2004

# Assessing the Accuracy of National Land Cover Dataset Area Estimates at Multiple Spatial Extents

Jeffrey W. Hollister
*University of Rhode Island*

M. Liliana Gonzalez
*University of Rhode Island*

John F. Paul

Pete V. August
*University of Rhode Island*, pvaugust@uri.edu

Jane L. Copeland



## Citation/Publisher Attribution

Hollister, J. W., Gonzalez, M. L., Paul, J. F., August, P. V., & Copeland, J. L. (2004). Assessing the Accuracy of National Land Cover Dataset Area Estimates at Multiple Spatial Extents. *Photogrammetric Engineering & Remote Sensing, 4,* 405-414. https://doi.org/10.14358/PERS.70.4.405
Available at: https://doi.org/10.14358/PERS.70.4.405

# Assessing the Accuracy of National Land Cover Dataset Area Estimates at Multiple Spatial Extents

## Publisher Statement

Jeffrey W. Hollister and Paul V. August are affiliated with the Department of Natural Resources Science.

M. Liliana Gonzalez is affiliated with the Department of Computer Science and Statistics.



## Creative Commons License

This work is licensed under a Creative Commons Attribution 4.0 License.

# Assessing the Accuracy of National Land Cover Dataset Area Estimates at Multiple Spatial Extents

Jeffrey W. Hollister, M. Liliana Gonzalez, John F. Paul, Peter V. August, and Jane L. Copeland

## Abstract
*Site-specific accuracy assessments evaluate fine-scale accuracy of land-use/land-cover (LULC) datasets but provide little insight into accuracy of area estimates of LULC classes derived from sampling units of varying size. Additionally, accuracy of landscape structure metrics calculated from area estimates cannot be determined solely from site-specific assessments. We used LULC data from Rhode Island and Massachusetts as reference to determine the accuracy of area measurements from the National Land Cover Dataset (NLCD) within spatial units ranging from 0.1 to 200 km$^2$. When regressed on reference area, NLCD area of developed land, agriculture, forest, and water had positive linear relationships with high $r^2$, suggesting acceptable accuracy. However, many of these classes also displayed mean differences (NLCD − REFERENCE), and linear relationships between the NLCD and reference were not one-to-one (i.e., low $r^2$, $\beta_0 \neq 0$, $\beta_1 \neq 1$), suggesting mapped area is different from true area. Rangeland, wetland, and barren were consistently, poorly classified.*

## Introduction
Measurements of landscape pattern are linked with many key ecological processes and are often used to model a variety of parameters such as species diversity or water quality (Flather *et al.*, 1992; Levine *et al.*, 1993; Hunsaker and Levine, 1995). These measurements are typically made from digital land-use/land-cover (LULC) data and are either calculated across an entire landscape or within sampling units of varying size such as a full watershed or a buffer around a sampling station (Comeleo *et al.*, 1996). The accuracy of the estimates of landscape composition and pattern are dependent upon the accuracy of the source data. The standard method of assessing accuracy involves comparing the land-cover class derived from the LULC data at various locations with the land-cover class occurring in a reference dataset, then generating error matrices that show the degree to which land-cover classes are correctly identified in the LULC dataset (Stehman and Czaplewski, 1998; Congalton and Green, 1999). These robust methods are effective at estimating LULC accuracy for an entire image. However, many ecological applications utilize measurements of the total area or percent of the total area of land-cover classes for a range of spatial extents such as watersheds, bio-reserves, or political divisions such as states or countries (McGarigal and McComb, 1995; Comeleo *et al.*, 1996; Jones *et al.*, 1997; Paul *et al.*, 2002). The accuracy of area measurements across a range of spatial extents cannot be fully inferred from accuracy assessments created using standard point-to-point methods, especially when site-specific accuracy is found to be poor.

The National Land Cover Dataset (Vogelmann and Wickham, 2000) is an LULC data product for the contiguous United States developed by the Multi-Resolution Landscape Characteristics Consortium, a consortium of federal agencies, from Landsat Thematic Mapper Imagery acquired from 1990 to 1993 (http://landcover.usgs.gov/natllandcover.asp, last accessed 25 December 2003). Using an unsupervised classification algorithm, along with ancillary data, a total of 21 thematic classes were derived (Lunetta *et al.*, 1998; Vogelmann *et al.*, 1998a; Vogelmann *et al.*, 1998b; Vogelmann and Wickham, 2000). The National Land Cover Dataset (NLCD) promises to serve as a rich source of information on landscape composition and pattern for ecological assessments. The accuracy of the NLCD within the New England region has been assessed using point-to-point comparisons with reference data, and the results of these studies are currently available (Yang *et al.*, 2001; USGS, 2003). However, our use of this accuracy information was limited because the accuracy of the NLCD within subsets of the New England region and the accuracy of LULC area measurements within sampling units of varying extents has not been measured.

We addressed these limitations with three objectives. Our first objective was to examine the site-specific accuracy of the NLCD within Rhode Island and Massachusetts and identify which classes exhibited poor site-specific accuracy. Second, we wanted to evaluate area estimates of NLCD classes using an accuracy assessment at multiple spatial extents and with a variety of analysis techniques. Third, we wanted to determine if the accuracies changed as a function of spatial extent.

J.W. Hollister and P.V. August are with the Department of Natural Resources Science, University of Rhode Island, Kingston, RI 02881 (jeff@edc.uri.edu;pete@edc.uri.edu).

M.L. Gonzalez is with the Department of Computer Science and Statistics, University of Rhode Island, Kingston, RI 02881 (liliana@cs.uri.edu).

J.F. Paul was with the Atlantic Ecology Division, U.S. Environmental Protection Agency, Narragansett, RI 02882; he is presently with the National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency (Mail Drop B205-1), Research Triangle Park, NC 27711 (paul.john@epamail.epa.gov).

J.L. Copeland is with Computer Sciences Corporation, Narragansett, RI 02882 (copeland.jane@epamail.epa.gov).

Photogrammetric Engineering & Remote Sensing
Vol. 70, No. 4, April 2004, pp. 405–414.

0099-1112/04/7004–0405/$3.00/0

TABLE 1. AREA ($KM^2$) AND PERCENT OF TOTAL AREA OF ANDERSON LEVEL 1 CLASSES FOR RHODE ISLAND GIS (RIGIS) AND MASSACHUSETTS GIS (MASSGIS) REFERENCE DATA AND NLCD DATA

| Anderson Level 1 | RIGIS and MassGIS | | NLCD | |
|---|---|---|---|---|
| | Area ($km^2$) | % Total | Area ($km^2$) | % Total |
| Developed | 5612.3 | 23.6 | 4116.1 | 17.3 |
| Agriculture | 1861.7 | 7.8 | 2053.4 | 8.7 |
| Rangeland | 46.3 | 0.2 | 16.3 | 0.1 |
| Forest | 13635.5 | 57.4 | 14698.7 | 61.9 |
| Water | 808.5 | 3.4 | 943.3 | 4.0 |
| Wetland | 1004.3 | 4.2 | 1670.3 | 7.0 |
| Barren | 788 | 3.3 | 232.8 | 1.0 |
| TOTAL | 23756.6 | 100.0 | 23730.9 | 100.0 |

## Study Area and Data

Our study area consists of the states of Massachusetts and Rhode Island in southern New England. This region has four large urban centers (Boston, Massachusetts; Springfield, Massachusetts; Worcester, Massachusetts; and Providence, Rhode Island), and is dominated by forest, developed land, agriculture, wetlands, and water. Barren land, orchards/vineyards, and rangeland encompass a small portion of the overall landscape (Table 1). Our choice of study area was driven by availability of large-scale, photointerpreted reference data to compare to the NLCD.

