Error Analysis of Impervious Surface Satellite Imagery in North Carolina

Travis Pate

May 3, 2009

Abstract

This paper outlines the procedure and results of a pilot project to determine the error in satellite imagery used to estimate the percentage of impervious surfaces in North Carolina. This analysis is the final step in a Master's project titled Watershed Assessment in North Carolina. The first section of this paper provides an overview of that Master's project. The latter portion of this paper explains the methods and results of the error analysis. This pilot project determined that the impervious surface data used in the Watershed Assessment project is preliminary accurate enough to proceed with its intended purposes.

Several methodologies have been developed for estimating impervious cover in watersheds.[1] Several methodologies involve remote sensed data (aerial photographs or satellite imagery). Other methodologies calculate impervious cover from population data. The EPA also proposed combining methodologies to make a more accurate estimate of impervious cover.[2] Each techniques has its own strengths and weaknesses, depending on how the information is to be used. Table 1 summarizes several different techniques used to calculate impervious cover.

The literature search for a method that met these criteria led to an EPA report on estimating impervious cover for the southeastern United States.[3] In this report the EPA evaluated a few methodologies that used data from the Census, and the National Land Cover Database. Both of these data are free to the public and updated at regular intervals. Research into the National Land Cover Database revealed that an additional satellite image existed that determined the percentage of impervious cover for all of the United States. It was decided that population, land cover, and percent impervious cover data would meet the criteria for this project.

[1] In reference section: McMahon (2007), Yang (2003), Exum (2005)

[2] Exum, L. et al. (2005). Estimating and Projecting Impervious Cover in the Southeastern United States. United States Environmental Protection Agency. Washington DC.

[3] Ibid

Method	Relative Accuracy	Relative Cost (to obtain)	Use
Ground	Highest	High	Small area planning, Stormwater control siting
Digitizing planimetric maps	High	High	Small area planning, Community planning
Digital photographs	High	High	Small area planning
Satellite Imagery	Moderate	Low	Community planning, Watershed Planning
Population Data	Moderate-High	Low	Community Planning, Watershed Planning
Combination of methods	Moderate-High	Low	Community Planning, Watershed Planning

Table 1 - Methodologies for estimating impervious cover. Source: Stocker, J. (1998)

In this paper, a pilot project is established to determine the accuracy of the impervious cover data. The impervious surface data obtained from the MRLC will be compared to detailed planimetric data for the Town of Chapel Hill. A detailed methodology for the pilot project is outlined as well as the results from the analysis. Similar analysis could be conducted when data becomes available in order to estimate the MRLC imagery over the span of the whole state. At least the three major regions of North Carolina, the mountains, piedmont, and coast.

Methodology

Impervious Surface Satellite Imagery

The Multi-Resolution Land Characteristics Consortium (MRLC) produced a database of impervious surfaces for the entire United States from 2001 satellite imagery.[4] These data consists of 30m by 30m resolution raster data. Each pixel of raster data represents a

[4] Urban Imperviousness. (2001) Multi-Resolution Land Characteristics Consortium http://www.mrlc.gov/nlcd_multizone_map.php

percentage of impervious surface covering the land area. Figure 2 provides a close up of impervious surface data in Chapel Hill. Roads, buildings, urban centers, and neighborhoods can be seen with the impervious cover data.

Figure 1 - Impervious Surface data for Chapel Hill, NC

The *Watershed Assessment* project relies on these impervious cover data to calculate the percentage of impervious surfaces within watersheds, municipal boundaries, and county jurisdictions. The procedure involves averaging the total 30m by 30m data for each natural and political boundary. Figure 3 shows a representation of how this was done for a 12 digit watershed and enclosed municipal boundaries. The data show this watershed is 8.5% impervious. The area of Chapel Hill within the watershed is 12.6% impervious.

