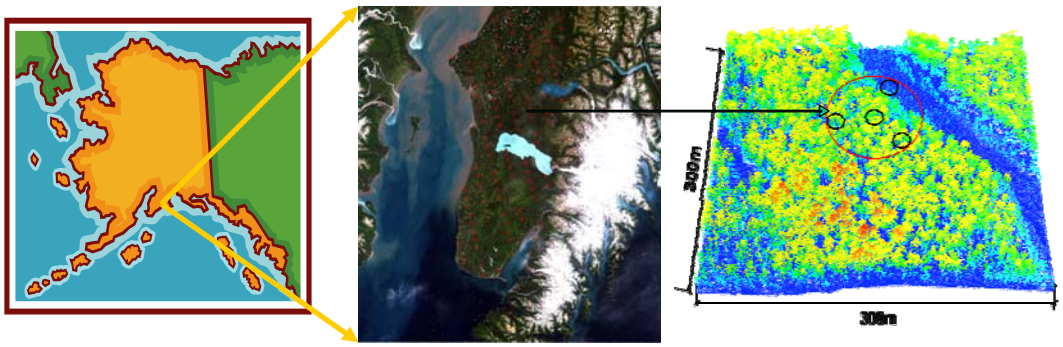


Figure 1: Map of study area. Picture in the middle is LANDSAT ETM+ image for the study area and red circles indicate field plot locations. Picture in the right is the LiDAR coverage over one example field plot and colored by height.

Method

Four aggregated plot-level heights (plot tree height and stand height from field measurements, LiDAR plot mean and LiDAR 90th percentile height) from 105 plots were assumed to be a partial realization of a stationary Gaussian process. That is $\{Z(s): s \in D \subset \Re^2\}$, $Z = (Z(s_1), ,\ldots Z(s_n))^T$ has a multivariate normal distribution, where Z(s) represents aggregated plot-level height at location s, D is

a fixed subset of 2-dimensional Euclidean space; $D \subset \Re^2$ contains spatial coordinates s={s_1,…,s_n} and s_i is the longitude and latitude coordinates at location i. n is the number of locations, 105 in our case. Stationary means that for any set of n sites {s_1,…,s_n} and any $h \in \Re^2$, the distribution of (Z(s_1),…, Z(s_n)) is the same as that of (Z(s_1+h),…, Z(s_n+h)), which implies that the joint distribution doesn't change when shifted in space. Further, an isotropic process was assumed, which means that the semivariogram function depends upon the separation vector h only through its length ||h||. For the sake of simplicity, the Gaussian process was assumed to have a constant mean, that is $Z(s) = \mu + \omega(s) + \varepsilon(s)$, where μ is the mean component of the model, and ω(s) is a zero-centered stationary Gaussian spatial process, which captures the residual spatial association, and the ε(s) is an uncorrelated pure error term. The ω(s) introduces the partial sill and range parameter and ε(s) adds the nugget effect (Banerjee et al 2004).

Empirical semivariograms of plot-level heights were first fitted by four theoretical parametric models: Gaussian, exponential, Matern and Spherical class. Model parameters were estimated by restricted maximum likelihood methods. For detailed differences between theoretical semivariogram models, please refer to Banerjee et al (2004). The theoretical models allow us to calculate semivariance values for any h that are necessary for other geostatistical calculations and analyses such as kriging. Finally ordinary kriging was applied and maps of predicted height were produced over the entire region along with its standard error. All computations were conducted in the geoR package in R (Ribeiro Jr. and Diggle 2001).

Results

Empirical semivariogram model fitting

Figure 2 shows empirical semivariogram and its fitting by four theoretical models for both field-measurement-based and LiDAR-based plot-level heights. The semivariogram is the function describing the degree of spatial dependence of aggregated plot-level heights and the empirical semivariogram is a nonparametric estimate of the semivariogram. The empirical semivariance for a vector of separation h is derived by calculating one-half the average squared difference in plot-level height for every pair of plots locations separated by h. These values are then plotted again the distances between data pairs. Field plots in our sample were spread over the western Kenai region with the maximum distance of about 163,500 m. It is common to not compute the empirical semivariogram up to the largest possible distance due to the fact that shrinking number of available pairs for larger distances increases the variability of the empirical semivarogram. A general recommendation is to compute the empirical semivariogram up to about one half of the maximum separation distance in the data (Schabenberger and Gotway 2005). In addition, since field plots don't fall on a regular grid, the distances between pairs are all different. The distance considered need to be divided into regular bins and the distance value represented by the bin midpoint.

At least 30 pairs per bin were used to calculate empirical semivariogram (Banerjee et al 2004).