We acquired the NLCD data for the states of Rhode Island and Massachusetts. The NLCD had not been filtered and exhibited a "salt and pepper" effect which is often seen in satellite-derived LULC data (Loveland *et al.*, 1999; USGS, 2003). It is common practice to use a small focal majority window to filter isolated pixels of land cover and merge them into the surrounding matrix cover (Lillesand and Kiefer, 1995; DeMers, 1997; Burrough and McDonnell, 1998). Additionally, isolated pixels are smaller than the effective minimum mapping unit of satellite-derived LULC data (Loveland *et al.*, 1999). Following the recommendations of the U.S. Geological Survey (USGS), we filtered the NLCD with a 3- by 3-pixel majority window with ties left unchanged, producing an LULC product with an 0.8-ha (2-acre) minimum mapping unit (MMU). All of our analyses were conducted on the resultant filtered data (USGS, 2003).

Reference data were acquired from the Massachusetts GIS (MassGIS, http://www.state.ma.us/mgis, last accessed 25 December 2003) and Rhode Island GIS (RIGIS, http://www.edc.uri.edu/rigis, last accessed 25 December 2003) Programs. The MassGIS LULC data were photointerpreted from 1:25,000-scale aerial photography into a total of 37 classes, with acquisition dates ranging from 1985 to 1997 and with a minimum polygon size of 0.4 ha (1 acre). The RIGIS LULC data consisted of 40 land-use/land-cover classes that were photointerpreted from 1:12,000-scale USGS digital orthophoto quadrangles (DOQ) obtained in 1995. The RIGIS data had a minimum polygon size of 0.2 ha (0.5 acre). Each dataset met National Map Accuracy Standards and underwent extensive quality assurance/quality control procedures (http://www.state.ma.us/mgis, last accessed 25 December 2003; http://www.edc.uri.edu/rigis; last accessed 25 December 2003. David W. Goodwin, personal communication). The RIGIS and MassGIS data were converted to the same format as the NLCD (raster format with 30-m pixels), recoded to Anderson Level 1 classification, and appended together to create a single, seamless reference LULC data set (Anderson *et al.*, 1976). The NLCD was recoded to Anderson Level 1 to be consistent with the Massachusetts and Rhode Island reference data (Table 2).

## Methods

### Site-Specific Accuracy Assessment

We assessed the NLCD within Rhode Island and Massachusetts to identify poorly classified classes, and to assess overall accuracy. This site-specific accuracy assessment gave us a basis for comparing the multiple-extents accuracy assessment. Our reference data represented complete coverage of our study area; therefore, it was unnecessary to sample and the assessment represents a complete census of the NLCD. All NLCD pixels within our study area were converted to points (points represent the center location of the pixel). For each point, the Anderson Level 1 LULC class was recorded for both the reference data and the NLCD. Agreement and disagreement between the two datasets were recorded in an error matrix; then user's accuracy, producer's accuracy, and overall accuracy were calculated (Stehman and Czaplewski, 1998; Congalton and Green, 1999).

### Multiple-Extents Accuracy Assessment Sampling Scheme

To test the accuracy of NLCD-derived area measurements at multiple extents, we created 35 randomly chosen, non-overlapping circular sampling units with radii ranging from 0.178 to 7.98 kilometers (area of 0.1 $km^2$ to 200 $km^2$). We accomplished this by placing a random point within our study area and delineating a circle with the specified radius. If the resulting circle overlapped another, it was discarded. This process was repeated until 35 non-overlapping circles were selected for each of 40 radii for a total of 1,400 circles (40 radii $\times$ 35 circles). Each group of 35 circles is collectively referred to as "spatial extents." Each of the 35 circles of the 40 spatial extents was overlaid on the majority filtered NLCD and reference LULC data. Total area of the LULC classes within each circle was calculated, resulting in a total of 245 area measurements within each spatial extent (35 circles and seven LULC classes).

It has been suggested that the NLCD not be used to characterize areas smaller than approximately 25 $km^2$ (USGS, 2003). Our range of spatial extents was intentionally chosen to span this suggested minimum functional area. We felt that the accuracy of area estimates from the NLCD may be acceptable within areas smaller than 25 $km^2$, and we wanted to test this. Additionally, prior analysis indicated that accuracy stabilizes at larger extents. Consequently, we chose to evaluate more of the smaller spatial extents, examining 25 spatial extents smaller than 25 $km^2$. All 40 (25 less than 25 $km^2$ and 15 greater than 25 $km^2$) spatial extents were included in the final analyses.

### Multiple-Extents Analysis 1: Testing for Differences in Mean Area

Because of the nature of the data (map and reference land-cover measurements on the same sampling units), it is natural to consider using the well-known paired $t$ test to test for a difference in the mean of the map and reference areas for each

Table 2. Modified Anderson Level 1 Classification Scheme Used to Standardize the Cover Classes between the NLCD and the Reference Data

| Anderson Level 1 | RIGIS Classes | MassGIS Classes | NLCD Classes |
|---|---|---|---|
| 100-Developed | All Residential Classes, Commercial, Commercial/Industrial Mixed, Institutional, Airports, Other Transportation, Power Lines, Railroads, Roads, Waste Disposal, Water and Sewage Treatment, Developed Recreation, Cemeteries, Vacant Land | All Residential Classes, Commercial, Industrial, Transportation, Waste Disposal, All Recreation Classes, Urban/Open | Low Intensity Residential, High Intensity Residential, Commercial/Industrial/ Transportation, Urban/Recreational |
| 200-Agriculture | Pasture, Cropland, Confined Feeding Operation, Idle Agriculture, Orchards, Groves, Nurseries | Cropland, Pasture, Woody Perennial | Pasture/Hay, Row Cross, Fallow, Orchard/Vineyards, Other |
| 300-Rangeland | Brushland | -NA- | Shrubland |
| 400-Forest | Deciduous Forest, Evergreen Forest, Mixed Deciduous Forest, Mixed Evergreen Forest | Forest | Deciduous Forest, Evergreen Forest, Mixed Forest |
| 500-Water | Water, Salt Water | Water, New Ocean | Open Water |
| 600-Wetlands | Wetland | All Wetland Classes | Wetlands |
| 700-Barren | Beaches, Sandy Areas, Rock Outcrops, Mines, Quarries and Gravel Pits, Transitional Areas, Mixed Barren Areas | Mining, Open Land | Bare Rock/Sand/Clay, Quarries/Strip Mines/Gravel Pits, Transitional |

LULC class within each spatial extent. This test quantifies the magnitude and direction (i.e., positive or negative) of the bias of the area estimates derived from the NLCD. But the paired *t* test cannot be used in isolation to evaluate accuracy. While equality in means implies no bias, it does not assure good accuracy. Large disagreements between the NLCD area and reference area may still exist because positive and negative differences may counteract and produce a mean difference near zero (i.e., similar mean but different variance). We address this limitation with our final analysis.

