Evaluation of a Moderate Resolution, Satellite-Based Impervious Surface Map Using an Independent, High-Resolution Validation Dataset.

John W. Jones[1] and Taylor Jarnagin[2]

Abstract: Given the relatively high cost of mapping impervious surfaces at regional scales, substantial effort is being expended in the development of moderate-resolution, satellite-based methods for estimating impervious surface area (ISA). To rigorously assess the accuracy of these data products high quality, independently derived validation data are needed. High-resolution data were collected across a gradient of development within the Mid-Atlantic region to assess the accuracy of National Land Cover Data (NLCD) Landsat-based ISA estimates. Absolute error (satellite predicted area - "reference area") and Relative Error [satellite (predicted area – "reference area")/"reference area"] were calculated for each of 240 sample regions that are each more than 15 Landsat pixels on a side. The ability to compile and examine ancillary data in a geographic information system environment provided for evaluation of both validation and NLCD data and afforded efficient exploration of observed errors. In a minority of cases, errors could be explained by temporal discontinuities between the date of satellite image capture and validation source data in rapidly changing places. In others, errors were created by vegetation cover over impervious surfaces and by other factors that bias the satellite processing algorithms. On average in the Mid-Atlantic region, the NLCD product underestimates ISA by approximately 5%. While the error range varies between 2 and 8%, this underestimation occurs regardless of development intensity. Through such analyses the errors, strengths, and weaknesses of particular satellite products can be explored to suggest appropriate uses for regional, satellite-based data in rapidly developing areas of environmental significance.

CE database subject headings: Remote Sensing, Satellites, Land Usage, Hydrologic Properties, Accuracy, Standardization

[1] Research Geographer, U.S. Geological Survey, Eastern Geographic Science Center, 521 National Center, Reston, VA 20132. (corresponding author). E-mail: jwjones@usgs.gov
[2] Research Ecologist, Environmental Protection Agency, Landscape Ecology Branch, USEPA/ORD National Exposure Laboratory, Mail Drop E243-05, 109 T.W. Alexander Dr., Research Triangle Park, NC 27711. E-mail: jarnagin.taylor@epa.gov

INTRODUCTION

Increases in watershed impervious surface area (ISA) have long been known to affect the quantity and timing of watershed runoff (Hammer 1972; Dunne and Leopold 1978; Jennings and Jarnagin 2002). Studies of aquatic ecology and water quality have shown that increases in watershed ISA adversely impact stream biological integrity (Schueler 1994). Given these linkages, ISA has been proposed (Arnold and Gibbons 1996) and widely adopted (CWP 2003) as an important, integrative environmental indicator. While popular for this purpose, periodically mapping and quantifying ISA over large areas is expensive. This expense has resulted in a large body of research regarding the use of satellite remote sensing to map impervious surfaces (Slonecker, et al. 2001). Generally, two approaches have been employed: the assignment of ISA values to land cover classes using coefficients, and the extraction of pixel or sub-pixel estimates of ISA directly from satellite data. When reported, assessments of the accuracy of satellite-based ISA mapping algorithms have varied in approach and rigor. In some cases, data used for satellite data processing algorithm development and implementation are resampled for accuracy assessment (Yang, et al. 2003; Yang 2005; Canters, et al. 2006). In others, a set of validation data were collected for comparison against satellite-generated results (Slonecker and Tilley 2004; Lu and Weng 2006). Evaluation of ISA estimate impacts on hydrologic modeling and water quality monitoring efforts has only recently begun to be reported in the literature (Canters, et al. 2006; McMahon 2007). But before a complete understanding of the implications of ISA accuracy in these contexts can be achieved, a clear understanding of ISA source data accuracy across a variety of physiographies and places with different development histories is needed. This research was aimed at the development and testing of generally applicable, rigorous, but cost effective ISA accuracy assessment techniques. To provide an objective, independent, and standardized means of evaluating the accuracy and appropriate uses of satellite-based estimates of ISA, the U.S. Geological Survey (USGS) developed a protocol for high-resolution ISA/land use data collection and analysis in a collaborative fashion (Jones, et al. 2003).

