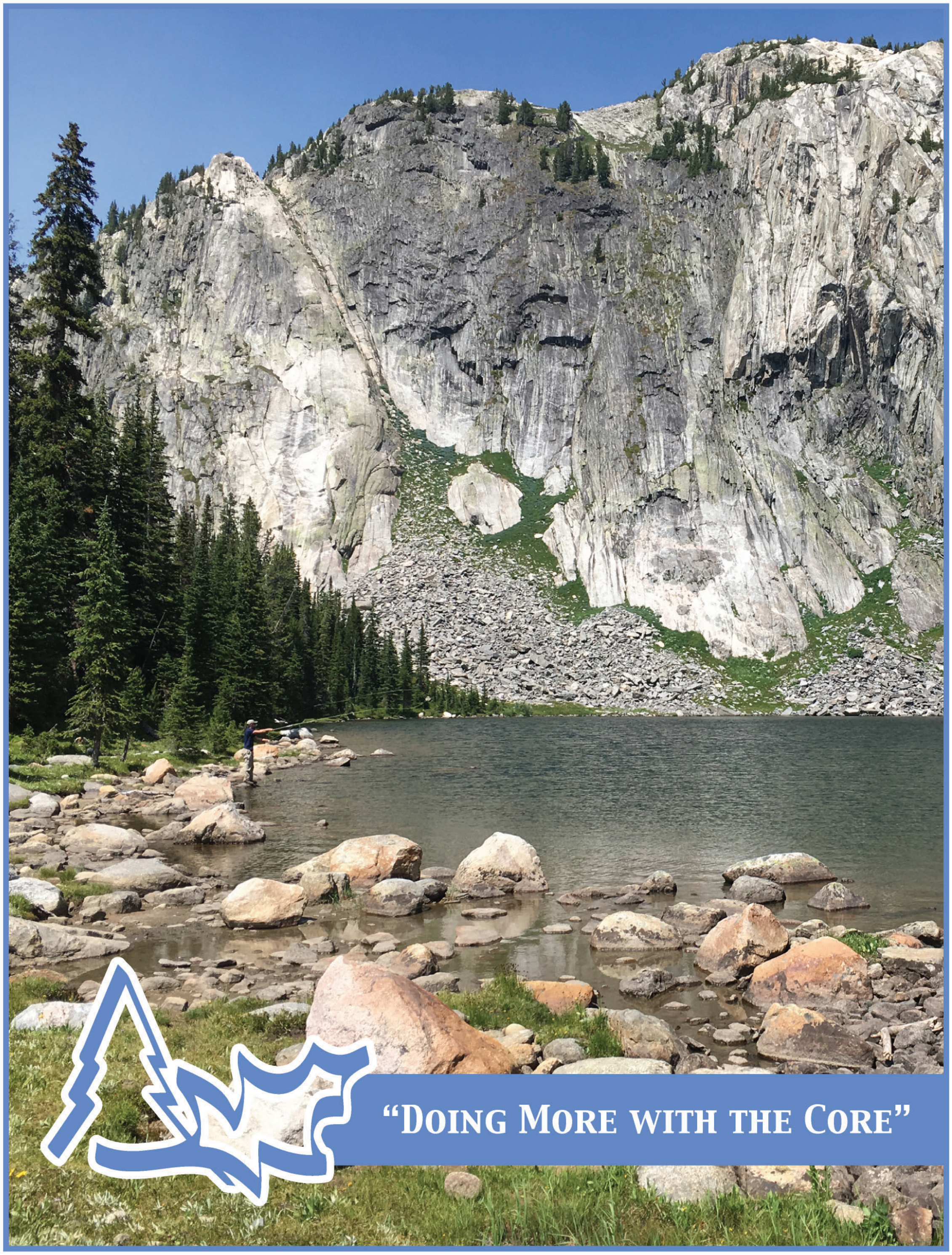
United States Department of Agriculture

Doing More With the Core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, Utah



Forest Service Rocky Mountain Research Station Proceedings RMRS-P-75 (Revised Edition) October 2017

Healey, Sean P.; Berrett, Vicki M., comps. 2017. **Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT**. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

Abstract

The Forest Service's Forest Inventory and Analysis Program (FIA) is the primary source of information about our forests' status and trends. A network of nationally consistent field observations forms FIA's core, and active collaboration with clients and peer organizations ensures that the resulting inventory remains agile, comprehensive, and relevant. An FIA Science Stakeholder Meeting was held October 24-26th, 2017, to bring key FIA personnel together with important partners and peers. The theme for this meeting, "Doing More with the Core," targeted innovation around how FIA analyzes, augments, and delivers information derived from its central data collection operations. Meeting topics included: 1) maximizing the value of the "annual" sample design adopted by the Program approximately 15 years ago; 2) developing better delivery tools that meet client needs and that take advantage of ever-expanding options for data visualization and communication; 3) assuring data continuity that supports a variety of land manager needs; 4) integrating more powerful and efficient monitoring technologies and statistical techniques into established analysis lines; and 5) supporting FIA's expansion in areas targeted by the most recent Farm Bill. This proceedings document includes details from selected papers presented at the meeting.

Keywords: statistics, estimation, sampling, modeling, remote sensing, forest health, data integrity, environmental monitoring, cover estimation, international forest monitoring

Compilers

Sean P. Healey is a research ecologist and **Vicki M. Berrett** is a program specialist (retired), U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 507 25th St., Ogden, UT 84401.

Copies of this and previous proceedings are available at
http://www.fia.fs.fed.us/symposium/proceedings/.

Photo credit: Rachel Simons, Forestry Technician, RMRS Forest Inventory and Analysis, Missoula, MT

Steering Committee

Vicki Berrett
Mike Clutter
Sean Healey
Greg Liknes
Randall Morin
Christopher Oswalt
Steve Prisley
Greg Reams
Todd Schroeder
Sharon Stanton
Michael Wilson

Acknowledgments

The Committee wishes to thank the National Council for Air and Stream Improvement (NCASI) for co-hosting the 2017 Stakeholder Meeting. We also thank the Rocky Mountain Research Station Publishing Services Group for excellent work in producing this proceedings document.

The following paper was added to this revised proceedings: A Forest Grid of México: Building Capacities for the Analytical Synthesis of the Mexican Forest and Soils Inventory, by Sergio Armando Villela Gaytán, Mario Guevara, Julian Equihua [et al]. Papers published in these proceedings were submitted by authors in electronic media. Editing was done for readability and to ensure consistent format and style. Authors are responsible for content and accuracy of their individual papers and the quality of illustrative materials. Opinions expressed may not necessarily reflect the policies and opinions of the U.S. Department of Agriculture. The use of trade or firm names in this publication is for reader information and does not imply endorsement by the U.S. Department of Agriculture of any product or service.

Foreword

The FIA (Forest Inventory and Analysis) Science Stakeholder Meeting brings together a cross-section of those involved with FIA's science delivery mission. Key clients, who use FIA data to support both public- and private-sector management decisions, interact with the FIA analysts and data delivery specialists responsible for ensuring that assessments derived from FIA's national forest inventory remain relevant and readily available. In turn, these groups exchange ideas with FIA scientists and their collaborators about new techniques and applications.

The inclusiveness of this meeting reflects longitudinal integration that has occurred within FIA over the last decade or so. Research into remote sensing, statistical techniques, and alternative data delivery platforms has become much less separate from FIA's operations. As a result, FIA has become more nimble in addressing emerging priorities; more adept at making use of ancillary data; more precise and more local; and accessible to a wider range of people.

At the same time, it is universally recognized that FIA's core field measurements comprise a globally unique information resource that becomes more valuable with every passing year of measurement continuity. The theme of this year's meeting, "Doing More with the Core," suggests the importance of building improved tools around FIA's sample to both increase the sample's relevance and to extend its value to new applications. The papers compiled here represent a variety of stages in the process of developing and delivering meaningful resource information to FIA clients, and they provide a snapshot of FIA's expanding integration of research and delivery.

Contents

"Doing More With the Core"—A Valuable Idea for All National Forest Inventories

Thomas J. Brandeis[1]

Abstract—The concept of "doing more with the core"—expanding the utility of a national forest inventory beyond its traditional stakeholder base into new areas—is particularly applicable in developing countries. Many of these countries are planning and implementing national forest inventories for the first time in response to the promise of climate change-related international aid; these inventories tend to focus on carbon estimation as the core information need. However, a national forest inventory has the potential to deliver much more information of value to stakeholders, and broadening its focus will allow countries to expand their client base and strengthen support for their efforts.

Keywords: forest inventory, international technical transfer, stakeholder engagement

Introduction

"Doing more with the core," the theme of this year's meeting, is a valuable idea that can be applied beyond the U.S. National Forest inventory realm. In the Forest Inventory and Analysis (FIA) context, our core mission has been the quantification of forest resources in ways useful to our original stakeholders: land managers, policymakers, and forest industry. These stakeholders require reliable estimates of the area of forest land, or more frequently timberland from which forest products could be harvested, as well as the volume of wood that could be found there. The FIA program was built around the need to deliver that important, but narrowly focused, information.

Outside the United States, particularly in developing countries, national forest inventories (NFI) are being designed around a new core information need: estimation of carbon status and trends. This focus on carbon estimation is directly tied a country's participation in programs designed to mitigate climate change by incentivizing forest carbon sequestration. The Reducing Emissions From Deforestation and Forest Degradation (REDD+) mechanism is a major driver behind increased interest in forest inventory and monitoring around the world. REDD+ "incentivizes developing

[1] USDA Forest Service, Southern Research Station, Forest Inventory and Analysis, 4700 Old Kingston Pike, Knoxville, TN, 37919 tjbrandeis@fs.fed.us

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

countries to keep their forests standing by offering results-based payments for actions to reduce or remove forest carbon emissions" (UN-REDD Programme 2017). Key to this program is the estimation of current forest carbon stocks and monitoring their rates of change; thus, participation requires some form of NFI. The UN-REDD Programme and donor nations have provided technical assistance to countries establishing either their first NFI or modifying existing ones to better assess forest carbon pools. But this narrow focus on forest carbon estimation neglects opportunities to build the capacity to "do more with the core."

Implementing an NFI means balancing the needs for statistical robustness within the limits of the resources we can commit to it. One way to increase the resources that can be made available to the NFI is by addressing the information needs of as many stakeholders as possible, thus garnering more support from potential funders. A critical step in NFI design is scoping information needs by engaging with a wide range of current and potential forest stakeholders. Typically, in countries where stakeholders are engaged and questioned, a broad range of information needs emerge beyond carbon estimation. These outreach activities build social acceptance and political support for committing often considerable resources to the effort.

The FIA program is familiar with this ever-present tension between producing the best possible estimates within the limits of the budget. This tension is amplified in developing countries. There is often a lack of forest inventory "tradition" and the understanding of what can be asked of an NFI; thus, there is little understanding of its value. Resources in many countries—including well-trained, financial staff and functioning institutions—are often severely limited. "Doing more with the core" under such conditions becomes more than a favored option; rather, it is critical to establishing and sustaining NFIs. Experience has shown commonalities in stakeholders' information needs beyond climate change-related carbon dynamics. When planned in advance, a well-designed NFI can address many additional questions with little or no additional resource commitment, and without compromising its ability to produce reliable, repeatable, and robust forest carbon estimates.

Examples of Expanding Beyond Carbon Estimation

Sampling Intensity

Starting with sampling, most phases of a forest inventory encounter pitfalls when the design is too narrowly focused on national level carbon estimation. Careful consideration must be given so that the sampling design delivers statistically valid estimates not only at the national scale, but also at the geographic scales at which management and planning takes place, such as within subnational political units. Are there additional domains to be considered for which resource managers need estimates (e.g., areas with intensive forest management, threatened forest habitat types, especially high biodiversity, or a higher-than-normal incidence of illegal logging)? Designing the NFI to answer these non-traditional questions, or even building in the flexibility to modify sampling as future needs arise, can greatly increase the utility of the NFI.

Land Use and Land Cover Change

In addition to on-the-ground NFI plots, supplemental Land Use/Land Cover (LULC) monitoring systems using remotely sensed data potentially deliver wall-to-wall, spatially explicit estimates of carbon loss and gain (IPCC 2003). These same systems can quantify and locate LULC changes due to disturbances like wildfire, pest and disease outbreaks, urbanization, illegal logging, etc., that are relevant to forest policy and management. Focusing solely on their utility for carbon monitoring misses opportunities to leverage the investment and provide information for a broader user group.

Data Collected

When viewed from the U.S. perspective, it is ironic that the quantification of our traditional core information need, timber volume, is typically seen as less important than that of carbon storage in many countries currently planning their first NFI. Much more focus is given to the compilation and development of allometric models for biomass estimation than for merchantable volume estimation. Processing system plans need to include volume estimation and field data collection needs to consider the inclusion of additional variables relevant to local wood products, such as tree quality or grade, stem defects, and/or cull proportion.

Non-timber forest products (both on and off the NFI plot) are frequently important to local economies; therefore, information on their status and trends is often needed by forest resource stakeholders. Can their inventory and monitoring be incorporated in the larger sampling scheme? Related to this, the United Nations' Food and Agriculture Organization (FAO) has designed and included socioeconomic surveys of forest resource users and the surrounding communities. Typically these surveys go beyond forest resource questions and delve into broader rural social and economic issues.

The FIA program recognizes the need for additional monitoring of forest health indicators, wildlife, nonnative invasive species, and botanical surveys with varying degrees of intensity and focus. Countries in the tropics frequently express interest in using their NFI network to assess the high levels of biodiversity in their forests. An NFI provides a statically robust framework within which additional data collection can be incorporated, while always being mindful of the additional costs and complexities involved.

Conclusions

From the outset, care must be taken to avoid designing the NFI with limitations. Broadening the focus of an NFI to deliver valued products to a wider range of stakeholders becomes especially critical in countries where resources are limited. Justifying the considerable expenditure necessary to implement an NFI when the only selling point is the possibility of future payments from international organizations can be a difficult argument to make to all stakeholders, from local forest communities to national politicians. Comprehensive, ongoing stakeholder engagement from needs assessment, through planning and implementation, and finally during results dissemination, helps create and maintain a client base for NFI information. This client base can express their desire to decisionmakers that the NFI receive support.

Acknowledgments

Thanks to the reviewers, Andy Lister and Sarah Banks, and Southern Research Station technical editors.

References

Intergovernmental Panel on Climate Change (IPCC). 2003. Good practice guidance for land use, land-use change, and forestry. Hayama, Japan: IPCC. Online at: http://www.ipcc-nggip.iges.or.jp/public/gpglulucf/gpglulucf_files/GPG_LULUCF_FULL.pdf

United Nations Reduced Emissions From Deforestation and Degradation (UN-REDD) Programme. Page last modified: 2017. Rome, Italy: United Nations Food and Agriculture Organization UN-REDD Programme Collaborative Online Workspace. http://www.unredd.net/. [Accessed 6 April 2017].

Using the Core in Lieu of Adding More: A Forest Health Case Study

Kerry Dooley[1]

Abstract—Evolving requests of Forest Inventory and Analysis (FIA) data may prompt additional data (i.e., new variables) to be collected. Before implementing changes, careful consideration should be given to whether FIA data are a good source for the information; whether the data fit well within the FIA framework; and whether the information could be gleaned from information already being collected by the FIA program or other sources. In this evaluation, data relating to forest health are used to illustrate limitations as well as highlight capabilities of the FIA program to meet evolving data needs.

Keywords: forest inventory and analysis, forest health, cross-agency collaboration, data utility

Introduction

The U.S. Department of Agriculture, Forest Service's Forest Inventory and Analysis (FIA) program is the largest available source of forest data across ownerships in the United States. The program is national in scope and has a long history; thus, there is an inherent interest in getting maximal information from the FIA plot system. The interest may be sparked by users of FIA data or may be required by mandates, such as The Agricultural Act of 2014 (H.R. 2642; Pub. L. 113-79, also known as the 2014 Farm Bill). Sometimes the interest in gaining more information leads to the addition or modification of collected data. There will be occasions where addition of variables is the only option, but it will often be preferable to use data already being collected. Here, the 'damage agents' dataset is evaluated in this context and then compared with other possible FIA sources of forest health data.

[1] Forester, USDA Forest Service, Southern Research Station, Forest Inventory and Analysis, 4700 Old Kingston Pike, Knoxville, TN, 37919, (865) 862-2098, kjdooley@fs.fed.us

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

Evaluation of the Damage Agent Dataset

Is FIA the Best Source for This Information?

For the damage agent dataset, data are collected on factors that have a direct negative effect on a tree's survival, growth, or marketable products. During regular plot visits, individual trees are examined for threshold levels of select agents. The list of agents includes biotic and abiotic factors, and up to three agents may be identified per tree.

Forest health data, from both ecologic and economic perspectives, are often time sensitive. Over time, biotic agents will become established, making eradication less likely and more costly. Therefore, the most useful collection and monitoring methods are those that can be quickly deployed and focused at a time and place of concern. FIA plots, however, are spread miles apart and are revisited at 5- to10- year intervals.

With the exception of group-level damage codes, FIA damage agents are very specific. While this may be a good match for what users want, FIA cruisers' training in entomology and plant pathology varies greatly and, by the nature of the job, FIA cruisers must be forestry generalists. As such, requiring crews to identify damage agents to the species level is an unrealistic expectation. In addition, some of the damage agent definitions are confusing, and some are redundant (either with other damage agents or with other FIA measurements). Finally, because FIA plot visits are not scheduled specifically to match potential damage agent cycles, the agent in question—and in some cases, the part of the tree affected (i.e., foliage)—may not be present. These factors all affect the ability of FIA crews to consistently and accurately identify damage by specific agents.

Is Collection of This Information Appropriate for FIA?

Maintaining data continuity is an important goal of the FIA program. Any addition or modification of variables will hinder this objective as these changes will cause current survey cycles to differ from past cycles. In addition, data types that are dynamic or those that are especially complex may lead to issues of continuity with future survey cycles.

Invasive pests and pathogens are constantly being introduced to the United States, making the list of damage agents a dynamic metric. For example, even if the FIA program had been collecting information on damage agents since its inception, emerald ash borer would be a relatively new addition to the list, as it was not discovered in the United States until 2002 (http://www.emeraldashborer.info). Unfortunately, it is almost certain that introductions of forest pests and pathogens will continue. Therefore, these variables will likely need to be continually modified, creating a challenge to temporal continuity.

The FIA program is nationally administered but regionally implemented. As such, the various regions all collect the same 'core data.' Often nested within these core data are more specific core-optional data, which each region may or may not collect. The damage agent dataset is an example of this, with group-level agents being core, and specific agents being core-optional. These core-optional damage agents cause breaks in data continuity at regional borders. Additionally, as outlined in the previous section, several factors may influence an FIA cruiser's ability to confidently determine the presence of a specific damage agent at a particular site. This may lead to the same observation being categorized differently by different crews, which in turn will cause spatial discontinuity (figs. 1 and 2).