#### Multiple-Extents Analysis 2: Linear Regressions

We assessed accuracy of the NLCD at multiple spatial extents using linear regression. We fit a simple linear regression equation for each of the seven classes in each of 40 spatial extents. We used the NLCD area as the dependent variable and the reference data area as the independent (NLCD = $\beta_0 + \beta_1$REFERENCE). An NLCD class can be considered accurate relative to the reference data when the regression equation displays a one-to-one relationship between the NLCD and the reference data (i.e., $\beta_0 = 0$, $\beta_1 = 1$, and high $r^2$) and indicates that an increase in area of the reference data results in an equivalent increase in area for the NLCD. We used these three values to determine the accuracy of the classes at all spatial extents and, in addition to the standard test for a regression equation ($H_0$: $\beta_1 = 0$), we explicitly tested $\beta_0$ and $\beta_1$ separately to determine if the intercept was equal to zero ($H_0$: $\beta_0 = 0$) and if the slope was equal to one ($H_0$: $\beta_1 = 1$). Additionally, it is possible to have a curvilinear relationship that explains much of the variation in the NLCD area measurements; however, this situation would indicate that agreement was poor over portions of the range of values of the reference data. To ensure that this was not impacting our results, we constructed quadratic regressions for each class at each spatial extent, and tested for the significance of the quadratic terms.

## Results

#### Site-Specific Accuracy Assessment

Our complete census for Anderson Level 1 classes resulted in an overall accuracy of the NLCD within Rhode Island and Massachusetts of 71.4 percent (18,821,600 correct pixels/ 26,367,557 total pixels). Only the producer's accuracy for the forest and water classes had an accuracy nearing 85 percent (Table 3). The poor accuracy of the rangeland class is likely a function of its rarity.

#### Multiple-Extents Accuracy Assessment: Testing for Differences in Mean Area

Forest and wetland classes showed significant, positive differences for mean area in more than half of the spatial extents, indicating higher mean values for the NLCD than for the reference data, and the percent mean difference ranged from near 0 to 5 percent (Figures 1 and 2). Mean area of developed and barren showed a significant, negative mean difference for most (approximately 93 percent) of the spatial extents, and the percent mean difference ranged from near 0 to −1 percent. These were the only LULC classes to consistently show lower mean values for the NLCD (Figures 1 and 2). Water and agriculture classes did not consistently display a significant difference between the two datasets. The magnitudes of the percent mean difference for all classes never exceeded 10 percent and rarely exceeded 5 percent (Figure 2). Statistical tests on rangeland may provide confusing and unreliable results because no rangeland was found in many of the samples. Thus, paired *t* test and regression results for the rangeland class will not be presented.

#### Multiple-Extents Accuracy Assessment: Linear Regression Results

Significant linear relationships were found in 100 percent of the spatial extents for developed, agriculture, forest, and water, and the linear component for these classes accounted for a significant portion of the variation (Figure 3). Wetlands

TABLE 3. SITE-SPECIFIC (I.E., PIXEL-TO-PIXEL) ERROR MATRIX FOR NLCD IN RHODE ISLAND AND MASSACHUSETTS. BOLD ENTRIES INDICATE AGREEMENT BETWEEN THE TWO DATASETS AND OFF-DIAGONAL VALUES INDICATE DISAGREEMENT

| Classified Data—MRLC | Reference Data—MassGIS and RIGIS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Developed | Agriculture | Rangeland | Forest | Water | Wetland | Barren | Row Total |
| Developed | **3,387,120** | 221,426 | 7,561 | 710,965 | 20,159 | 89,266 | 136,936 | 4,573,433 |
| Agriculture | 749,086 | **1,038,488** | 3,833 | 323,435 | 3,596 | 40,797 | 122,280 | 2,281,515 |
| Rangeland | 1,283 | 1,972 | **101** | 4,144 | 38 | 510 | 10,010 | 18,058 |
| Forest | 1,616,502 | 647,596 | 28,187 | **13,128,860** | 52,162 | 467,195 | 391,409 | 16,331,911 |
| Water | 64,651 | 11,014 | 772 | 95,509 | **758,576** | 96,976 | 20,557 | 1,048,055 |
| Wetland | 331,274 | 118,056 | 4,625 | 849,219 | 58,822 | **405,289** | 88,582 | 1,855,867 |
| Barren | 78,190 | 25,415 | 953 | 33,210 | 3,618 | 14,166 | **103,166** | 258,718 |
| Column Total | 6,228,106 | 2,063,967 | 46,032 | 15,145,342 | 896,971 | 1,114,199 | 872,940 | 26,367,557 |

**Overall Accuracy** = 71.4%, **Producer's Accuracy** Developed = 54.4% Agriculture = 50.3% Rangeland = 0.2% Forest = 86.7% Water = 84.6% Wetland = 36.4% Barren = 11.8%, **User's Accuracy** Developed = 74.1% Agriculture = 45.5% Rangeland = 0.6% Forest = 80.4% Water = 72.4% Wetland = 21.8% Barren = 39.9%.

had significant linear relationships for only 45 percent of the spatial extents, accounting for, on average, 29.0 percent of the variation (Figure 3). Barren land had significant linear relationships for 72.5 percent of the spatial extents and accounted for an average of only 27.0 percent of the variation (Figure 3).

For all classes except wetland and barren, significant curvilinear relationships existed for only a few spatial extents. For those that were significant, the inclusion of a quadratic term only accounted for a small amount of additional variation. Many spatial extents for the barren and wetland classes had significant curvilinear relationships, which accounted for nearly 20 percent of the variation in some cases. The low number of significant linear regressions and higher amount of variation accounted for by the quadratic term implies that a strong linear relationship does not exist for the wetland and barren classes. These classes appear to be non-linear (i.e., have poor accuracy), and assessment of the slope and intercept would not yield additional information about accuracy; therefore, slope and intercept results for barren and wetland will not be discussed.