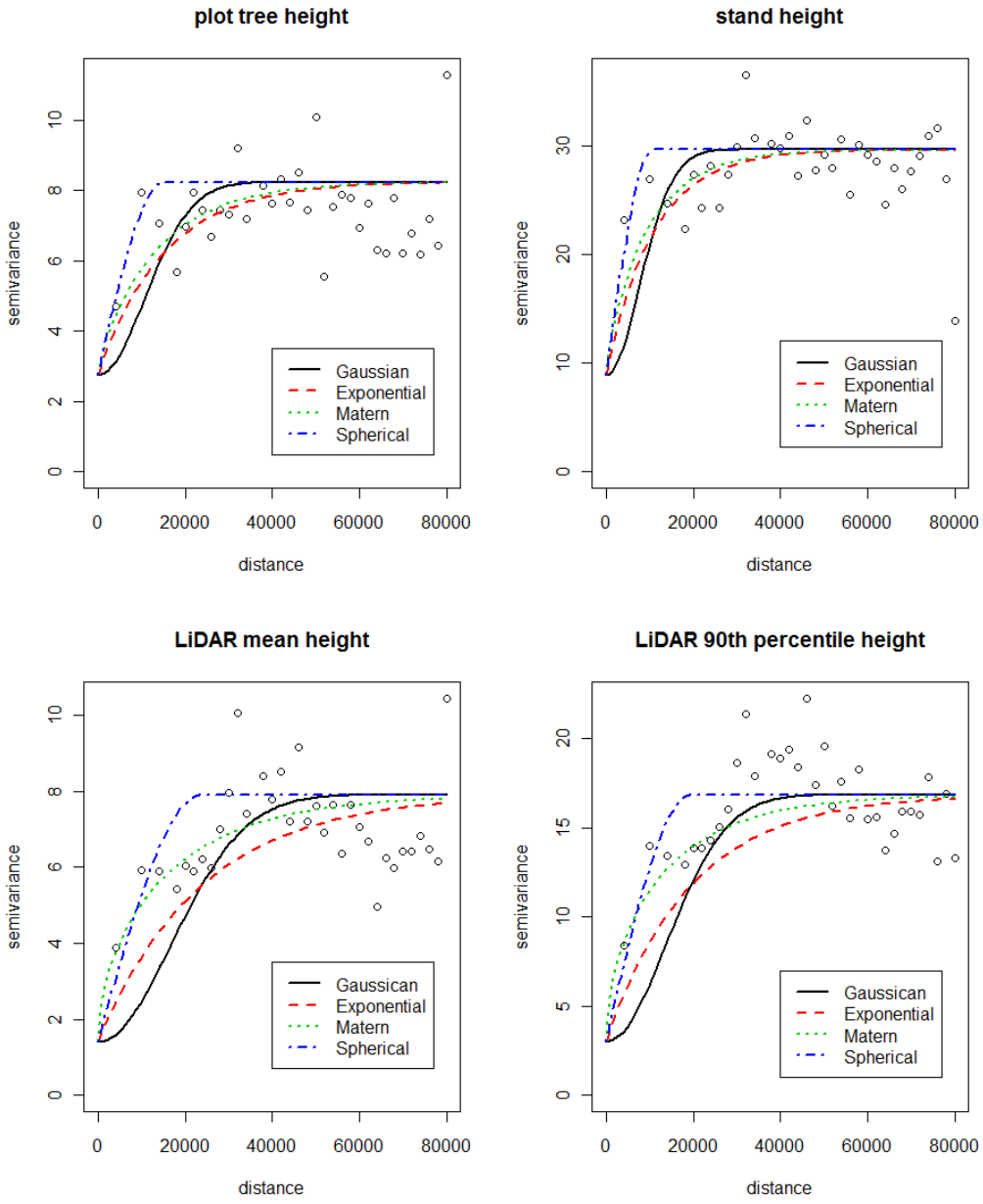


Figure 2: Empirical semivariogram fitting of four aggregated plot-level height

Figure 2 clearly shows that semivariance of aggregated plot-level heights has a similar pattern along distance. All semivariograms rise to a distance around 40,000 m then level off or decrease, which implies that aggregated plot-level heights from two plots may not be correlated when their distance is beyond 40,000 m. No semivarigrams pass through the origin, which suggests that the nugget effect is not zero for all cases. However, estimated sills are not the same. The estimated sill values are about 8, 30, 8, 17 for plot tree height, stand height,

LiDAR mean height and LiDAR 90th percentile height respectively. The estimated sill is the sum of total variation explained by the spatial structure and nugget effect. It seems stand height has more variation across the area than plot tree height in which only trees are considered. LiDAR 90th percentile height appears to have more variation than LiDAR mean height.

Four different semivariogram models - Gaussian, exponential, Matern and spherical model were fit to empirical semivariograms. The main differences among these theoretical models lie on curve smoothness and whether sill can be reached or not. The smooth parameter is infinity for Gaussian model, 1 for Matern model and 0.5 for exponential model. These models were fit interactively "by eye" and curves based on the best fitting model parameters were drawn in Figure 2. Within small distances, the spherical curve rises quickly and reaches the plateau in short distance. The curvature of Gaussian curve changes sign within a short distance. There is not much difference between exponential (red dash line) and Matern (green dot line) models. From visual examination, no models fit well. The better fitting - Matern model was finally chosen to be the covariance function.

Spatial prediction

Using the Matern covariance model, ordinary kriging was applied and height prediction and standard error over the region were computed at 300m by 300m pixel resolution. Contour maps of predicted height and standard error are displayed in Figure 3 and summary statistics are shown in Table 2. Empirical cumulative distribution functions and probability density functions of predicted plot-level height are plotted in Figure 4. As expected, predicted plot tree height is higher than predicted stand height and predicted LiDAR 90th percentile height is higher than predicted LiDAR mean height. The mean of predicted plot tree height is very similar to the mean of predicted LiDAR 90th percentile height, but predicted plot height has much less range than predicted LiDAR 90th percentile height. This is confirmed by distribution curves in Figure 4 in which the predicted LiDAR 90th percentile height represented by blue line spreads more widely than the predicted plot tree height represented by black line. Predicted stand height has similar mean and range as predicted LiDAR mean height. In fact their empirical distributions (green and red lines in Figure 4) seem very close. But predicted stand height has much larger kriging standard error (5.05-5.37 m) than predicted LiDAR mean height (1.94 to 2.78 m).