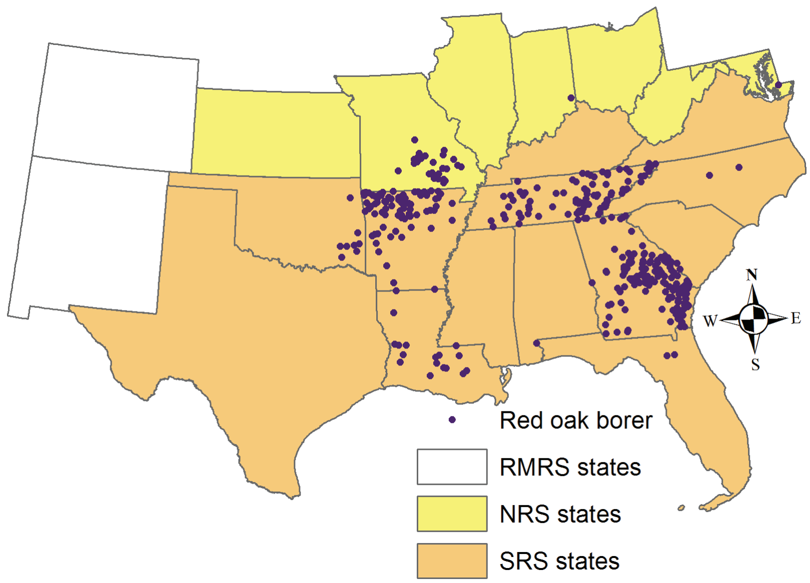


Figure 1—Fuzzed locations of FIA plots with red oak borer damage recorded; plots are located in Southern Research Station states and bordering states.

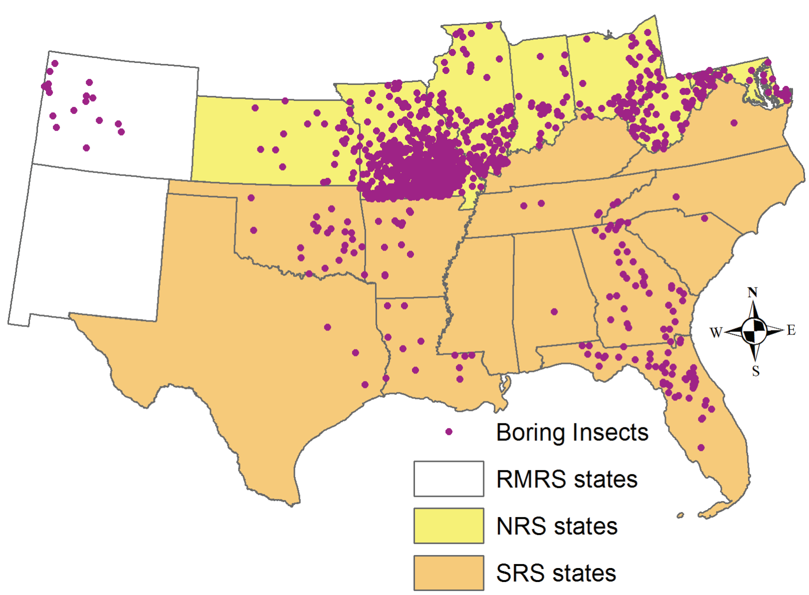


Figure 2—Fuzzed locations of FIA plots with non-specific boring insect damage recorded; plots are located in Southern Research Station states and bordering states.

Alternative FIA Sources of Forest Health Information

Analysis of group-level (core) damage agents can provide interesting insights. This is particularly true when these data are coupled with other data and information, as demonstrated by Morin et al. (2016) where the authors collapsed the damage agent data into broad categories and focused on the 20 most prevalent tree (host) genera.

Several other forest health indicators are part of the core FIA measurements, including crown condition measurements and site or stand disturbances. Randolph (2007) demonstrated the power of combining these measurements with other FIA (basal area by species, tree and plot notes) and non-FIA (aerial sketch maps) data.

Finally, the most basic FIA measurements are some of the most useful. For example, identification of tree species is crucial for damage-agent-host distribution mapping and evaluation. This can include analyses of the impact of known, current pests or pathogens (see range and intensity evaluations by Morin et al., 2005); predictive assessments of pests and pathogens seen as potential threats, as illustrated in the Forest Health Technology Team's oak splendor beetle map, reviewed by Vogt and Koch (2016); or spreading threats such as the predicted redbay ambrosia beetle extent map developed by Koch and Smith (2008).

As an alternative to the FIA program collecting specific-level damage data, FIA's core data can be coupled with complementary datasets, such as those acquired by the Forest Health Monitoring (FHM) program under Forest Health Protection, State and Private Forestry. These data are not sample-based, but document actual forest-health incidents in near real-time and are spatially explicit, damage-agent specific, and collected annually by forest health professionals in State and Federal Governments.

Conclusions

FIA's real strength comes from the consistently measured, long-term, and geographically broad data it collects and provides. FIA must allow for some adaptability to meet legislative mandates and user requests, but addition or modification of variables should be carefully considered and alternatives, such as collaboration with other programs and connections with other data, should be sought when possible.

Acknowledgments

Gratitude to Anita Rose and Chris Asaro for reviewing this paper, and to the conference steering committee and editors of the General Technical Report for their efforts.

References

Emerald Ash Borer Information Network. 2017. East Lansing, MI: Michigan State University and U.S. Department of Agriculture Forest Service. Online: http://www.emeraldashborer.info [Accessed 3 May 2017].

Koch, Frank H.; Smith, William D. 2008. Spatio-temporal analysis of *Xyleborus glabratus* (Coleoptera: Circulionidae: Scolytinae) invasion in eastern U.S. forests. Environmental Entomology. 37(2): 442-452.

Morin, Randall S.; Pugh, Scott A.; Steinman, Jim. 2016. Mapping the occurrence of tree damage in the forests of the Northern United States. Gen. Tech. Rep. NRS-GTR-162. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 19 p.

Morin, Randall S.; Liebhold, Andrew M.; Luzader, Eugene, R.; [et al.]. 2005. Mapping host-species abundance of three major exotic forest pests. Res. Pap. NE-726. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 11p.

Randolph, KaDonna C. 2007. A comparison of tree crown condition in areas with and without gypsy moth activity. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; McWilliams, William H., eds. Proceedings, Seventh Annual Forest Inventory and Analysis Symposium; 2005 October 3-6; Portland, ME. Gen. Tech. Rep. WO-77. Washington, DC: U.S. Department of Agriculture, Forest Service: 107-113.

Vogt, James T.; Koch, Frank H. 2016. The evolving role of Forest Inventory and Analysis data in invasive insect research. American Entomologist. 62(1): 46-58.

Development and Evaluation of Multi-Species, Cross-Regional Stem Taper and Bark Thickness Models for Predicting Merchantable Volume Across the United States

Jereme Frank[1], Aaron Weiskittel[2], David Walker[3], Phil Radtke[4], James Westfall[5], David MacFarlane[6], David L.R. Affleck[7], and Hailemariam Temesgen[8]

Abstract—Tree-level models used in large-scale inventories necessitate flexible modeling structures that can accommodate multiple species across varying sites and regions. Presently the United States Forest Service, Forest Inventory and Analysis Program (USFS-FIA) utilizes over 40 volume models of different forms specific to 22 geographic regions. Recently much of the data used to create these volume models has been compiled into a database including nearly 250,000 trees. This database provides the opportunity to formally assess tree-level variation in volume and taper among and within eco-physical regions and at varying taxonomic levels. In this analysis, we developed national-scale, non-linear, mixed effects outside bark taper and bark-thickness models to calculate merchantable volume in a consistent manner between regions and taxonomic groups. We assessed these models in terms of accuracy and precision and compared them to merchantable volume models currently used by FIA. The mixed effects taper-based approach performed well when compared to the conventional approach for estimating merchantable wood volume. Using a single, wide scale volume modeling system should lead to improved estimates of volume for some species particularly where little data is available.

[1] Research Assistant, School of Forest Resources, University of Maine, Orono, ME, 04469-5755, jereme.frank@maine.edu

[2] Associate Professor, School of Forest Resources, University of Maine

[3] Research Associate, Department of Forest Resources and Environmental Conservation, Virginia Tech University

[4] Associate Professor, Department of Forest Resources and Environmental Conservation, Virginia Tech University

[5] Research Forester, USDA Forest Service, Northern Research Station

[6] Associate Professor, Department of Forestry, Michigan State University

[7] Associate Professor, W.A. Franke College of Forestry and Conservation, University of Montana

[8] Professor, Department of Forest Engineering, Resources, and Management, Oregon State University

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

Introduction

Presently, to estimate merchantable wood volume (from a 1-foot stump to a 4-inch top diameter), the USFS-FIA uses over 40 volume models across 22 different regions (Woodall et al. 2011). These models vary in terms of the extent and number of trees sampled within a region, and in the model form. Furthermore, some of these models are based on volume tables rather than actual data, compromising error estimates. Much of the actual data used to develop these models have recently been compiled into a national legacy tree database (Radtke et al. 2016) and augmented with additional data collected as part of an ongoing effort to validate the current USFS-FIA approach for estimating volume and biomass across the United States. Recent work showed that depending on the region, FIA volume models may under-predict by as much as 19.2 percent (Radtke et al. 2017). Preliminary efforts showed that a broad-scale multi-species taper model could 1) provide merchantable volume estimates on par with current regional volume models; 2) provide a unified, compatible framework across the eastern United States; and 3) allow for volume estimates to a flexible stump and top limit (Weiskittel et al. 2016a). In this analysis, we expand our scope from the eastern United States to the entire United States.

Methods

The legacy database contains taper measurements from nearly 250,000 trees, 118,403 of which were measured for outside bark diameter and suitable for fitting an outside bark taper function. These data span 17 eco-divisions (fig. 1) across the United States and Canada, with the highest representation occurring in the subtropical division

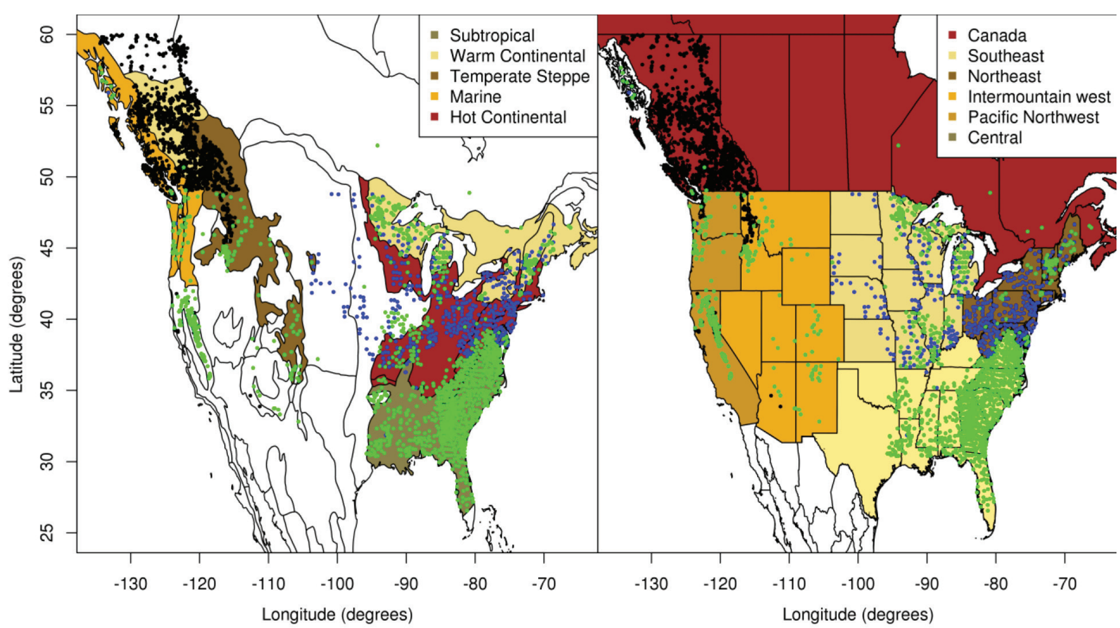


Figure 1—Spatial distribution for all sampled trees where a) inside bark only measurements are available (black circles); b) outside bark only measurements are available (blue circles); and c) where both inside and outside bark measurements are available (green circles). The panel on the right shows Canada along with five broad regions in the United States, while the left panel shows all eco-divisions (Bailey 1973). The five eco-divisions with the greatest number of sampled trees are illuminated.

(Bailey 1983). Fewer merchantable trees (70,493) were available with both diameter inside and outside of the bark measurements (table 1) and suitable for developing a bark thickness model and merchantable volume verification. These trees spanned 26 families with the number of species per family ranging from one to 35 in the pine family (Pinaceae). The most prominent species with both inside and outside bark measurements was loblolly pine (FIA species code 131) with 12,710 trees.

Table 1—Summary statistics. Total number (no.) of trees and taper measurements for 19 of the 26 families (those with over 100 observations) along with average (avg.) tree d.b.h., height, and merchantable (merch.) volume. Number of species and number of trees for the most prominent species (FIA species [spp.] codes as presented in Appendix F in the FIA database) are also given for each family.

Family	Avg.	SD	Avg.	SD	Avg.	SD	No. of trees per family	No. of taper meas. per family	No. of spp.	FIA spp. code	No. of trees per spp.
	d.b.h. - - *inches* - -		*Height* - - -*feet* - - -		*Merch. volume* - - - *cu. ft.* - - -						
Aceraceae	11.3	4.9	64.3	15.0	20.8	23.4	2,042	22,947	6	316	1,538
Altingiaceae	11.4	4.8	71.6	19.3	24.0	27.4	2,494	39,160	1	611	2,494
Betulaceae	10.6	4.5	65.3	12.7	19.0	20.9	1,272	11,791	6	370	363
Cannab.	13.9	5.5	69.7	15.0	30.6	27.4	183	2,966	2	460	163
Cornaceae	11.7	4.7	63.0	17.7	20.8	19.8	1,757	26,145	4	694	889
Cupress.	9.3	3.6	43.9	15.1	9.7	10.6	550	4,631	4	68	232
Fabaceae	10.4	4.2	62.9	18.2	14.6	15.0	385	4,899	3	901	380
Fagaceae	12.5	5.2	65.7	16.9	25.5	26.9	12,131	178,120	24	802	2,727
Hippocast.	12.5	5.4	66.3	17.7	27.0	26.8	104	1,540	1	330	104
Jugland.	12.2	4.7	69.9	17.8	25.3	26.4	1,994	29,684	6	400	1,727
Magnoli.	12.9	5.0	79.4	20.0	31.9	31.9	2,788	48,498	4	621	2,452
Oleaceae	12.4	5.3	74.6	17.6	27.7	27.9	901	9,322	4	541	397
Pinaceae	10.8	4.8	61.8	19.2	21.7	39.4	40,465	502,737	35	131	12,710
Platanaceae	12.4	4.4	76.1	17.3	26.9	21.0	332	5,735	1	731	332
Rosaceae	13.1	6.6	70.6	24.0	38.6	46.5	175	2,158	2	762	173
Salicaceae	9.7	3.5	68.0	14.2	15.5	16.5	962	8,779	7	746	761
Taxodiaceae	12.4	5.8	70.4	23.2	26.3	36.3	567	8,762	3	222	318
Tiliaceae	14.8	5.7	81.1	16.0	45.4	40.3	417	5,425	2	950	352
Ulmaceae	12.4	5.3	66.7	17.0	26.2	27.5	432	6,664	3	970	421
All species	11.3	4.9	64.2	19.1	22.9	34.6	70,493	924,882	-	-	-

We used the legacy database to fit an outside bark variable exponent taper model in a hierarchical framework with random effects on variables $\propto_0$ and β_3 (Li et al. 2012):

$$dob = (\propto_0 + Random_0)\, D^{\propto_1} H^{\propto_2} X^{\left(\beta_1 z^4 + \beta_2\left(1/e^{D/H}\right) + (\beta_3 + Random_3) X^{0.1} + \beta_4 (1/D) + \beta_5 H^Q + \beta_6 X\right)},$$

where

$X = 1 - z^{1/3} / 1 - 1/3,^{1/3},$

$Q = 1 - z^{1/3}$,

p = 1.3/H ,

z = h/H ,

dob = diameter outside bark,

D = diameter at breast height (d.b.h.),

h = height within the tree where diameter was measured, and

H = total tree height.

The fullest model used the following hierarchy: 1) evergreenness (conifer or deciduous); 2) taxonomic family; 3) taxonomic genus; 4) taxonomic species; 5) eco-division (Bailey 1983); and 6) tree. Five additional models were built dropping the lowest level in each subsequent model. The best performing model in terms of AIC was used to estimate stem diameter outside bark and converted to stem diameter inside bark using a simple bark thickness power function (Weiskittel et al. 2016a). Section volumes were calculated using Smalian's formula on the taper-estimated and measured diameters and summed to provide taper-estimated and measured merchantable volumes. The latter we used as a baseline for error comparisons by region and species in terms of root mean square error (RMSE), and mean bias.

Results

The best performing model in terms of AIC, RMSE, and mean bias was the fullest model that included a tree-level random effect. However, since it is generally impractical to predict at the tree level without taper measurements, we chose the second highest performing model at the level of species and eco-division for further analysis. When compared to the baseline Smalian's volume estimate, mean bias was -1.3 percent using the FIA approach and 0.8 percent using the taper model, while RMSE was 25.8 percent using the FIA models and 24.5 percent using the taper model (table 2). The taper model performed as well or better for all regions in terms of RMSE. The FIA models had lower mean bias in the southern states and the Pacific Northwest while the taper function performed better in the Northeast, Central States, and Intermountain West.

Overall, model performance was similar, with both methods generally predicting below 5 percent mean bias for each species with higher within species variation (fig. 2).

Table 2—Sample size, mean volume, taper-estimated volume and FIA-estimated volume and associated mean bias and RMSE for five regions across the United States.

Region	Mean merch. volume	Mean bias		RMSE		No. of trees
	Actual	Taper	FIA	Taper	FIA	
	- - - - - - - - - - - - - - - - - - Cubic feet- - - - - - - - - - - - - - - - -					
Central	23.6	1.1	-3.0	5.3	6.1	2,898
Intermountain West	48.3	-1.6	-2.2	11.2	13.7	3,360
Northeast	18.2	0.0	0.3	3.5	3.9	3,342
Pacific Northwest	45.5	-2.4	-0.9	14.8	15.0	2,786
Southeast	20.6	0.4	-0.1	4.2	4.2	58,107
All data	22.9	0.2	-0.3	5.6	5.9	70,493

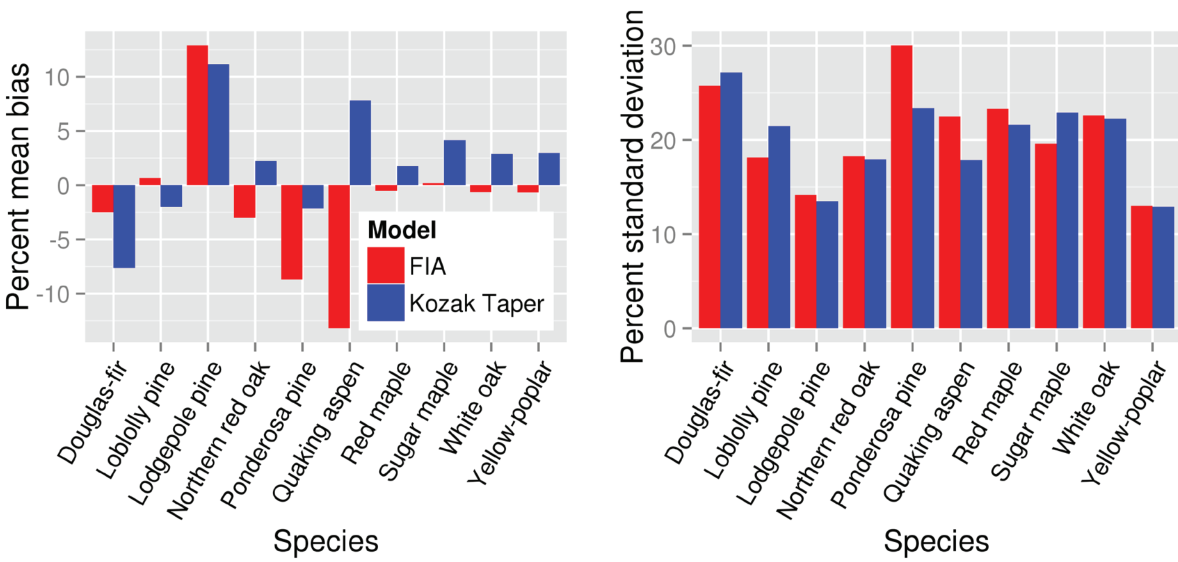


Figure 2—Comparison between the current FIA and the Kozak taper merchantable volume estimates in terms of mean bias and standard deviation for 10 prominent species.