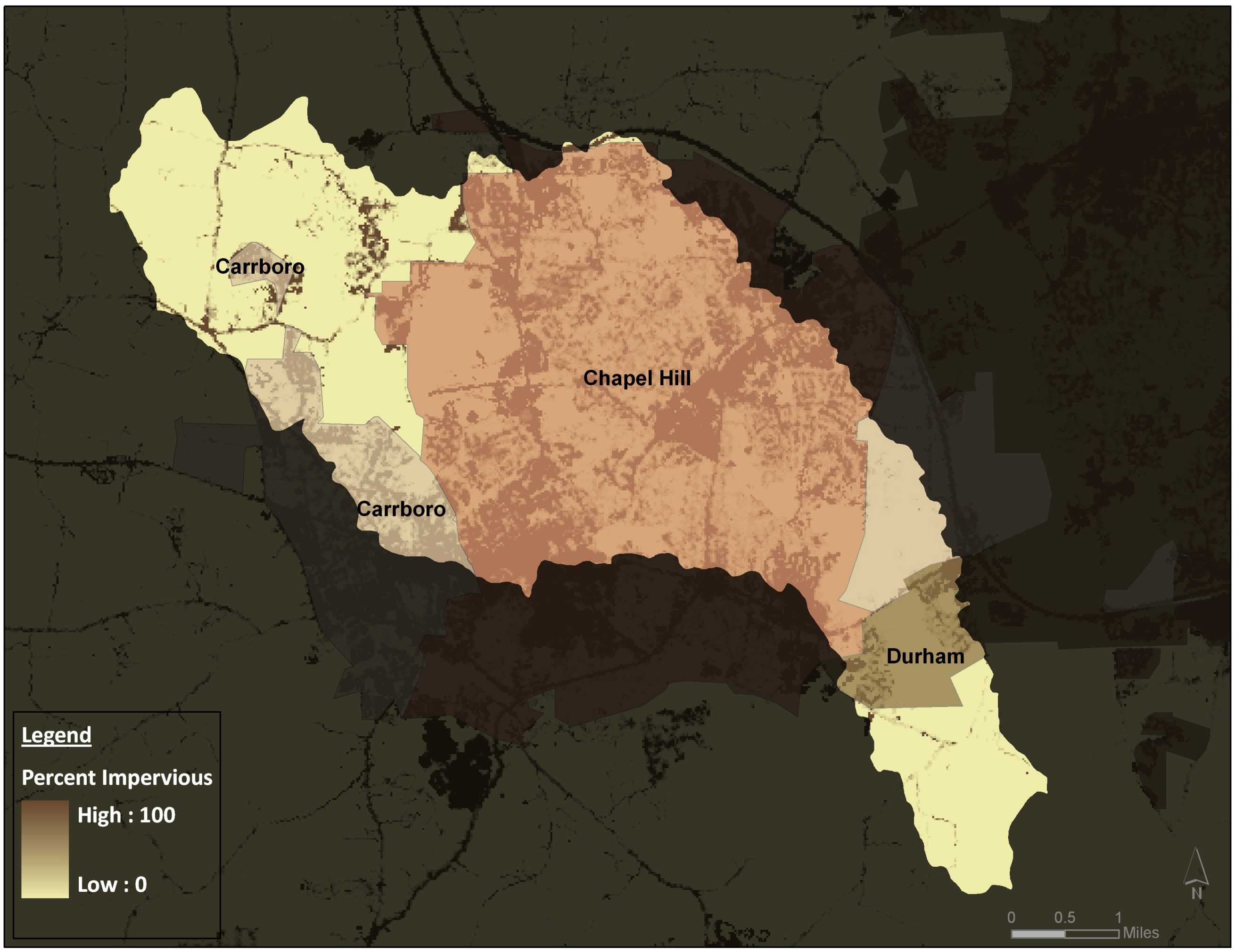

Figure 2- Impervious surface data broken up by natural and political boundaries.