Table 1: Summary of predicted plot-level height

	Mean (m)	Median (m)	Minimum (m)	Maximum (m)
Plot tree height	12.34	12.41	10.12	14.62
Stand height	7.66	7.72	4.62	10.96
LiDAR mean height	7.37	7.49	4.12	11.25
LiDAR 90th percentile height	12.00	12.22	6.05	17.18

Contour maps shown in Figure 3 reveal similar spatial patterns for height predicted from field measurements and LiDAR data. A circular area of low height

is shown in the north-east of the Kenai Peninsula. Maps of kriging standard error also show the same pattern among different types of plot-level heights except that standard error of predicted stand height is a slightly larger. As expected, all standard error maps indicate that standard error is small near the location of the observed points.

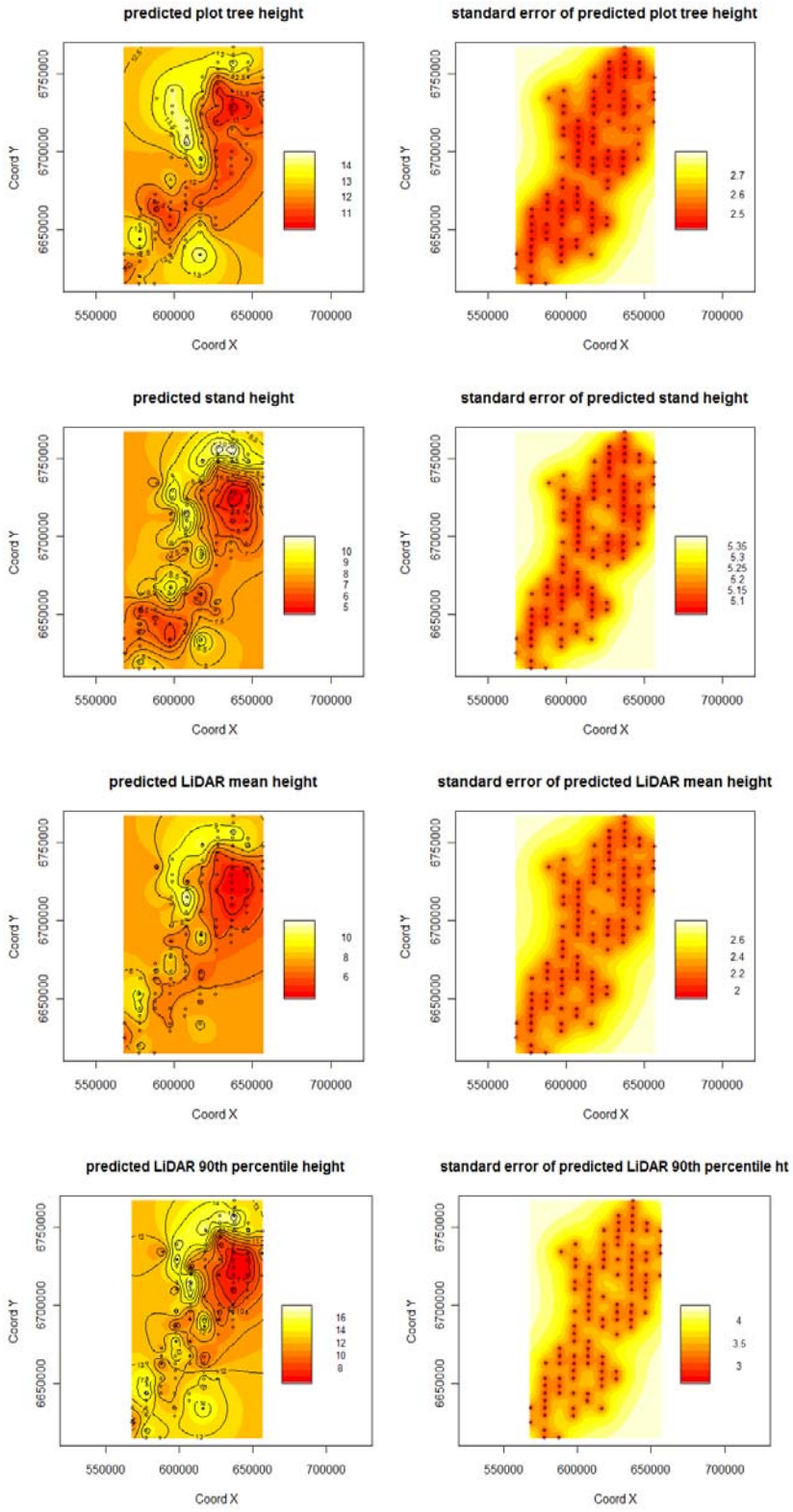


Figure 3: Maps of predicted plot-level heights along with their standard error estimates