Examining 10 species, the Kozak taper function improved precision for all species except loblolly pine (*Pinus taeda* L.), Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco), and sugar maple (*Acer saccharum* Marsh.). In terms of mean bias, the FIA approach predicted better for species such as loblolly pine and Douglas-fir. However, the taper function approach improved estimates for species such as ponderosa pine (*Pinus ponderosa* Dougl. Ex Laws.) and quaking aspen (*Populus tremuloides* Michx.).

Discussion

The results presented here suggest that a broad-scale mixed species model may perform as well as separate models for species and region. However, the taper function used a hierarchical fitting approach on new data to develop new coefficients, while the FIA models that we verified were not refit and relied on previously fit coefficients generated using least squares regression. Since the taper function was fit to the verification dataset, subsequent analyses will require cross validation or bootstrapping to improve upon the within taper model verification presented here. We also note that for a single prominent species (e.g., Douglas-fir for which the FIA model performed best), a species-level equation may still be more appropriate. One additional limitation in this analysis is the relatively limited data for certain species and regions. Their influence will be similar to gaps described in biomass sampling efforts (Weiskittel et al. 2016b).

References

Bailey, Robert G. 1983. Delineation of ecosystem regions. Environmental Management. **7**: 365-373.

Li, Rongxia; Weiskittel, Aaron; Dick, Adam R.; [et al.]. 2012. Regional stem taper equations for 11 conifer species in the Acadian region of North America: Development and assessment. Northern Journal of Applied Forestry. 29(1): 5–14.

Radtke, P.J.; Walker, D.M.; Weiskittel, A.R.; [et al.]. 2016. Legacy tree data: A national database of detailed tree measurements for volume, weight, and physical properties. In: Stanton, S.M.; Christensen, G., eds. General Technical Report PNW-GTR-931. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 25-30.

Radtke, Philip J.; Walker, David M.; Frank, Jereme, Weiskittel; [et al.]. 2017. Improved accuracy of aboveground biomass and carbon estimates for live trees in forests of the eastern United States. Forestry.90(1), pp. 32-46.

Weiskittel, A.R.; Frank, J.; Walker, D.M.; [et al. 2016a.]. Advancing individual tree biomass prediction: Assessment and alternatives to the component ratio method. In: Stanton, S.M.; Christensen, G., eds. General Technical Report PNW-GTR-931. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station125-132.

Weiskittel, A.R.; Frank, J.; Walker, D.M.; et al. [2016b.]. Gaps in sampling and limitations to tree biomass estimation: A review of past sampling efforts over the past 50 Years. In: Stanton, S.M.; Christensen, G., eds. General Technical Report PNW-GTR-931. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 31-35.

Woodall, C.W.; Heath, L.S.; Domke, G.M.; Nichols, M.C. 2011. Methods and equations for estimating aboveground volume, biomass, and carbon for trees in the U.S. forest inventory. Gen. Tech. Rep. NRS-88. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 30 p.

Number of Points Needed for Image-Based Change Estimation

Tracey S. Frescino[1] and Paul L. Patterson[1]

Abstract—The Image-based Change Estimation (ICE) project is an ancillary data collection project to enhance the USDA Forest Service Forest Inventory and Analysis (FIA) program's field-based inventory for monitoring changes on our Nation's lands. The ICE is designed to capture change at a more frequent interval than FIA's current remeasurement schedule by taking advantage of the USDA Farm Service Agency National Agriculture Imagery Program (NAIP)'s frequently acquired, high-resolution digital photography. The ICE sample design includes a point sample grid of 45 points within an approximate 1-acre circle surrounding each FIA plot center. All 45 points are interpreted on plots where a change is observed between 2 different years, and only 5 points are interpreted on plots where no change is observed. We examine the consequences of reducing these 5 points to 1 point for characterizing land use at the state and county level.

Keywords: photo-based interpretation, change estimation, land use, land cover

Introduction

The Image-based Change Estimation (ICE) project is an alternative data collection mechanism to enhance the USDA Forest Service, Forest Inventory and Analysis (FIA) program's monitoring of our Nation's lands. ICE emerged in response to several requests for FIA to expand its capacity to produce timely estimates of land use, land cover, and agent of change. ICE is designed to capture change at a more frequent interval than FIA's current remeasurement schedule by taking advantage of the extensive and freely available resource of digital ortho photography provided by the USDA Farm Service Agency National Agriculture Imagery Program (NAIP). The NAIP acquires high-resolution aerial imagery across the continental United States during the peak growing seasons, every 2 to 3 years (https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/).

[1] USDA Forest Service, Rocky Mountain Research Station, Forest Inventory and Analysis, 507 25th St, Ogden, UT 84401

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

The current ICE methods and estimation strategies were developed based on a photo-based inventory pilot for the state of Nevada (Frescino et al. 2009; Patterson 2012). The ICE sample design includes a point sample grid of 45 points within an approximate 1-acre circle surrounding each FIA plot center. Interpreters observe all FIA plots during the growing season in 2 separate years and determine if a change has occurred within the photo-plot extent. If there is no change, 5 of the 45 points are classified with a land use and land cover class. If a change is observed, all 45 points are classified with a land use and land cover class as well as an agent of change. In the interest of improving efficiency and reducing cost, we consider here whether there is sufficient information gathered by interpreting only 1 point instead of 5 points for plots where no change occurred. We specifically examine the consequences for the amount of information captured at the state level and county level and for the precision of state-wide population estimates.

Data and Methods

For each FIA plot location, using 1-meter resolution aerial photography, a 144 foot (1.449 acre) co-located photo-plot is established on both the Time 1 (T1) and Time 2 (T2) images. At each plot location, the T1 and T2 areal images are compared to see if any change has occurred. If change has occurred, 45 point locations are interpreted on both T1 and T2 plots (fig. 1). For each point, a land use and land cover is recorded.

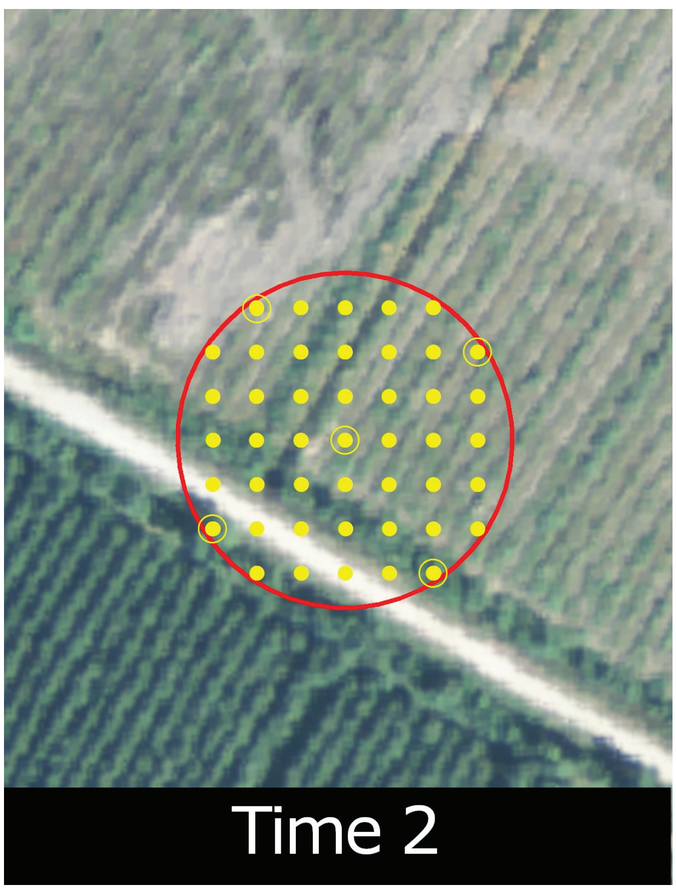


Figure 1—Shows the 5 points (with circles around them) that are a subset of the 45 points. When interpreting only 1 point, the center point is used.

For points where change is observed, an agent of change is recorded. If no change is observed on the overall plot, a subset of 5 of the 45 (fig. 1) points are interpreted for land use and land cover. In this study, we used ICE data from completed states of Utah (UT) with 9,170 plots, New Hampshire (NH) with 992 plots, and Vermont (VT) with 1,037 plots. The dates of the NAIP imagery are 3 years apart for Utah, with T1 acquired in 2011 and T2 in 2014. For NH and VT, the imagery was from 2012 and 2014.

We used the R statistical programming software (https://cran.r-project.org/) to summarize the interpreted data by state and county and an R package, FIESTA, (Frescino et al. 2012) to generate population estimates for all three states. The sample design and associated estimators are based on an infinite sampling paradigm where a simple random sample of points is interpreted within a support region around each point in an independent stratified sample of plot centers (Patterson 2012). Under the current ICE protocols and methods, estimates are based on the 5 points for the plots with no change and 45 points for the plots with change. In this study, we generate a second set of estimates and associated estimated standard errors by using only one point for all plots with no change. The two sets of estimates of percent cover for all land use categories and associated estimated standard errors were compared.

Results

Table 1 shows the percent difference between the number of plots containing each land use category using 5 points compared to using only 1 point. In all three states, the number of plots that included Non-census Water and Right of Way land use categories increased by more than 50 percent when 5 points were interpreted. In New Hampshire and Vermont, the number of plots on which Wetland/Riparian, Cultural classes, and Farmland categories were observed

Table 1—The percent difference between the number of plots containing each land use category using 5 points compared to using only 1 point. Since the 1 point is a subset of the 5 points, the percent difference will always positive. NA values represent classes that were not observed in the state.

Land use	Utah	New Hampshire	Vermont
Forest	6	7	5
Wetland/riparian	21	43	35
Non-census water	59	73	70
Census water	5	6	3
Farmland	9	36	26
Agricultural woody cropland	0	NA	100
Windbreak/shelterbelt	40	100	0
Cultural	14	42	50
Right-of-way	58	68	63
Recreation	8	0	60
Strip mines/quarries/gravel pits	17	0	0
Rangeland	6	NA	NA
Other non-vegetated land	6	NA	NA

increased by 25 to 50 percent when 5 points were interpreted. In Vermont, interpretation of 5 points resulted in a 100-percent increase in the number of plots on which Agricultural Woody Cropland was observed. Little difference was seen for Forest and Census Water categories.

At a county level, figure 2 shows the difference between the number of unique categories captured using 5 points, and the number of unique categories from data based on just 1 point. In both Utah and Vermont, the use of data based on 1 point instead of 5 points resulted in missed land use categories in roughly two-thirds of the counties; in New Hampshire, one third of the counties had a difference. The categories missed for all three states combined are: Right of Way, Wetland/Riparian, Census Water, Windbreak/Shelterbelt, Cultural, Other Non-vegetated Land, Recreation, and Strip Mines/Quarries/Gravel Pits, Farmland, and Agricultural Woody Cropland.

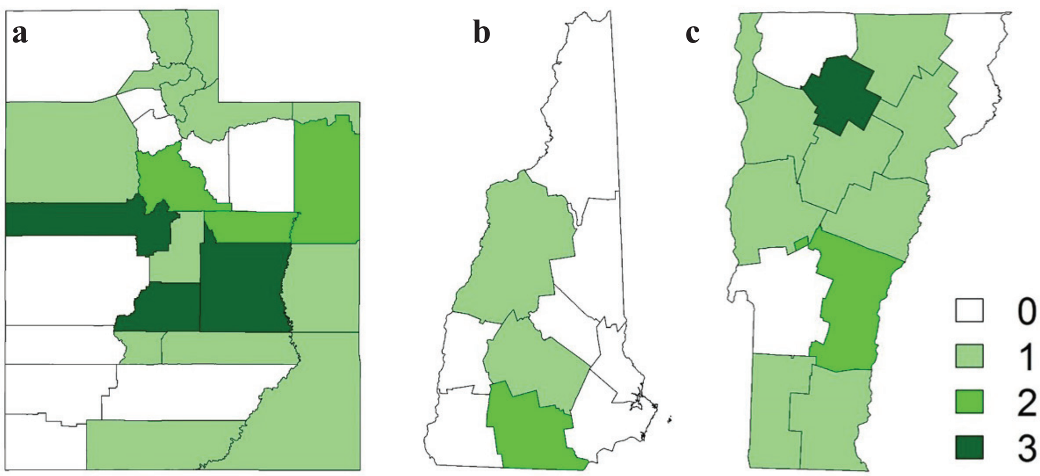


Figure 2—The difference between the number of unique categories from summarized data using 5 points where no change was observed and summarized data based on 1 point where no change was observed: a. Utah; b. New Hampshire; c. Vermont.

Table 2 presents the estimates and standard errors using 5 points and the difference between the estimate and standard errors with 1 point and 5 points for Utah and New Hampshire. A positive difference means the estimate for 1 point was larger and similarly for the standard error. The estimates generated by 5 points and 1 point are close to each other, resulting in small differences ("Est Diff" field in table 2), although for the less prevalent categories, the relative difference is larger. For most of the categories, the difference in the standard errors is non-negative, meaning the estimates using 5 points appear to be slightly more precise. The results for Vermont were similar.

Table 2—The estimates and standard errors of all land for Utah and New Hampshire. “Diff” values represent the difference of the results using 1 point for lands with no change from the results using 5 points. Positive values for SE Diff indicate higher precision for estimates based on 5 points.

	Utah				New Hampshire			
Land Use	**Est 5pt**	**Est Diff**	**SE 5pt**	**SE Diff**	**Est 5pt**	**Est Diff**	**SE 5pt**	**SE Diff**
Forest	38.79	0.02	0.497	0.011	81.74	0.46	1.079	0.125
Wetland/riparian	0.46	0.02	0.065	0.006	2.24	-0.12	0.411	0.045
Non-forest chaparral	0.00	0.00	0.000	0.000	0.00	0.00	0.000	0.000
Non-census water	0.14	-0.04	0.026	0.005	0.35	-0.04	0.097	0.078
Census water	3.13	0.00	0.179	0.002	3.12	0.20	0.533	0.036
Farmland	4.05	-0.03	0.200	0.005	2.71	-0.04	0.446	0.058
Agricultural woody cropland	0.02	0.00	0.014	0.001	0.00	0.00	0.000	0.000
Windbreak/shelterbelt	0.01	0.01	0.004	-0.002	0.02	0.02	0.020	-0.020
Cultural	1.49	0.00	0.121	0.005	7.67	-0.12	0.697	0.130
Right-of-way	0.54	0.01	0.053	0.021	1.69	-0.36	0.259	0.091
Recreation	0.11	0.00	0.034	0.000	0.26	0.04	0.152	0.022
Strip mines/quarries/gravel pits	0.22	0.01	0.048	0.001	0.20	0.00	0.139	0.000
Rangeland	47.51	-0.03	0.517	0.014	0.00	0.00	0.000	0.000
Other non-vegetated land	4.53	0.02	0.214	0.004	0.00	0.00	0.000	0.000

Discussion and Conclusion

The ICE project is moving from prototype to production stage in FIA as an ancillary dataset for monitoring change of land use and land cover. There are several considerations and decisions to make with regards to the balance between the efficiency of the interpretation work and meeting strategic goals for monitoring change of land use and land cover in a timely manner. The interpretation of 5 points versus 1 point for plots with no change between 2 points in time is part of this discussion.

We examined the effect of reducing the number of points interpreted on plots with no observable change on the precision of statewide estimates of land use and, at the county level, the number of land use categories missed to determine the magnitude of the procedure change. In this analysis, we found that some of the more rare and linear land use categories, including Wetland/Riparian and Right of Way, are being missed when using only 1 point. It appears that 1 point is insufficient to characterize the photo-plot. The amount of time-saving achieved by only interpreting 1 point has to be balanced against missing some rare and linear feature attributes.

Our analysis is based on the current sample strategy and does not address the effect of changing the sample strategy through intensification. In the case of intensification, further analysis is needed to determine the information captured and precision gain versus cost of interpreting more plots and the issue of 5 points vs. 1 point would be germane to that discussion.

References

Frescino, Tracey S.; Patterson, Paul L.; Freeman, Elizabeth A.; Moisen, Gretchen G. 2012. Using FIESTA, an R-based tool for analysts, to look at temporal trends in forest estimates. In: Morin, Randall S.; Liknes, Greg C., comps. Moving from status to trends: Forest Inventory and Analysis (FIA) symposium 2012; 2012 December 4-6; Baltimore, MD. Gen. Tech. Rep. NRS-P-105. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. [CD-ROM]: 74-78.

Frescino, Tracey S.; Moisen, Gretchen G.; Megown, Kevin A.; [et al.]. 2009. Nevada Photo-Based Inventory Pilot (NPIP) photo sampling procedures. Gen. Tech. Rep. RMRS-GTR-222. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 30 p.

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

A Forest Grid of México: Building Capacities for the Analytical Synthesis of the Mexican Forest and Soils Inventory

Sergio Armando Villela Gaytán[1], Mario Guevara,[2] mguevara@udel.edu, Julian Equihua[3], Rafael Mayorga Saucedo[1], Carlos Isaias Godinez Valdivia[1], Oscar Ignacio Martinez Diaz[1], Esteban Alberto Suarez Muro[1], Rubi Angelica Cuenca Lara[1], Mayra Ramírez Salgado[1], Carlos Arroyo-Cruz[3], Johny Romero Correa[1], Miguel Angel Rosas Mayon[1], Carina Edith Delgado Caballero[1], Jacinto Samuel García Carreón[1], Raul Rodríguez Franco[1], Juan Carlos Leyva Reyes[1], Enrique Serrano Gálvez[1], Rafael Flores Hernández[4], Salvador Sánchez Colón[5], Rodrigo Vargas[2]

Abstract—We present a national framework for building analytical and institutional capacities for the management and synthesis of the Mexican National Forest and Soils Inventory (Inventario Nacional Forestal y de Suelos de la Comision Nacional Forestal, INFyS-CONAFOR). The main objectives were to 1) improve the efficiency in data management and processing, 2) develop a standardized and curated database based on the best information available, and 3) predict and validate forest variables across non-response sites and inaccessible plots. Based on data-sharing among institutions, transparent methods, and the use of open source, we improved data management efficiency. Error measurements were identified based on expert knowledge along with land use/cover type maps. We provided a first curated version of the INFyS-CONAFOR database that included more than 2.5 million registers of more than 100 forest and site variables. This database represents 19,820 plots (and three replicates) systematically distributed across México. Tree height, tree density, and tree diameter were selected to compare different approaches of modeling and to quantify their spatial variability. Gridded environmental information (for example, climate, soils, remote sensing, and topography) was used to predict these variables using machine learning (random forests). Continuous maps were generated across México in a 1x1 km grid to populate non-response sites, which make up almost 10 percent of the national inventory.