A method was developed for testing the accuracy of the MRLC satellite imagery. A series of grids were created to provide a sampling scheme for the impervious surface raster data. Grids were used at several resolutions to provide a scale of reference. The satellite raster data was sampled at the 30m by 30m, 90m by 90m, 300m by 300m, and 900m by 900m scales.

Planimetric data from the Town of Chapel Hill was also sampled with the different scale grids. The planimetirc data is created from digitizing site plans and is the most accurate source of impervious cover data available. Figure 4 shows the two data types for the same area. The samples were then compared to determine accuracy of the MRLC impervious cover data.

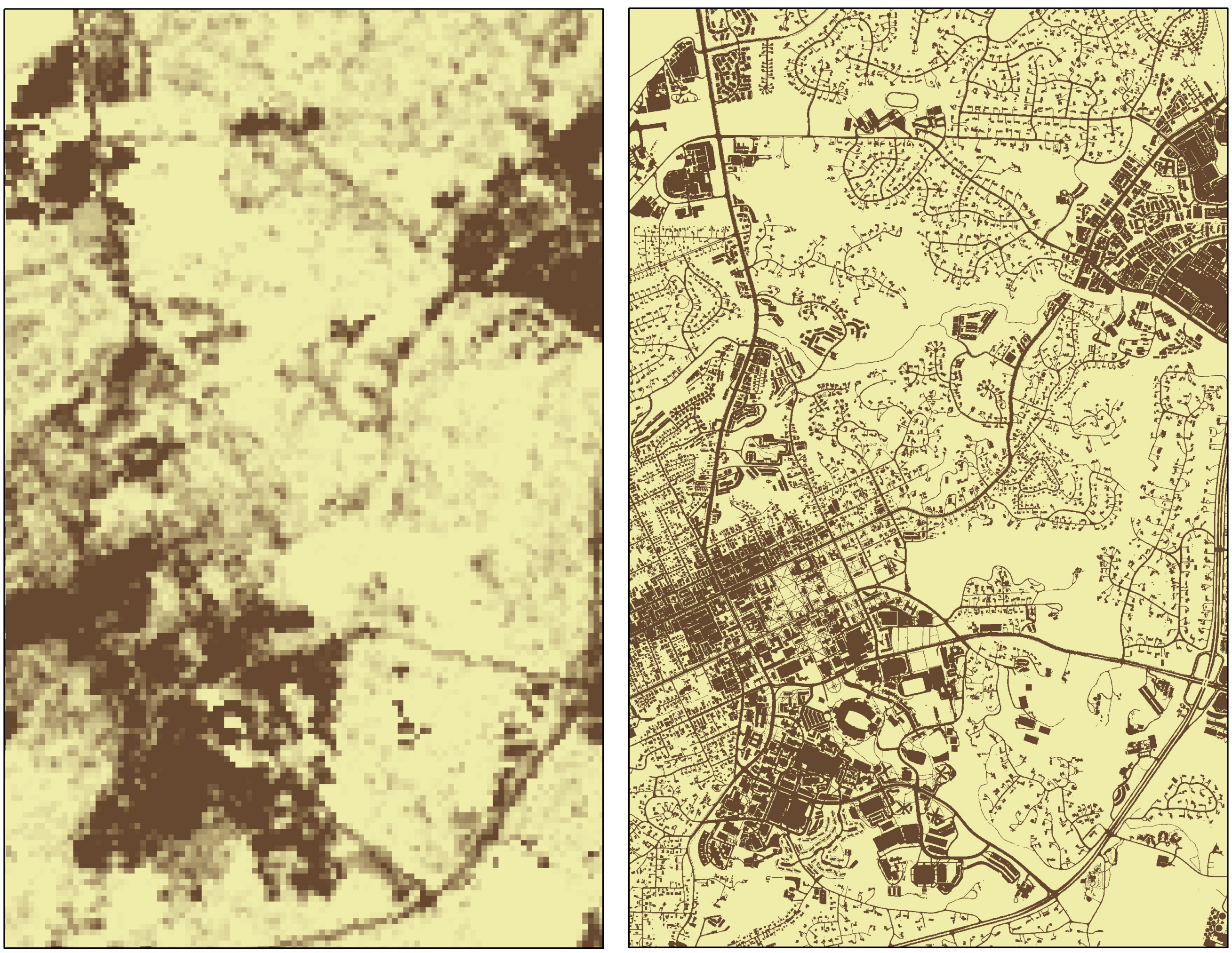

Figure 3 - A comparison of the satellite imagery with planimetirc data

Results

A statistical analysis was conducted on the samples to determine the accuracy of the MRLC satellite imagery. It was determined that the relationship between the MRLC and the planimetric data was present (R^2=0.74). Samples that deviated more than 2 times the standard deviation of the residuals were removed from the analysis. This is was due to the assumption that these samples were the result of the time difference between the MRLC (2001) and the planimetric data (2008). The remaining data were used to determine the mean error and mean absolute error of the samples. The mean error is the average of the difference between the planimetric data and the MRLC data. This number indicates the systematic error between the data sets. Specifically, it describes any over-estimation or under-estimation of the impervious surface percentage.[5] The mean absolute error is the average absolute error between the data sets and indicates the overall accuracy of the estimation. Table 1 summarizes the analysis performed on the data sets.