For the purpose of protocol development, the term impervious surface is applied to any earth surface cover that, on a permanent basis, prevents water (i.e., rainfall and runoff from adjacent surfaces) from infiltrating into the soil directly below. To allow flexibility, the protocol includes a hierarchical classification system for impervious surfaces that allows for maximum information content where possible and the creation of uniform "common denominator" data where necessary. At the highest level of the hierarchy, a surface is either pervious or impervious. At the next, it is either natural, or anthropogenic. At the third level, impervious surface features are coded by type (e.g., rooftop, driveway, parking lot, sidewalk, etc.). And at the next level, they are attributed with a land use class (e.g., industrial, residential, commercial). Finally, an opportunity exists for interpreters to note attributes of individual impervious surface polygons as needed. This is helpful when interpreters encounter particular features or land uses with unusual characteristics that prove troublesome for some satellite-based mapping algorithms, such as green-colored asphalt tennis courts. To create a consistent data product regardless of study area, source data that meet common data standards must be used. At a minimum, digital orthophotography (DO) that meets USGS standards is employed. Impervious surfaces visible in DO within areas with ground dimensions of 500 meters on a side are delineated using the classification scheme. Each 500 meter by 500 meter area is referred to as a "chip". For each chip, a variety of standard products are produced. The impervious surface layer, including the type and use classifications is created with all appropriate topological encoding necessary to conduct geographic information system (GIS) analysis. For each chip, a MS Word file is also created to store sufficient information to populate Federal Geographic Data Committee compliant metadata records. In addition, each metadata file contains important information such as the date of photography for the source DO, date of DO production, interpreter ID code, start and end dates for compilation, ancillary data sets used, and miscellaneous notes regarding particular validation chips. This level of detail was very useful when unraveling observed inconsistencies in results, pushing mapping methods to their limits, or understanding the strengths and weaknesses of particular methods.

The ISA calibration/validation protocol consists of sampling (described in the next section), data collection, quality assessment and assurance, metadata creation, and analysis specifications. It was applied to the evaluation of USGS National Land Cover Database (NLCD) ISA data for a sub-region of the Mid-Atlantic to assess the NLCD data's accuracy and demonstrate its utility. For environmental monitoring purposes, the amount of impervious surface over areas such as a watershed, zoning areas, or other planning units are of interest. Therefore, for this analysis, we focus on assessing the over or under-prediction of the sum of ISA over individual chips. We term this measure "total impervious area" or TIA.

METHODS

Input Data

The District of Columbia is the geographic center of the study area that covers portions of the Piedmont and Coastal Plain physiographic provinces in the Mid-Atlantic where 240 sample points were randomly selected (Fig. 1).

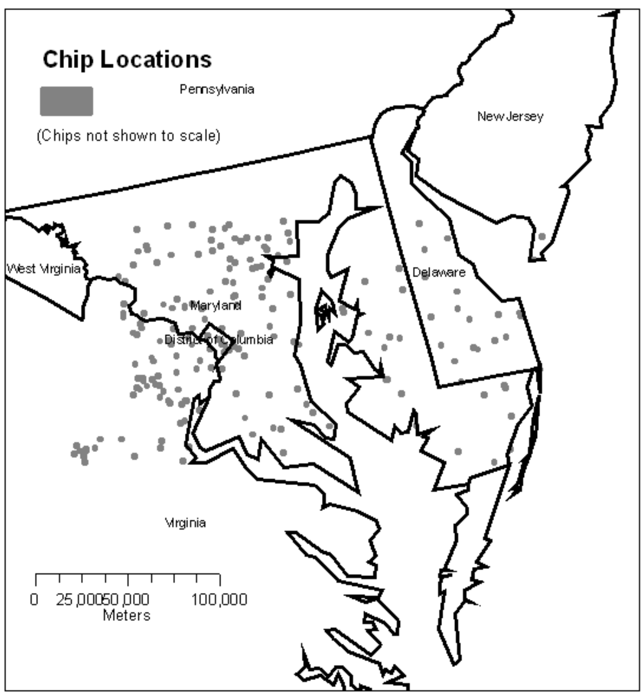

Fig. 1. Study area and sample locations.

As dictated by the protocol, chips (i.e., previously mentioned polygons with ground dimensions of 500 meters on a side) were delineated around each of these points to form the boundaries of the regions for which impervious surfaces would be hand delineated and attributed with land use information. Areas 450 meters on a side would include 225 Landsat 30 meter pixels – a sufficient size to create statistical power in the analysis of Landsat data for each individual chip. The 500-meter dimension was designated as the chip standard to ensure that interpreters had additional context during interpretation and to account for issues of satellite data registration with each ground area. DO with spatial resolution of 1 meter or better was assembled for each sample chip. Every effort was made to obtain digital orthophotography with source image dates close to those of the Landsat data used for NLCD production. However, this was not always possible. A majority of the DO was created from imagery collected circa 2000, but source imagery dates for the DO source ranged from 1988 through 2002. Any impervious surface feature wider than 1 meter with a minimum area of 10 square meters was captured in the ISA database (Fig. 2). For a subset of chips, sidewalks and other features narrower than 1 m have also been collected for research on the impact these small features may have on remote sensing and hydrology. However, that analysis is not reported here and sidewalks were not included in this analysis. Additional details of the protocol, data collection, metadata production, software tools used and QA/QC are provided elsewhere (Jones, 2008). For the satellite derived ISA estimates, the NLCD product was used (Yang, et al. 2003). NLCD ISA data for a period around 2000 are available for the conterminous United States and have a nominal spatial resolution of 30 m for which sub-pixel impervious surface area is estimated using regression tree analysis of multi-season Landsat data (Yang, et al. 2003). All NLCD data layers are created on a mapping zone basis and data from the USGS Mapping Zone 60, which were created from Landsat data spanning the period of July 1999 - April 2001 were used for this analysis (Fig. 3).

