Integrating TIGER, DOQ, and Forest Inventory Field Calls

Joseph M. McCollum

Abstract:

Historically, FIA has employed photointerpreters to determine whether a given dot was one of four classes: forested land, nonforested land, census water, or noncensus water. Plot center was a subset of the dot grid. Field crews would then use the sample plot center to determine a land use class. The field classification was compared with the photointerpretation, with the assumption that the field call was correct. A proportion of the land base that was in each class was calculated for each county using conditional probability. The author has written a program to allow photointerpreters to call dots forest, nonforest, census water, or noncensus water. The dots are overlaid on TIGER files to see what the Census Bureau has actually called them. This activity will add clarification in the final results. It is recommended that TIGER files be used as truth when determining whether a given plot is in a particular county, but that the field crew's call be considered truth when determining whether a given plot is in land or water.

Introduction

Last year at this very conference, Osborn and Cunningham presented “Adjusting U.S. Census TIGER/Line Blocks to Register With Local Geography.” That paper was aimed at urban planners. It has occurred to the author that a strategy is needed for more rural applications – specifically, to resolve differences among TIGER/Line geography, digital orthophotographs (or other digital imagery), and field crew calls.

Forest Inventory and Analysis (FIA) is the program within the USDA Forest Service responsible for estimating the amount of the nation’s forests, in terms of the land area of the forests as well as the volume and the biomass of the timber supply, among other things. Reports on each of the states are produced, as required by 16 U.S.C. 1642(e). (Law of Revision Counsel 2001). The reader unfamiliar with FIA may wish to look at a sample report, such as Rosson (2002).

The first task that FIA must do is to estimate the amount of forested land. This is done in two phases. The first phase consists of photo-interpreters making a call on aerial photographs – historically either photos have been scored with dots or a mylar overlay scored with dots has been placed on the photo. The photo-interpreter must call each dot as forested land, non-forested land, census water, or non-census water. The second phase consists of field crew visiting a subset of those dots – called plots – and making the same call. Additionally, the field crew may call the plot inaccessible. If the plots are forested land, the field crew take measurements of trees that appear on the plot.

Plots are laid out on a grid of hexagonal cells, in a manner outlined by Brand (1999). Prior to that date, in much of the country, plots were laid out on a grid of squares, and in other parts of the country, haphazardly. The Brand rules favored retaining previously existing plots, and then if certain hexagons remained empty, generating new plots. Occasionally, there has been dispute over which county a plot is really in. The author makes the observations that FIA uses the census bureau’s gazetteer to obtain area estimates, which in turn have been generated from analysis of TIGER files. Now raw TIGER is in latitude and longitude, and while the Census Bureau does not reveal its exact projection methodology, experience with the data show that standard Albers parameters (standard parallels at 45º 30’ N and 29º 30’ N, with a projection point of 96º W 23º N, come fairly close.) The author’s recommendation is that TIGER be the ultimate arbiter in determining what county a plot is in.

Census Water

At this point, the reader may ask, “What is census water?” Until recently, hard copy aerial photographs have been used in FIA’s photo-interpretation, and it has been impractical to issue maps to field crew with census water marked on them. As a result, a heuristic rule for identifying census water had to be developed. According to the FIA Field Guide (USDA 2003), census water is “Rivers and Streams that are more than 200 feet wide and bodies of water greater than 4.5 acres in size.”

This rule arose in the following way: According to the “Standards for 1:100,000-Scale Quadrangle Maps, Part 3: Feature Specifications and Compilation,” (USGS 1985) lakes and ponds were to be digitized if they were at least 0.06 inch in the shortest dimension. Observe that 0.06 inch at the 1:100,000 scale is 6000 inches, or 500 feet. A circular pond with a diameter of 500 feet is about 196,350 square feet, or 4.5 acres. The same manual also stated that all perennial streams that were at least 0.024 inches wide were to be mapped. Observe that 0.024 inches at the 1:100,000 scale is 2400 inches, or 200 feet.

FIA is aware that this definition is not perfect. The USGS manual also contains the following caveats:

(a) “In predominantly wet or well-watered regions, …. Features of small areal extent may be omitted, especially if they tend to impair the legibility of the more important features.”

(b) “In arid and semiarid areas, the presence and location of water is important as a means of orientation. In these areas, as many hydrographic features as possible should be shown.”

(c) “In small localized areas where hydrographic features are too numerous to show to scale, no attempt is made to show every feature. Instead a representative pattern of the symbols is shown that covers the localized area.”