All data	Mean Error	Mean Absolute Error	R^2	Sample size
Cell Size				
30m	-5.5	8.6	0.74	61,687
90m	-6.6	7.9	0.81	9,199
300m	-7.4	7.8	0.82	771
900m	-8.5	8.6	0.83	72
Subbasin (Ave. 561,600 m^2)	-6.0	7.9	0.53	118

Table 1 - Statistical analysis for the different levels of sampling of the MRLC data.

The mean error of the analysis shows that the MRLC slightly under-estimates the percentage of impervious surface at all scales. The MRLC has a 7.8 to 8.6 mean estimation accuracy error for the different scales. The accuracy of the MRLC in this analysis is

[5] Canters, F. et al. Effects of different methods for estimating impervious surface cover on runoff estimation at catchment level. 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.

consistent with the USGS report on the accuracy of the data.[6] Since the focus of the data is to identify watersheds with between 0 and 30 percent impervious surface, this analysis included a sub-sampling of the data in this range. Table 2 provides the results of the sub-sample analysis. Both the mean error and mean absolute error are slightly higher when focusing on areas with less that 30 percent impervious cover.

Impervious Surface <=30	Mean Error	Mean Absolute Error
Cell Size		
30m	-5.7	7.9
90m	-6.9	7.9
300m	-7.6	7.9
900m	-8.5	8.6
Subbasin (Ave. 561,600 m^2)	-6.6	8.0

Table 2 - Statistical analysis for the MRLC data where actual impervious cover is below 30%.

Finally, a detailed analysis was performed to determine the relative accuracy of the estimates at exact levels of impervious surface under 30 percent. Appendix I shows the results of this analysis. Figure 4 shows the results of the detailed analysis of the 30m sample. The analysis of the other sample levels demonstrate similar trends.