Sampling distribution assessment

Every aspect of the project, that is data sampling, collection, and analysis, was conducted in the GIS framework. This allowed for various spatial manipulations and examination of sample and error distributions. To quickly assess whether a representative sample of the region's development gradient was obtained, the NLCD land cover data were aggregated for each chip. Pixels classified in the NLCD as "developed" that is, categories 21, 22, 23, and 24 for the 2001 NLCD (Homer, et al. 2004) were summed and divided by total number of cells per chip for a rough estimate of developed area. Each chip was then classified into 1 of 6 development classes based on the following aggregated developed area expressed as a percent (Table 1, first column). In addition, each chip was tagged with the EPA ecoregion designation (i.e., “Piedmont” or “Coastal”) in which it is found.

Fig. 2. Examples of delineated reference data for Annandale, Va. Areas interpreted as impervious are surrounded with white lines. USGS submeter resolution, digital orthophotography are shown as source.

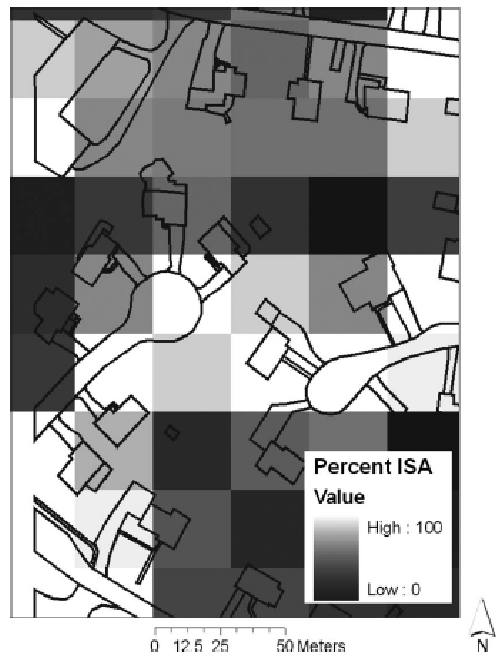

Fig. 3. Delineated reference data (black lines) displayed over 2001 NLCD ISA data for same area depicted in Fig. 2. Individual NLCD impervious surface data pixels are visible.

Table 1. Counts of Number of Chips Classified into Each Development Category Based on Grouping of NLCD Land Cover Data into Six Development Density Categories and the Number of Samples per Development Category as Function of EPA Ecoregion (i.e., Piedmont or Coastal Plain).

Development Category	Samples	Piedmont	Coastal Plain
None (0)	46	23	23
Rural (0< - <10%)	41	25	16
Exurban (10 - <30%)	35	19	16
Suburban (30 - <50%)	33	22	11
Dense_Suburban (50 - <80%)	39	32	7
Urban (80 – 100%)	46	22	24
Total:	240	143	97

Error Calculation

We focused on the accuracy of NLCD ISA estimates over each sample chip area (TIA). A GIS was used to aggregate Zone 60 NLCD pixel-based estimates of percent impervious and visually delineated impervious surfaces to chip total impervious area (i.e., TIA) from each source. Then, the Actual Error was calculated:

$$\mathrm{AE} = P - R \qquad (1)$$

where AE = Actual Error of a given chip, P = NLCD-estimated TIA (“predicted”) and R = high-resolution, chip-estimated TIA (“reference”). Relative Error was also calculated for each chip:

$$\mathrm{RE} = (P - R) / R \qquad (2)$$

where RE = Relative Error of a given chip, and P and R are as previously defined. Finally, the absolute value of Actual Error was calculated for each chip and ranked across all chips:

$$\mathrm{AAE} = |\mathrm{AE}| \qquad (3)$$

where AAE = the absolute value of the Actual Error (i.e., AE defined above) calculated for each chip.

Thresholds for Outlier Identification

Because it would be difficult to visually examine every chip for discrepancies, as a first cut, a threshold of 20% error was used to easily and objectively identify outliers. Error at this level is more likely to result in chip misclassification among development categories shown in Table 1. Then, chips with unusual differences between predicted and reference ISA, such as a predicted ISA of 10% and a measured reference ISA of 0, were added to the outlier list for further scrutiny. A 10% threshold of watershed imperviousness has been suggested for the categorization of impaired stream ecological integrity (Arnold and Gibbons 1996; CWP 2003). Finally, some relatively undeveloped areas where minor development was known to have occurred were also flagged for further examination. Causes for objectively identified outlier chips were then explored through visual examination of the ISA delineation, associated metadata and the impervious surfaces mapped in the NLCD data set. Since the analysis was conducted within the GIS environment, additional ancillary information could be assembled for outlier chips, sometimes providing explanations for gross errors that are not readily observed in the database itself. Through such analyses, chips with large error caused by reference imagery/NLCD temporal differences (i.e., chips where impervious surface area had changed between the date of chip source orthophotography and satellite imaging) were eliminated before the minimum, maximum, average and spread of the actual and relative differences were generated across the remainder of the chip set. Summary assessment of NLCD TIA accuracy was then possible for this region.

