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

Brock Stewart[1], **Chris J. Cieszewski**[2], **and Michal Zasada**[3,4]

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

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

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

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

Problem Definition and Objective

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

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

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

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

Definitions

Timberland

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

[1] Graduate Student Assistant, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602.
[2] Associate Professor, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602. Phone: 706–542–8169; fax: 706–542–8356; home page: www.growthandyield.com/chris.
[3] Postdoctoral Fellow, D.B. Warnell School of Forest Resources, University of Georgia, Athens, GA 30602.
[4] Assistant Professor, Department of Dendrometry and Forest Productivity, Faculty of Forestry, Warsaw Agricultural University, Nowoursynowska 159, 02-776, Warsaw, Poland.

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

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

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

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

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

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

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

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

Volume of All Live Trees

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

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

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

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

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

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

Growing Stock Volume

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

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

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

For the three above variables, we developed the equations below to make calculations on the condition-level (for area estimates) or tree-level (for volume estimates) data sets and summed over conditions or trees meeting the requirement filters below.

Equations

For TIMBERLAND, we used the following equation:

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

(The number of acres a condition represents) =

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

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

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

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

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

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

Requirements (Filters)

For TIMBERLAND area, we tested the following two filters:

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

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

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

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

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

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

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

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

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

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

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

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

Results

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

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

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

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

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

Table 2.—*Results obtained for various TIMBERLAND algorithms.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Discussion

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

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

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

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

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

Literature Cited

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

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

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