Intercepts ($\beta_0$) for the developed class were less than zero for 60.0 percent of the spatial extents. Agriculture and forest intercepts were greater than zero in more than half of the spatial extents and water rarely had intercepts significantly different from zero. Having an intercept different from zero implies a bias in the estimate of the NLCD, indicating that agriculture and forest were often overestimated while developed area was underestimated. Although testing slightly different parameters, these results and the results of the paired t-tests corroborate one another and show similar patterns of over- and underestimation. Because of this it is not necessary for us to include a graph of the intercepts, because it illustrates the same pattern as seen in Figure 1.

The final parameter to be tested was slope of the regression line for each of the spatial extents. Slopes ($\beta_1$) were significantly less than one in 52.5 percent of the spatial extents for developed, in 70 percent of the spatial extents for agriculture, and in 42.5 percent for forest (Figure 4). The slope for area of water was greater than 1 in 12.5 percent of the spatial extents and less than 1 for 7.5 percent of the spatial extents (Figure 4).

## Discussion and Conclusions

### NLCD Area Accuracy

Two classes—barren and wetland—were shown to have low $r^2$ values, significant differences in the mean, and non-linear relationships; therefore, the area measurements for these classes could be considered to have poor accuracy. The water class had high $r^2$ values, a slight difference in mean and intercept, and consistently exhibited a one-to-one relationship with the reference data. Thus, water appears to be well classified for area across most of the spatial extents. The remaining three classes—developed, agriculture, and forest—had significant linear relationships, suggesting acceptable levels of accuracy, but also showed differences in the mean, intercepts not equal to zero, and a slope often not equal to one. Specifically, both the mean and intercept for developed area, as shown by the t-test and test of the intercept, were consistently underestimated in the NLCD, whereas agriculture and forest classes tend to be overestimated. Making a determination of poor accuracy for these classes is not entirely straightforward. Several of the tests do imply poor accuracy, but inherent differences in the two datasets (see **Sources of Error**) make it difficult to determine if the lack of a one-to-one relationship is due to misclassification or other errors.

In general, area measurements of the developed, agriculture, forest, and water classes are fairly accurate. Area of rangeland, barren, and wetland are poorly classified, and users of these data should either be cautious when using area estimates for these classes or, if necessary, explore ancillary sources of data. As with all accuracy assessments, making a final determination of the accuracy of area measurements must be done with caution. The accuracy and relevance of the reference data can have a profound impact on the determination of accuracy, and it is always important to consider the use of the data and the level of detail in the classification scheme when making decisions regarding acceptable levels of misclassification (Congalton and Green, 1999; Loveland *et al.*, 1999).

### Relationship between Accuracy and Spatial Extent

One of the objectives of our study was to determine if accuracy changes as a function of spatial extent. Such a relationship would suggest that a minimum size for sampling units (i.e., a minimum functional extent) may exist, below which the accuracy of the NLCD would reach unacceptable levels; therefore, LULC area estimates based on those sampling units should be used cautiously. Our analyses proved effective at exploring such a relationship. The $r^2$ values suggest that accuracy increases as spatial extent increases, with the $r^2$ leveling off at a radius between about 1.5 to 2.5 km (approx. 7- to 20-km$^2$ spatial extents) for developed, agriculture, forest, and water (Figures 3). However, $r^2$ is not ideal for assessing this relationship because of the difficulties in determining an "acceptable" accuracy value for $r^2$ as well as the considerable variation in $r^2$ from similar spatial extents due to sampling and other errors (see **Sources of Error**). If slope also displays a

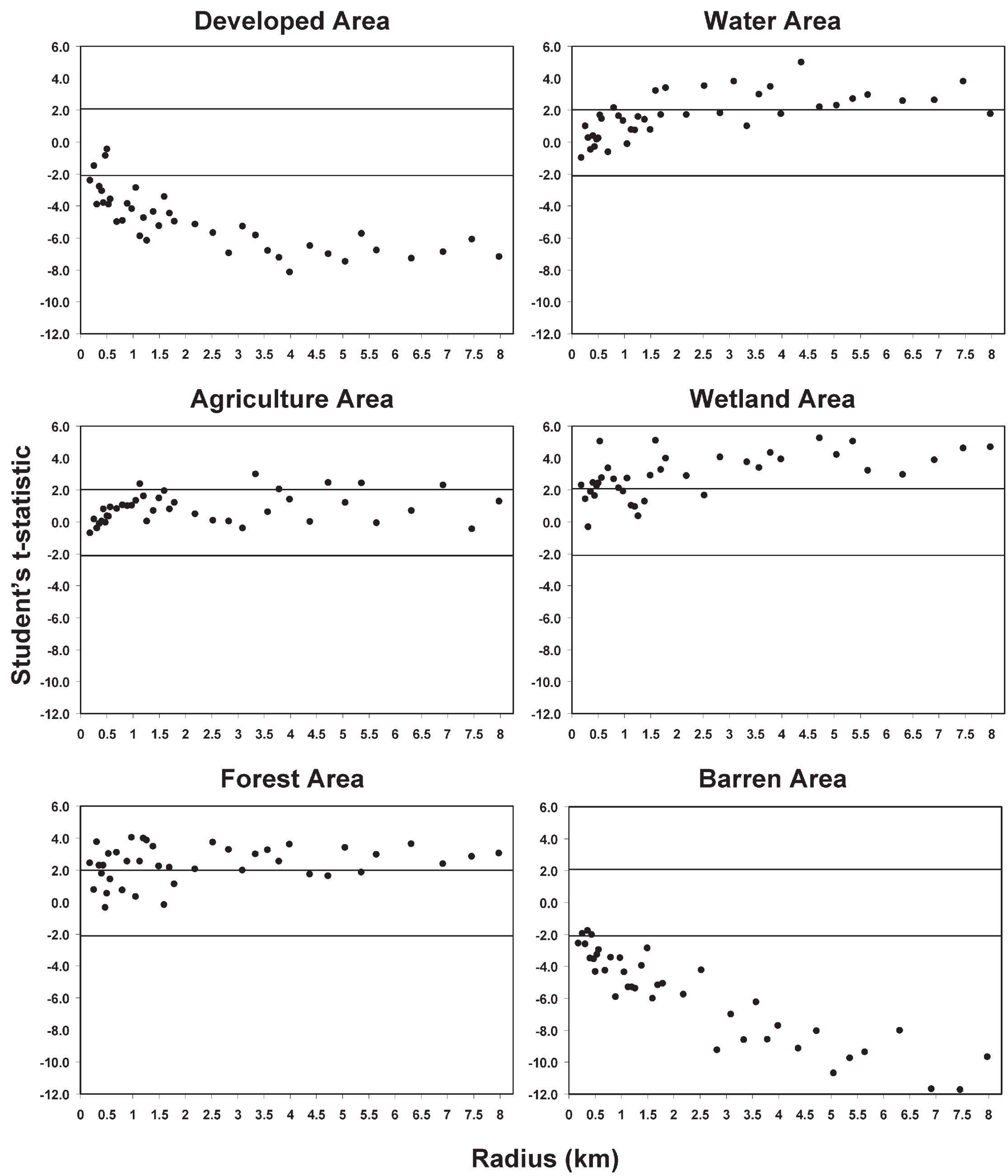


Figure 1. Student's t-statistic for the paired t-test testing difference between the NLCD and the Reference data plotted as a function of radius of the spatial extent. Reference lines represent critical value for $t(\alpha = 0.05, n = 35)$. Values outside of reference lines represent extents with significant mean differences.