RESULTS and DISCUSSION

Sample distribution and calculated errors

Table 1 shows the number of sample chips in each development class that resulted from the post-selection stratification of the random samples. A sufficient number of chips were delineated for each development category and a fairly even distribution of chips was achieved among development categories. This gave us confidence that no particular level of development within the overall study region was underrepresented and that our summary assessment of NLCD TIA accuracy would not be biased by our set of chip test areas. Table 1 also shows the total number of samples in each ecoregion as well as the number of samples by development class. In this case, statistical power was insufficient to allow comparisons across development classes within ecoregions. For this reason, we report summaries and comparisons for development classes across the entire study area and for total observations within ecoregions.

Table 2. Small Excerpt from Chip Database Showing Variables Used in Analysis.

Chip identifier	Predicted TIA %	Reference TIA%	AE %	RE %
manassas_va_se_c	4.63	8.38	-3.76	-44.83
waterford_va_sw_a	4.21	6.01	-1.80	-29.93
falls_church_va_nw_a	3.08	10.14	-7.06	-69.60
germantown_md_ne_a	5.38	30.70	-25.32	-82.48

Note: "Chip" denotes a 500 meter by 500 meter test area. "TIA" refers to total impervious area in the chip. "AE" refers to the Actual Error, i.e., predicted minus the reference total impervious area [Eq. (1)]; "RE" refers to the Relative Error, the predicted minus the reference divided by reference total impervious area [Eq. (2)].

A sample of chip identifiers, predicted TIA, reference TIA, as well as actual and Relative Errors by chip is shown in Table 2. Relative Error indicates how error varies as a percent of actual impervious area within the chip. Because of its implications for hydrologic modeling and for planning when TIA thresholds are being used, actual error is the measure of greatest importance in understanding the utility of the data. Summary statistics on Actual Error by ecoregion are provided as Table 3.

Table 3. Tabulation of Summary Statistics on TIA Actual Error as Function of Ecoregion for Entire Data Set (N = 240).

Summary statistic AE	Piedmont %	Coastal Plain %
Max under-prediction	-54.46	-25.85
Max over-prediction	38.89	36.98
Mean error	-5.83	-1.53
Error standard deviation	9.98	6.58
Mean absolute error	8.01	4.35

The Piedmont ecoregion exhibits higher maximum under-prediction and average error than the Coastal ecoregion and the means of their Actual Error are statistically significantly different (p-value = 0.0002). The Piedmont is an area of more complex relief, mixed land uses, and mixed hardwood tree canopies. These factors constitute challenges for moderate resolution remote sensing of ISA. Statistics on Actual Error for the entire data set are provided in the middle column of Table 4.

Table 4. Tabulation of Summary Statistics on TIA Actual Error for Entire Data Set (Middle Column) and when Outliers Caused by Temporal Miss-matches (i.e., Reference Chips out of Date Compared to NLCD imagery) are Removed (Right Column).

Actual Error	Value in percent (*n* = 240) (%)	Value in percent (*n* = 226) (%)
Max underprediction	-54.46	-25.85
Max overprediction	38.89	36.98
Mean	-4.09	-4.41
Median	-3.22	-3.28
Error standard deviation	9.00	6.62
Mean absolute	6.53	5.67

The absolute values of the Actual Errors were ranked (Fig. 4) to gain an understanding of their frequency distribution. Across the entire study region, our thresholds of 10 and 20% error were exceeded by 49 and 12 of the 240 chips or approximately 20 and 5% of the time, respectively.

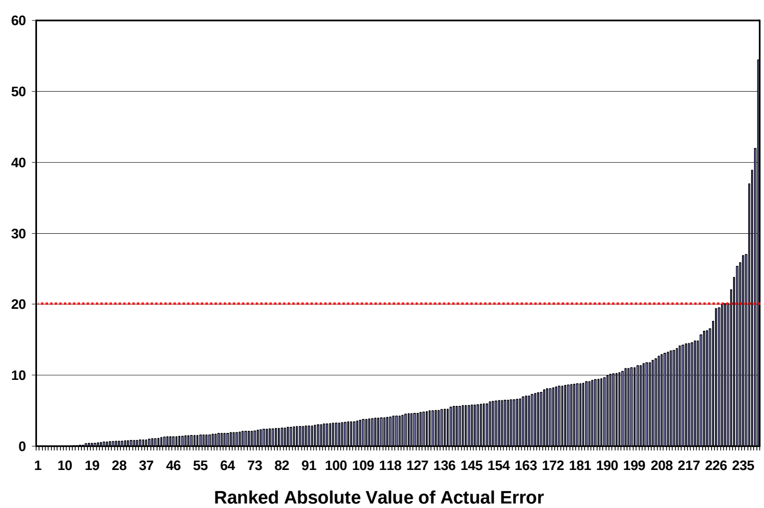

Fig. 1 Absolute value of chip actual error sorted by rank. The red line demarks chips that surpass the 20% error threshold.

