Spatially Locating FIA Plots from Pixel Values

Greg C. Liknes[1], Geoffrey R. Holden[1], Mark D. Nelson[1], and Ronald E. McRoberts[2]

Abstract.—The USDA Forest Service Forest Inventory and Analysis (FIA) program is required to ensure the confidentiality of the geographic locations of plots. To accommodate user requests for data without releasing actual plot coordinates, FIA creates overlays of plot locations on various geospatial data, including satellite imagery. Methods for reporting pixel values associated with FIA plots that reduce the likelihood of inadvertent release of plot locations were tested. The number of plots that can be correctly located using only pixel values was reduced by perturbing image band values, averaging band values of neighboring pixels, and reducing the number of image bands.

In the FY2000 Consolidated Appropriations Bill (PL 106-113), Congress included language that modified the Food Security Act of 1985 (7 U.S.C. 2276(d)) to add FIA data collection to a list of items requiring confidential treatment. As a result, the FIA program must ensure the confidentiality of coordinates of field plots, maintaining sample integrity and protecting the privacy of landowners granting FIA access to their lands.

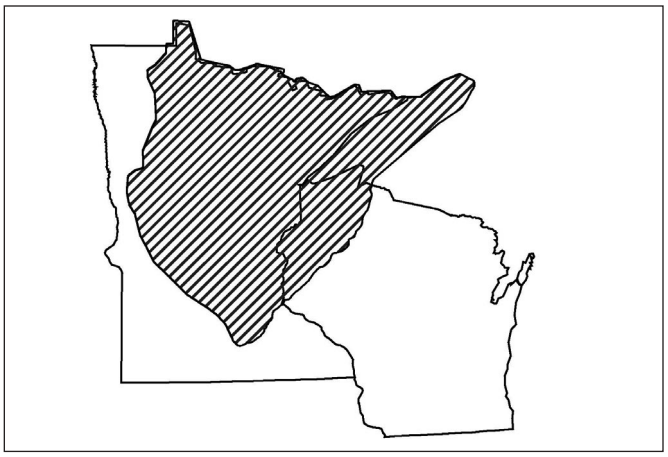
Researchers interested in using FIA plot data as a source of training data for remote sensing classification do not necessarily need plot location information. Rather, FIA staff could extract spectral values from satellite imagery at plot locations and provide the spectral information along with requested FIA plot data (e.g., percent forest cover, basal area), but without plot coordinates. This method of providing FIA data, however, may inadvertently reveal plot location information when a pixel associated with a plot has a combination of spectral values unique to that pixel. In this study, we examined the probability of determining a plot's location from spectral information.

Data

We used data from the Multi-Resolution Land Characterization (MRLC) Consortium National Land Cover Dataset 2000 (NLCD 2000) mapping zone 41 (Homer and Gallant 2001). Zone 41 encompasses part of Minnesota, a portion of western Wisconsin, and Isle Royale in Michigan (fig. 1). The data set is comprised of multiple Landsat Thematic Mapper (TM) and Landsat Enhanced Thematic Mapper+ (ETM+) scenes at 30-m pixel resolution.

Thermal bands and Landsat-derived data, including tasseled cap transformations (greenness, soil brightness, and wetness), and textural information, were available for the area of interest. Additionally, digital elevation data and State Soil Geographic (STATSGO) (http://www.ftw.nrcs.usda.gov/stat_data.html) soil data (available water content, carbon content, quality) were available, as well as a data band containing the date of the input image used for each pixel. In total, 24 bands of data (19 continuous and 5 categorical) were used in this study (table 1). Of the 5,939 measured FIA plots falling within zone 41 cloud-free areas, 200 were randomly selected for testing.

Figure 1.—*MRLC NLCD 2000 mapping zone 41 (hatched area) covering a large portion of Minnesota, some of Wisconsin, and Isle Royale in Michigan.*

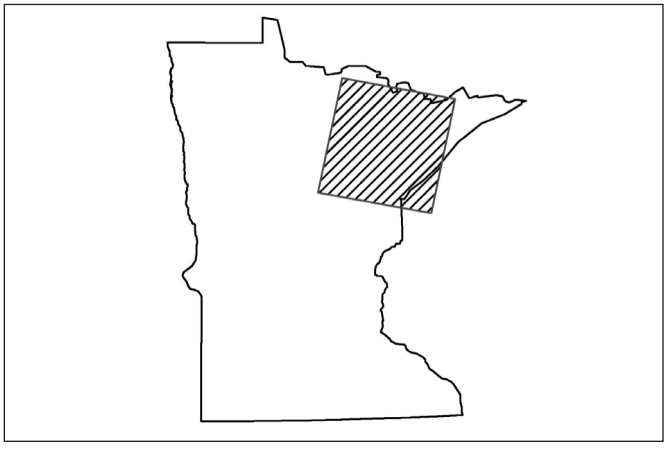


[1] GIS/RS Specialist, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108.
[2] Mathematical Statistician, U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN 55108.