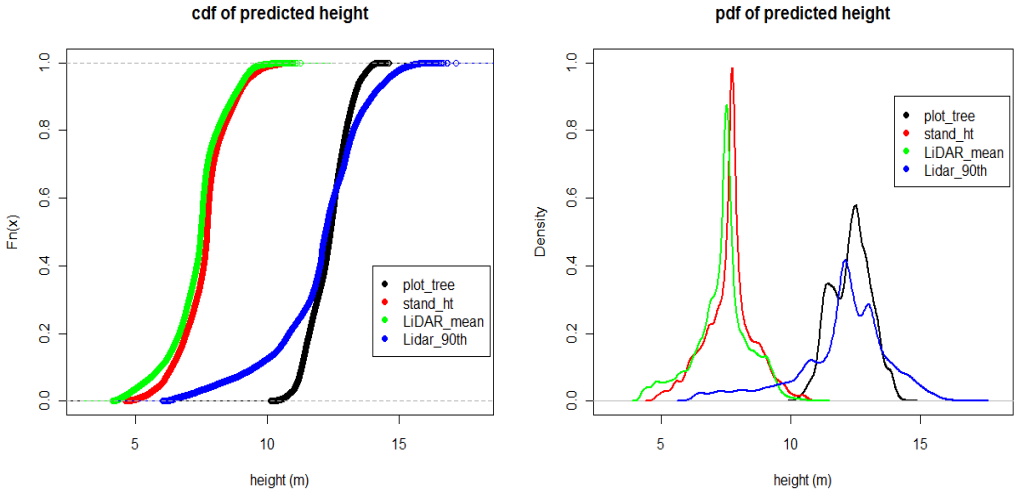


Figure 4: Empirical cumulative distribution function and kernel density function of predicted plot-level heights

Difference of predicted plot-level heights between field-based measurements and LiDAR-based measurements

Predicted plot tree height and predicted LiDAR mean height, predicted plot tree height and predicted LiDAR 90th percentile height, and predicted stand height and predicted LiDAR mean height were compared. Maps of the differences are shown in Figure 5. On average, predicted plot tree height is much higher than predicted LiDAR mean height with a mean difference 4.97m. The differences between predicted plot tree height and predicted LiDAR 90th percentile height, and between predicted stand height and predicted LiDAR mean height, are very small. For the majority of grids, these differences are within 1m as shown in Figure 6. On average, predicted plot tree height is higher than predicted LiDAR 90th percentile height by 0.34m and predicted stand height is higher than predicted LiDAR mean height by 0.28m.

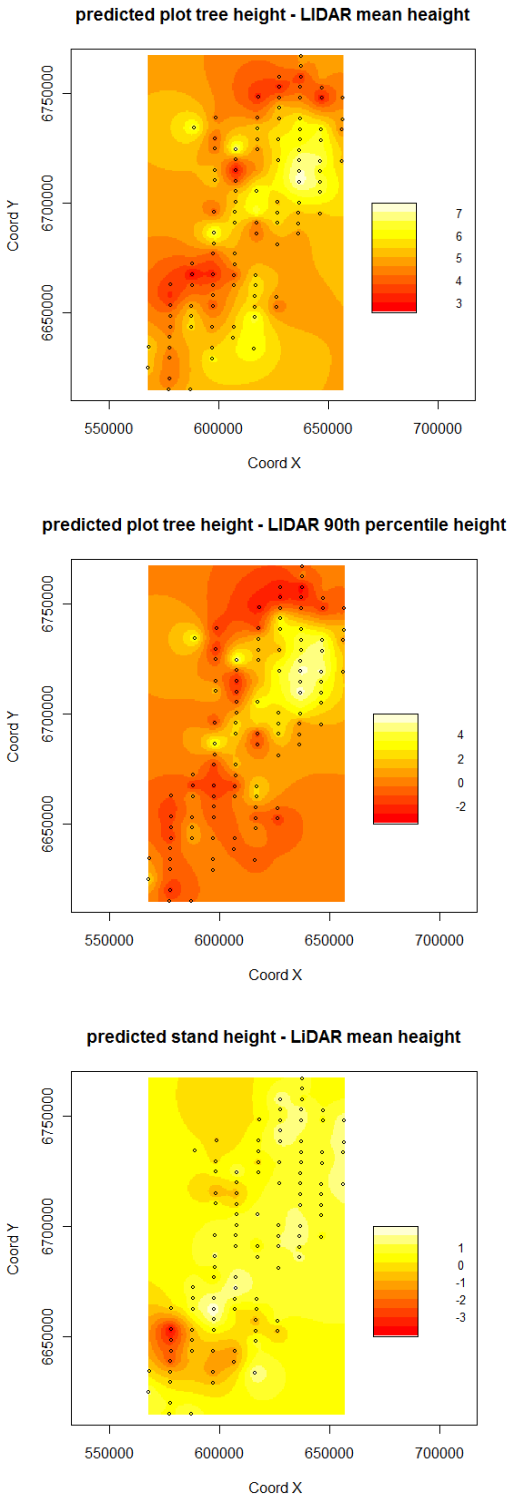


Figure 5: Differences of predicted plot-level heights between field-based measurements and LiDAR-based measurements