A plot of reference vs. predicted TIA by chip in which automatically and subjectively selected outliers are identified is provided as Fig. 5. A larger number of predicted over-estimations of TIA were selected for visual examination than were for the opposite case (reference percentages higher than predicted), in juxtaposition to the nature of the error on the whole. Predicted TIA generally is below that measured in the validation data set, but these under-predictions are typically small in magnitude.

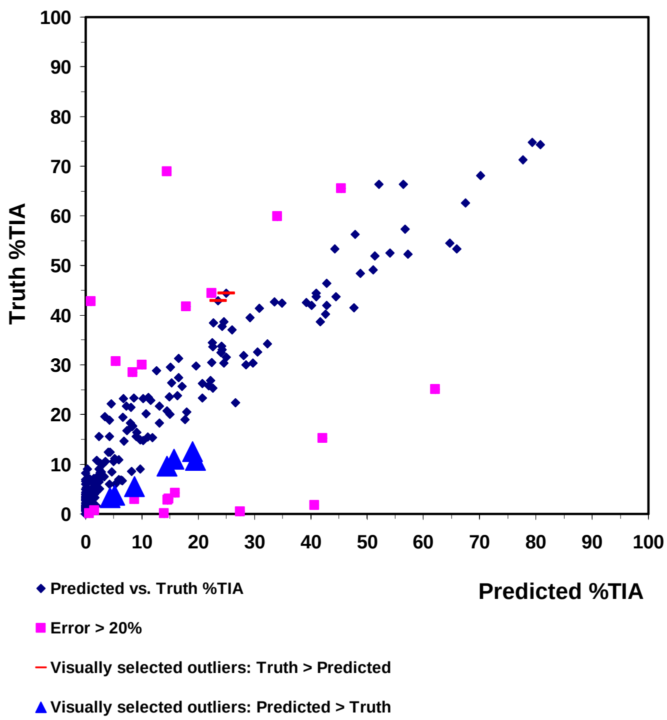

Fig. 2. Scatter plot of predicted verses truth TIA estimates by chip. Chips with errors exceeding the 20% threshold are shown as squares. Chips added on the basis of visual analysis where truth exceeded predicted and predicted exceeded truth are shown by horizontal bars and triangles, respectively.

Close examination of high error cases

The source imagery, delineated "reference" impervious surface GIS data, and the satellite predicted ISA were each visually scrutinized along with additional information such as available (unrectified) airborne imagery, cadastral data, or other collateral information in an effort to determine why large errors were measured. However, in 4 of the 240 original cases, no additional information was available to determine whether the reference or predicted ISA estimate appeared to be the source of the large error value. These chips were therefore eliminated from the analysis. In 10 cases, examination of ancillary data confirmed that changes had occurred within the chip between validation source image and satellite capture and the only available orthorectified source imagery for reference ISA delineation was out of date when compared to the date of the NLCD image collection. These chips were also eliminated. In one unusual case (the chip identified as “Frederick_md_se_a”) a predicted value more than 20% lower than the reference estimate led to the discovery that bright parking lots nearby construction sites that were visible in 1988 source imagery were eventually replaced by grass and shrub – leading to an “under prediction” by the NLCD data. This was one of the reference chips that was eliminated from the analysis. In 3 of 8 cases ancillary data (including current unrectified airborne imagery) indicated that ISA had increased and the NLCD correctly predicted a higher amount of ISA. These cases are noted by an “X” in the reference column and “H” in the “higher/lower” column of Table 5. This would be expected since impervious surface area rarely decreases as time progresses. Unfortunately, time-intensive update of the ISA delineation would be needed to further quantify the error for these chips. However, for five reference chip problem cases the ancillary data showed an increase in ISA while the NLCD prediction was lower than the obviously out-of-date reference ISA estimate. These chips are noted by an “X” in the reference column and “L” in the “higher/lower” column of Table 5. In these cases, measured underestimation errors would actually increase with an update of the reference data. And for 12 of the outlier cases, the reference data were correct but the NLCD prediction was certainly in error. Demarked by an “X” in the predicted column of Table 5, the nature of the errors, that is over or under-prediction are noted in the “higher/lower” column by an “H” or “L”, respectively. Various explanations apparent at the chip level are possible for each result. For example, another of the greatest outliers was the Sparrow Point MD chip, centered on a peninsula at the mouth of the Patapsco River and Baltimore Harbor. The bright reflectance of the sand and gravel substrate in this coastal and industrial area resulted in an overestimation of ISA by the NLCD. A detailed QA/QC process was applied to every chip. Examination of contemporary airborne imagery for the chip showed that it remains mostly bare ground or lightly vegetated – creating a source of confusion for satellite based ISA mapping algorithm and resulting in an overestimation of ISA.

Table 5. Evaluation of Potential Discrepancies between Dates of Reference Source Imagery and NLCD Satellite Imagery.