[1] The National Forest Commission (CONAFOR), Zapopan, Jalisco, México
[2] Department of Plant and Soil Sciences, University of Delaware, Newark, DE
[3] The National Commission for the Knowledge and Use of Biodiversity (CONABIO), México DF
[4] USDA Forest Service, International Programs, Zapopan, Jalisco, México
[5] Natural Resources Advisor, United States Agency International Development, Washington D.C.

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

The explained variance (R^2) varied from 0.54 ±0.04 for tree height, 0.40 ±0.05 for tree density, and 0.27 ±0.05 for tree diameter, suggesting a moderate to low prediction capacity. Therefore, the uncertainties of these maps were also mapped (represented by the standard deviation of the full conditional distribution of the response variables to its predictors). These maps revealed spatial knowledge gaps that will inform future sampling strategies and modeling approaches. The results will improve accuracy and spatial detail in future predictions, while informing decision support systems for the protection and conservation of Mexican forest resources.

Keywords: Mexican National Forest and Soils Inventory, INFyS-CONAFOR database, Mexican forest resources

Introduction

To support land degradation mitigation strategies and combat climate change, there is an increasing demand for information about forest distribution and dynamics. We present a data-driven approach to support the establishment of a measurement, reporting, and verification system of forest resources. In México, the INFyS-CONAFOR (Inventario Nacional Forestal y de Suelos de la Comision Nacional Forestal, or the Mexican forest and soils inventory) is the main source of information about the variability and structure of forest resources. Dynamic creation of appropriate organizational knowledge (Nonaka 1994) is required to continuously translate tacit knowledge into explicit knowledge and to facilitate the development of institutional capacities for the management and synthesis of INFyS-CONAFOR data. Organizational learning will enhance CONAFOR's capacity to provide the best information available on Mexican forests and to avoid interoperability barriers (Vargas et al. 2017) to enable the monitoring of forest conditions and carbon storage capacity.

This study focuses on the development and synthesis of a curated version of the INFyS that would be useful for nationwide modeling efforts. Forest reports would aid in policy decisions about Mexican forest resources (for example, carbon and water). Predictive modeling could address the increasing levels of inaccessibility to sampling plots due to their remote locations and land ownership issues or insecurity. It is important to fill these information gaps because they can lead to differences in estimates of forest carbon stocks and services (Domke et al. 2014).

The main objectives here are to: 1) improve the efficiency in data management and processing levels of INFyS-CONAFOR, 2) develop a standardized and curated database based on the best information available, and 3) predict and validate forest variables across non-response sites and inaccessible plots. This effort will soon provide a user friendly and curated database that can be used for several applications (for example, environmental assessments). The new knowledge will strengthen CONAFOR's capacity to provide estimates in places where no information is available at reasonable time frames and spatial resolution.

Methods

All processes were migrated to open source code (R Core Team 2016). All methods were shared and previous versions of databases were archived. These efforts were based on principles such as data-sharing and transparent methods. The raw databases were cleaned by forest experts from CONAFOR, and finger errors and/or unreliable observations were removed. Discussions among experts were required to define maximum and minimum acceptable values. Databases were standardized, with consideration of the rare extreme but reliable values. Environmental predictors for forest variables were gathered and harmonized to a 1x1 km regular grid. We used climate information (precipitation and temperature, Zavala et al. 2010), the MODIS enhanced vegetation index (Broxton et al. 2014), soil organic carbon and clay content (Hengl et al. 2017), a soil depth map (Guerrero et al. 2014), and a USGS geological age map provided by the WorldGrids initiative (http://worldgrids.org/doku.php/start). Tree height, tree diameter, and tree density were modeled using a quantile regression form of random forests (Meinshausen 2006), which is a data-driven method that allows for the estimation of uncertainty. The uncertainty in this case is represented by the range of the full conditional distribution of each variable as a function of the aforementioned environmental factors. Continuous maps of these variables were estimated using the mean and median value of this distribution. Because these were calculated by pixel, it made implementation a computational challenge. Therefore, the range of the distribution estimated for each pixel captures the uncertainty of the estimate.

Results

We provide forest data and forest covariates. A first curated version of the INFyS-CONAFOR database includes more than 2.5 million registers of more than 100 forest and site variables. A total of 2.8 million trees were measured between 2004 and 2014. This database represents 19,820 plots (and three replicates) systematically distributed across México. We also provide a set of environmental factors to contextualize the spatial variability of forest structural variables and for further synthesizing applications and ecological interpretations.

We provide models and spatial predictions of forest structural data in places where no information is available (for example, non-response). The internal out-of-bag validation of the quantile regression forest showed that the explained variance (R^2) varied from 0.54 ±0.04 for tree height, 0.40 ±0.05 for tree density, and 0.27 ±0.05 for tree diameter, suggesting a moderate to low prediction capacity. According to CONAFOR field experts, these predictions are very useful in different forest scenarios.

Figure 1 shows a visual example of the spatial integration of the three predicted variables. The forest grid (represented by tree height, tree density, and tree diameter) is represented by pixels of 1x1 km. A red, green and blue (RGB) composite was generated to illustrate the spatial variability of forest parameters across México. This figure also shows a close-up of a fragment of the southeastern part of the country to illustrate the spatial detail that was achieved by mapping forest structural variables at the spatial resolution of 1x1 km. The red values indicate a dominance of high diameter values, green values indicate the dominance of a higher tree density, and blue values indicate where the trees are taller. Light areas indicate a higher dominance of higher values of all three variables while darker areas indicate a dominance of lower values. Note that across the southeastern proximity with Central America, the spatial detail is achieved by mapping forest characteristics at a spatial resolution of 1x1 km.

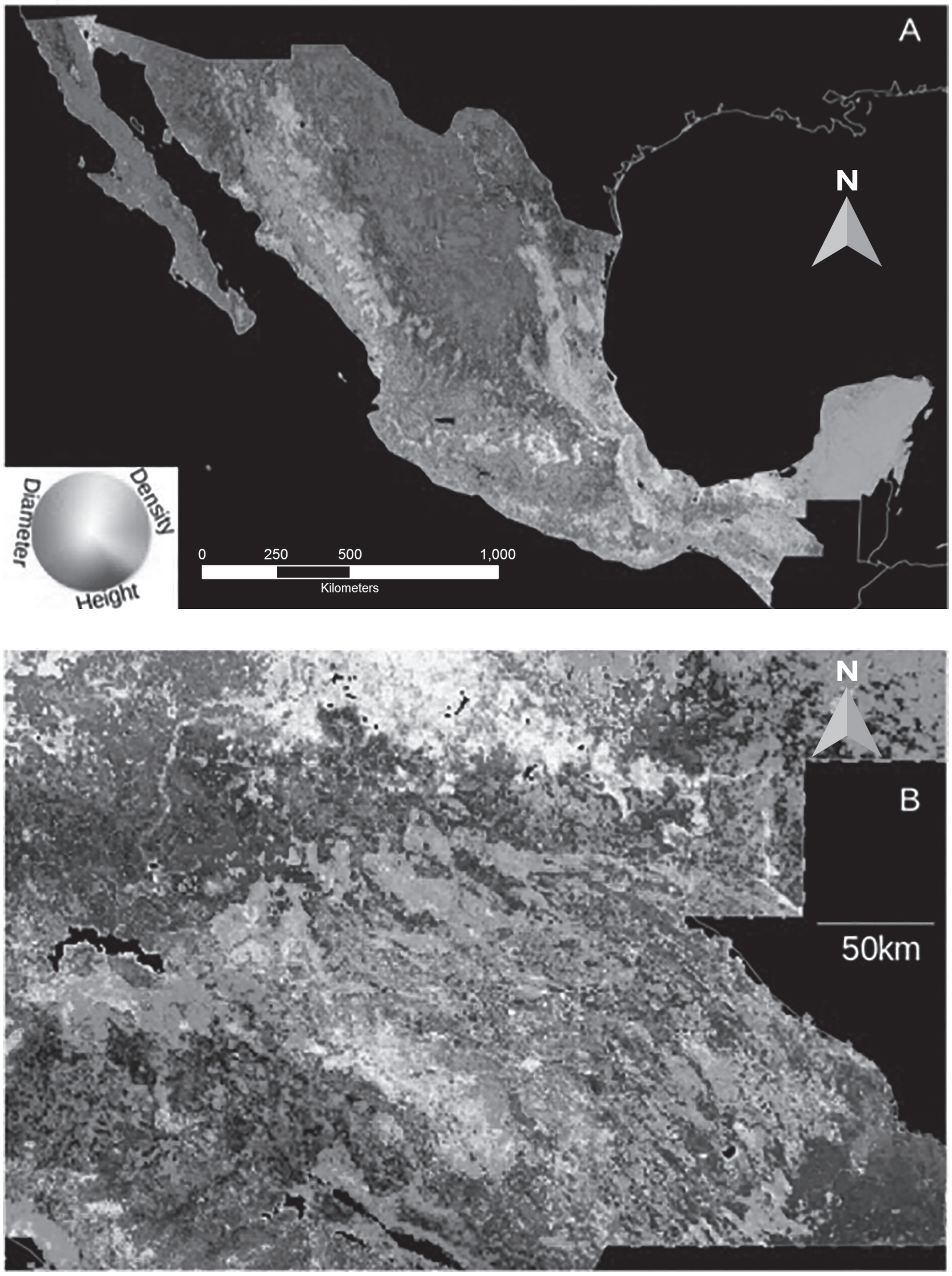


Figure 1—The forest grid of Mexico. This figure shows an integrated visualization of the three mapped variables. An RGB composite was created to illustrate the spatial variability of forest parameters across (a) Mexico, and (b) a close-up of a fragment of the southeast part of the country. Red values indicate a dominance of diameters, green values indicate the dominance of a higher tree density, and blue values indicate where the trees are taller. Light areas indicate a higher dominance of higher values on the three variables while darker areas indicate a dominance of lower values. Note in (b) the spatial detail that is achieved by mapping forest characteristics at 1x1 km of spatial resolution.

We provide a spatially explicit measure of uncertainty for testing the reliability of predictions (fig. 2). These uncertainties are different between maps, suggesting a different spatial variability for each variable. There are large areas of the country where information is less available (for example, in non-forested plots). By separating the uncertainties estimated for each variable, figure 2 better illustrates and delineates well-represented and poorly represented areas.

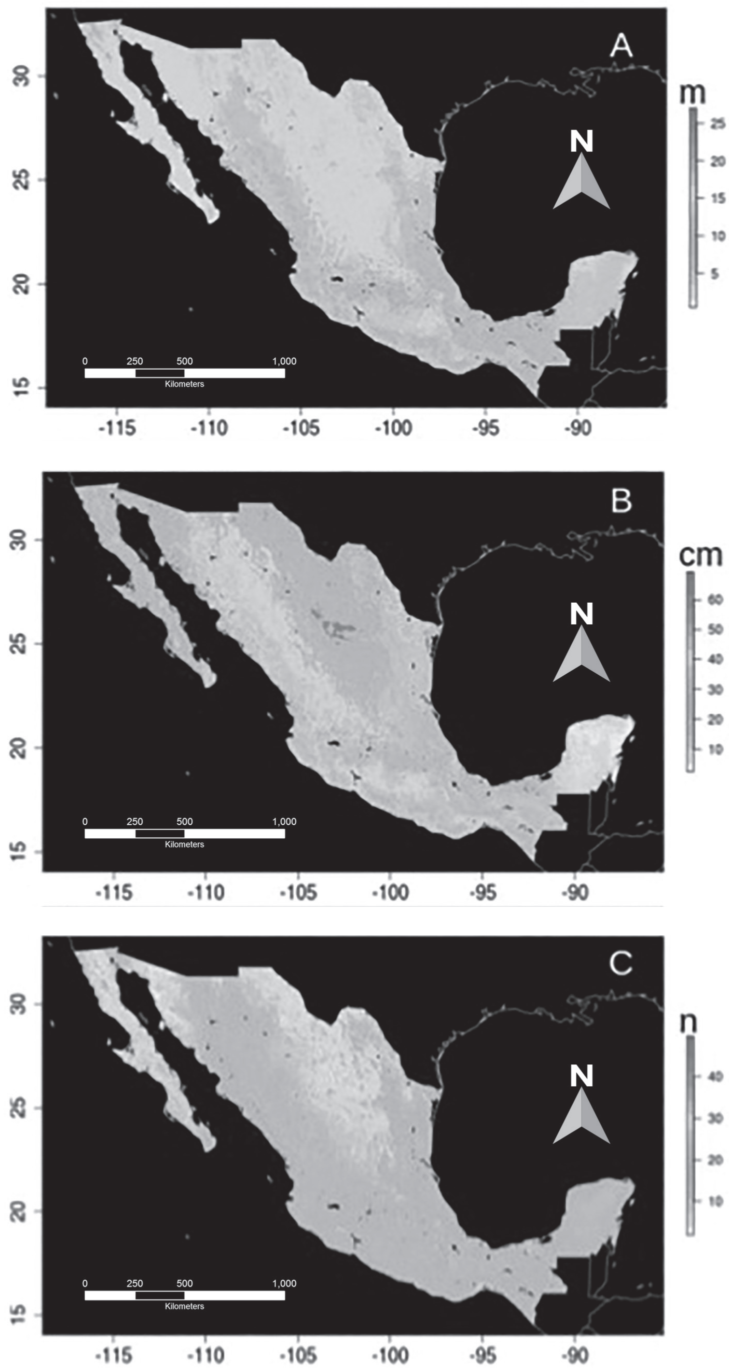


Figure 2—Uncertainty of spatial predictions of forest variables. (a) tree height, (b) tree density, and (c) tree diameter. These uncertainties were derived by estimating the full conditional distribution of these variables as a response of the predictors used. Note that the higher uncertainty of these is not the same on each map. To have spatially explicit measures of uncertainty allows for future planning of sampling strategies.

Discussion

We provide a set of data, methods, and analytical and visualization tools to 1) enhance the data management capacities and synthesis activities of the INFyS and 2) facilitate the generation of socially relevant science about Mexican forests. This is a dynamic effort and it is very likely that the dataset and the national-level predictions will improve and newer versions will be released. We argue that this database represents the best information currently available for validating and increasing the accuracy of national-to-global modeling efforts. It includes land use change (Hansen et al. 2013; Gebhardt et al. 2014), tree density (Crowther et al. 2015), and tree height (Simard et al. 2011). We argue that data-sharing principles and reproducible scientific research led by CONAFOR (assisted by academic groups) will reduce interoperability barriers for successfully mitigating climate change (Vargas et al. 2017).

These principles will also allow for the establishment of a functional and accurate measurement system for reporting and monitoring forest resources. Accurate forest estimates are required for successful implementations of climate change mitigation policies and decision making related to land use evaluation and planning (Saatchi et al. 2011). Country-specific capacities are important for reducing uncertainties from global models. Although our modeling performance is slightly lower than previous efforts for mapping vegetation parameters (Walker et al. 2007; Simard et al. 2011), we provide a spatial measure of uncertainty that will support decisions for improving future field sampling strategies and modeling efforts.

Conclusion

We provide forest estimates and a reproducible framework across México. We present a first official version of the INFyS dataset that includes more than 2.5 million registers of more than 100 forest and site variables from a total of 2.8 million measured trees. We predict forest structural parameters and highlight important levels of uncertainty across large areas of the country. Data-sharing among institutions, transparent methods, and organizational learning activities led by mandated institutions will enhance our capacity to reduce this uncertainty and enable the monitoring of our forest resources.

References

Broxton P.D.; Zeng, X.; Scheftic, W.; Troch, P.A. 2014b. A MODIS-Based 1 km maximum green vegetation fraction dataset. Journal of Applied Meteorology and Climatology. doi: http://dx.doi.org/10.1175/JAMC-D-13-0356.1.

Crowther, T.W.; Glick, H.B.; Covey, K.R.; [et al.]. 2015. Mapping tree density at a global scale. Nature. 525: 201-205.

Domke, G.M.; Woodall, C.W.; Walters, B.F.; [et al.] 2014. Strategies to compensate for the effects of nonresponse on forest carbon baseline estimates from the national forest inventory of the United States. Forest Ecology and Management. 315: 112-120.

Gebhardt, S.; Wehrmann, T.; Ruiz, M.; [et al.]. 2014. MAD-MEX: Automatic Wall-to-wall land cover monitoring for the Mexican REDD-MRV program using all Landsat data. Remote Sensing. 6: 3923-3943.

Guerrero, E.; A. Pérez, A.; Arroyo, C.; [et al.]. 2014. Building a national framework for pedometric mapping: Soil depth as an example from México. In: Arrouays, Dominique; McKenzie, Neil; Hempel, Jon; [et al.], eds. GlobalSoilMap Basis of the global spatial soil information system. Boca Raton, FL: CRC Press. Pages 103-108 doi: 10.1201/b16500-22.

Hansen, M.C.; Potapov, P.V.; Moore, R.; [et al.]. 2013. High-resolution global maps of 21st-century forest cover change. Science. 342: 850-853.

Hengl, T.; Mendes de Jesus, J.; Heuvelink, G.B.M.; [et al.]. 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE. 12(2): e0169748. doi: https://doi.org/10.1371/journal.pone.0169748.

Meinshausen, N. 2006. Quantile regression forests. Journal of Machine Learning Research. 7: 983-999.

Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. Organization Science. 5: 14-37.

R Core Team. 2016. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf

Saatchi, S.S.; Harris, N.L.; Brown, S.; [et al.]. 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. Proceedings of the National Academy of Sciences. 108: 9899-9904.

Simard, M.; Pinto, N.; Fisher, J.B.; Baccini, A. 2011. Mapping forest canopy height globally with spaceborne LiDAR. Journal of Geophysical Research: Biogeosciences. 116: G04021. doi:10.1029/2011JG001708.

Vargas, R.; Alcaraz-Segura, D.; Birdsey, R; [et al.]. 2017. Enhancing interoperability to facilitate implementation of REDD+: Case study of México. Carbon Management. 8: 57-65.

Walker, W.S.; Kellndorfer, J.M.; LaPoint, E.; [et al.]. 2007. An empirical InSAR-optical fusion approach to mapping vegetation canopy height. Remote Sensing of Environment. 109: 482-499.

Zavala-Hidalgo, J.; Fernández-Eguiarte, A.; Romero-Centeno, R.; Zavala-Romero, O. 2010. Digital atlas of Mexico provides accessible climate information. Eos. 91(14): 125-132. doi: 10.1029/2010EO140001.

Quality Assurance in National Forest Inventories: Lessons Learned From International Partnerships

Sara A. Goeking[1], Kerry Dooley[2], Heather L. Hayden[3], Dana Lambert[4], and Andrew Lister[5]

Abstract—National forest monitoring systems ensure data quality by means of quality assurance and quality control (QA/QC) programs, which strive not only to maintain high data quality, but also to provide objective assessments of data quality in reports and estimates. In response to international incentives to monitor forest carbon, numerous countries have recently implemented or are currently implementing, national forest monitoring systems that include QA/QC processes to varying degrees. Forest monitoring specialists from throughout the U.S. Forest Inventory and Analysis (FIA) program have provided technical assistance regarding QA/QC processes in several countries. This paper summarizes lessons learned from international technical transfer, where these lessons can be applied not only in other countries, but also within the FIA program.

