CONSEQUENCES OF DATA REDUCTION IN THE FIA DATABASE: A CASE STUDY WITH SOUTHERN YELLOW PINE

Anita K. Rose[1], James F. Rosson, Jr[2]., and Helen Beresford[3]

ABSTRACT.--The Forest Inventory and Analysis Program strives to make its data publicly available in a format that is easy to use and understand most commonly accessed through online tools such as EVALIDator and Forest Inventory Data Online. This requires a certain amount of data reduction. Using a common data request concerning the resource of southern yellow pine (SYP), we demonstrate how results may vary depending on the particular data reduction approach employed. We used the loblolly-shortleaf pine forest-type group (FTG) and the longleaf-slash pine FTG as surrogates for SYP. The volume of our four target species (loblolly pine, shortleaf pine, slash pine, and longleaf pine) in these two FTGs was 98.2 billion cubic feet, but this was only 84 percent of the total volume for the target species in all FTGs. In addition, many unwanted species were included by using total volume in the two FTGs as a surrogate for SYP.

The Forest Inventory and Analysis (FIA) Database represents measurements made on a large number of sample unit variables. The FIA Program strives to make these data publicly available in a format that is easy to use and understand most commonly accessed through online tools such as EVALIDator and Forest Inventory Data Online (FIDO). This requires a certain amount of data reduction (collapsing data into understandable formats; examples are forest type and species groups). However, users may not be aware of the precise data reduction algorithms in these online tools, and this may result in output that falls short of what the user needed or expected, many times without their knowledge. In addition, the actual query a user wants may not be available, forcing them to use a surrogate that may or may not reflect what is desired. An example of this would be the use of forest type as a surrogate to determine the spatial extent of a species.

Issues in data reduction also arise from definition confusion. Forestry terms are sometimes informal, vague, or change over time, and this causes problems when applied in a precise manner. For example, the term southern yellow pine (SYP) has evolved over the last 100 years from originally meaning only slow-growing (ring-width restricted) longleaf pine (*Pinus palustris* Mill.) to various combinations of southern pines (USDA Forest Service 1936). Furthermore, formal definitions for commercial lumber are set by the American Society for Testing and Materials and may change in reaction to market conditions. Adding to the confusion is the fact that commercial timber definitions may not coincide with botanical descriptors such as forest type. Therefore, it is important that these definitions are documented for users. Using a common data request concerning the resource of SYP, we demonstrate how results may vary depending on the particular approach used with the data.

METHODS

Our target species for SYP were loblolly pine (*P. taeda* L.), slash pine (*P. elliottii* Engelm.), shortleaf pine (*P. echinata* Mill.), and longleaf pine. Because spatial extent of an individual species is not currently an option in the online tools, we assessed the area

[1] Anita K. Rose; Research Ecologist; USDA Southern Research Station, Forest Inventory and Analysis; 4700 Old Kingston Pike; Knoxville, TN 37919; anitarose@fs.fed.us; 865-862-2029

[2] James F. Rosson, Jr.; Research Forester; USDA Southern Research Station, Forest Inventory and Analysis; 4700 Old Kingston Pike; Knoxville, TN 37919; jrosson@fs.fed.us; 865-862-2067

[3] Helen Beresford; IT Specialist; USDA Southern Research Station, Forest Inventory and Analysis; 4700 Old Kingston Pike; Knoxville, TN 37919; hberesford@fs.fed.us; 865-862-2091

and volume of SYP in the 13 Southern States using the loblolly-shortleaf pine forest-type group (FTG) and the longleaf-slash pine FTG as surrogates for SYP. In order to duplicate the choices and selections available to and made by many users of FIA data in FIDO or EVALIDator, we made the following assumptions for this analysis: 1 – some users would accept that all the volume in the two FTGs was SYP volume, and 2 – some users would further filter the results from the two FTGs to only include volume for the target species. The loblolly-shortleaf pine FTG includes the following detailed forest types: loblolly pine, shortleaf pine, Virginia pine, sand pine, Table Mountain pine, pond pine, pitch pine, and spruce pine. The longleaf-slash pine FTG includes the following detailed forest types: longleaf pine and slash pine. Forest type is a classification of forest land based upon and named for the tree species that forms the plurality of live-tree stocking. Forest-type groups are combinations of forest types that share closely associated species or site requirements.