Chip Identifier	P	R	No information	Higher/Lower
Annandale_va_se_d_2002	X			L
Arcola_va_ne_a_2002	X			H
Buckeystown_md_nw_a_1988			X	(H)
Damascus_md_nw_a_1988		X		H
Deale_md_se_a_1994			X	(H)
Ellicott_city_md_se_a_1994		X		L
Fairmount_md_sw_a_1992			X	(H)
falls_church_va_nw_b_2002	X			L
Finksburg_md_se_a_1995		X		H
Frederick_md_se_a_1988	X			L
gainesville_va_se_b_1994		X		H
germantown_md_ne_a_2001	X			L
independent_hill_va_se_a_1994		X		L
independent_hill_va_se_b_1994		X		L
Laurel_md_nw_b_2002	X			L
Leesburg_va_se_b_1995	X			H
Nokesville_va_ne_a_1994		X		H
piscataway_md_sw_a_2002	X			L
Rockville_md_ne_c_2002	X			L
Sparrows_point_md_nw_a_2002	X			H
Stafford_va_nw_a_1994			X	(H)
Urbana_md_se_a_2001	X			L
walkersville_md_sw_b_1988			X	(H)
washington_west_dc_sw_a_2002	X			L
White_marsh_md_sw_a_1994		X		H
Total:	12	8	5	8H/12L

An "X" under "P": confirmed predicted TIA error; An "X" under "R": reference source imagery out-of-date. An "X" under "No information": no concurrent/newer ancillary data are available; "L": prediction underestimates reference ISA; "H": predicted overestimates ISA; "(H)": predicted value likely but unconfirmed.

Importantly however, in the majority (10 of 12) of predicted error outlier cases the NLCD under-predicted TIA. The greatest under-prediction errors often occurred in the oldest developed areas where tree canopy and other vegetation are relatively mature. This reflects the additional challenges that mixed vegetation/impervious surface satellite pixels pose for satellite-based ISA mapping algorithms – even when, as in the NLCD case, multi-season satellite data are used to create the ISA estimates.

Results with problem reference chips removed

Changes to the sample sizes for each development category given the removal of "problem chips", that is, chips for which the reference was determined to be incorrect or unverifiable given available ancillary data are provided as Table 6.

Table 6. Original and Reduced Sample Sizes within Development Categories Following Screening Process to Remove Chips where Reference Data were Out-of-Date.

Development Category	Original sample size	Samples dropped	Reduced sample size
None (0)	46	1	45
Rural (0< - <10%)	41	0	41
Exurban (10 - <30%)	35	1	34
Suburban (30 - <50%)	33	7	26
Dense_Suburban (50 - <80%)	39	4	35
Urban (80 - 100%)	46	1	45
Total:	240	14	226

Note: The "Suburban" and "Dense Suburban" categories appropriately saw the greatest reductions as the places where most development has recently occurred.

As noted, the total sample size was decreased to 226. The two categories for which the greatest reduction in samples occurred were the suburban and dense suburban development classes. This was appropriate as it is logical to expect that these would be the areas of greatest change in the sometimes short time frames between reference and satellite imagery. Summary statistics on the difference between predicted and reference TIA following the removal of the problem chips is provided in Table 4. Even with the removal of "problem chips", the summary statistics on the errors were relatively unaffected. The variance of the error decreased slightly. Because some chips suggesting large over-prediction (that is, high positive Actual Error) were removed through the screening process, the negative average and median differences between predicted and reference actually increased in magnitude slightly (Table 4), nudging the average difference closer to -5%. This means that on average, the NLCD ISA data under-predict TIA by 5%. This average error is considerably lower than the approximately 10% that has been typically reported in the literature (Canters, et al. 2006). Our analysis design also allows us to determine whether the average 5% under-prediction is consistent across a gradient of development.

Error as a function of NLCD-derived development class

Error was examined as a function of development class to determine whether NLCD TIA estimates are affected by relative abundance of impervious surfaces. Statistics regarding Actual Error (equation 1) by NLCD development class are shown in Table 7, Table 8, and Figure 6.

Table 7. Actual Error [Eq. (1)] statistics as a function of development class for the reduced sample set (N = 226).

Actual Error	None	Rural	Exurban	Suburban	Dense Suburban	Urban
Max under-prediction	-8.24	-8.8	-25.32	-16.53	-23.8	-25.85
Max over-prediction	0	0.41	1.03	4.87	12.66	36.98
Mean	-1.67	-3.07	-5.72	-7.94	-3.07	-2.16
Median error	-1.04	-2.92	-5.03	-6.63	-8.57	-3.05
Standard deviation	2.02	1.89	5.46	5.51	8.42	9.07
Mean Absolute	1.67	3.09	5.82	7.29	9.94	6.66

The presence of an under-estimation bias for the "None" category is explained by the assignment of development class using the NLCD land cover data. In the NLCD land cover production process the ISA product is used as a mask in processing (Collin Homer EROS data center, personal communication 2006). Therefore, it is possible that chips with preliminary classifications of no development using the NLCD land cover data would show an under-estimation of TIA given the relationship among NLCD products. The average of Actual Error by development class was negative for all classes (Table 7). This occurred even in the presence of gross over-predictions for individual chips (e.g., the urban class with a maximum of nearly 40%). Average Actual Error was highest in magnitude for the dense suburban and suburban classes while it was lowest for the undeveloped and highly developed classes (i.e., none, rural, and urban). Sensitivity to large errors at the individual chip level is reflected in the Mean Absolute Error (equation 3) by development class: the urban class' third-best ranking for mean Actual Error contrasts with third-worst given Mean Absolute Error. This is also reflected in the variance of Actual Error, as it was greatest for the dense suburban and urban categories. Pair-wise comparisons of average Actual Error by NLCD-calculated development class produced mixed results (Table 8).