Keywords: forest inventory, international technical transfer, quality assurance, quality assessment, quality control

[1] USDA Forest Service, Rocky Mountain Research Station, Forest Inventory and Analysis, 507 25th Street, Ogden, UT 84401, sgoeking@fs.fed.us

[2] Forester, USDA Forest Service, Southern Research Station, Forest Inventory and Analysis, 4700 Old Kingston Pike, Knoxville, TN 37919, kjdooley@fs.fed.us

[3] USDA Forest Service, Pacific Northwest Region, Forest Biometrics Program, 1220 SW 3rd Avenue, Portland, OR 97204, hhayden@fs.fed.us

[4] Supervisory Ecologist, USDA Forest Service, Rocky Mountain Research Station, Forest Inventory and Analysis, 216 N. Colorado St., Gunnison, CO 81230, dlambert@fs.fed.us

[5] USDA Forest Service, Northern Research Station, Forest Inventory and Analysis, 11 Campus Boulevard, Suite 200, Newtown Square, PA 19073, alister@fs.fed.us

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

Introduction

The goal of quality assurance in national forest inventories (NFIs) is to maximize data quality to the extent that is economical and logistically feasible (FAO 2015) and to provide transparency in reporting (Penman et al. 2000). Quality assurance and quality control (QA/QC) processes allow quantification of uncertainties in measurements of forest attributes and suggest areas of improvement in NFI documentation, training, field protocols, automation, or analysis.

The United Nations' Reducing Emissions from Deforestation and forest Degradation (REDD+) program has motivated several countries to implement NFIs that include QA/QC processes (UN-REDD 2017). At the request of the U.S. Forest Service's International Programs, several Forest Inventory and Analysis (FIA) specialists have collaborated with NFI partners in other countries as they develop QA/QC programs. This paper seeks to summarize challenges and lessons learned from these outreach partnerships, including those that may apply to the U.S. NFI.

QA/QC Challenges

Many NFIs implement QA/QC programs as linked yet separate processes from data collection and data flow, and thus support for QA/QC is often sought after NFI data collection has begun. While this sequence of steps may allow for post-hoc quality assessment of field data quality, it neither integrates quality control within all NFI processes nor provides necessary feedback to NFI documentation and training. Given that most countries strive to not only estimate current forest attributes but also quantify change over time, it is important to consider that the lessons learned during first-cycle data collection can improve data quality in future inventory cycles (or even in latter parts of the first cycle).

While QA/QC is often implemented as a feedback to NFI data collection, quality control processes are relevant to all phases of an NFI, including planning, implementation, feedback, and reporting. The planning phase includes development of a field manual and possibly a QA manual. The implementation phase includes field testing and training; logic checks that occur during data collection if a digital data recorder is used; data collection and analysis from QA/QC verification plots; and validation that databases, analyses, and estimation procedures are operating as planned. The feedback phase consists of using the information gathered during the implementation phase to improve inventory processes, e.g., by revising the field manual, improving training, or modifying the data model or compilation process. The reporting phase typically consists of publication of the results of quality assessment, which represent the uncertainty in NFI data based on quality control measures conducted.

Once QA/QC has begun, the next challenge is to identify sources of error and thus suggest process improvements that may minimize future errors. A prime example of the need to close the feedback loop between quality assessment and field data collection is the challenge in relocating plot centers and individual trees, not only during blind checks, but also for future remeasurement of permanent plots. In some countries, asking the question, "Why is it difficult to relocate plots and individual trees?" has led to identification of unclear protocols in the field manual and crew training, which in turn has led to manual revisions, improved training, and modified compliance standards for plot and tree location data. In this scenario, protocol improvements might include: 1) using durable, non-rotting markers for plot centers (e.g., concrete or rebar); and 2) clearly specifying the point on each tree--particularly on plot reference trees--to which crews should measure distances and azimuths.

A final challenge is the difficulty in planning for QA/QC processes within the analysis and estimation workflows of an inventory in the absence of data. Even when NFI and QA coordinators recognize the importance of integrating QA/QC throughout all inventory phases, actual data often present unforeseen analytical challenges.

Lessons Learned: Application to the U.S. NFI

The lessons learned in other countries also apply to FIA's QA/QC program. First, transparency could be increased by making quality assessment data available in the public FIA database (FIA is currently developing this capability). Second, ongoing national-scale QA reporting (e.g., Pollard et al. 2006; Westfall 2009) would improve upon the consistency and transparency of the U.S. NFI; currently most quality assessments are conducted at the state or regional level (e.g., Gormanson et al. 2017). Finally, a national mechanism to incorporate feedback from national QA results (e.g., quality assessment results provide feedback for targeted field manual revisions and training improvements) may further improve FIA's data quality.

Conclusions

The main lessons learned from international technical transfer work are that QA/QC processes should be integrated early, often, and in all phases of the NFI workflow; that identifying and quantifying errors, and determining where and why they occur, are equally important; and finally, quality control of data storage, analysis, and reporting remain challenging, especially for countries that do not yet have data. Most of these challenges exist not only in first-cycle NFIs, but also ongoing inventories such as FIA.

Acknowledgments

Many thanks to our collaborators in multiple countries, and also to Thomas J. Brandeis, Jeremy Rogers, and Paul Patterson for their insightful comments on this paper.

References

Food and Agriculture Organization (FAO). 2015. Voluntary guidelines on national forest monitoring. http://www.fao.org/forestry/fma/84322/en/. Rome, Italy: United Nations Food and Agriculture Organization. [Accessed 20 April 2017].

Global Forest Observations Initiative (GFOI). 2016. Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and guidance from the Global Forest Observations Initiative, Ed. 2.0. Rome, Italy: Food and Agriculture Organization. http://www.gfoi.org/methods-guidance/ [Accessed 20 April 2017].

Gormanson, D.D.; Pugh, S.A.; Barnett, C.J.; Miles, P.D.; Morin, R.S.; Sowers, P.A.; Westfall, J.A. 2017. Statistics and quality assurance for the Northern Research Station Forest Inventory and Analysis Program, 2016. Gen. Tech. Rep. NRS-166. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 23 p.

Penman, J., Kruger, D., Galbally, I.; [et al.], eds. 2000. Good practice guidance and uncertainty management in national greenhouse gas inventories. Hayama, Japan: Intergovernmental Panel on Climate Change (IPCC) National Greenhouse Gas Inventories Programme Technical Support Unit. http://www.ipcc-nggip.iges.or.jp/public/gp/english/. [Accessed 20 April 2017].

Pollard, J.E.; Westfall, J.A.; Patterson, P.L.; [et al.]. 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.

United Nations Reduced Emissions from Deforestation and Degradation Workspace (UN-REDD). UN-REDD Programme Collaborative Online Workspace. Rome, Italy: United Nations Food and Agriculture Organization. Online: http://www.unredd.net/. [Accessed 20 April 2017].

Westfall, J.A., ed. 2009. FIA national assessment of data quality of forest health indicators. Gen. Tech. Rep. NRS-53. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 80 p.

Developing Remotely Sensed Methods for Estimating Tall Shrub Biomass in Forest and Subalpine Communities: Linking Plot-Level Measures to LiDAR

Eric Lewis-Clark[1], Dr. Roman Dial[1], and Bethany Schulz[2]

Abstract—Tall shrub expansion is a hallmark of recent vegetation change across subalpine, boreal, and arctic ecosystems, offering both conservation challenges and potential resource opportunities. The lack of allometric equations for individual shrubs makes biomass estimates at larger scales difficult. We developed estimates for tall shrub biomass in two areas of south-central Alaska for two tall alder species (*Alnus* sp.). These estimates were based on 1) development of individual shrub allometric models giving dry-weight as a function of simple field measurements; 2) application of the allometric equations to Forest Inventory and Analysis (FIA) design sample plots to estimate subplot-scale tall shrub biomass; 3) development of LiDAR-based models of subplot-scale biomass; and 4) application of the LiDAR models to estimate tall shrub biomass across the landscape. We determined that diameter at root collar (d.r.c.) measurement was sufficient to produce allometric equations and that different species required different equations, but not for the different areas sampled. LiDAR models developed to estimate landscape level biomass estimations worked well for one area, but not the other.

Introduction

Biomass estimates of tall (> 4.5 feet [1.37 m]) shrubs are necessary for accurate carbon accounting in forest and subalpine ecosystems and anticipating shifts in vegetative community dominance. Tall shrub expansion in south-central Alaska has been documented (Dial et al. 2016; Klein et al. 2005). However, methods for including shrub species, which often approach tree size but are not considered "tally" trees in FIA protocols, are poorly developed. Scaling up from the FIA plot to the landscape scale to estimate and forecast shrub biomass is a necessary but undeveloped modeling effort. We provide an example from a well-sampled region with historic aerial imagery and an Alaska Native village forest where shrub biomass could provide a renewable resource.

[1] Alaska Pacific University, Anchorage, AK

[2] USDA Forest Service, Pacific Northwest Research Station, Resource Monitoring and Assessment, Anchorage, AK

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

Methods

We used two study sites in south-central Alaska. One site was near Port Graham on the Kenai Peninsula in the Chugach-St Elias Coastal Rainforest Ecoregion about 20 miles southwest of Homer, Alaska. The other was on Joint Base Elmendorf-Richardson (JBER) in the Cook Inlet Basin Ecoregion about 2-5 miles north of downtown Anchorage. Alders at Port Graham were sampled from subalpine coastal thickets with nearby forest dominated by Sitka spruce (*Picea sitchensis*). Alders (*Alnus* sp.) from JBER were sampled from lowland coastal thickets with nearby forest dominated by white spruce (*P. glauca*) and Alaska paper birch (*Betula neoalaskana*). LiDAR was flown in 2009 across the JBER site with a nominal spacing of 2.9 feet (0.8 m; Dial et al. 2016) and across the Port Graham Site in 2016 at a nominal spacing of 1.1 feet (0.35 m).

We sampled JBER in May and June 2016 from five FIA-style plots with four subplots each. Three subplots lacked tall shrubs, here defined as shrubs with d.r.c. > 0.8 in and height > 4.5 feet. From two of the 20 total FIA-style subplots we destructively sampled a total of 16 Sitka alders (*Alnus viridis* ssp. *sinuata*) and one thinleaf alder (*A. incana* ssp. *tenuifolia*). In October 2016 we destructively sampled 54 Sitka alders and 18 thinleaf alders from six FIA subplots near Port Graham. The d.r.c. and height were measured before the individual tall shrub was cut and weighed. Sample stem “cookies” were cut, weighed, dried, then weighed again to estimate the dry-weight fraction. Previously we found that d.r.c. and height are sufficient variables to identify weight ($R^2 > 0.9$).

Linear models were tested with biomass as the dependent variable, study site, species, height, and d.r.c. as independent variables using R (v 3.3.2). We chose not to apply non-linear fitting because the error distribution was heteroscedastic without log-transformation. While we fit linear models with log-transformed variables, we report the equivalent power functions. To determine if models differed by species, we used a likelihood ratio test (LRT) of two biomass models that each included covariates log-height and log-d.r.c.; however, one model included species in an interaction with each covariate. To determine if study site mattered, we used LRT comparing biomass models for Sitka alder with and without study site. We treated study site as a fixed effect included as an interaction with covariates. To find the best model we used AIC. If ΔAIC < 3 when comparing models, then we selected the model with fewer variables, because such a small difference in AIC suggests similar likelihoods.

In a preliminary analysis, we applied the allometric equations to estimate tall shrub biomass for 753 alders among 20 FIA-style subplots, each of area 1,809 feet^2 at JBER and 862 alders among 18 subplots at Port Graham. After summing the total tall shrub biomass in each subplot, we regressed biomass against standard LiDAR metrics to best predict the tall shrub biomass in each subplot. We refer to subplots used to construct models as “training subplots.” For each training subplot we calculated a handful of likely candidate LiDAR metrics: 95 percent height, cover above one meter, cover above 0.5 meter, cover above the mean height, and mean height less than 23 feet (maximum alder height in training subplots). We used a backwards stepping model approach to find a combination of these metrics to best explain variability in biomass among the training subplots. We did this separately for JBER and Port Graham. Using the best fit landscape model, we first predicted the total tall shrub biomass across all pixels with LiDAR at a patch size of 1,819 feet^2 (13 m pixel), as this was the approximate area of the training subplots, then aggregated across pixels by summing over a 5 pixel by 5 pixel area (1.044 acres) to arrive at an estimate of shrub biomass in pounds (lbs) acre^{-1}.

Results

We found significant differences between the allometric model predicting log-biomass that included species as a fixed factor interaction term and one that did not in (LRT, $X^2 = 84.8$, df = 6, $p < 0.001$), suggesting different allometric models for each species. There was no significant difference between biomass models for Sitka alder with and without study site as an interaction term (LRT, $X^2 = 2.3$, df = 3, p = 0.51); similar results held when including study site as an additive term (LRT, $X^2 = 0.6$, df = 1, p = 0.43), suggesting one allometric equation for biomass is sufficient for Sitka alder at both study sites. Insufficient sample size for thinleaf alder at the JBER site prevented a similar comparison. The addition of the height variable did not substantially improve the model for either species over the single variable d.r.c.-model (table 1).

Table 1—Allometric equations for pounds of biomass by species using DRC (inches) with and without height (h) (feet), and their R^2 and ∆AIC values.

Species	Equation	R^2	∆AIC
Sitka alder	$B = 0.86\ DRC^{2.41}$	0.92	2.5
(n = 70)	$B = 0.33\ DRC^{2.31}h^{0.42}$	0.93	0.0
thinleaf alder	$B = 0.033\ DRC^{2.11}$	0.88	2.4
(n = 19)	$B = 0.028\ DRC^{1.63}h^{1.31}$	0.90	0.0

Overall, subplots at JBER had less tall shrub biomass and more variability than subplots at Port Graham (JBER: $\overline{B}$ =15,924 lbs acre^{-1}, sd = 16,019 lbs acre^{-1}, n = 17 subplots; Port Graham: $\overline{B}$ = 20,2548, lbs acre^{-1}, sd = 9,387 lbs acre^{-1}, n = 19 subplots). Moreover, we found that LiDAR better predicted tall shrub biomass at JBER than at Port Graham (table 2) and did so with two variables: cover and log-transformed mean height 3.28 feet (1 m) above ground level. In contrast, Port Graham used only mean height below 23 feet. By applying the JBER LiDAR-biomass equation to forested regions across 40,000 acres of forest land on Joint Base Elmendorf-Richardson and averaging, we arrived at a mean point estimate of 12,856 lbs acre^{-1} of shrub biomass with standard deviation across all pixels of 11,550 lbs acre^{-1}.

Table 2—Models for estimating tall shrub biomass in pounds per 1,819.1 feet^2 pixels using LiDAR.

Study site	Landscape biomass model	R^2	Training sample size (subplots)
JBER LiDAR Estimation	$B = 4{,}220.7 – 1{,}580.1 \ln H_1 + 1{,}491.1C$ where H_1 = mean height above 3.28 feet in feet C = fraction representing the number of first returns in a 3.28 ft resolution LiDAR higher than 3.28 ft out of all pixels.	0.82	$n = 20$
Port Graham LiDAR Estimation	$B = 44.67 + 84.48H_{22.96}$ where $H_{22.96}$ = mean height below 22.96 ft.	0.59	$n = 18$

Discussion

Tall shrubs have been documented as increasing in cover and abundance in south-central Alaskan landscapes (Dial et al. 2016). Tall shrubs may also be increasing in forests; however, to date we are unaware of any study to address this phenomenon. In particular, FIA-style plots are not currently well-suited to detect such an increase.

We found that d.r.c. > 0.8 in generates satisfactory allometric equations for estimating tall alder biomass in two tall shrub species common in south-central Alaska. Models fit young, small alders well; however, as alders age their history of damage and growth reduces model predictive ability. We also found that while species specific equations were necessary, the two study areas 250 miles apart and in adjacent but different ecoregions did not differ in their allometric equations when estimating biomass as a function of d.r.c. for Sitka alder. In addition, we found that these allometric equations could be applied to an alder census in FIA-style subplots to estimate total tall alder biomass. Total alder biomass regressed against LiDAR metrics using linear models provided a model applied at the landscape scale. Models differed between the two study areas, one of which was a lowland forest and the other a subalpine upland. We applied the lowland forest model to estimate mean biomass density of tall shrubs across 40,000 acres of forest.

Acknowledgments

Research was supported by the Forest Service and Alaska Pacific University. We thank Nathan Lojewski, Shelby Faulconer, Bobbi Sue McMullen, Dannielle Malchoff, and Naomi LeeAnn McMullen.

References

Dial, R.J.; Smeltz, T.S.; Sullivan, P.F.; [et al.]. 2016. Shrub-line but not treeline advance matches climate velocity in montane ecosystems of south-central Alaska. Global Change Biology. 22:1841-1856.

Klein, E.; Berg, E.; Dial, R. 2005. Wetland drying and succession across the Kenai Peninsula lowlands, southcentral Alaska. Canadian Journal of Forest Research. 35:1931-1941.

Automated Change Detection With High-Resolution Image Chips in Support of Image-Based Change Estimation (ICE)

Greg C. Liknes[1] **and James Garner**[2]

Abstract—The USDA Forest Service Forest Inventory and Analysis (FIA) program and the National Forest System have implemented Image-based Change Estimation (ICE) to improve FIA estimates of land use and land cover change. The ICE project involves human photo-interpretation of permanent locations on a design-based sample grid of FIA plots. Interpreters use high-resolution aerial photography from the National Agriculture Imagery Program (NAIP). Because such a small proportion of the landscape experiences change in any given year, huge gains in efficiency could be realized by screening plots for potential changes in an automated fashion. We tested four approaches to change detection (mean square error, normalized root mean square error, structural similarity index, and hashing) using NAIP image pairs (2012, 2014) for FIA locations in Vermont. Results indicate that none of the methods result in clear differences between changed and unchanged locations. However, additional investigations into change detection approaches are warranted because of the large potential cost savings to the FIA program. We describe limitations of current approaches and imagery, and we recommend additional analyses be conducted whether using NAIP or other high-resolution imagery.

Keywords: machine vision, structural similarity index, change detection, perceptual hashing, land use change, land cover change

Introduction

With Landsat imagery freely available via the internet and increased access to high-performance computing, there has been a proliferation of forest canopy disturbance mapping over large geographic extents (e.g., Hansen et al. 2013). These mapping efforts highlight harvest, fire, and other disturbances and lead to high profile reports of forest loss across the United States. While disturbances do temporarily remove forest cover, the land area impacted by permanent forest land use change is much smaller.

[1] USDA Forest Service, Northern Research Station, Forest Inventory and Analysis, (651) 649-5192, gliknes@fs.fed.us

[2] USDA Forest Service, Northern Research Station, Forest Inventory and Analysis

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

To address ongoing inquiries about changes to land use and land cover, the USDA Forest Service Forest Inventory and Analysis (FIA) program and the National Forest System have embarked on the Image-based Change Estimation (ICE) project.

The project involves human photo-interpretation of permanent plot locations on FIA's design-based sample grid. Interpretation proceeds using high-resolution aerial photography from the National Agriculture Imagery Program (NAIP). The protocol applied at each plot location is estimated to take 10 minutes on average. Summed over hundreds of thousands of plot locations, the ICE project represents a large investment in human image interpretation. Once image interpretation for all plots has been completed, no determination of land use and land cover will need to be made on plots going forward, unless they have experienced land use or land cover change. As such, updates in the ICE project using new imagery should proceed more quickly but would still require that an interpreter examine each new image pair for change.

Because such a small proportion of the landscape experiences change in any given year, huge gains in efficiency could be realized by screening plots for potential changes in an automated fashion. For example, estimates from ICE indicate that less than 3 percent of the Vermont land area experienced change between 2012 and 2014. However, NAIP imagery presents challenges to change detection due to inconsistencies in view angle, collection dates, camera characteristics, and so on. As such, traditional remote sensing change detection methods are not well-suited for use with NAIP. There are, however, a growing number of techniques in the realm of machine vision designed for use with high-resolution video and imagery that compensate for small differences from frame-to-frame or image-to-image.