RESULTS AND DISCUSSION

Area and volume for the loblolly-shortleaf pine FTG and the longleaf-slash pine FTG together in the South was 70.4 million acres of forest land, and 115.5 billion cubic feet of volume (regardless of species) (Table 1). Volume per acre for these two FTGs averaged 1,641.6 cubic feet per acre and ranged from a low of 1,152.1 cubic feet per acre in Florida to a high of 2,033.5 cubic feet per acre in Virginia. Georgia, Alabama, and Mississippi ranked as the top three States in the South for both area and volume of the two FTGs combined (Table 1). Rankings changed slightly when the FTGs were considered separately. Florida ranked first for area and volume of the longleaf-slash FTG, the majority of which was accounted for by slash pine. Our target species (loblolly pine, slash pine, shortleaf pine, and longleaf pine) were, not surprisingly, the top four species for volume in the two FTGs. They accounted for 98.2 billion cubic feet (85 percent of the total) (75.6, 11.8, 6.7, and 4.1 billion cubic feet, respectively)

Table 1.—Area and volume of all-live trees in the loblolly-shortleaf pine and the longleaf-slash pine forest-type groups (as surrogates for southern yellow pine) by State and forest-type group, 2012

			Forest-type group			
	Both groups		Longleaf-slash pine		Loblolly-shortleaf pine	
State	Forest land	Volume	Forest land	Volume	Forest land	Volume
	thousand acres	*million cubic feet*	*thousand acres*	*million cubic feet*	*thousand acres*	*million cubic feet*
Alabama	9,639.8	14,496.5	1,055.8	1,593.8	8,584.0	12,902.7
Arkansas	5,668.0	9,902.3	—	—	5,668.0	9,902.3
Florida	7,478.3	8,615.7	5,800.7	6,510.2	1,677.6	2,105.5
Georgia	11,114.9	18,407.8	3,691.8	5,055.5	7,423.1	13,352.3
Kentucky	191.2	320.2	—	—	191.2	320.2
Louisiana	5,927.6	9,580.3	812.7	1,192.9	5,114.9	8,387.4
Mississippi	8,080.6	13,692.6	822.2	1,413.0	7,258.3	12,279.6
North Carolina	5,825.0	10,489.5	323.6	519.4	5,501.4	9,970.1
Oklahoma (east)	1,102.8	1,437.3	—	—	1,102.8	1,437.3
South Carolina	6,067.9	11,184.5	549.7	701.7	5,518.2	10,482.9
Tennessee	973.2	1,546.5	—	—	973.2	1,546.5
Texas (east)	5,350.6	9,849.0	144.2	270.6	5,206.5	9,578.4
Virginia	2,950.2	5,999.2	—	—	2,950.2	5,999.2
All States	70,369.9	115,521.5	13,200.7	17,257.0	57,169.2	98,264.5

— = no value for the cell.

Table 2.—Volume of live trees (≥5.0 inches diameter at breast height; where total volume ≥100.0 million cubic feet) in the loblolly-shortleaf pine and the longleaf-slash pine forest-type groups by species and State, 2012