Table 8. NLCD-Based Development Class Mean Errors verses "0" (First Column/First Row) and Pair-Wise between Development Classes (All Other Off-Diagonals).

	0	None	Rural	Exurban	Suburban	Dense Suburban	Urban
0		Y	Y	Y	Y	Y	Y
None	0		Y	Y	Y	Y	N
Rural	0	0.01		N	N	Y	N
Exurban	0	0.0	0.16		N	N	N
Suburban	0	0.0	0.10	0.71		N	N
Dense Suburban	0	0.0	0.02	0.29	0.49		Y
Urban	0	0.32	0.57	0.12	0.09	0.02	

Statistically significant differences (threshold 0.05) are noted by "Y" above the diagonal. Insignificant differences are noted by "N" above the diagonal. Resulting probability values are shown below the diagonals.

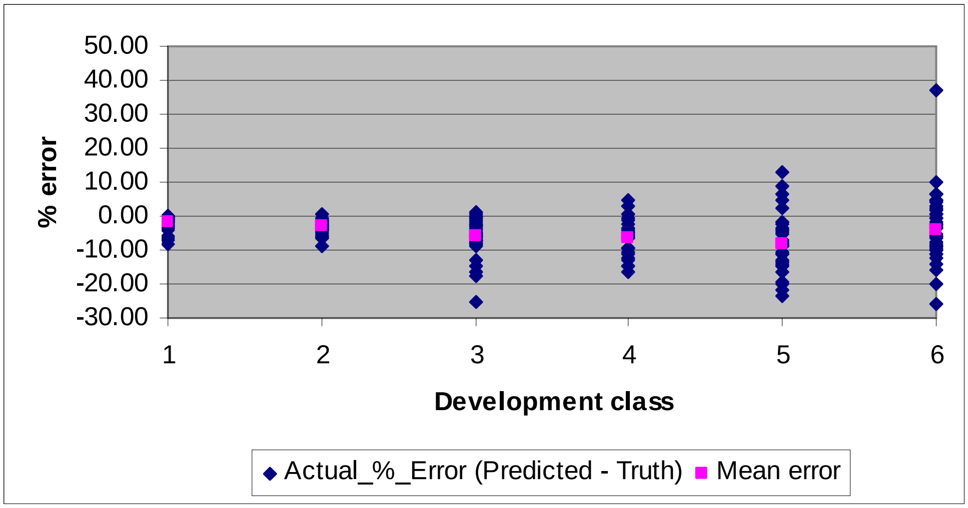

Fig. 3. Error and mean error as functions of development class. 1= None, 2 = Rural, 3 = Exurban, 4 = Suburban, 5 = Dense Suburban, and 6 = Urban.

The mean Actual Error for "none" was significantly different for all but that of the other tail end of the distribution (i.e., urban). With the exception of the most extreme mean Actual Error obtained in the dense suburban case (significantly different from the means for none, rural, and urban), most other mean errors were not significantly different from one another. However, the mean of every class was both negative (Table 7) and statistically significantly different from 0 at the 0.01 level (Table 8). These finding contradict analyses of other moderate-resolution satellite impervious surface estimates, in which under-prediction occurred at the low development classes and over-prediction was measured for highly developed or high percent impervious land cover areas (Canters, et al. 2006; Lu and Weng 2006).

Table 9 Relative Error [Eq. (2)] Statistics as a Function of Development Class for Reduced Sample Set (*N* = 226).

Relative Error	None	Rural	Exurban	Suburban	Dense Suburban	Urban
Max under-prediction	-100.00	-99.75	-85.13	-72.73	-66.70	-69.85
Max over-prediction	0	237.04	125.00	57.29	81.38	147.12
Mean	-78.26	-76.42	-47.80	-35.76	-24.73	-7.76
Median	-100.00	-89.10	-64.00	-49.93	-33.45	-7.70
Standard deviation	41.7	52.30	42.74	34.06	33.36	30.06
Mean Absolute	77.78	89.61	59.96	44.64	36.48	16.58

Both the mean and mean absolute values for Relative Error as a function of development class (Table 9) show inverse relationships with levels of development. That is, the largest (under-prediction) errors occur at the lowest levels of development and decrease as impervious surface area increases within chips. It is important to note that while some and unusual variation occurs as a function of development class, the average errors generated by the NLCD data are all comparatively low.