relationship with spatial extent and levels off around a slope of 1, then a minimum functional extent can be said to exist at that point. For the four classes with the highest accuracy, there is little variation in the slope, and finding a point below which the slopes tended to be significantly different from 1 was not possible (Figure 4). Percent mean difference provides perhaps the best indication of a relationship between accuracy and spatial extent. Percent mean difference showed greater variation at smaller spatial extents and leveled off at a radius around 1 to 2 km (approx. 3- to 12-km$^2$ spatial extents) for all classes except water, which remained constant across most spatial extents (Figure 2). These results corroborate the suggestion by the USGS to exercise caution when using the NLCD within small, local areas; however, they also imply that the accuracy of the NLCD may be acceptable at extents smaller than the 25 km$^2$ originally suggested by the USGS (2003).

#### Sources of Error

There are four principal reasons, besides misclassification, for the poor correspondence (i.e., low r$^2$, biases in the mean, $\beta_0 \neq 0$, $\beta_1 \neq 1$) between the NLCD and reference data and the large variation in accuracy sometimes shown between spatial extents of similar sizes. First, temporal discrepancies among the acquisition dates of the NLCD and reference data may add error

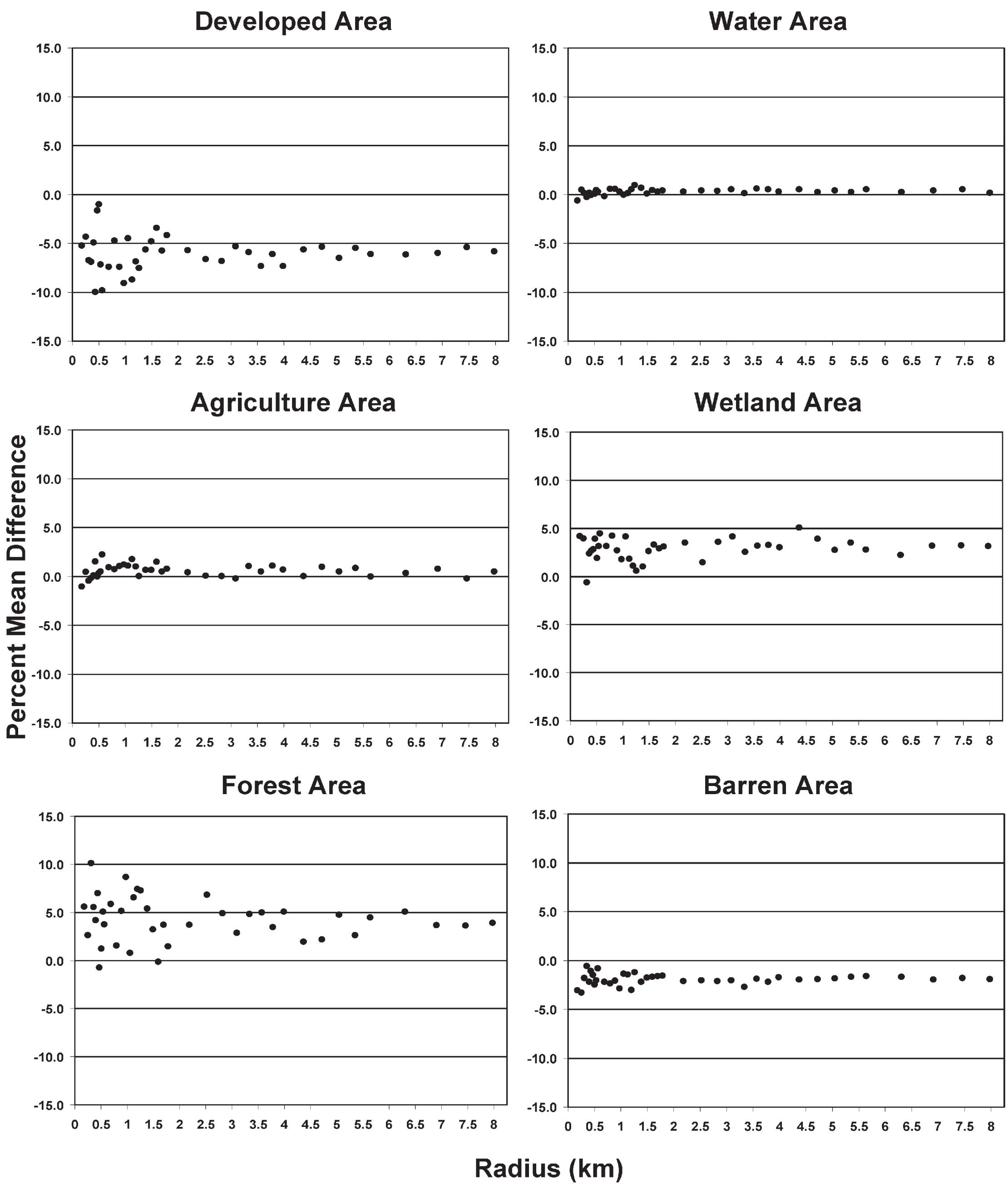

Figure 2. Percent mean difference ((NLCD − REFERENCE)/Area of Spatial Extent) between the NLCD and the Reference data plotted as a function of radius of the spatial extent.

(Congalton and Green, 1999). NLCD data were obtained between 1990 and 1993. MassGIS data were created from imagery obtained between 1985 and 1997 and RIGIS data were obtained in 1995. Assessing the accuracy of a dataset using reference data from slightly different time periods may result in lowered accuracies due to landscape change, not misclassification (Congalton and Green, 1999). It is also common, as is the case with the NLCD, to have internal temporal inconsistencies in source data of regional scale LULC datasets (Vogelmann *et al.*, 1998b; Loveland *et al.*, 1999). The rates of landscape change are variable across classes, across periods in history, and across regions. Thus, it is difficult to predict the effect that temporal discrepancies will have on estimation of classification accuracy (Dunn *et al.*, 1990). Discrepancies such as these are, as Loveland *et al.* (1999) have suggested, an unavoidable part of characterizing landscapes over broad regions.