Species	All States	AL	AR	FL	GA	KY	LA	MS	NC	east OK	SC	TN	east TX	VA
							million cubic feet							
All species	115,521.5	14,496.5	9,902.3	8,615.7	18,407.8	320.2	9,580.3	13,692.6	10,489.5	1,437.3	11,184.5	1,546.5	9,849.0	5,999.2
Loblolly pine	75,565.7	10,593.5	6,278.6	1,123.4	11,008.2	89.1	7,159.4	10,228.6	7,224.3	625.8	8,891.3	829.4	7,329.7	4,184.5
Slash pine	11,769.2	653.4	—	5,075.1	4,133.2	—	658.0	774.3	110.4	—	144.7	—	220.3	—
Shortleaf pine	6,715.8	440.7	2,601.3	24.1	383.3	16.7	272.3	585.5	276.8	656.6	146.5	147.2	1,034.3	130.7
Longleaf pine	4,135.5	743.8	—	909.1	543.2	—	453.5	444.9	445.0	—	548.2	—	47.7	—
Sweetgum	2,882.4	407.4	231.0	30.7	424.7	6.0	276.2	356.5	350.4	8.2	286.3	30.0	322.7	152.3
Virginia pine	2,338.5	330.5	—	—	291.0	108.5	—	0.2	547.4	—	169.2	245.7	—	646.2
Water oak	1,333.1	219.2	41.3	73.4	268.6	—	142.7	174.3	68.2	3.2	168.6	—	166.4	7.2
Yellow-poplar	1,127.4	193.1	—	3.3	187.6	23.9	7.2	141.3	222.0	—	79.2	47.1	—	222.7
Southern red oak	881.8	111.3	74.8	9.1	114.1	1.4	127.0	140.1	45.1	3.8	49.2	23.5	127.6	54.7
White oak	867.7	81.9	201.7	—	86.3	10.8	73.5	106.4	76.5	19.4	55.5	26.1	59.9	69.5
Red maple	742.2	71.5	30.1	15.5	110.8	6.1	30.1	47.7	202.4	3.0	69.3	27.0	21.4	107.3
Pond pine	730.5	—	—	86.2	94.2	—	—	—	432.6	—	114.9	—	—	2.6
Sand pine	726.1	—	—	719.1	7.0	—	—	—	—	—	—	—	—	—
Post oak	586.1	51.8	101.3	1.2	33.8	0.7	67.7	79.3	22.2	65.9	25.2	8.1	120.1	8.9
Blackgum	420.7	51.4	38.2	13.2	43.0	0.7	56.6	87.1	17.5	3.4	32.2	5.8	56.7	15.1
Eastern redcedar	307.1	36.0	36.7	5.0	31.4	9.7	3.2	25.6	30.1	5.8	48.0	23.2	16.3	36.1
Laurel oak	305.9	55.8	0.4	87.4	73.5	—	4.6	21.6	9.0	—	33.7	—	19.6	0.5
Winged elm	263.0	23.5	32.5	1.0	38.9	0.6	19.2	35.6	9.6	5.2	30.8	4.9	57.2	4.0
Black cherry	260.9	37.3	17.3	14.4	44.3	3.7	16.0	42.7	26.7	0.2	26.5	8.3	5.5	18.0
Swamp tupelo	209.6	12.8	—	48.8	82.1	—	1.2	12.4	12.2	—	38.9	—	—	1.0
Willow oak	204.9	13.2	20.6	0.5	16.1	—	15.2	30.0	17.4	0.2	32.7	0.5	26.3	32.2
Cherrybark oak	202.0	15.4	21.4	0.2	2.1	0.1	33.0	31.5	12.6	—	22.3	0.2	59.6	3.7
Mockernut hickory	174.8	23.9	20.1	1.9	19.5	1.6	18.8	19.9	12.6	5.5	21.5	8.3	12.8	8.3
Sweetbay	173.4	33.6	6.2	50.2	10.5	—	9.6	52.5	3.2	—	0.5	—	6.5	0.6
Spruce pine	159.3	36.6	—	34.0	17.4	—	28.2	41.7	—	—	1.5	—	—	—
Pond cypress	131.6	1.0	—	68.7	52.8	—	2.3	3.9	0.2	—	2.8	—	—	—
Loblolly bay	122.1	—	—	37.9	22.3	—	—	—	56.0	—	5.9	—	—	—
Scarlet oak	118.2	13.8	—	—	23.0	0.1	—	2.0	22.1	—	9.8	16.3	—	31.1
Northern red oak	113.9	10.6	41.6	—	11.0	0.4	—	5.7	16.4	10.8	5.6	0.5	—	11.2
Pitch pine	111.5	—	—	—	8.6	19.6	—	—	36.5	—	0.1	3.4	—	43.3
Live oak	111.2	3.9	—	61.5	31.9	—	1.4	2.4	0.7	—	9.3	—	0.1	—
Pignut hickory	109.5	26.1	0.2	0.9	27.9	3.3	2.0	13.8	10.5	0.1	6.6	10.2	0.3	7.7
Black oak	106.5	13.7	18.2	—	7.0	—	4.3	7.0	11.1	3.2	6.2	6.3	1.5	28.0
Sourwood	104.8	18.9	—	—	23.3	1.2	0.8	16.9	21.7	—	5.7	6.0	—	10.4

— = no value for the cell.

of the volume in the two FTGs (Table 2). Mostly nonpine species accounted for the remaining volume, especially sweetgum (*Liquidambar styraciflua* L.) and water oak (*Quercus nigra* L.). There were a total of 130 species (not including unknowns) tallied in these two FTGs across the South. Virginia pine (*P. virginiana* Mill.) accounted for a significant percentage of the volume in the two FTGs in southeastern States like Tennessee, Virginia, North Carolina, and Kentucky. In addition, using the total volume in the two FTGs as a surrogate for SYP overestimated the volume of the four target species in several States and underestimated it in other States (compare totals in Table 1 to totals in Table 3). Kentucky had the highest percentage overestimate. Volume of SYP in Kentucky was estimated at 320.2 million cubic feet (Table 1) using the FTG method. However, the actual total volume of the four target species, regardless of FTG, was only 215.3 million cubic feet in that State (Table 3). This is primarily due to the amount of Virginia pine and yellow-poplar included in the FTGs.