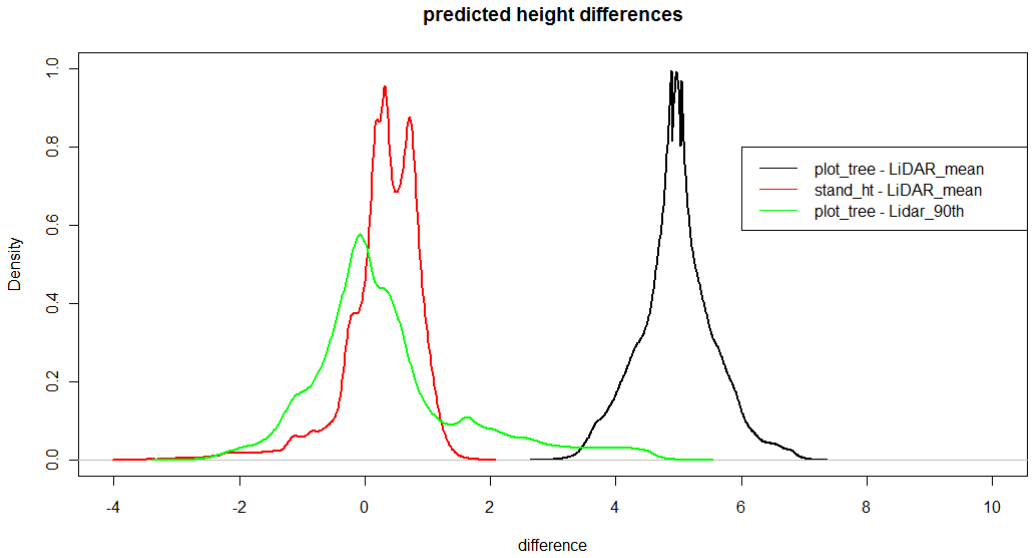


Figure 6: Empirical probability density function of the differences of predicted plot-level heights

Discussion

Semivariogram results indicate that aggregated plot-level heights in this dataset seem to spatially correlate until the distance between locations exceeds about 40,000m. However, since few pairs are located within short distances due to the fact that FIA plots are established based on an array of approximately 6,000-acre hexagons with each hexagon containing one plot (Bechtold and Patterson 2005), results may be different if field plots have a different distribution pattern.

Spatial prediction results show that at 300m by 300m pixel resolution, the distribution of predicted stand height is comparable to the distribution of predicted LiDAR mean height with a mean difference of only 0.28m, but predicted plot tree height is much higher than predicted LiDAR mean height with a mean difference of 4.97m. As described earlier, stand height is calculated from trees, saplings and seedlings, while plot tree height is calculated from trees only. In the literature, mean tree height measured on the ground is often reported to be higher than laser canopy height averaged over the sample plots due to the fact that the majority of laser returns would miss tree tops and would be reflected from the side of the crowns of dominant and co-dominant trees. The magnitude of difference depends on forest conditions and the LiDAR acquisition specifications used and it may vary from study to study, but is usually within 3m (Næsset et al 2004). The big difference between predicted plot tree height and predicted LiDAR mean height in our results is probably because forests in the western Kenai region have very low stand density (the mean is 66 trees per acre), low height and relatively open canopies, the laser can easily pass through the upper canopy so some lasers returns are indeed reflected from saplings and seedlings. This also explains why average height from trees, saplings and seedlings is very similar to the predicted LiDAR mean height. In addition, field plot height is the weighted

average of tree height from four surveyed subplots while LIDAR mean height is the average of the canopy return heights within the big plot containing all four subplots.

The mean of predicted plot tree height is comparable to the mean of predicted LiDAR 90th percentile height, but predicted plot tree height tends to have smaller standard error and range than predicted LiDAR 90th percentile height. Both field-based plot-level height and LiDAR-based height display similar spatial patterns across the whole region.

The choice of the covariance function impacts the kriging prediction. Since our primary interest is spatial prediction, the correctness of covariance model is important. Unfortunately the selected parametric Matern model doesn't fit the empirical semivariogram very well even though cross validation results indicate it is acceptable, it should be noted that spatial prediction may not be very accurate. Nevertheless, kriging surface maps produced in this study provide a visual display describing the spatial distribution of height, which is very useful information for forest inventory and monitoring. For the sake of simplicity, a constant mean model of Gaussian process was assumed. Considered the large area coverage, adding some covariant variables, such as weather parameters and site conditions, may improve prediction precision.

Conclusions

Assuming constant mean Gaussian process, spatial explicit maps of predicted plot-level heights are produced from field inventory and LiDAR data for western Kenai peninsula, Alaska. General spatial trends of predicted plot-level height are similar between field measurements and LiDAR measurements. The distribution of predicted stand height is very similar to the distribution of predicted LiDAR mean height, and the mean of predicted plot tree height is comparable to the mean of predicted LiDAR 90th percentile height, but the distribution of predicted LiDAR 90th percentile height has much heavier tails.

Reliable tree height mapping is useful to support forest inventory and monitoring. Most vegetation mapping today is conducted by manual photo-interpretation or satellite imagery combined with field surveys. The manual photo interpretation technique is costly and the results are dependent on the interpreter. Mapping based on optical satellite imagery requires that the area of interest is cloud-free. Nearly persistent cloud cover precludes acquisition of useful optical satellite images for a particular time period in Alaska. A remote measurement of forest structure that is rapid, reproducible and that provides reasonable spatial resolution is needed. As a rapidly-growing remote sensing technology, LiDAR offers great potential to capture canopy structure. However, due to high costs, LiDAR data are primarily acquired over specific project areas that are typically much smaller than the spatial extent at which most satellite image datasets are routinely acquired. In addition, it is unusual to have accurately georeferenced

field plots available over large regions. These factors may limit the operational use of LiDAR. In this study, instead of developing regression models assuming accurate field plot location, we develop a new approach that uses discontinuous LiDAR coverage and spatial models. We produced estimates of plot-level height over a large region using discontinuous LiDAR data that are comparable to those obtained using field inventory. The results are particularly useful for remote areas like Alaska where field work is expensive and optical satellite imagery is not easy to obtain. This approach could save time when greater accuracy is not needed, but quick assessment is necessary.