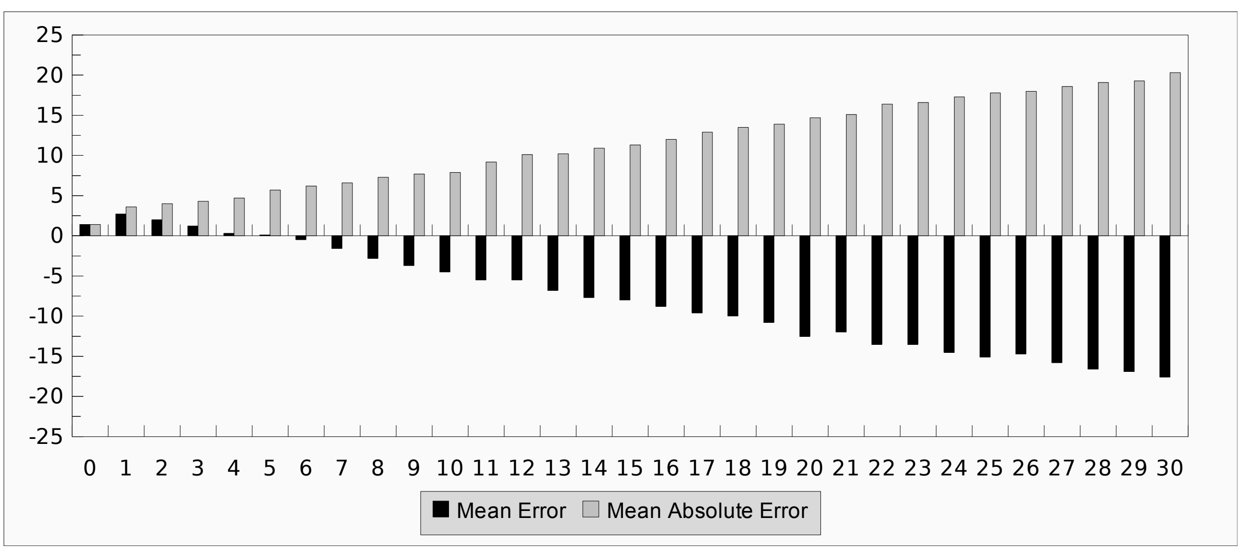

Figure 4 – The level of error at exact percentages imperviousness of the 30m sample.

[6] Yang, L. et al. An approach for mapping large-area impervious surfaces. United States Geological Survey.

Data Sources

Multi-Resolution Land Characteristics Consortium. 2001. Urban Imperviousness. http://www.mrlc.gov/nlcd_multizone_map.php

Town of Chapel Hill. 2006. Chapel Hill Subbasins. From UNC-Chapel Hill Library, for student use.

Town of Chapel Hill. 2008. Impervious Surfaces. From UNC-Chapel Hill Library, for student use.

Procedure

1. Produced four grids for sampling at 30mX30m, 90mX90m, 300mX300m, and 900mX900m sizes using Hawth's Tools, which are available for download.
2. Selected subbasins that contained entire coverage of Town of Chapel Hill's impervious surface layer (TOCH_IS). This resulted in 127 subasins for the study.
3. Selected grid squares from the four grid layers that were completely within the subbasin study area. This created four grid layers that could be used to sample both the TOCH_IS and MRLC data at different scales.
4. Converted the MRLC data from a raster layer to a shapefile layer.
5. Intersected the MRLC layer with the four grid layers. This created four MRLC layers with the scaled grid identifiers.

For the 30mX30m sample grid the following procedure was used:

6. Intersected the TOCH_IS layer with the sample grid layer.
7. Recalculated the area of the TOCH_IS layer in meters.
8. Created a summary statistics table for the TOCH_IS layer which summed the impervious area by grid identifier.
9. Intersected the MRLC shapefile with the grid layer.

10. Used Hawth's Tools to create a random selection of 65,500 reticules. This is due to Microsoft Excel's limitation for number of records it can process.
11. Created a summary statistics table for the MRLC layer which identified the impervious cover percentage for each grid.

Each of the remaining sample grid layers went through the follow procedure:

12. Intersected the TOCH_IS layer with the grid layer.
13. Recalculated the area of the TOCH_IS layer in meters.
14. Created a summary statistics table for the TOCH_IS layer which summed the impervious area by grid identifier.
15. Intersected the MRLC shapefile with the grid layer.
16. Created a summary statistics table for the MRLC layer which calculated the mean of the MRLC impervious cover data by grid identifier.