We examined the prospect of using both traditional and alternative approaches to detect changes in NAIP imagery with the aim of improving the efficiency of the ICE project. Specifically, the overall goal is to develop a change detection method for NAIP image pairs that can identify plots that have experienced change without any omission errors and with as few commission errors as possible. That is, we want to correctly identify all change plots while excluding as many non-change plots as possible. In this first phase of the research, we developed a workflow to process the image chips and to calculate a number of metrics and have evaluated results qualitatively.

Methods

Using NAIP image chips (88 m x 88 m) that were prepared for the state of Vermont from two image dates, 2012 and 2014, we calculated multiple change metrics using four different approaches (fig. 1). The first, mean square error (MSE), is an accumulation of per-pixel differences and is an intuitive, straightforward change detection approach. The second is a slight modification of MSE for which images are first normalized prior to the calculation of root MSE; we will refer to this approach as NRMSE.

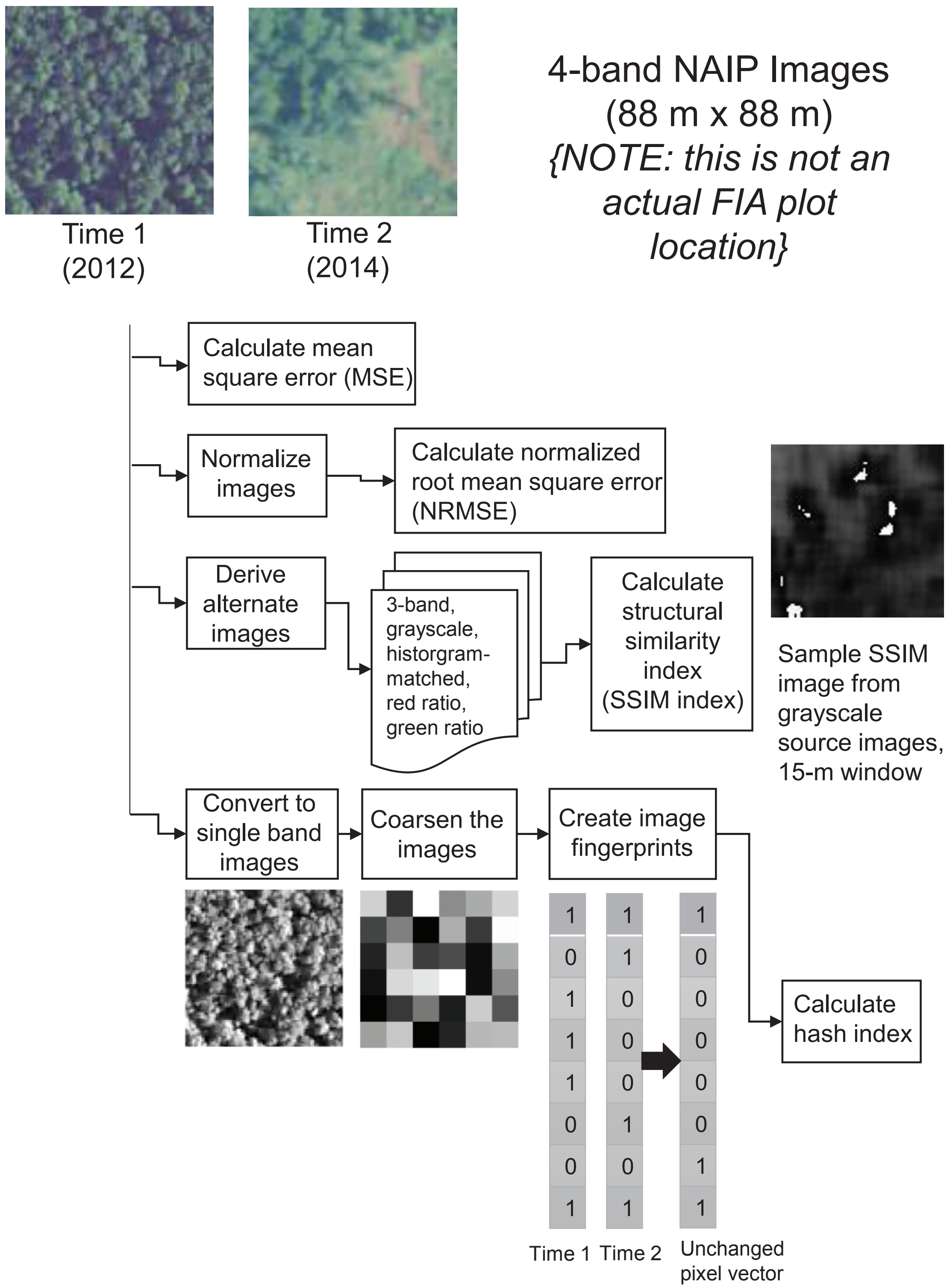


Figure 1—Flowchart for four change detection methods applied to NAIP image chips.

The third approach comes from machine vision literature and was designed to evaluate the quality of an image compared to a reference. The structural similarity index (SSIM index) (Wang et al. 2004) compensates for differences in illumination between image pairs and focuses on contrast and the structure of objects in the image. The SSIM index was calculated for a range of pixel window sizes and using four bands of imagery for each pair, as well as a number of derived images (table 1).

Table 1—Change detection approaches applied to NAIP image chips from Vermont, USA, to find differences between images taken in 2012 and 2014. The first column indicates the statistic or index that was calculated using each window size listed in the second column. An "X" in a cell indicates which band combinations or derived images were used for each approach.

Approach	Window sizes (pixels)	4-band	3-band	Grayscale	Histogram-matched	Red ratio	Green ratio
MSE	1	X		X		X	
NRSME	1	X		X		X	
SSIM index	3, 5, 7, 9, 11, 13, 15, 21, 33	X	X	X	X	X	X
Hashing	3, 5, 7, 9, 11, 13, 15			X		X	

The fourth method, hashing, is derived from the fields of cryptography, multimedia, internet security, and others (see Menezes et al. 1996). Hashing is a technique that involves simplifying an image and developing a "fingerprint" for the image in order to facilitate rapid comparison with a database of images. We created a custom implementation of this concept that involved 1) converting four-band color imagery to a single band; 2) downscaling to a coarser version of the image; 3) tracking each pixel's relationship to its neighbor to the right and also to its neighbor below; 4) comparing the neighbor relationships between time 1 and time 2; and 5) determining the proportion of neighbor relationships that were maintained through time. The downscaling proceeded using a range of window sizes.

For all four approaches, calculations were made using the Python scripting language and the scikit-image library (van der Walt et al. 2014). The approaches are summarized in table 1 along with image details and window sizes.

The values or indices calculated for each approach were then qualitatively compared for locations that experienced change and those without change, as assessed by human image interpretation applying a well-defined ICE protocol. Values were also split out by the interpreted land use class at time 1 (2012) in order to determine if better separation between change and non-change plots could be achieved in particular landscapes. We included only forest, agriculture, and developed land uses since other classes were not land (e.g., water, uninterpretable), were not represented in the sample (e.g., rangeland), or had very small sample sizes (e.g., natural/semi-natural).

Results

Because of the large number of change indices tested, we present only general observations of the results. None of the approaches tested provided good separation between locations that experienced change between 2012 and 2014 and those without change (fig. 2). More specifically, only 0–4 percent of non-change plots fell outside the range of change plots for any given metric, with little difference between simple pixel difference approaches (MSE, NRMSE) and more complicated approaches (SSIM index, hashing). In general, the indices provided better separation between change and non-change plots in agricultural land use compared to those in forest and developed land uses, with the exception of hashing, which provided slightly better separation in forest land use. The number of image bands used had minimal effect; for example, results using four-band or single band grayscale imagery produced similar SSIM index outputs. The window sizes tested for both the SSIM index and hashing did have an impact on results, and in general, better separation between change and non-change locations occurred in middle ranges, with best results at 13-m window size for hashing, and 15-m window size for SSIM, which are the results displayed in figure 2.

Conclusions

The goal of this research is to develop an automated method that can identify all change plots from image pairs while minimizing the number of non-change plots that would need to be viewed by an image interpreter. The proportion of non-change plots eliminated from manual interpretation could be quite small, and a time savings could still be realized because of the comparatively small amount of time required for the automated screening process. (The Python script that calculated all indices for the analysis in this study takes less than 3 minutes to execute for over 1,000 image pairs.)

Because of the properties of the NAIP image collection (inconsistent image acquisition dates, differing view angles, etc.), it is challenging to distinguish actual land change from image changes that are due to image anomalies. These initial results indicate little difference between simple pixel difference approaches (MSE, NRMSE) and more complicated machine vision approaches such as the SSIM index. Because the potential time savings via automated screening is so large, additional investigation into automated change detection approaches should continue. New high-resolution data sources are becoming available, and they may be alternatives to NAIP for change detection.

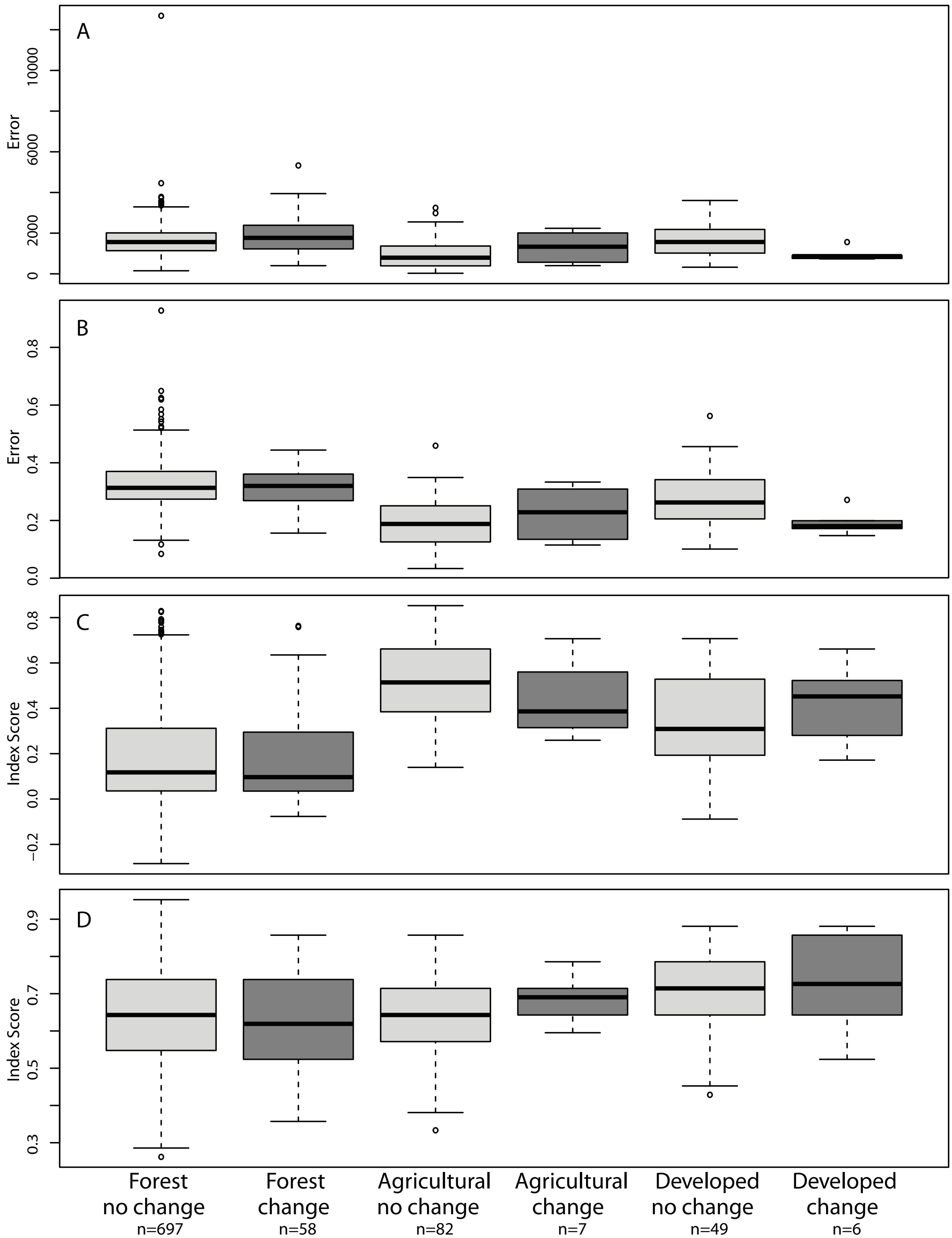


Figure 2—Box-and-whisker plots for four change detection methods applied to image chips from Vermont, USA, for three land use categories (forest, agricultural, developed) for locations with change between 2012 and 2014 (in dark gray) and those without change (light gray). For mean square error (A) and normalized root mean square error (B), lower values indicate more similarity between images. For structural similarity index (C) and hash index (D), higher values indicate more similarity between images. Horizontal dark lines in each box indicate the median value, the boxes span the interquartile range, and the horizontal error bars represent an inner fence at 1.5 times the interquartile range.

References

Hansen, M.C.; Potapov, P.V.; Moore, R.; [et al.]. 2013. High-resolution global maps of 21st-century forest cover change. Science. 342(6160): 850-853.

Menezes, A.J.; Oorschot, P.C.V.; Vanstone, S.A. 1996. Handbook of applied cryptography. Boca Raton, FL: CRC Press, Inc. 780 p.

van der Walt, S.; Schönberger, J.L.; Nunez-Iglesias, J.; [et al.]. 2014. Scikit-image: Image processing in Python. PeerJ. 2:e453. doi: https://doi.org/10.7717/peerj.453.

Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. 2004. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing. 13(4): 600-612.

A Sentinel Satellite-Based Forest Ecosystem Change Detection System

Andrew J. Lister[1] **and Laura P. Leites**[2]

Abstract—Forest monitoring using traditional remote sensing methods requires significant overhead in terms of data management and processing expertise. As an alternative to traditional methods, we propose a new approach to monitoring using point sampling with high-resolution, frequently updated Sentinel satellite imagery. By classifying the probability of different land cover transitions using spectral indices and machine learning techniques at specific locations, analysts can prioritize areas for change monitoring and other activities. The advantages of this approach include the use of existing forest inventory sampling tools and photointerpretation procedures, automated classification, and integration with existing forest inventory monitoring workflows.

Keywords: forest inventory, photointerpretation, Sentinel monitoring, forest change detection, forest monitoring

Background and Objectives

Analysts need forest change monitoring systems that are more cost-effective and timely than existing options. Traditionally, vegetation monitoring has been done at the landscape scale using remote sensing (RS), ground inventory, or a combination. In monitoring that uses RS-based models, satellite imagery is often combined with ground data in a geographic information system (GIS), and models are built based on the relationship between the ground, RS, and other GIS data. The model is then applied at all locations in order to create maps of the resource of interest. Maps can be made from either uni- or multi-temporal imagery. Current multi-temporal RS methods include a suite of options derived from time series of Landsat images (Hansen et al. 2013; Moisen et al. 2016; Schroeder et al. 2014). Resulting map products are used as tools for land management—either as graphical aids that focus attention on areas of interest, or as sources of estimates created by GIS summaries of pixel values for a given geographic area such as a watershed. Olofsson et al. (2013) present methods for producing valid error estimates of these map-based estimates.

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

[2] The Pennsylvania State University, 312 Forest Resources Bldg., University Park, PA 16802

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

There are pros and cons of each monitoring strategy, but certain fundamental principles guide monitoring choices. First, there must be a well-defined criterion for successful monitoring. Generally, success is defined as the generation of information with which a manager feels confident to make a decision, such as estimates of total or average amounts of the forest resource, change, and uncertainty metrics such as confidence intervals. Second, monitoring systems need to provide this information in a cost-effective way. There is little value to collecting more information than is needed to make decisions, and similarly, underestimating data requirements will not lead to successful monitoring. Finally, monitoring systems need to be scientifically defensible. For example, RS maps or tabular inventory outputs produced with designs that violate statistical assumptions provide information that appears useful, but since the underlying design is flawed, that information is not suitable as a basis for management decisions.

Technology associated with the recently launched Sentinel satellites (Drusch et al. 2012 can help meet monitoring needs better than existing RS data such as Landsat. The Sentinel spacecraft are sun-synchronous Earth observation satellites with 13 spectral bands and spatial resolutions of 10 m (four visible and near-infrared bands), 20 m (six red-edge/shortwave-infrared bands) and 60 m (three atmospheric correction bands). They present a great opportunity to make RS-based monitoring, used in the context of point sampling, much more efficient. On the one hand, Sentinel imagery offers a higher resolution than alternatives like Landsat (10- to 20-m vs. 30-m pixels). Secondly, it offers a more frequent revisit time for each location within its orbital path (5 to 10 vs. 14 days), creating more options for obtaining cloud-free views. Finally, Sentinel satellite data, like those from Landsat, are free, and can be accessed via a unique cloud-based image retrieval and processing platform that allows a user to submit requests to a server for a polygon of interest (POI) and retrieve summary statistical information. Using this service and the open source statistical software R (R Core Team 2015), summarized RS data can be easily converted to information like that shown in figure 1, which depicts a temporal distribution of values for a Sentinel-derived vegetation index (NDVI - normalized difference vegetation index) within the indicated POI. This process eliminates the need for downloading and processing imagery, something that introduces significant overhead to forest monitoring tasks and limits opportunities for automation and operationalization of the technology. Sentinel data can thus be very useful for applications in research and monitoring where efficiency and automated procedures are valued.

Including NDVI, there are over 60 vegetation indices provided with the Sentinel Web services, and a total of at least 250 that could be generated (Henrich et al. 2009. We plan to use these Sentinel products, along with ground- and photo-based POIs for model calibration and validation, to create a cost-effective, POI-based change probability monitoring system. For each POI, a vector of pixel-based histogram metrics will be generated, including values for chosen percentiles, measures of central tendency, and variance. Kernel density estimates (i.e., non-parametric, nonlinear models) of the probability density functions of pixel values will also be generated for each sampling location. These data will serve as the predictor variables for change monitoring models. Forest change will be evaluated or modeled in various pilot areas following one of two methods:

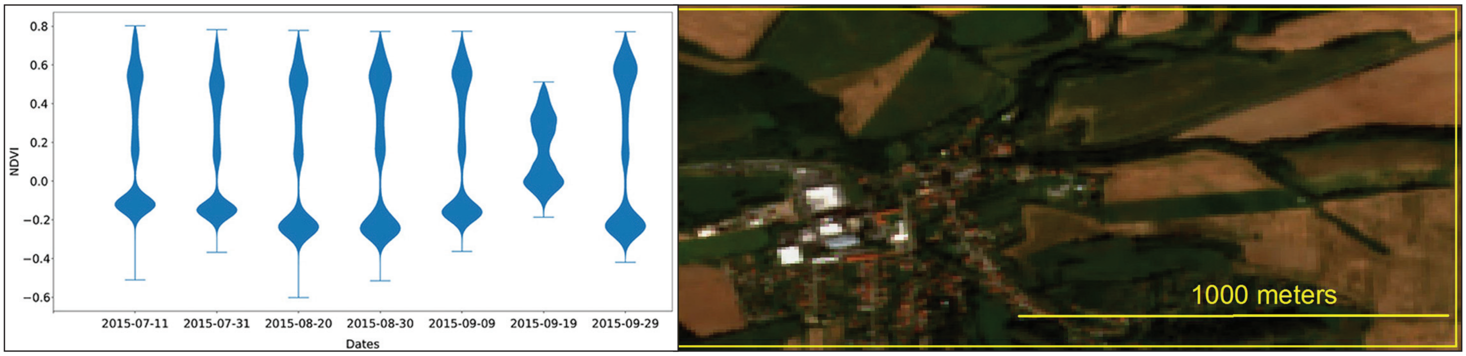


Figure 1—Multi-temporal violin plots representing histograms of normalized difference vegetation index (NDVI) values for the polygonal area depicted. Violin plot from http://www.sentinel-hub.com/apps/fis-request. Right-hand image is from Sentinel natural color image from July 2015.