Table 1.—*Available data bands and their data types for MRLC NLCD 2000 mapping zone 41*

Data band	Band name	Data type
1	Texture (band 1—leaf-on)	continuous
2	Texture (band 4—leaf-on	continuous
3	Texture (band 7—leaf-on)	continuous
4	Greenness (spring)	continuous
5	Brightness (spring)	continuous
6	Wetness (spring)	continuous
7	Greenness (leaf-on)	continuous
8	Brightness (leaf-on)	continuous
9	Wetness (leaf-on)	continuous
10	Greenness (leaf-off)	continuous
11	Brightness (leaf-off)	continuous
12	Wetness (leaf-off)	continuous
13	Thermal (spring)	continuous
14	Thermal (leaf-on)	continuous
15	Thermal (leaf-off)	continuous
16	Elevation	continuous
17	Slope	continuous
18	Aspect	categorical
19	Soil quality	categorical
20	Soil carbon	continuous
21	Soil available water content	continuous
22	Spring date	categorical
23	Leaf-on date	categorical
24	Leaf-off date	categorical

Figure 2.—*Location of Landsat path 27 row 27 (P27R27), Minnesota.*

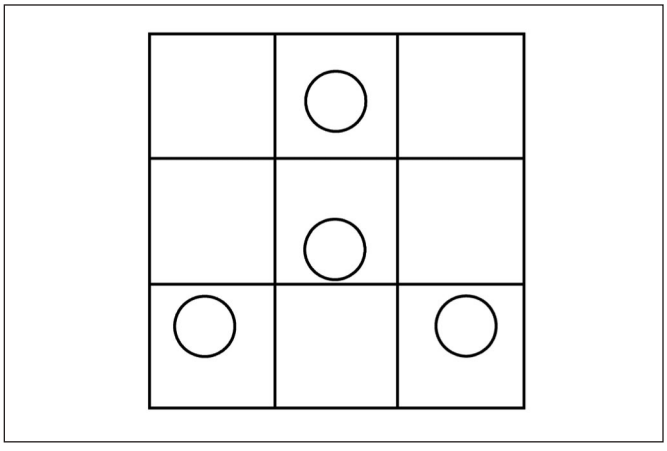


A second study was performed using the data from a single Landsat ETM+ scene, path 27 row 27 (P27R27) (fig. 2), which is one of the scenes contained in zone 41. For this single scene, imagery from three different dates (31 May 2000, 5 July 2001, and 5 November 1999) was available. The six reflective bands (1-5, and 7) for each date were used, totaling 18 bands for this scene. Of the 1,277 measured FIA plots falling within the scene, 200 were randomly selected for testing.

Methods

To determine if providing spectral information reveals plot location information, pixel values at each of the 200 plot locations were extracted for all the data bands in each study. Each field plot consists of four 7.31-m (24-foot) radius circular subplots. The subplots are configured as a central subplot and three peripheral subplots with centers located 36.58 m (120 ft) and azimuths of 0°, 120°, 240° from the center of the central subplot. The plot/pixel arrangement shown in figure 3 is the most probable arrangement and therefore was used in this study. The arrangement of pixels relative to subplots can change slightly depending on where the center subplot falls within a pixel. Three methods for extracting pixel values were considered: (1) values associated with the pixel containing the central subplot, (2) the average of the pixel values associated with the four subplots (fig. 3), and (3) the average of the pixel values in a 3x3 window centered on the pixel associated with the center subplot. For methods 2 and 3, new images were generated using