Second, comparing metrics derived from analog sources (i.e., aerial photos) to metrics from digital sources (i.e., satellite imagery) can be problematic (Loveland *et al.*, 1999). The resultant land-cover datasets derived from these different sources represent the landscape in two distinct ways, vector data from photointerpretation of aerial photos and raster from the classification of satellite imagery (DeMers, 1997; Burrough and McDonnell, 1998). Raster data suffer from a "stair-stepped" effect, noticeable along polygon perimeters, whereas vector data represent perimeters with clean, smooth lines (Congalton, 1997). This will produce variation in the

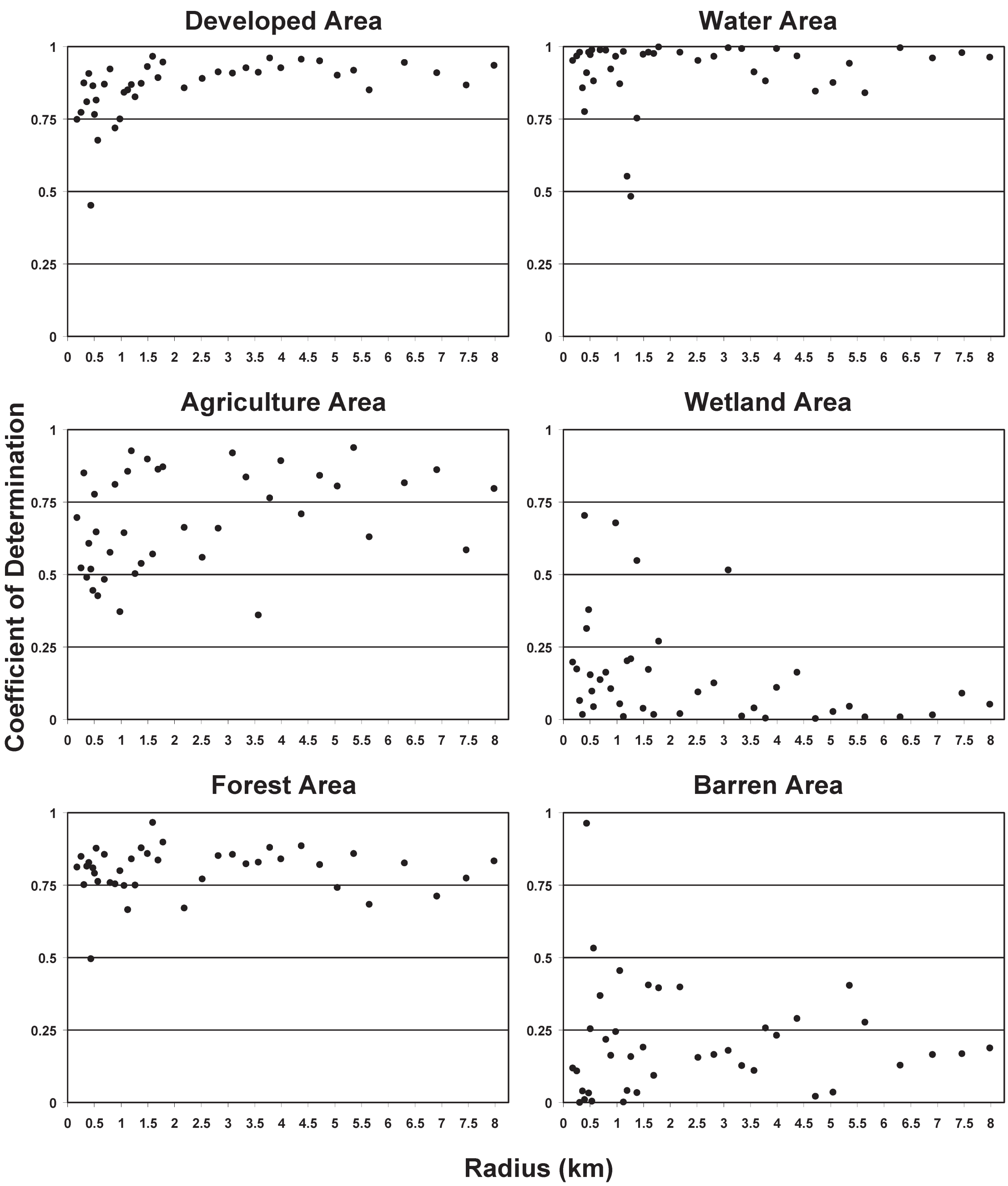


Figure 3. Coefficient of determination ($r^2$) of the regression for area ($\text{NLCD} = \beta_0 + \beta_1\text{REFERENCE}$) plotted as a function of radius of the spatial extent.

estimation of area totals based on data format and not necessarily misclassification.

Third, variation in the MMU of the three datasets may contribute to error in area estimation. The MMU for the three datasets ranged from 0.2 to 0.8 ha. The result of this is variation in detail, creating a different number of features for each reference dataset. Similarly, the minimum width of included features also differs. Although no specific mention of a consistent minimum width is included with any of the three datasets, the inclusion of roads (greater than 61 m (200 ft) for the RIGIS data) and power lines (greater than 30.5 m (100 ft) for the RIGIS data) in the reference datasets suggests that linear features smaller than the MMU of the NLCD were included. Because there are currently no raster sources of reference data that are larger scale than NLCD, encompass a large area, and have relevant LULC classes, it is necessary that we use data of differing initial format and varying MMU.

Fourth, Congalton and Green (1999) have suggested that quantitative accuracy assessments not be conducted with existing sources of reference data because of additional error introduced from differing classification schemes. Typically, reference data are collected through field visits or with aerial photograph interpretation using the same classification protocols used to develop the LULC data being assessed and are assumed to represent "true" values (Congalton, 1991; Congalton and Green, 1999). Because our study was initiated several years after the completion of the NLCD, field data collection was not possible. Additionally, our method requires complete