1) Change detection models: Training data for stables areas and for areas that have changed between 2015-2016 or 2016-2017, will be used to develop a change probability model that will use the vegetation indices and their histogram metrics as predictor variables in a machine learning algorithm such as Random Forests or Support Vector Machines.

2) Land classification models: A sample of locations will be classified into land classes using 2015 as the baseline and will serve to train a model that will predict land class. The vector of histogram metrics for the 60 vegetation indices will serve as predictors in a classification algorithm using a machine learning technique. The model will then be used to predict land class at the second time period (2016). The difference in class probability between the baseline and that from the second time period will be used to label change likelihood. For each of these techniques, traditional error metrics like confusion matrices with user's, consumer's, and overall accuracies will be computed under different change probability threshold scenarios.

The change detection procedures described here are a small subset of the myriad traditional RS change detection methods that could be applied. We will discuss the feasibility of using this technology for change monitoring and for identifying plots that have changed between time periods, as well as various practical ways that forest monitoring systems like FIA and other countries' national forest inventories can use this approach for forest monitoring.

References

Drusch, M.; Del Bello, U.; Carlier, S.; et al. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment. 120: 25-36.

Hansen, M.C.; Potapov, P.V.; Moore, R.; et al. 2013. High-resolution global maps of 21st-century forest cover change. Science. 342(6160): 850.

Henrich, V.; Jung, A.; Götze, C.; et al. 2009. Development of an online indices database: motivation, concept, and implementation. 6th EARSeL Imaging Spectroscopy SIG Workshop Innovative Tool for Scientific and Commercial Environment Applications; 2009 March 16—18; Tel Aviv, Israel. http://www.indexdatabase.de/info/credits.php.

Moisen, G.G.; Meyer, M.C.; Schroeder, T.A.; et al. 2016. Shape selection in Landsat time series: A tool for monitoring forest dynamics. Global Change Biology. 22(10): 3518-3528.

Olofsson, P.; Foody, G.M.; Stehman, S.V.; Woodcock, C.E. 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sensing of Environment. 129: 122-131.

R Core Team. 2015. Vienna, Austria: R Foundation for Statistical Computing. R: A language and environment for statistical computing. https://www.R-project.org/. [Accessed 17 April 2017].

Schroeder, T.A.; Healey, S.P.; Moisen, G.G.; et al. 2014. Improving estimates of forest disturbance by combining observations from Landsat time series with U.S. Forest Service Forest Inventory and Analysis data. Remote Sensing of Environment. 154: 61-73.

The FIA Analytical Hypervolume: Exploring New Dimensions with New Tools

Andrew Lister[1], Bonnie Ruefenacht[2], and Rachel Riemann[3]

Abstract—A New Workflow For Combining Geographic Information Systems (GIS) and forest inventory data from the USDA Forest Service's Forest Inventory and Analysis Program (FIA) will be presented. The workflow exploits a recently released data distribution tool called the Analysis Tool for Inventory and Monitoring (ATIM). The goal of the project is to increase the analytical capabilities of FIA staff and data consumers to make fuller use of the FIA dataset to answer management, planning, and policy questions.

Keywords: forest inventory data analysis, ATIM, Spatial Intersection Tool, GIS analysis

Background and Objectives

The Forest Service's Forest Inventory and Analysis Program (FIA) conducts the National Forest inventory of the United States by gathering information on forest attributes from approximately 300,000 plots. This dataset contains information not only on trees, but also on land type attributes. These land attributes consist of information such as owner class, forest type and type group, topographic variables, geographic location, land cover, and land use. Generally, this information is assigned to plots by the field data collection teams during a ground visit to the plot or, for plots not visited, by photo interpreters who examine the plots using high resolution imagery provided by the National Aerial Imagery Program (NAIP) in a geographic information system (GIS). Data are organized, stored in a corporate Oracle database, and summarized and distributed in a combination of online tools, annual update reports, and comprehensive inventory-cycle reports.

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

[2] USDA Forest Service, Geospatial Technology and Applications Center (GTAC), 2222 West, 2300 South, Salt Lake City, UT 84119

[3] USDA Forest Service, Northern Research Station, c/o USGS, 425 Jordan Road, Troy, NY 12180

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

FIA analysts create many diverse products using the standard FIA database. For example, analysts commonly create cross-tabulation summaries, which can be joined with GIS data layers such as county boundaries that are cross-referenced to county identifiers found in the database. Other examples include tables and figures describing the total number of trees by species by slope class and ownership class, or total forest tree volume by forest type and county. Land type attribute summaries can include total area by land cover class, land use class, ownership class, and all combinations of these. The classification variables used to construct these types of tables and figures are referred to as domain variables (Bechtold and Patterson 2005), which partition the overall estimate into subsets or domains. It is these domain variables that provide richness to the FIA analyses by turning data into information that FIA data consumers can use to make management, planning, and policy decisions.

Over the past two decades, both computing power and the availability of large area GIS datasets have experienced explosive growth. Previously, GIS data layers were distributed in a piecemeal fashion, i.e., county-by-county or state-by-state. This was primarily due to computer processing, storage, and Internet speed limitations. Currently, these limitations have been surmounted and agencies have responded by providing increasingly detailed datasets to users via either direct download or Web services platforms streamed directly into GIS applications. There has not been commensurate growth of exploitation of these new datasets by many Federal agencies due mainly to the challenges associated with integration with legacy database and summary tools.

To bridge this gap, FIA has developed a new data distribution system called the Design and Analysis Tool for Inventory and Monitoring (DATIM) (USDA Forest Service 2017). The DATIM analysis tool, ATIM, allows users, via a simple Web interface, to generate traditional cross-tabulation summary tables as previously described. Something that differentiates ATIM from previous reporting tools is a new utility called the Spatial Intersection Tool (SIT). SIT allows users to intersect plot data with GIS data layers to assign new classification variables to each FIA plot, and then it makes these classification variables available to ATIM for use in table-making. For example, by intersecting the FIA plots with a GIS data layer consisting of a habitat quality category, FIA data users can generate summaries such as “total biomass by species group and habitat quality class,” and use the information to answer questions that inform management decisions. The general workflow for this process is shown in figure 1.

To take full advantage of this new capability, FIA needs standardized, national-scale GIS and remote sensing derived data layers to conduct consistent, multivariate analyses at both regional and national scales. A major challenge, however, is obtaining, preprocessing, and standardizing the ancillary GIS and remote sensing data layers in a way that makes them useful and easily accessible to analysts who are not GIS or database experts. To address this need, FIA and the Forest Service’s Geospatial Technology and Applications Center (GTAC) have begun to assemble GIS data layers that meet the aforementioned requirements. The set of layers, termed the “analytical hypervolume” because it is a multidimensional space within which new analyses can be created, are

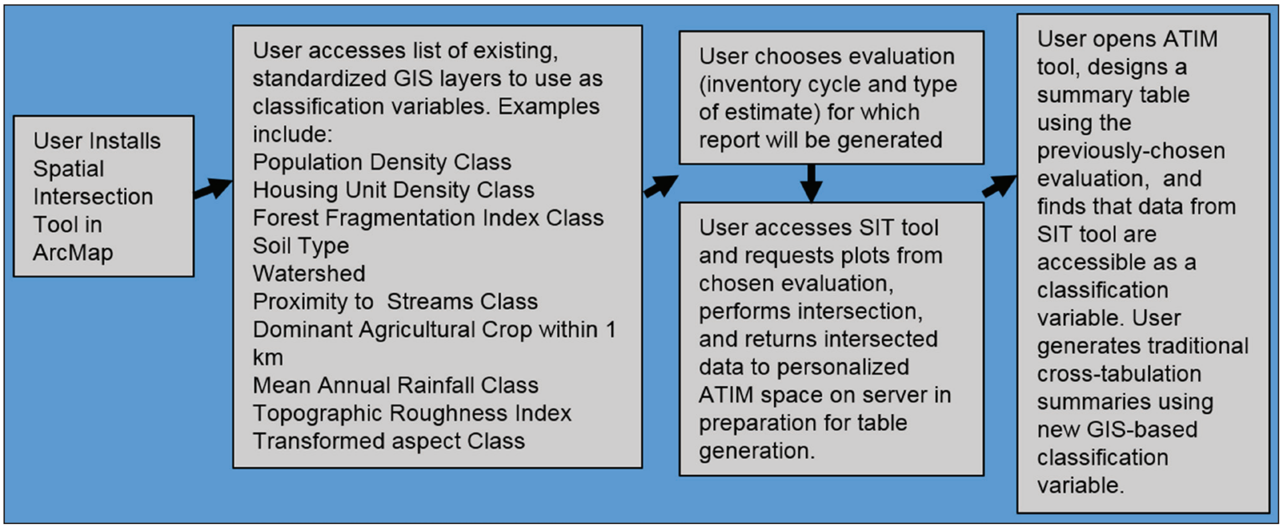


Figure 1—Workflow for use of the SIT and ATIM with GIS data layers.

stored in a corporate Oracle database, giving FIA staff access to a large suite of new categorical variables (or continuous variables that can be broken into categories) to perform different types of analyses. The GIS layers available as of spring 2017 are shown in table 1. In addition, online map services that provide Web coverage services or Web feature services capabilities allow users to stream GIS data directly into GIS tools and thus perform SIT operations without needing local copies of the GIS data. All of these new capabilities will expand creative, national-scale reporting capacity by allowing FIA to use nonstandard GIS data layers, including those from other agencies. Furthermore, it will allow FIA to respond more quickly to emerging national-scale data requests and analysis priorities.

Table 1—List of GIS data layers available as of spring 2017 in the Forest Inventory and Analysis analytical hypervolume GIS dataset.

Dataset	**Description and reference**
Population	Landscan population data, which are disaggregated population estimates derived by Oak Ridge National Laboratory (http://web.ornl.gov/sci/landscan/landscan_documentation.shtml)
Soils	Gridded Soil Survey Geographic Database (gSSURGO), consisting of summarized soil attributes from the NRCS Soil Survey Geographic Database (https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2_053628)
Topography	Topographic data based on 30-m DEM, including aspect, cosine of aspect, sine of aspect, and slope in degrees
Cropland	The Cropland Data Layer, depicting agricultural land cover types of the United States (http://nassgeodata.gmu.edu/CropScape/)
Fires	Monitoring Trends in Burn Severity (MTBS) burn perimeters for fires > 1,000 acres (in the West) and > 500 acres (in the East), for fires between 1984-2014
Forest Change	Forest cover change classification between 2000-2014, developed by the University of Maryland (http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.2.html)
Fragmentation	Several landscape fragmentation pattern maps derived by staff at the Eastern Forest Environmental Threat Assessment Center (https://forestthreats.org/research/tools/landcover-maps/lcm)
NLCD	The National Land Cover Dataset, including land cover, impervious surface and canopy cover from 2001, 2006, and 2011 (https://www.mrlc.gov/faq_lc.php)

The two goals of the current project are to identify the benefits of this approach to increasing analysis capacity and to develop the SIT workflow within ArcGIS for conducting the analysis. Project components include the distribution of a user guide for the process, the generation of analysis examples, and development of a data management strategy for keeping datasets current.

References

Bechtold, W.; Patterson, P., 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.

USDA Forest Service. 2017. Page last modified 2017 July 10.Washington, DC: U.S. Department of Agriculture, Forest Service. Design and Analysis Toolkit for Inventory and Monitoring (DATIM). https://www.fs.fed.us/emc/rig/DATIM/index.shtml [Accessed 2017 April 1].

Calibration Estimators in South Texas

Joseph M. McCollum[1] and Dennis M. Jacobs[2]

Abstract—The Forest Inventory and Analysis' (FIA's) estimates for South Texas are presented. Various cover types such as water, cropland, urban, and wetland-beach are clearly overestimated. A possible solution is presented involving calibration estimators.

Keywords: calibration estimators, nonresponse

Introduction

The 4th survey unit of Texas is over 26 million acres—larger than 14 of the 50 states, including South Carolina and Kentucky. This unit has a nonresponse rate of 30 percent, the highest in the Southern Station's FIA region, thereby subject to nonresponse bias in the program's estimation procedures. The purpose of this paper is to demonstrate the overestimation errors in FIA's area estimates, which are caused by nonresponse bias in this survey unit, and then to propose a solution.

Figure 1 shows the dominant land use in the neighborhood of each plot as calculated by a Thiessen expansion of the plurality condition of each plot, and table 1 shows estimates of surface area, some diagnostics, and the results of calibration.

Standard errors were computed from FIA's own estimates, under the assumption that standard errors for individual strata are independent. Other nonforest land consists of land that is most likely to mimic forest: orchards, Christmas tree plantations, wildlife openings, recreation, and low canopy tree land.

The estimate for cropland is taken from the Census of Agriculture (U.S. Department of Agriculture 2014). On one hand, this estimate includes "pastured cropland," but on the other hand, FIA's estimate includes "idle farmland" and "general agricultural."

[1] Information Technology Specialist, USDA Forest Service, Southern Research Station, Forest Inventory and Analysis, 4700 Old Kingston Pike, Knoxville, TN 37919, jmccollum@fs.fed.us

[2] Research Forester, USDA Forest Service, Southern Research Station, Forest Inventory and Analysis, 4700 Old Kingston Pike, Knoxville, TN 37919

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

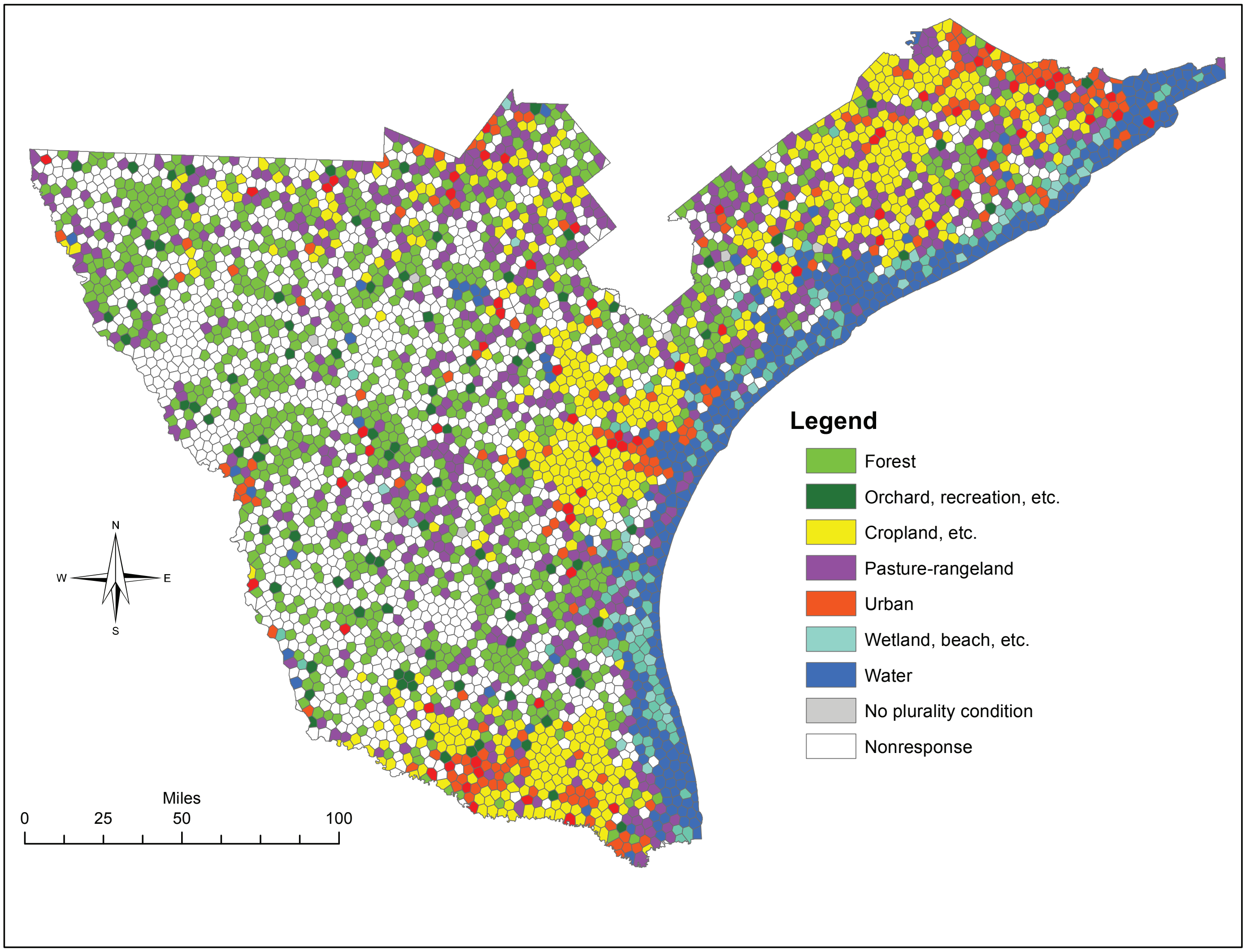


Table 1—Estimates of various land cover for South Texas (in acres).

	FIA estimate	Standard error	Outside source	FIA calibrated
Forest	8,031,773	200,794		9,482,773
Pasture/rangeland	5,633,549	204,355		5,727,545
Cropland, etc.	5,348,733	192,917	4,078,777	4,676,097
Urban	2,423,654	13,4029	2,027,942	2,147,641
Wetland/beach	1,147,055	9,6941		997,686
Other nonforest	783,946	78,630		933,047
Surface water	3,256,897	154,512	2,556,438	2,660,818

The estimate for urban land cover is taken from the Census Bureau's PLACES layer in its TIGER files (U.S. Census Bureau 2012). These are city boundary lines. On one hand, other land types, such as wetland, forest, or water, can occur within a city. On the other hand, urban-like features can occur outside a city—including a right of way or an industrial park in an otherwise rural area. Finally, the estimate for surface water is taken from the Census Bureau's water layer in its TIGER files. The National Hydrography Database gives a similar estimate (U.S. Geological Survey 2009).

For cropland, urban, wetland-beach, and surface water, FIA's estimates are clearly too high. Take a close look at the map (fig. 1). Nearly all the water is in the eastern part of the unit. However, the nonresponse plots are in the western part of the unit. There is very little surface water there. Wetland and beach often adjoin surface water, so this stratum is overestimated as well.

The major urban centers of this unit are also in the east. Exurban Houston is in the northeast, Corpus Christi is in the east central, and Brownsville-McAllen is in the southeast. The western part of the unit is much more rural. What cities there are, such as Laredo in the west central and Eagle Pass in the northwest, have no nonresponse plots. Exurban San Antonio in the north central contributes some urban plots but relatively few nonresponse plots. The nonresponse plots are not in urban areas.

Near each of the major urban centers is a cluster of cropland. The climate becomes drier to the west, and according to the Census of Agriculture, the major agricultural product in the western part of the unit is beef cattle, not crops.