Figure 3.—*Location of FIA subplots relative to Landsat TM 30-m pixels.*



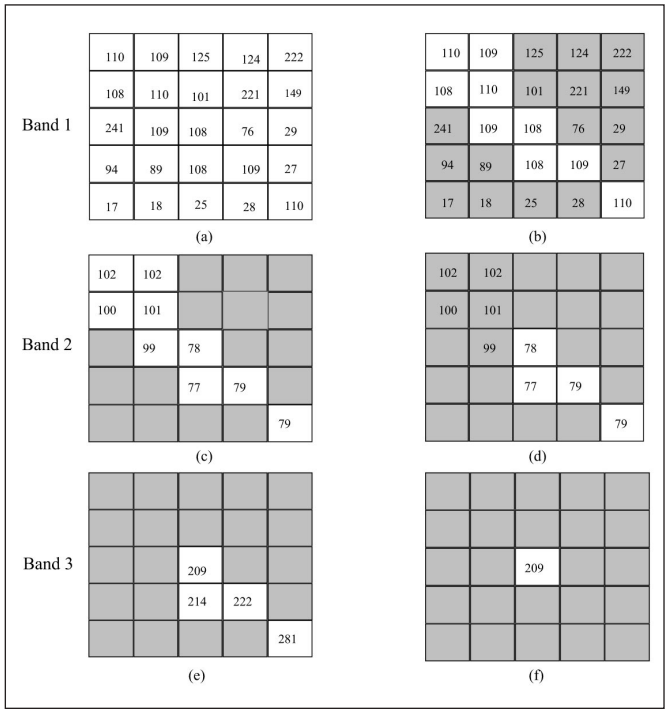
focal analysis techniques, with the averaged values applied at every pixel.

To test the effect of alterations of pixel values on a user's ability to determine correct and unique plot locations, extracted values were perturbed for each of the three methods by an integer randomly selected from the following intervals: [0], [-1, 1], [-2, 2], and [-3, 3]. The combination of the four perturbation intervals with the three methods resulted in 12 tests for both the zone 41 and the P27R27 studies. Additionally, the 12 tests were repeated for a single date (leaf-on) of P27R27.

For each test, the extracted values for each plot were compared to all image pixel values in respective bands. To illustrate, consider a 25-pixel, 3-band image (fig. 4), and suppose band values were extracted at the location of a plot falling within the image. Values for bands 1, 2, and 3 were 108, 79, and 209, respectively. Furthermore, a perturbation of [-1, 1] was then applied to the extracted values, resulting in a set of band values of 109, 78, and 209. The data from band 1 are shown in figure 4(a). All pixels with values not within ±1 of the perturbed value (109) are grayed out in figure 4(b). The data from band 2 are displayed in figure 4(c). Note that all gray pixels no longer need to be checked against the perturbed values because they failed to match the band 1 perturbed value. The remaining data that fell within ±1 of the band 2 perturbed value (78) are shown in white in figure 4(d), and the non-matching pixels are now shown in gray. The same process is applied for band 3 data (perturbed value 209) in figures 4(e) and 4(f). In the end, a single pixel remains that matches all three perturbed values. If Landsat TM pixels are represented in figure 4, this process reveals the location of an FIA plot to within 30 m (the resolution of a Landsat pixel), assuming the true ground position of each pixel is known.

For each test, if only a single pixel matched the extracted values for all bands, its location was compared with the FIA plot locations to determine if the plot had been correctly "located" using only spectral information. All processing was performed using ESRI Grids in ArcInfo and automated using an Arc Macro Language (AML) script.

Figure 4a–f.—*The sequence of steps involved in determining from which pixel an extracted set of values came.* Extracted values with [-1, 1] perturbation applied: band 1 = 109, band 2 = 78, band 3 = 209. *In this example, pixels are grayed out if they don't fall within ±1 of the perturbed values for the corresponding band. Grayed pixels are then ignored for each subsequent band.*



Results and Discussion

For zone 41, all of the 200 FIA plots were correctly located using only spectral (pixel) information for the single pixel and 4-subplot average methods with no perturbation, while 98 percent were located for the 3x3 average method (table 2). As the amount of perturbation increased, the percentage of plots correctly located decreased. This is because the number of potential matches for each pixel increased as the interval around each extracted value increased, thereby reducing the likelihood of uniquely identifying a pixel. Although the percentage of plots correctly located decreased with increased perturbation, more than half of the plots were located for each of the three methods at all perturbation levels.