Moreover, 1:100,000 Digital Line Graphs were not the only source of TIGER data. Other sources included 1:24,000 USGS paper maps (whose standards were 0.025 inch wide for streams, and 0.024 inch in the shortest dimension for other bodies of water), and the Census Bureau’s own DIME (Dual Independent Map Encoding) maps.

In short, census water is whatever has been digitized in the TIGER database as water. Noncensus water includes single-line streams in the TIGER database and other hydrographic features that have not been digitized.

Historically, FIA photo-interpreters have not had access to the TIGER database. From looking at the above caveats in the USGS manual, we can see that without access to TIGER, the photo-interpreter cannot be certain whether a given body of water is census water or not.

Area Estimates

Osborn and Cunningham gave some ideas on conflating or rubber sheeting TIGER data to other digital maps. For FIA, conflation would not be a good idea since surface area estimates are taken from the Census Bureau's gazetteer and summary files.

The field crews visit the plots. They do not visit every dot, but rather a set of 1/6-acre plots. There is approximately one plot per 6000 acres of surface area. The field crews make the same call that the photo-interpreter does: forested land, non-forested land, census water, or non-census water. They may also call the plot inaccessible if a landowner denies access or it is too dangerous to get to the plot.

Historically, the field crew call has been assumed correct, and the photointerpreters' estimates have been corrected based on those calls, using a Bayesian adjustment.
See McCollum 2003 for details.

For instance, there might be a call of:

		Photointerpreter call			
		Plots		Dots	
		Forest	Nonforest	Forest	Nonforest
Field	Forest	$n_{11} = 45$	$n_{21} = 3$	$d_1 = 1200$	$d_2 = 1300$
Call	Nonforest	$n_{12} = 5$	$n_{22} = 57$		

The phase 1 estimate of forest would be $P_F = \frac{d_1}{d_1+d_2} = 0.48$, while the phase 2 estimate would be adjusted as $P_F \cdot \frac{n_{11}}{n_{11}+n_{12}} + P_{NF} \cdot \frac{n_{21}}{n_{21}+n_{22}} = 0.4580$.

How might triple sampling be done?

The dots, historically laid out on a square grid – at times uniformly throughout the landscape, at other times clustered around the plots – this time were laid out on a hexagonal grid just as the plots were. The author wrote a C program which in turn writes Arc Macro Language (AML) scripts to decompose 2400 hectare hexagons into 88.89 hectare hexagons. A square may be elegantly decomposed into n^2 smaller squares, where n is an integer. A hexagon, on the other hand, is decomposed into $T = h^2 + hk + k^2$ smaller hexagons, where h and k are integers. There are always partial hexagons that add up to whole hexagons. A threefold (h = 1, k = 1, T = 3) decomposition is easily performed by selecting the center point of the hexagon plus its six nodes, then performing a Thiessen polygon expansion on all of those points. This procedure was repeated three times, resulting in a 27-fold reduction in the size of the hexagons.

The center of each small hexagon in the resultant grid was chosen as a dot. This type of dot network minimizes the amount of spatial autocorrelation between dots.

Some preliminary data has been collected for South Carolina.

T=

		Photointerpreter Call Forest Land	Nonforest Land	Census Water	Noncensus Water	No Photo
TIGER Call	Land	51719	34745	908	275	0
	Water	496	566	3791	16	688

Let

T_{ij} be the ith row and jth column of matrix **T**.
$T_{i\bullet}$ be the sum of the ith row of matrix **T**.

It is expected that the T_{15} entry will always be zero. If it is not, then FIA will consider getting other photography, such as Landsat Thematic Mapper (TM) data. Areas in South Carolina where there is no aerial photography are generally territorial sea, but there may be an occasional barrier island.

The Census 2000 gazetteer cited A_1 = 77,983,161,188 square meters of land for South Carolina, and A_2 = 4,948,771,376 square meters of water. The question is now how to distribute that area among the dots.

There are several approaches, but the one the author likes best is to divide all the TIGER land dots into the TIGER land estimate, and all the TIGER water dots into the water estimate.