While the four target species accounted for 85 percent of the volume in these two FTGs, this was only 84 percent of the total volume for these four species in the South (Table 3). So, while using the two FTGs as a surrogate for SYP resulted in the inclusion of unwanted species, it also resulted in the exclusion of 16 percent of the total volume for the species of interest. And in States like Kentucky, Tennessee, and east Oklahoma, a significant percentage of the volume of loblolly and shortleaf pine (no slash or longleaf pine were recorded in these States) was excluded (51, 30, and 25 percent, respectively) by using the FTG method. This is most likely due to these species occurring in mixed oak-pine stands.

While we recognize that for the Southern States as a whole the estimate of SYP volume using the FTG method was very close to the actual volume for the four species of interest regardless of FTG (115.5 (estimated) – Table 1 versus 116.5 (actual) billion cubic feet – Table 3) this does not take into account the inaccuracies at the State level, especially for States like Kentucky. In addition, while total volume

Table 3.—Actual volume of loblolly pine, shortleaf pine, slash pine, and longleaf pine by State and forest-type group, 2012

State	All groups	Forest-type group		
		Longleaf-slash pine	Loblolly-shortleaf pine	All other groups
		million cubic feet		
Alabama	15,082.2	1,375.5	11,055.9	2,650.8
Arkansas	10,881.4	—	8,879.8	2,001.5
Florida	8,514.2	5,981.2	1,150.5	1,382.5
Georgia	18,761.8	4,650.6	11,417.2	2,694.0
Kentucky	215.3	—	105.7	109.6
Louisiana	9,767.9	1,132.4	7,410.7	1,224.8
Mississippi	13,968.6	1,206.6	10,826.7	1,935.2
North Carolina	9,704.8	493.8	7,562.7	1,648.3
Oklahoma (east)	1,719.9	—	1,282.4	437.5
South Carolina	11,319.3	635.7	9,094.9	1,588.6
Tennessee	1,400.7	—	976.7	424.0
Texas (east)	9,862.6	247.7	8,384.3	1,230.6
Virginia	5,265.9	—	4,315.2	950.8
All States	116,464.6	15,723.5	82,462.8	18,278.4

— = no value for the cell.

may not be terribly overestimated or underestimated using this method in many States, other analyses may be significantly (and negatively) affected. For example, including a large amount of Virginia pine in an analysis of SYP may negatively impact growth and removal estimates or detailed diameter-class analyses.

One option for a slightly more accurate assessment of SYP using online tools would be to use detailed forest types rather than FTGs. This would (hopefully) eliminate some of the unwanted species (like Virginia pine). Further analyses into this issue will include an examination of the actual spatial extent of the four target species. This is accomplished through the creation of a specialized pseudoforest type by the user (in a programming language like Oracle) based upon the user's specifications, as well as by an investigation of the effect of including unwanted species on the resulting detailed analysis of SYP in the South.

The most accurate assessment would be for the user to select the species to include in what they perceive as the SYP group. However, current capacity of the online tools often eliminates that prospect because an estimate of forest area is usually needed along with volume. This often forces the user to rely on the SYP volume in the loblolly-shortleaf and longleaf-slash pine FTGs. Users need to be aware of the intricacies that exist and the inaccuracies introduced when forest type (or FTG) is used as a surrogate for species. The FIA Program needs to clearly document the definitions of the tool algorithms and encourage increased flexibility in online capabilities so users can accurately retrieve the information they need.

ACKNOWLEDGMENTS

The authors would like to acknowledge the many people involved in data collection, editing, and processing as well as the reviewers and editors involved in improving this manuscript.

LITERATURE CITED

U.S. Department of Agriculture Forest Service. 1936. Southern yellow pine. Technical Note 214. Madison, WI: U.S. Department of Agriculture Forest Service, Forest Products Lab. 8 p.