For the P27R27/multiple date study (18 bands), all of the 200 FIA plots were correctly located using only spectral (pixel) information for each of the three methods with no perturbation (table 3). As was the case for zone 41, the percentage of plots located in P27R27 decreased as the amount of perturbation increased. For P27R27, the percentage of plots located is not

Table 2.—*Percentage of FIA plots located from MRLC NLCD 2000 mapping zone 41 pixel information for various methods and perturbation levels*

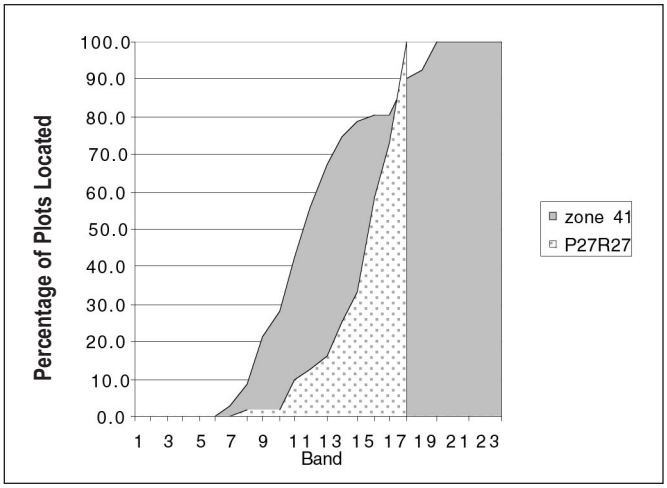
Method	Perturbation	[0]	[-1,1]	[-2,2]	[-3,3]
Single pixel / center subplot		100	94	89	72
3x3 average		98	88	72	53
Four subplot average		100	91	76	58

Table 3.—*Percentage of FIA plots located from Landsat TM path 27 row 27 pixel information for various methods and perturbation levels*

Method	Perturbation	[0]	[-1,1]	[-2,2]	[-3,3]
Single pixel / center subplot (three dates)		100	97	63	32
3x3 average (three dates)		100	71	31	14
Four subplot average (three dates)		100	84	41	19
Single pixel / center subplot (summer only)		17	0	0	0
3x3 average (summer only)		6	1	0	0
Four subplot average (summer only)		7	0	0	0

nearly as high as in zone 41 for the [-3,3] perturbation. In figure 5, the cumulative percentage of extracted values that identify single pixels (single pixel method, [-3,3] perturbation) is graphed against the band that has been processed. In both the zone 41 and P27R27 tests, a high percentage of the plots were located before all bands were processed. Thus, the difference is not simply a matter of the number of bands (18 vs. 24). All zone 41 and P27R27 bands were 8-bit data (256 possible values), with the exception of elevation (zone 41 only), which had values ranging from 177 to 701. The greater variability in this band may have contributed to the higher percentage of plots located in zone 41 in the single pixel method, [-3,3] perturbation test. Also, the elevation band was processed first for all zone 41 tests, which may have significantly impacted at which step unique pixels were identified in the process (fig. 5).

Figure 5.—*Cumulative percentage of FIA subplots correctly located as a function of sequentially adding bands, MRLC NLCD 2000 mapping zone 41 (24 bands) and Landsat TM path 27 row 27 (18 bands).*



For the P27R27/single date study (6 bands), few plots were correctly located. For example, 17 percent of plots were correctly located in the single pixel/no perturbation test. In this case, the noticeable decrease in correctly located plots is most likely due to the small number of bands considered. In both the zone 41 study (24 bands) and the P27R27/multiple date study (18 bands), most plots could not be found until after the sixth band (fig. 5).

Conclusions

If the method for extracting pixel values at plot locations is known and no perturbation is applied to the pixel values, it is possible to locate correctly all or nearly all of the plots with 18 or 24 bands of data. For a single-date, 6-band image, few plots can be correctly located if no perturbation is applied, and

almost no plots if a perturbation is applied. Increasing the perturbation reduces the number of plots that can be correctly located but also reduces the usefulness of the pixel information for some users' applications.

In the future, if spectral data are requested for FIA plot locations, the method described in this paper could be used to screen plots for possible inadvertent disclosure of plot information. Currently, the greatest drawback to the procedure is the time involved in processing an array of extracted values. Future work may include optimizing this processing or develop alternative processing procedures.

Literature Cited

Homer, C.G.; Gallant, A. 2001. Partitioning the conterminous United States in mapping zones for Landsat TM land cover mapping. USGS Draft White Paper.