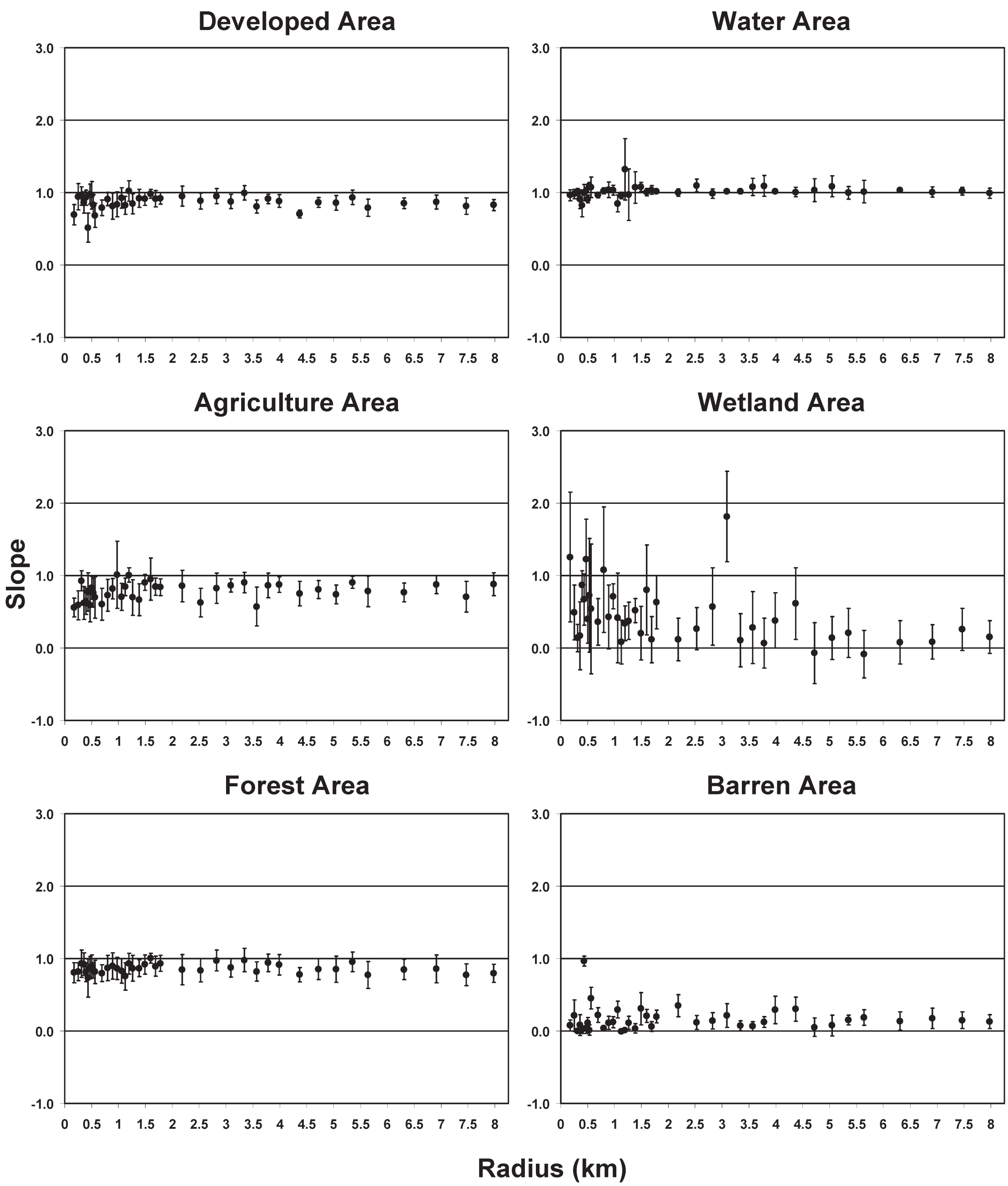


Figure 4. Slope of the regression for area (NLCD $= \beta_0 + \beta_1$REFERENCE) plotted as a function of radius of the spatial extent. Error bars represent 95 percent confidence limits. Slopes not significantly different from 1 represent high accuracy.

photointerpretation of large areas. Aerial photograph acquisition and interpretation for this research was not feasible. We followed Congalton and Green's (1999) recommendation for our site-specific accuracy assessment and only looked at producer's, user's, and overall accuracy. For the multiple-extent accuracy assessment, our need to have an estimate of the accuracy of area measurements, and the value added to our understanding of NLCD accuracy, overrides the potential problems associated with using existing reference data sources.

#### Poorly Classified Classes

The low accuracy of areas measured across multiple spatial extents for wetlands, barren, and rangeland is probably due to their low classification accuracy in the NLCD. These classes tend to have low overall classification accuracies in quality assurance testing conducted by the USGS (2003) and relative to our reference data (Table 3). The spectral signatures for wetland vegetation may be similar to forest (for forested wetlands) and rangeland (for shrub/scrub wetlands). This contributes to a high rate of misclassification for the wetland classes derived from the digitally processed NLCD. Our results and prior accuracy assessments suggest that it may not be advisable to use the NLCD to estimate wetland area. Although used as an ancillary source for labeling of clusters in the development of the NLCD, the National Wetlands Inventory (NWI, http://www.nwi.fws.gov, last accessed 25 December 2003) may provide better information as a sole source of data on wetlands because it is less likely to reflect the spectral confusion of the NLCD. The remaining poorly depicted classes, barren and rangeland, make up only 1.1 percent of the total study area. The rarity of these two classes likely contributes to the poor classification results.

The confusion identified by the site-specific accuracy assessment between the developed, forest, and agriculture classes is likely due to differences in classification methodologies of the reference and NLCD data and not necessarily due to misclassification. For instance, LULC data derived from satellite imagery often identify individual trees within an urban or agricultural setting as forest. However, photointerpretation of those same sites will include individual trees within the overall developed or agricultural land-use class and, thus, add to misclassification within the error matrix as well as the overestimation indicated in Figure 2 (Y.Q. Wang, personal communication, 2002).

Finally, our study only considers NLCD area accuracy within Massachusetts and Rhode Island. To fully understand the accuracy of NLCD-derived metrics within other geographic settings, similar multiple-extent assessment methods should be conducted in other regions. Published site-specific accuracy assessments show similar overall and individual LULC accuracies across a range of geographic locations; thus, we may expect a multiple-extent accuracy assessment to find similar results in these other regions as well (Yang *et al.*, 2001; USGS, 2003). It is also reasonable to expect that the suite of poorly classified and correctly classified classes could change. For example, the rangeland class of the NLCD is exceedingly rare in Massachusetts and Rhode Island; however, in many western states rangeland would be a much more common class and it is possible that the NLCD would accurately depict it. The fact that rangeland was poorly depicted in this study reflects its rate of occurrence in the New England landscape.

### Multiple-Extents Accuracy Assessment Methodology

Because we use areal sampling units (i.e., spatial extents), our accuracy evaluation protocol is similar to the non-site-specific accuracy assessments which preceded the quantitative, site-specific assessments used today (Stehman and Czaplewski, 1998). An important difference between our methodology and non-site-specific assessments is our use of linear regression. Regression analysis allows us to utilize well developed statistical approaches for testing hypotheses and determining how well one classification agrees with another (Zar, 1999). For instance, it would be possible to perform our analysis, even at a single extent, and compare two datasets developed with different classification methodologies to a common reference source and determine the relative performance of those two methods by assessing linearity, intercept, and slope. These analyses typically have not accompanied previous non-site-specific assessments.

Furthermore, a site-specific assessment might identify several classes as being poorly classified, but those same classes might be variable in their ability to correctly estimate area. Several classes do exhibit poor classifications with the site-specific accuracy assessment (Table 3). However, the ability of these classes to accurately depict area as indicated by the multiple-extent accuracy assessment is quite different. For instance, agricultural and wetland both displayed poor user's and producer's accuracies (Table 3), but area of agricultural lands was accurately estimated across many extents, while wetland area was poorly depicted (Figures 1 through 4).