Acknowledgements:
The author would like to thank Robert McGaughey and Hans-Erik Andersen for their advice and helpful review on the manuscript.

References

Anderson, H-E., Reutebuch, S. E., and McGaughey, R. J. 2006. A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. Canadian Journal of Remote Sensing, Vol. 32 (5): p355-366

Banerjee, S., Carlin, B. P. and Gelfand, A. E. 2004. Hierarchical modeling and analysis of spatial data. Chapman & Hall/CRC

Bechtold, W. and Patterson, P. 2005. The enhanced forest inventory and analysis program- national sampling design and estimation procedures. USDA Forest Service, Southern Research Station, General Technical Report SRS-80. 85p.

Biging, G. S. and Dobbertin, M. 1995. Evaluation of competition indices in individual tree growth models. Forest Science. Vol 41: p360-377.

Field procedures for the coastal Alaska inventory. 2003. US forest service. 182p.

Gaveau, D., and Hill, R. 2003. Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. Canadian Journal of Remote Sensing. Vol 29 (5): p650-657.

Hoppus, M. L. and A. Lister (2006). The Status of Accurately Locating FIA Plots Using GPS. USFS Northeastern Research Station, Forest Inventory and Analysis White Paper.

Hyyppä, J., Pyysalo, U., Hyyppä, H., and Samberg, A. 2000. Elevation accuracy of laser scanning-derived digital terrain and target models in forest environment. In proceedings of EARSel-SIG-workshop on LIDAR. June 16-17, 2000, Dresden, Germany. FRG, Dresden.

Lappi, J. 2001. Forest inventory of small areas combining the calibration estimator and a spatial model. Canadian Journal of Forest Research Vol 31: p1551–1560.

Maltamo, M., Mustonen, K., Hyyppä, J., Pitkänen, J., and Yu, X. 2004. The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. Canadian Journal of Forest Research, Vol 34(9): p1791–1801.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment. Vol 80: p88-99.

Næsset, E, Gobakken, T., Holmgren, J. Hyyppä, H,. Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å ., Söderman, U. 2004. Laser scanning of forest resources: the Nordic experience. Scandinavian Journal of forest research. Vol 19: p482-499.

Persson, Å., Holmgren, J. and Söderman, U. 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogrametric engineering and Remote Sensing, Vol 68: p925-932.

Pretasch, H. 1997. Analysis and modeling of spatial stand structures. Methodological considerations based on mixed beech-larch stand in Lower Saxony. Forest Ecology and Management. Vol 97: p237-253.

Ribeiro Jr., P.J. and Diggle, P.J. 2001, geoR: A package for geostatistical analysis. R-news, Vol 1(2): p15-18. ISSN 1609-3631.

Schabenberger, O. and Gotway, C.A. 2005. Statistical methods for spatial data analysis. Chapman & Hall/CRC. 488p

Stoyan, D. and Stoyan, H. 1998. Non-homogeneous Gibbs process models for forestry-a case study. Biometrical Journal. Vol 40: p521-531.

Stoyan, D. and Penttinen. 2000. Recent application of point process methods in forestry statistics. Statistical Science. Vol 15(1): p61-78

Yu, X., Hyyppä, J., Hyyppä, H, and Maltamo, M. 2004. Effects of flight altitude on tree height estimation using airborne laser scanning. International Archives of Photogrammetry, Remote sensing and spatial information science. Vol XXXVI 8 / W2

Zawadzki, J., Cieszewski, C.J., Zasada, M. and Lowe, R.C. 2005. Applying geostatistics for investigations of forest ecosystems using remote sensing imagery. Silva Fennica. Vol 39(4): p599–617.

Calculation of Upper Confidence Bounds on Not-sampled Vegetation Types using a Systematic Grid Sample: An Application to Map Unit definition for Existing Vegetation Maps

Paul L. Patterson[1], Mark Finco[2]

Abstract: *This paper explores the information FIA data can produce regarding forest types that were not sampled and develops the equations necessary to define the upper confidence bounds on not-sampled forest types. The problem is reduced to a Bernoulli variable. This simplification allows the upper confidence bounds to be calculated based on Cochran (1977). Examples are provided that demonstrate how the resultant equations are relevant to creating mid-level vegetation maps by assisting in the development of statistically defensible map units.*

Keywords: Map unit, remote sensing, FIA, dominance type, grid sampling, confidence bounds, mid-level map

Background

Mid-level vegetation maps are important information sources for forest planning, habitat management, and many other resource management activities (Brohman and Bryant 2005). One of the first and most important steps in creating any map is defining the classes, or map units, which will be mapped.

In the Forest Service, several of the Regions have adopted similar map unit design processes that integrate regionally defined dominance type definitions with Forest Inventory & Analysis (FIA) data. Very generally, the first step of the process estimates dominance type composition over an area of interest using the Regional Dominance Type Classification (RDTC).

The second step of the process aggregates the RDTC classes into map units. This is generally done by setting a minimum abundance criterion (e.g., must comprise at least 3% of the area being mapped) for any map unit and hierarchically aggregating (i.e., create ecologically rational groups) the RDTC classes into map units, so that the map units meet the specified minimum abundance criterion. Some Forest Service Regions take the process a step further

[1] Mathematical Statistician, Interior West FIA, Ft. Collins, CO
[2] Senior Analyst, Remote Sensing Applications Center, Salt Lake City, UT

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

by considering "mappability" or "separability" of the map units in the context of the imagery and other geospatial data being used in the mapping process.