Eighty-four percent of the surface area falls on private land and the National Land Cover Database (NLCD) 2001 zero percent canopy (Homer et al. 2007). However, plots in this unit are not identically distributed, nor does nonresponse occur randomly. The assumption that plots are identically distributed has led to overestimates of several land use classes, as described above, and anything that follows from those inaccurate results. For example, the northeastern part of the unit has high volume live oak (*Quercus virginiana*), while the rest of the unit has low volume honey mesquite (*Prosopis glandulosa*). Unfortunately, there are fewer forested acres in the northeastern part of the unit than FIA claims, so consequently, volume is off considerably. Another problem is that if a user's request does not coincide with evaluation group lines, he or she can get an answer that does not make sense.

For instance, FIA claims a surface area of 4.35 million acres for the 28th Congressional District of Texas within the 4th survey unit of Texas, with a sampling error of 4 percent. This district is larger than 33 of the Southern Research Station's 66 survey units. Examination of a shapefile shows this district to be 6.0 million acres, with 5.9 million acres within the 4th unit. Thus, FIA is nearly 9 standard deviations away from reality. This should happen less often than one in every quintillion (10^{18}) attempts. At the other extreme is the 27th District. FIA estimates it to be 6.91 million acres, with a sampling error of 3 percent. In reality it is 5.43 million acres, which is over 9 standard deviations off the mark. Perhaps rare events have been witnessed, but a better explanation is that FIA has used an inappropriate formula for computing the estimate and its error.

Goeking and Patterson (2013) offer a solution to a similar problem in New Mexico. However, a major difference between South Texas and New Mexico is that South Texas has very little public land, about 2 percent, and even that has a nonresponse rate of 9 percent, higher than the 5 percent goal sought in Goeking-Patterson. If the Goeking-Patterson solution is applied to South Texas, the forest area of the unit increases to nearly 9.1 million acres. Surface water decreases to 2.49 million acres. Yet, the 28th Congressional District is still far too small; its recomputed size is 4.63 million acres. The 27th District is still too large at 6.50 million acres.

Calibration estimators allow for satisfaction of several constraints at once. Among the first papers on calibration estimators was Deming and Stephan (1940). Their method was eventually called "raking," because it resembles raking a field in one direction and then the other. For instance, the raking variables might be county and stratum. First, the stratum is divided by the number of observed plots in the stratum. But county sizes are shown to be incorrect. Expansion factors are adjusted to make the county sizes correct. But then the stratum sizes are off. They are adjusted. County sizes are again incorrect, but not as bad as they were on the previous iteration. The process continues until convergence occurs. It is possible that convergence will not occur.

Under Deming's method, expansion factors can become very large. Deville and Särndal (1992) offer a variation on raking based on logistic regression, as well as several other possible variants. One of those variants involves linear regression.

The regression equation is:

$$\mathrm{Ax} = \mathrm{b} \,,$$

where

A is a p x n matrix of zeros and ones, or even fractions (to account for partially observed plots or unequal inclusion probabilities), where each row corresponds to a plot, and each column corresponds to a constraint;

$\boldsymbol{x}$ is an n x 1 vector of expansion factors; and

b is a p x 1 vector corresponding to the size of each constraint.

A disadvantage to this approach is that it is possible that an expansion factor for a plot might be negative or very large. To deal with this problem, there is a variation called "truncated linear." Like the logistic, this one has a maximum and minimum value. Of course, it is possible that this one might not have a solution at all, and therefore an FIA statistician or analyst must evaluate which approach has a reasonable solution.

Calibrators

There are a number of classification variables in EVALIDator (Miles 2016) that could be used as calibrators. Among these are: Congressional District, County, Distance to Road, Ecozone, Elevation, EMAPHEX, HUC, and Public Ownership. An EMAPHEX is 160,300 acres, but this number could vary slightly due to the choice of projection. Due to border effects, the size of a remnant EMAPHEX could be much smaller. However, there is no way that an EMAPHEX could be as large as 230,000 acres, which is what the current processing system reports the largest EMAPHEX in the 4^{th} unit of Texas to be. On average, a complete EMAPHEX will contain 27 plots, which may be too small to report on.

Ecozone and Hydrologic Unit Code are hierarchical. The smallest HUC8 within Unit 4 is 12040203, at 287 acres. No plot falls in this polygon, so calibration will fail. One solution could be combining this polygon with a neighbor and reporting the HUC6 as 120402xx, even if HUC8 is requested. Similarly, Ecozone could be reported at only the section or province level for certain plots, even if subsection is requested.

Elevation is reported in EVALIDator at 1,000 foot contours. There are several publicly available Digital Elevation Models, and while they do not all agree, they are similar. However, in this case, the entire unit is in the 0-1,000 foot interval.

There are several publicly available road layers. Distance to road is estimated by the field crews rather than calculated from a road layer. Also, the method should produce a similar answer whether public or true coordinates are used.

Results

With County, Congressional District, and Public Ownership as calibrators, the amount of forest increased by 7 standard deviations to 9.48 million acres, up from the current estimate of 8.03 million acres. The estimate of census water decreased nearly 4 standard deviations, from 3.16 million acres to 2.56 million acres, while cropland and related decreased over 3 standard deviations from 5.35 million acres to 4.68 million acres.

References

Deming, W.E.; Stephan, F.F. 1940. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics. 11(4): 427-444.

Deville, J.C.; Särndal, C.E. 1992. Calibration estimators in survey sampling. Journal of the American Statistical Association. 87(418): 376–382.

Goeking, S.A.; Patterson, P.L. 2013. Stratifying to reduce bias caused by high nonresponse rates: A case study from New Mexico's forest inventory. Res. Note RMRS-RN-59. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 22 p.

Homer, C.; Dewitz, J.; Fry, J; [et al.]. 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogrammetric Engineering and Remote Sensing. 73(4): 337-341.

Miles, P.D. 2016. Forest Inventory EVALIDator web-application version 1.6.0.01. St. Paul, MN: U.S. Department of Agriculture, Forest Service, Northern Research Station. Online: https://apps.fs.fed.us/Evalidator/evalidator.jsph

U.S. Census Bureau. 2012. TIGER 2012. [Machine-readable datafiles] prepared by the U.S. Census Bureau, 2012. Online: https://www.census.gov/geo/maps-data/data/tiger-line.html. [Accessed 2017 March].

U.S. Department of Agriculture. 2014. 2012 Census of Agriculture. Part 43: Texas State and County Data. 1,228 p. Online: quickstats.nass.usda.gov. [Accessed 2017 March].

U.S. Geological Survey. 2009. National Hydrography Dataset feature catalog. Based on 1:24,000-scale USGS Topographic Map Content. 123 p. Online: http://nhd.usgs.gov/NHDFeatureCatalog.pdf. [Accessed 2017 March].

Adding New Classification Variables to Reports Via the Spatial Intersection Tool

Thomas A. Weber[1] and Andrew Lister[1]

Abstract—The EVALIDator Web application is a popular program used to generate population estimates and their associated sampling errors from data stored in the Forest Inventory and Analysis Database (FIADB). EVALIDator can produce estimates for a wide variety of forest attributes, allowing the user to select from a predetermined list of classification variables to use in output tables. For example, a user may choose to estimate the area of forest land in Pennsylvania by forest type and ownership group, which are two of several available classification variable choices provided in the tool. However, there is interest in adding new classification variables to FIA's table-making abilities. The current project details a new tool, called the spatial intersection tool (SIT), that offers this capability. An example of its ability to generate tables with non-standard classification variables is provided using the available water storage attribute within the gridded soil survey geographic database (gSSURGO) from the Natural Resources Conservation Service.

Project Background and Objectives

The analysis tool for inventory and monitoring (ATIM) is a module within the Design and Analysis Toolkit for Inventory and Monitoring (DATIM) that, similar to the EVALIDator, allows the user to produce estimates from data stored in the FIADB. However, ATIM provides users with a simple way to add new classification variables to an analysis through the spatial intersection tool (SIT). The SIT allows DATIM to intersect plot-based data with geospatial layers by providing a link between ATIM and spatial data through the SIT tool add-in for ArcMap (Esri 2017). These results are then stored in DATIM for analysis in ATIM.

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

In: Healey, Sean P.; Berrett, Vicki M.; comps. 2017. Doing more with the core: Proceedings of the 2017 Forest Inventory and Analysis (FIA) Science Stakeholder Meeting; 2017 October 24-26; Park City, UT. Proc. RMRS-P-75. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 63 p.

This paper presents a workflow for using SIT to add new classification variables to an analysis. In this example, we intersect data from the Pennsylvania 2010-2015 analysis with the gridded soil survey geographic database (gSSURGO) from the Natural Resources Conservation Service for the region (NRCS 2017). This dataset is publicly available (http://websoilsurvey.sc.egov.usda.gov), but it is also available to Forest Service employees via a corporate "T drive" (T:\FS\RD\Collaboration\NRSFIA_SHARED\Hypervolume_Datasets). The gSSURGO database is one of many useful spatial datasets in this hypervolume_datasets folder. All of the datasets located in this folder have been preprocessed and standardized in a way that makes them easily accessible to analytical staff members that are not GIS or database experts.

Methods

The SIT can be downloaded from http://apps.fs.usda.gov/DATIM/ArcMapAddin.aspx. Once installed, SIT will appear as a tool within the ArcMap toolbar. The SIT interface guides the user through two tasks (figure 1). The first task instructs the user to select an evaluation (identified by state and inventory years, in our example) from which a point layer will be created. The user will also select a GIS layer to define the geographic projection of the analysis point layer. This projection layer must be loaded into ArcMap before opening SIT. Once an evaluation and projection layer are selected, the "Create Point Layer" button is clicked and a point layer representing the plots in the Pennsylvania 2010-2015 analysis is created.

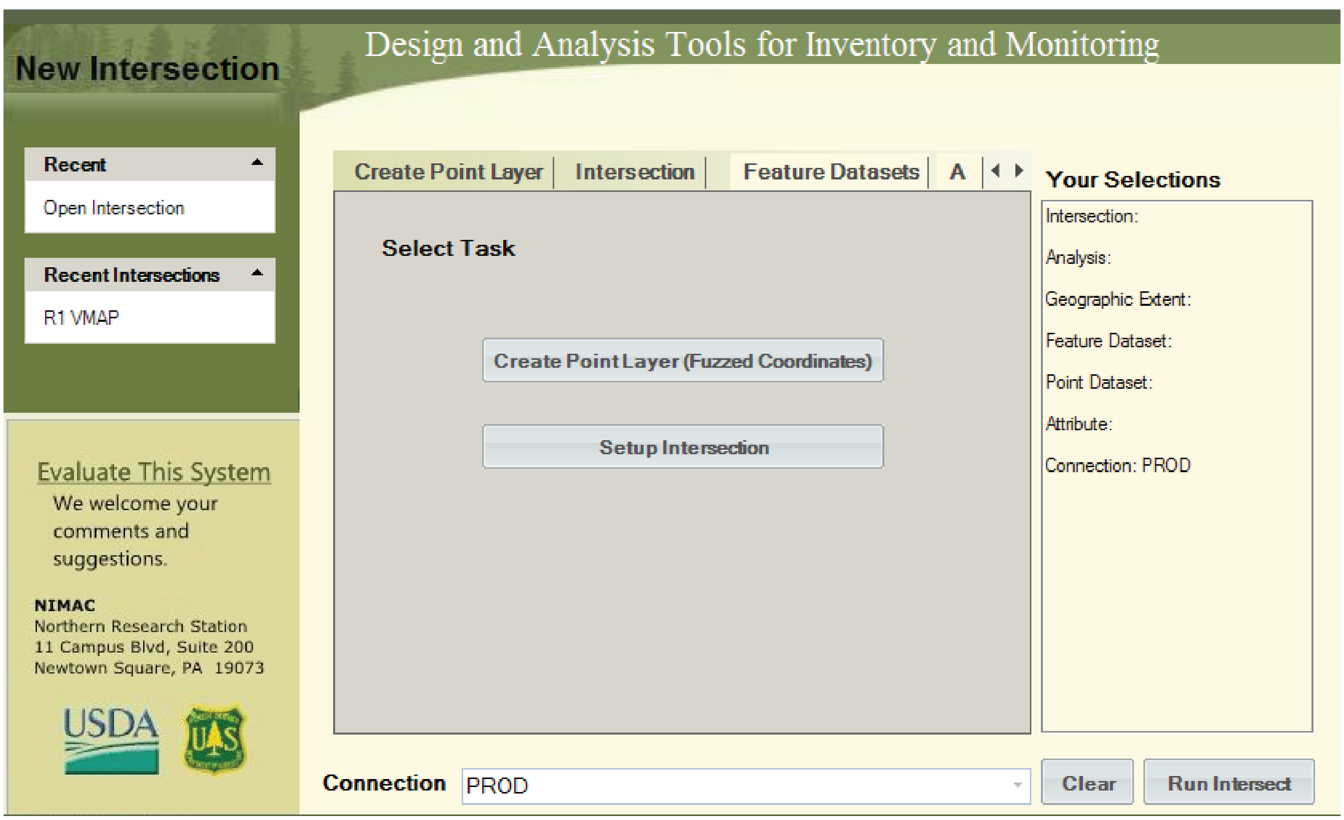


Figure 1—The SIT user interface.

The second task intersects the analysis point layer with a spatial dataset through the Setup Intersection button. In this example, the plots were intersected with the available water storage attribute within the gSSURGO dataset. After clicking the "Setup Intersection" button, SIT directs the user to 1) select an analysis from the dropdown menu (Pennsylvania 2010-2015); 2) create a name for the new ATIM classification variable (available water storage); 3) select the GIS layer that contains the attribute that will become the classification variable (gSSURGO layer in this example); 4) select the point dataset that was created in the first task; 5) select the attribute to use as the classification variable (available water storage); and 6) click the "Run Intersect" button and a new classification variable is created and loaded into ATIM. To create a report using this variable, login to ATIM and the variable will be available in the page, row, and column dropdown menu (figure 2).

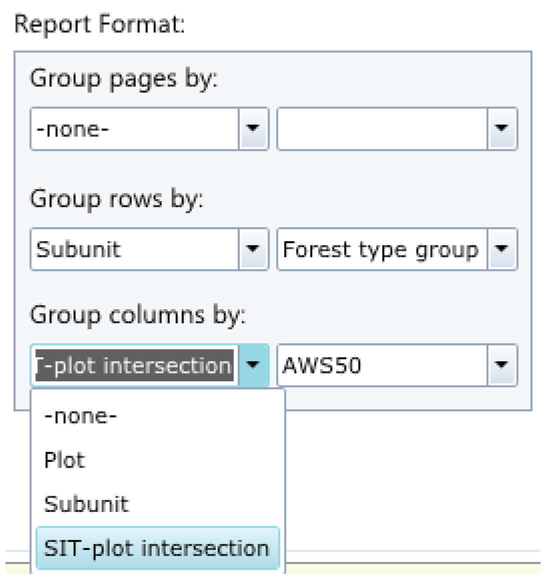


Figure 2—Classification variable created using SIT available within ATIM.

At the time of writing, SIT is not able to work with raster datasets. As a work around, we used a combination of tools within ArcMap to extract the raster values to a polygon layer. First, we buffered the analysis point layer using a 30-m radius buffer, then we used the lookup table tool to reclassify the gSSURGO raster, and lastly we used the zonal statistics as a table tool to add the soils raster values to the buffered point layer. Another thing to note is that for an attribute to work within ATIM as a classification variable, it must be a categorical variable. Therefore, we classified the available water storage into unique classes: 0-49.9 mm, 50-99.9 mm, 100-149.9 mm, 150-199.9 mm, 200-249.9 mm, and > 250 mm prior to doing the intersection.

Table 1—Area of forest land for major forest type groups by available soil water storage, Pennsylvania, 2010-2015 (acres).

Estimate in acres:

Forest type group	Available water storage (millimeters) 0-50	50-100	100-150	150-200	200-250	>250	**Total area**
Oak/hickory	350,657	3,735,371	4,392,890	51,5894	6,741	4,851	9,006,404
Maple/beech/birch	82,986	2,147,313	3,017,231	234,975	1,897	-	5,484,402
Elm/ash/cottonwood	5,171	93,450	147,382	112,711	11,253	17,942	387,910
White pine	8,988	171,621	169,117	31,704	-	1,574	383,003
Aspen/birch	60,312	113,100	106,117	42,005	-	-	321,535

Sampling error percent:

Forest type group	Available water storage (millimeters) 0-50	50-100	100-150	150-200	200-250	>250	**Total area**
Oak/hickory	12.85%	3.41%	3.12%	10.21%	89.73%	116.75%	1.73%
Maple/beech/birch	25.51%	4.83%	3.85%	15.15%	97.40%	-	2.63%
Elm/ash/cottonwood	99.88%	23.88%	19.54%	22.57%	76.32%	57.34%	11.84%
White pine	87.08%	17.78%	18.10%	45.75%	-	104.53%	11.97%
Aspen/birch	31.66%	21.54%	22.39%	38.29%	-	-	13.00%

Results and Discussion

For this example, we created a report estimating area of forest land by forest-type group and available soil water storage in Pennsylvania, 2010-2015 (table 1). Available water storage represents the volume of water that the soil can store and make available to plants. It can contribute to a species' physiological response to drought (Phillips et al. 2016). For example, most oak species in the eastern United States are considered drought tolerant based on their high abundance on xeric sites (Hanson et al. 2001), but many oak species demonstrate some of the highest sensitivity to drought in terms of reduced growth relative to other hardwoods on sites with a low water holding capacity (Maxwell et al. 2015; Orwig and Abrams 1997). A similar response has been documented for eastern white pine (Clark et al. 2011), while the opposite pattern has been documented for Virginia pine (Orwig and Abrams 1997). Breaking forest type out by available water storage helps identify how much forest area is likely to experience reduced physiological function during drought conditions. As this example shows, the SIT tool within ATIM will allow FIA analysts and other data users to expand their analysis capabilities by adding non-standard classification variables from GIS data to FIA data analyses.

References

Clark, J.S.; Bell, D.M.; Hersh, M.H.; Nichols, L. 2011. Climate change vulnerability of forest biodiversity: Climate and competition tracking of demographic rates. Global Change Biology. 17: 1834–1849.

Esri. 2017. ArcGIS, version 10.3. Redlands, CA: Esri.

Hanson, P.J.; Todd, D.E.; Amthor, J.S. 2001. A 6-year study of sapling and large-tree growth and mortality responses to natural and induced variability in precipitation and throughfall. Tree Physiology. 21: 345–358.

Natural Resources Conservation Service (NRCS). 2017. Gridded soil survey geographic (gSSURGO) database. Washington, DC: U.S. Department of Agriculture, Natural Resources Conservation Service. Online: http://websoilsurvey.sc.egov.usda.gov [Accessed 4 April 2017].

Orwig, D.A.; Abrams, M.D. 1997. Variation in radial growth responses to drought among species, site, and canopy strata. Trees--Structure and Function. 11: 474–484.

Phillips, R.P.; Ibáñez, I.; D'Orangeville, L.; Hanson, P.J.; [et al.]. 2016. A belowground perspective on the drought sensitivity of forests: Towards improved understanding and simulation. Forest Ecology and Management. 380: 309-320.

Maxwell, J.T.; Harley, G.L.; Matheus, T.J. 2015. Dendroclimatic reconstructions from multiple co-occurring species: a case study from an old-growth deciduous forest in Indiana, USA. International Journal of Climatology. 35: 860–870.

To learn more about RMRS publications or search our online titles:

www.fs.fed.us/rm/publications

www.treesearch.fs.fed.us