## Summary

Our primary goal for this study was to address the use of the NLCD as a source of landscape composition metrics by assessing the accuracy of areal summaries of LULC classes derived from the NLCD. Our general conclusion is that the NLCD provides reliable area estimates for LULC classes that are dominant and display unique spectral signatures. Additionally, sampling units used to subset the NLCD must be large enough (i.e., 10 km$^2$ or greater in area or radii greater than 1.5 to 2 km) to ensure accurate measurements of area. Although our intent was not to develop an areal-based accuracy assessment method, we were forced to do so. We discovered that, in cases where unacceptable accuracies are shown at fine scales with site-specific accuracy assessments, explicitly examining the accuracy of area measurements with a multiple-extents method is a logical additional step. Both site-specific and multiple-extents methods provide distinct information, which is important to the understanding of a dataset's thematic accuracy. Site-specific accuracy assessments provide unique information on estimates of overall accuracy (i.e., kappa statistics), user's and producer's accuracy, and sources of confusion among LULC classes (Stehman and Czaplewski, 1998; Congalton and Green, 1999). An accuracy assessment at multiple extents adds additional information regarding the accuracy of area measurements and, indirectly, the landscape metrics derived from those measurements.

## Acknowledgments

We thank Dr. Jim Heltshe, Anne Kuhn-Hines, Dr. Y.Q. Wang, Dr. Yong Wang, Steve Rego, and the anonymous reviewers for their time and effort in reviewing this manuscript. Dr. Steve Stehman deserves special mention for his insightful review of an earlier version and his valuable advice and guidance. We also thank the Rhode Island Agricultural Experiment Station for partial funding (RIAES Contribution No. 3991). JWH supported through Cooperative Agreement CT825802, Brian D. Melzian, Project Officer. JLC supported through USEPA contract 68-W-01-032, Larry Rossner, Task Order Project Officer. The U.S. Environmental Protection Agency has funded, wholly or in part, the research described in this article. This article has not been subjected to Agency review. Therefore, it does not necessarily reflect the views of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. Contribution No. AED-02-012 of the Atlantic Ecology Division, National Health and Environmental Effects Research Laboratory.

## References

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*, USGS Professional Paper 964, U.S. Geological Survey, U.S. Government Printing Office, Washington, D.C., 28 p.

Burrough, P.A., and R.A. McDonnell, 1998. *Principles of Geographical Information Systems*, Oxford University Press, New York, N.Y., 346 p.

Comeleo, R.L., J.F. Paul, P.V. August, J.L. Copeland, C. Baker, S.S. Hale, and R.W. Latimer, 1996. Relationships between watershed stressors and sediment contamination in Chesapeake Bay estuaries, *Landscape Ecology*, 11(5):307–319.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37(1):35–46.

———, 1997. Exploring and evaluating the consequences of vector-to-raster and raster-to-vector conversion, *Photogrammetric Engineering & Remote Sensing*, 63(4):425–434.

Congalton, R.G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data*, Lewis Publishers, New York, N.Y., 137 p.

DeMers, M.N., 1997. *Fundamentals of Geographic Information Systems*, John Wiley and Sons, New York, N.Y., 498 p.

Dunn, C.P., D.M. Sharpe, G.R. Guntenspergen, F. Stearns, and Z. Yang, 1990. Methods for analyzing temporal changes in landscape pattern, *Quantitative Methods in Landscape Ecology* (M.G. Turner and R.H. Gardner, editors), Springer-Verlag, New York, N.Y., pp. 173–198.

Flather, C.H., S.J. Bradley, and D.B. Inkley, 1992. Regional habitat appraisals of wildlife communities: A landscape-level evaluation of a resource planning model using avian distribution data, *Landscape Ecology*, 7(2):137–147.

Hunsaker, C.T., and D.A. Levine, 1995. Hierarchical approaches to the study of water quality in rivers, *BioScience*, 45(3): 193–203.

Jones, K.B., K.H. Riitters, J.D. Wickham, R.D. Tankersley, R.V. O'Neill, D.J. Chaloud, E.R. Smith, and A.C. Neale, 1997. *An Ecological Assessment of the United States Mid-Atlantic Region: A Landscape Atlas*, EPA/600/R-97/130, Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C., 104 p.

Levine, D.A., C.T. Hunsaker, S.P. Timms, and J.J. Beauchamp, 1993. *Geographic Information System Approach to Modeling Nutrient and Sediment Transport*, Report 6736, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 108 p.

Lillesand, T.M., and R.W. Kiefer, 1995. *Remote Sensing and Image Interpretation*, John Wiley and Sons, New York, N.Y., 750 p.

Loveland, T.R., A.L. Gallant, and J.E. Vogelmann, 1999. Perspectives on the Use of Land Cover Data for Ecological Investigations, *Issues in Landscape Ecology*, July-August, Snowmass Village, Colorado (International Association of Landscape Ecology, Guelph, Ontario, Canada), pp. 28–32.

Lunetta, R.L., J.G. Lyon, B. Guidon, and C.D. Elvidge, 1998. North American landscape characterization: Data set development and data fusion issues, *Photogrammetric Engineering & Remote Sensing*, 64(8):821–829.

McGarigal, K., and W.C. McComb, 1995. Relationships between landscape structure and breeding birds in the Oregon Coast Range, *Ecological Monographs*, 65(3):235–260.

Paul, J.F., R.L. Comeleo, and J.L. Copeland, 2002. Quantitative relationships between landscape metrics and sediment contamination for small estuarine systems in the Mid-Atlantic and Southern New England regions of the United States, *Journal of Environmental Quality*, 31(3):836–845.

Stehman, S.V., and R.L. Czaplewski, 1998. Design and analysis for thematic map accuracy assessment: Fundamental principles, *Remote Sensing of Environment*, 64:331–344.

USGS, 2003. *Accuracy Assessment of 1992 National Land Cover Data*, U.S. Geological Survey, http://landcover.usgs.gov/accuracy/, last accessed 24 December 2003.

Vogelmann, J.E., T. Sohl, and S.M. Howard, 1998a. Regional characterization of land cover using multiple sources of data, *Photogrammetric Engineering & Remote Sensing*, 64(1):45–57.

Vogelmann, J.E., T. Sohl, S.M. Howard, and D.M. Shaw, 1998b. Regional land cover characterization using Landsat Thematic Mapper data and ancillary data, *Environmental Monitoring and Assessment*, 51(1/2):415–428.

Vogelmann, J.E., and J.D. Wickham, 2000. *Implementation Strategy for Production of National Land Cover Data (NLCD) from the Landsat 7 Thematic Mapper Satellite*, EPA 600/R-00/051, Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C., 42 p.

Yang, L., S.V. Stehman, J.H. Smith, and J.D. Wickham, 2001. Thematic accuracy of MRLC land cover for the eastern United States, *Remote Sensing of Environment*, 76:418–422.

Zar, J.H., 1999. *Biostatistical Analysis,* Prentice Hall, Upper Saddle River, New Jersey, 663 p.

(Received 28 August 2002; accepted 17 February 2003; revised 01 May 2003)