Problem Statement

FIA data is a sample of forest characteristics, having approximately one sample per 6000 acres of forest land. Traditionally, FIA data are used to describe forest composition based on the dominance types that were sampled. Because FIA is a sample, however, minor components of the landscape (in terms of abundance, not necessarily management or ecological importance) may not be represented in the FIA sample. A good example of this is riparian areas, which are ecologically important, but generally cover a very small percentage of the landscape.

This paper asks a non-traditional question of the FIA sample (or any other spatially balanced sample). Specifically, this paper explores what information FIA data can produce regarding dominance types or map units that were not sampled. The paper shows how sample data can be used to define the upper confidence bounds on not-sampled dominance types. This allows the analyst to state their confidence ($(1-\alpha)100\,\%$) that the proportion of a not-sampled dominance type is less than some proportion (η) of the region of interest (R). By doing this, an upper confidence bound is determined for proportion of R in not-sampled dominance types.

Statistical Derivation

We will assume the dominance type classification is based on data gathered on a plot, that the plots are a uniform shape. We are not assuming the plot design is the national FIA plot design, only that the plot design is same for all plots. Our proposed method for constructing the upper confidence bound is in the context of the finite sampling paradigm. Our population unit will be a co-located square that contains the plot. The square is only used in the construction of the statistical model; the size of the square is based on two constraints; first it is big enough to contain the square and second the square is small enough so that it can classified based on the data gathered on the plot. For example, the FIA plot characterizes and is contained in a co-located 300x300 foot square, but the FIA plot in most landscapes will not characterize a co-located square that is a quarter mile on a side. Let R_{sq} denote the square. If the region R is large enough, it is reasonable to assume we can tessellate the region R with the square R_{sq}. We further assume that the sample can be viewed as a simple random sample of the tessellation. As pointed out by Milne (1959) and Ripley (1981) the periodicity of a systematic grid, such as used by FIA, coinciding with a systematic land feature is of little concern. So the FIA sample grid can be viewed as simple random sample. The population characteristic of interest is whether each square in the tessellation is

classified as being of the rare dominance types. This is a Bernoulli variable; 0 when R_{sq} is not one of the rare dominance types and 1 when R_{sq} is one of the rare dominance types.

With the above background, we now outline how an upper $(1-\alpha)100\,\%$ confidence bound can be calculated based on Cochran (1977) sections 3.4, 3.5, and 3.6.

An estimate of the proportion (P) of region (R) that is classified as being in one of the rare dominance types is given by $p = \frac{a}{n}$, where n is the number of plots in the sample and a is the number of plots in the sample that are classified as one of the rare dominance types. Since the population characteristic is a Bernoulli variable we can find the frequency distribution of the estimate p. If we assume the population size N is much larger than the sample size n, then we can use the binomial approximation for the frequency distribution of p (see Cochran section 3.4). The population size N is equal to A/A_{sq}, where A and A_{sq} are the area of R and R_{sq} respectively.

We can now calculate an upper $(1-\alpha)100\,\%$ confidence bound for the proportion of R that is classified as being in one of the rare dominance types. It is the largest p such that

$$\sum_{i=0}^{a} \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \geq \alpha \qquad [1]$$

If a is greater than zero, then one solves equation [1] for p using numerical techniques; the case $a > 0$ is not of interest in this paper. If no sample plots are classified as being one of the rare dominance types, then a is equal to zero and equation [1] has a simple form which has the following solution:

$$p^0 (1-p)^n \geq \alpha \Leftrightarrow p \leq 1-\alpha^{1/n} \qquad [2]$$

If no FIA plots are classified as being in the rare types, then a $(1-\alpha)100\,\%$ upper confidence bound for proportion of rare categories in R is $1-\alpha^{1/n}$. That is we are $(1-\alpha)100\,\%$ confident that the proportion of land in R that is classified as being in one of the rare types is less than $1-\alpha^{1/n}$.

A word of caution about the sample size should be noted when using this method. To illustrate using an example, suppose the region of interest (R) is the state of Utah, and the rare types are types that occur within forested lands, it may

not be appropriate to use the number of plots in the state of Utah for n. It may be more appropriate to restrict to the subpopulation of forested lands in Utah. It is okay to do so (Cochran, section 3.10), but we need the size of subpopulation N' (in this example forested lands) to be much larger than n', the number of sample elements in the subpopulation, and we need to use n' instead of n in equation [2]. Our $(1-\alpha)100\%$ upper confidence bound for proportion of rare categories in the sub-population is then $1-\alpha^{1/n'}$.

Application Example

In this example, assume we are mapping an 800,000 acre National Forest. The National Forest has approximately 200 FIA plots, of which 100 plots are in forested land. When the regional dominance type definitions are applied to the FIA data, we estimate the proportion of each sampled dominance type on the forest lands (note that the non-forest dominance types would not be relevant using FIA data). Based on the number of plots on forested lands we estimate there are 400,000 forested acres, so N' is much larger than $n'=100$. Using the formula [2], with $n'=100$ the following table is generated relating to the forest types that were not sampled.

Table 1: Upper estimate of rare types as a function of desired statistical confidence.