APPENDIX I – Detailed analysis of MRLC data

Table A. Analysis of 30m samples

% Imperviousness	Mean Error	Mean Absolute Error	Observations
0	1.4	1.4	28754
1	2.7	3.6	1066
2	2	4	799
3	1.2	4.3	689
4	0.3	4.7	632
5	0.1	5.7	616
6	-0.5	6.2	606
7	-1.6	6.6	600
8	-2.8	7.3	591
9	-3.7	7.7	635
10	-4.5	7.9	620
11	-5.5	9.2	612
12	-5.5	10.1	561
13	-6.8	10.2	551
14	-7.7	10.9	576
15	-8	11.3	533
16	-8.8	12	567
17	-9.6	12.9	543
18	-10	13.5	613
19	-10.8	13.9	666
20	-12.5	14.7	620
21	-12	15.1	681
22	-13.5	16.4	648
23	-13.5	16.6	669
24	-14.5	17.3	602
25	-15.1	17.8	659
26	-14.7	18	610
27	-15.8	18.6	651
28	-16.6	19.1	609
29	-16.9	19.3	546
30	-17.6	20.3	631

Table B. Analysis of 90m samples

% Imperviousness	Mean Error	Mean Absolute Error	Observations
0	0.6	0.6	2267
1	0.7	1.8	221
2	-0.2	2.2	199
3	-0.3	3.1	197
4	-2.3	3.3	152
5	-2.5	4.2	179
6	-3.1	4.8	193
7	-3.8	5.2	203
8	-3.8	6.4	171
9	-4.4	6.3	194
10	-6.3	7.5	178
11	-6.2	7.6	172
12	-6.5	8.5	173
13	-7.6	9.4	166
14	-7.6	9.1	159
15	-8.9	10.1	177
16	-10.3	11.2	168
17	-10.9	12	189
18	-10.9	11.9	176
19	-12	12.9	169
20	-11.1	12.7	189
21	-13.6	14.3	177
22	-12.4	13.6	173
23	-13.3	14.3	141
24	-14.6	12.8	163
25	-15	15.5	153
26	-14.4	15.8	143
27	-15.2	16	136
28	-14.7	15.3	131
29	-16.2	16.6	133
30	-16.4	16.5	144

Table C. Analysis of 300m samples

% Imperviousness	Mean Error	Mean Absolute Error	Observations
0	0.1	0.1	54
1	-0.3	1.2	18
2	-0.8	2	30
3	-1.5	2.9	21
4	-2.4	2.5	18
5	-2.8	3	15
6	-4	4.4	16
7	-4.6	5.7	19
8	-4	5.2	11
9	-5.6	7.2	14
10	-5.4	5.4	12
11	-6.4	6.6	22
12	-8.2	8.2	25
13	-7.5	8	21
14	-8.3	8.7	13
15	-9.6	9.6	23
16	-11.1	11.1	22
17	-9.8	9.8	18
18	-10.1	10.1	24
19	-9.8	9.9	23
20	-13.9	13.9	19
21	-11.5	11.7	29
22	-12	12.1	21
23	-11.7	11.8	33
24	-13.4	13.5	18
25	-13	13	15
26	-7.5	7.6	7
27	-11.7	11.7	20
28	-12	12	12
29	-10.9	10.9	9
30	-10.9	11.3	8

NOTE: It was necessary to analyze ranges of imperviousness at the 900m and Subbasin level due to smaller number of observations.

Table D. Analysis of 900m samples

% Imperviousness	Mean Error	Mean Absolute Error	Observations
0-5	-1.3	1.3	3
5-10	-4.5	4.5	9
10-15	-7	7.4	10
15-20	-9	9	14
20-25	-11	11	17
25-30	-10.1	10.1	7

Table E. Analysis of Subbasin samples

% Imperviousness	Mean Error	Mean Absolute Error	Observations
0-5	0.3	3.2	20
5-10	-2.4	4.3	16
10-15	-5.2	7	21
15-20	-9.1	10.6	22
20-25	-9.3	10.9	22
25-30	-7.8	8.1	5