This results in a preliminary Phase 1 estimate of

(1) $$\frac{T_{11}}{T_{1\bullet}} \cdot A_1 = \quad 46016533520.6$$ square meters of forested land,

(2) $$\frac{T_{12}}{T_{1\bullet}} \cdot A_1 = \quad 30914063635.7$$ square meters of nonforested land,

(3) $$\frac{T_{13}}{T_{1\bullet}} \cdot A_1 = \quad 807885157.0$$ square meters of land that photo-interpreters believe to be census

water,

(4)	$\frac{T_{14}}{T_{1\bullet}} \cdot A_1 =$	244678874.7	square meters of non-census water,
(5)	$\frac{T_{15}}{T_{1\bullet}} \cdot A_1 =$	0	square meters of land that is off the photos,
(6)	$\frac{T_{21}}{T_{2\bullet}} \cdot A_2 =$	441711463.5	square meters of water that photo-interpreters believe to be forested land,
(7)	$\frac{T_{22}}{T_{2\bullet}} \cdot A_2 =$	504049774.8	square meters of water that photo-interpreters believe to be forested land,
(8)	$\frac{T_{23}}{T_{2\bullet}} \cdot A_2 =$	3376064834.7	square meters of census water,
(9)	$\frac{T_{24}}{T_{2\bullet}} \cdot A_2 =$	14248756.9	square meters of census water that photo-interpreters believe to be non-census water, and
(10)	$\frac{T_{25}}{T_{2\bullet}} \cdot A_2 =$	612696546.1	square meters of census water not on the aerial photographs.

Cases (3), (6), and (7) present contradictions. It is unreasonable to expect that the area in these cases will all be zero, because it is not practical to digitize every point at which the shoreline of a lake or the bank of a river changes course. Moreover, rainfalls, droughts, and tides will all cause fluctuations in the water level.

It is hoped that the number of TIGER land points called census water will roughly balance the number of TIGER water points called land. In mathematical terms,

$T_{21} + T_{22} \approx T_{13}$. In the event that $T_{13} >> T_{21} + T_{22}$, chances are good that some large water bodies did not get digitized; caveat (a) may have been invoked, and there is more non-census water than the photo-interpreters see. In the event that $T_{13} << T_{21} + T_{22}$, it is possible that caveat (c) was invoked, and some of the non-census water is representative of the census water. At the same time, non-census water should not be negative, even for individual counties.

Although the following method is not perfect, the author recommends the following heuristic.

The amount of additional forest land should be:

$$\frac{T_{21}}{T_{21} + T_{22} + T_{14}} \cdot (T_{13} + T_{14})$$

The amount of additional non-forest land should be:

$$\frac{T_{22}}{T_{21} + T_{22} + T_{14}} \cdot (T_{13} + T_{14})$$

The amount of non-census water should be:

$$\frac{T_{14}}{T_{21} + T_{22} + T_{14}} \cdot (T_{13} + T_{14})$$

Thus, with these data, the final Phase 1 estimate should be:

46407013520.4 square meters (11467377 acres) of forested land,
31359651700.0 square meters (7749108 acres) of non-forested land,
216495967.6 square meters (53497 acres) of non-census water, and
4948771376.0 square meters (1222863 acres) of census water.

References

Brand, G. et al. 1999. The Hexagon/Panel System for Selecting Plots Under An Annual Inventory. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C. 2000. Proceedings of the first annual forest inventory and analysis symposium; 1999 November 2-3, San Antonio, TX. General Technical Report NC-213. St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station.

McCollum, J. 2003. Standard Errors in Forest Area. In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C.; Moser, John W. 2003. Proceedings of the third annual forest inventory and analysis symposium; 2001 October 17-19; Traverse

City, Michigan Gen. Tech. Rep. NC-230. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 208 p..

Office of the Law Revision Counsel. 2001. United States Code, 2000 Edition, Containing the General and Permanent Laws of the United States. Washington, DC: U.S. Government Printing Office.

Osborn, W. and Cunningham, J. 2003. Adjusting U.S. Census TIGER/Line Blocks to Register With Local Geography. Presented at the 2003 ESRI International User Conference. Available online at http://gis.esri.com/library/userconf/proc03/p0947.pdf, June 3, 2004.

Rosson, J. 2002. Forest Resources of Arkansas, 1995. Resource Bulletin SRS–78. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 82 p.

U.S. Census Bureau. 2000. Census 2000 TIGER/Line Files. [machine-readable data files]. Prepared by the U.S. Census Bureau, Washington, DC.

U.S. Department of Agriculture, Forest Service. 2003. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 2.0. U.S. Department of Agriculture, Forest Service, Washington Office. Internal report. On file with: U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis, 201 14th St., Washington, D.C., 20250.

U.S. Geological Survey. 1985. Standards for 1:100,000-Scale Quadrangle Maps. Part 3: Feature Specifications and Compilation.

Author information:

Joseph M. McCollum
Information Technology Specialist
Forest Inventory and Analysis
4700 Old Kingston Pike
Knoxville, TN 37919
Ph. 865-862-2065
FAX 865-862-2048
e-mail: jmccollum@fs.fed.us