Sample Size	Confidence	Upper Confidence Bound for Percentage of Rare Types
(n')	$(1-\alpha)100\%$	$(1-\alpha^{1/n'})100\%$
100	98%	3.8%
100	96%	3.2%
100	95%	3.0%
100	94%	2.8%
100	92%	2.5%
100	90%	2.3%

Using this table, we can state that we are 95% confident that the FIA sample did not miss a dominance type that comprised more than 3.0 % of the forested area. Actually, we can make a slightly stronger statement that we are 95% confident that all non-sampled dominance types combined comprise less than 3.0% of the forested area.

Conversely, we could use the same formulas to determine the number of plots (at the $(1-\alpha)100\,\%$ confidence level) required to not miss a dominance type that is at least $p100\,\%$ of the forest area. Solving the equation [2] for n we get

$$n = \frac{\ln(\alpha)}{\ln(1-p)} \qquad [3]$$

For example, if the requirement is to be 95% confident that a dominance type that comprises 1% (or greater) of the forest area was not missed, then n is 298. For the fictitious National Forest discussed above, this means that the standard FIA sample would have to be intensified to include an additional 198 plots in the forested area. Figure 1 illustrates sample size as a function of the minimum percentage of the dominance type of that is not be missed at a confidence level 95%. Note the rapid increase in sample size as the percentage of the region covered by the rare type decreases from 3.0%.

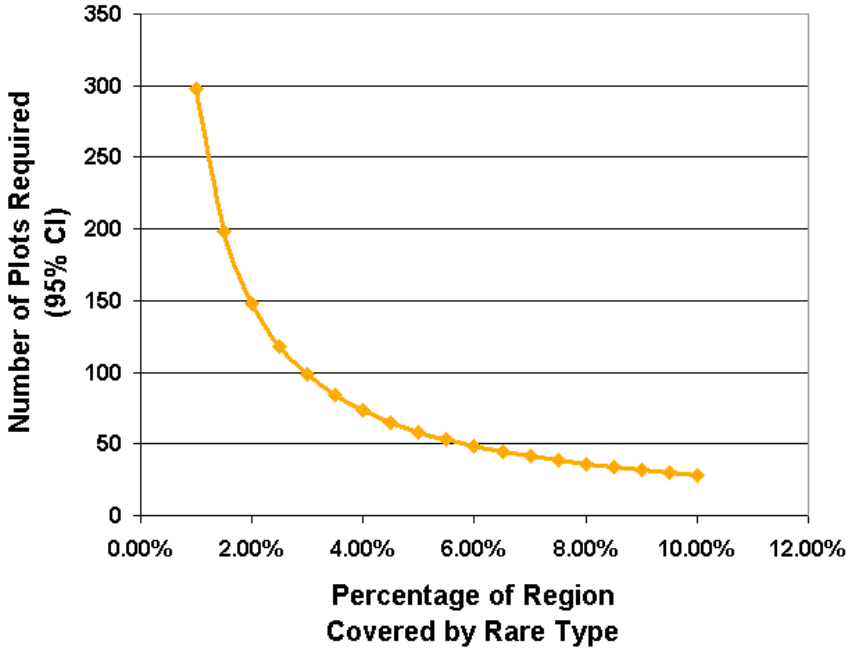


Figure 1. Sample size requirements as a function of percentage of region covered by the rare type.

Conclusions

The primary resource that land management agencies manage is existing vegetation. Vegetation impacts most resource management concerns and impacts issues ranging from timber harvest to threatened and endangered species. Therefore, it is critical for the Forest Service and other agencies to understand what vegetation exists, how it is spatially distributed on the landscape, and how much of it there is to effectively accomplish their land management responsibilities.

The idea that vegetation classification systems, existing vegetation maps, and inventory data need to be well integrated has long been recognized within some Forest Service regions and the rest are quickly adopting integrated classification, mapping and inventory programs. In all of these efforts, FIA and locally intensified inventory data have played a critical role. This paper provides insight into some previously unanswered questions that relate to the integration of existing vegetation maps and vegetation inventories.

References

Brohman, R.; Bryant, L. eds. 2005. Existing Vegetation Classification and Mapping Technical Guide. Gen. Tech. Rep. WO–67. Washington, DC: U.S. Department of Agriculture Forest Service, Ecosystem Management Coordination Staff. 305 p.

Cochran, W.G. 1977. Sampling Techniques. Ed. 3. Wiley, New York. 428 p.

Milne, A. 1959. The centric area-sample treated as a random sample. Biometrics. 15: 270-297.

Ripley, B.D. 1981. Spatial statistics. Wiley, New York. 252 p.



















The Rocky Mountain Research Station develops scientific information and technology to improve management, protection, and use of the forests and rangelands. Research is designed to meet the needs of the National Forest managers, Federal and State agencies, public and private organizations, academic institutions, industry, and individuals. Studies accelerate solutions to problems involving ecosystems, range, forests, water, recreation, fire, resource inventory, land reclamation, community sustainability, forest engineering technology, multiple use economics, wildlife and fish habitat, and forest insects and diseases. Studies are conducted cooperatively, and applications may be found worldwide.

Station Headquarters
Rocky Mountain Research Station
240 W. Prospect Road
Fort Collins, CO 80526
(970) 498-1100

Research Locations

Flagstaff, Arizona	Reno, Nevada
Fort Collins, Colorado	Albuquerque, New Mexico
Boise, Idaho	Rapid City, South Dakota
Moscow, Idaho	Logan, Utah
Bozeman, Montana	Ogden, Utah
Missoula, Montana	Provo, Utah

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 to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, DC 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

Federal Recycling Program Printed on Recycled Paper