Conclusions

A protocol was used to collect independent validation data with sufficient observations to evaluate NLCD ISA data across a gradient of development in the Mid-Atlantic Piedmont and Coastal Plains regions centered on the District of Columbia. Implemented in the GIS environment, ancillary data and personal knowledge were used to rigorously examine both the reference data and the NLCD ISA predictions. An unusual finding of this research was the average under-prediction of ISA regardless of development category. The impact that this under-estimation has on hydrologic modeling requires further study. However, from the standpoints of monitoring and development planning, this underestimation may have important implications. This may be particularly true when the maintenance of relatively low ISA percentages (i.e., less than 10%) in small watersheds is the goal. However, it is important to note that the average Actual Errors produced by the NLCD data for the study area are comparatively low and the consistent underestimation of ISA demonstrated by NLCD-derived development class suggests that some adjustment across classes may be feasible. Because the GIS environment in which the validation data collection protocol has been implemented allows for pixel-scale analysis, next steps in this research include pixel-level study of the NLCD data to better describe potential causes and possible adjustments for the underestimation. A primary goal in the creation of this validation data set was the capability to compare accuracy and utility of ISA estimates generated by different satellite processing algorithms. This research is also currently underway. Finally, similar validation datasets have been collaboratively collected using the protocol in areas of New England and Florida where significantly different building practices and construction materials are used and where impervious surfaces exist in very different background (e.g., soil) and overstory (i.e., vegetation) conditions. The evaluation of NLCD ISA data in these regions is also an area of current research.

ACKNOWLEDGEMENTS

The United States Environmental Protection Agency through its Office of Research and Development partially funded and collaborated in the research described here under assistance agreement number DW14939340-03 to the U.S. Geological Survey. It has been subjected to Agency review and approved for publication. Although this work was reviewed by EPA and approved for presentation, it may not necessarily reflect official Agency policy. Mention of trade names or commercial products does not constitute endorsement or recommendation by EPA for use. Gail Winter's careful QA/QC of contractor, student, and Federal Employee ISA delineations and the helpful efforts of several anonymous reviewers are gratefully acknowledged.

REFERENCES

Arnold, C. L. and Gibbons, C. J. (1996). "Impervious surface coverage: emergence of a key environmental indicator." Journal of the American Planning Association **62**(2): 243-258.

Canters, F., J., Chormanski, J., Van de Voorde, T., Batelaan, O. (2006). Effects of different methods for estimating impervious surface cover on runoff estimation at catchment level. 7th International Symposium on Spatial Accuracy in Natural Resources and Environmental Sciences.

CWP (2003). Impacts of Impervious Cover on Aquatic Systems. Ellicott City, Center for Watershed Protection**:** 142.

Dunne, T. and Leopold, L. B. (1978). Water in Environmental Planning. New York, W.H. Freeman and Company.

Hammer, T. R. (1972). "Stream enlargement due to urbanization." Water Resources Research **8**(6): 1530-1540.

Homer, C. C., Huang, L., Yang, L. Wylie, B., and Coan, M. (2004). "Development of a 2001 National Landcover Database for the United States." Photogrammetric Engineering and Remote Sensing **70**(7): 829-840.

Jennings, D. B., and Jarnagin, S. T. (2002). "Changes in antrhopogenic impervious surfaces, precipitation and daily streamflow discharge: a historical perspective in a mid-Atlantic subwatershed." Landscape Ecology **17**(5): 471-489.

Jones, J. W., Claggett, P., Jarnagin, S.T., and Jennings, D.B. (2003). Shared assessment of USGS and NGO impervious surface data sets for the Cheapeake Bay Watershed. Reston VA, USGS.

Jones, J. (in review). A protocol for concerted evaluation of satellite-based impervious surface estimates. Unpublished report undergoing USGS internal review.

Lu, D. and Weng, Q. (2006). "Use of impervious surface in urban land-use classification." Remote Sensing of Environment **102**: 146-160.

McMahon, G. (2007). "Consequences of land-cover misclassification in models of impervious surface." Photogrammetric Engineering and Remote Sensing **73**(12): 1343-1353.

Schueler, T. R. (1994). "The importance of impervious area." Watershed Techniques **1**(3): 100-111.

Slonecker, E. T., Jennings, D. B. , and Garofalo, D. (2001). "Remote Sensing of Impervious Surface Area: A Review." Remote Sensing Reviews **20**(3): 227-235.

Slonecker, E. T. and Tilley, J. S. (2004). "An evaluation of the individual components and accuracies associated with the determination of impervious area." GIScience and Remote Sensing **41**(2): 165-184.

Yang, L., Huang, C., Homer, C.G., and Wylie, B.K. (2003). "An approach for mapping large-scale impervious surfaces: Synergistic use of Landsat-7 ETM+ and high spatial resolution imagery." Canadian Journal of Remote Sensing **29**: 230-240.

Yang, X. (2005). "Estimating landscape imperviousness index from satellite imagery." IEEE Geoscience and Remote Sensing Letters: 1-4.