

IMPROVING ESTIMATES OF THE ACCURACY OF THEMATIC MAPS WHEN USING AERIAL PHOTOS AS THE GROUND REFERENCE SOURCE

Siamak Khorram, Xiaolong Dai, Joseph Knight, Halil Cakir, Hui Yuan, Zhiyan Mao

Siamak Khorram
Professor and Director
Center for Earth Observation / North Carolina State University
Box 7106
Raleigh, NC 27695-7106
Khorram@ncsu.edu

Joseph Knight
Research Associate
Center for Earth Observation
jfknight@unity.ncsu.edu

Cakir, Yuan, Mao
Graduate Students
Center for Earth Observation
hicakir, hyuan2, zmao - @unity.ncsu.edu

Xiaolong Dai
Applied Science Fiction
xldai@asf.com

Abstract

Thematic classifications based on remotely sensed data are frequently used in decision making for businesses, governments, and individuals. As such, it is very important to provide an estimate of the accuracy of these classifications. The standard point-for-point comparison method of accuracy assessment when using aerial photos as the reference source suffers from limitations which call into question its validity. Two of these limitations are: difficulty in choosing a single correct class from the reference photos and potential positional errors in locating the sample points on the photos. These problems may cause the accuracy of the classification to be under-estimated. The objective of this paper is to recommend a solution to these problems based on taking into account primary and alternate photo interpretation classes for the reference point and classified pixel values near the reference point. The study area for this project was the U.S. Southeastern States. NAPP 1:40000 aerial photos were used as the reference data source for a land cover classification of the study area. The accuracy assessment was composed of 1500 sample points which were assigned a primary and, if applicable, an alternate interpreted land cover class from the photos. The accuracy assessment procedure was in two parts: a standard error matrix analysis, and extended analysis that took into account both the primary and alternate interpreted land cover values and classified values in a three-by-three neighborhood of the sample points. This new procedure produced an accuracy assessment method which gives a more representative estimate of the accuracy of a thematic classification than the standard point-for-point method.

Introduction

Assessing the accuracy of thematic classifications based on remotely sensed data is critical to the utility of such classifications. A common source of ground reference data used in accuracy assessment of thematic maps is aerial photographs. A frequently used approach involves the selection of a set of randomly determined sample points which are examined on the reference photos and on the classified map. The classified value on the thematic map is compared with the interpreted class on the reference photos on a point-by-point basis. The results of this comparison are presented in an error matrix. The error matrix gives the overall percent accuracy of the classification and the omission and commission errors for each class.

Potential problems with using this approach are: 1) It is often difficult to choose only one correct class from the reference photos. For example, depending on the acquisition date and scale of the photos, it can be very difficult to choose between pasture and cropland areas. Similarly, site heterogeneity can also make choosing only one class problematic. 2) Accurate location of the sample points on the photos can be troublesome. In the case of standard hard-copy photos, location of the sample points is done by context. Even with newer soft-copy photogrammetric techniques, errors introduced in digitizing, georeferencing, and mosaicking aerial photos can cause uncertainty in the locations of sample points.

These issues call into question the validity of using the standard point-for-point comparison method of accuracy assessment. Difficulty in choosing a single correct class from the reference photos and potential positional errors in locating the sample points on the photos may cause the accuracy of the classification to be under-estimated. The objective of this paper is to recommend a solution to these problems based on taking into account primary and alternate photo interpretation classes for the reference point and classified pixel values near the reference point. In situations where the selection of only one correct class from the photos is difficult, an alternate class may be selected. The selection of primary and alternate classes provides a way to account for the natural heterogeneity present in many areas. Similarly, under-estimates of the accuracy of a thematic classification due to positional errors in locating the correct sample point on the photos can be improved by examining classified values in the neighborhood of the sample point. This new procedure produced an accuracy assessment method which gives a more representative estimate of the accuracy of a thematic classification than the standard point-for-point method.

Background

Assessing the accuracy of land cover maps generated from remotely sensed data is costly in both time and money. Particularly for large projects, collecting reference data from ground observations is often prohibitively expensive. Therefore using existed data such as aerial photos as a reference data source is widely accepted (Congalton, 1999; George, 1986). An important issue when using aerial photos as the reference source for accuracy assessment is that any errors in photo interpretation would be misrepresented as errors in the classification, thereby lowering the estimate of the classification accuracy (Congalton, 1991). A study by Biging and Congalton (1991) showed that photo-interpretation may not be suitable as a verification data source when highly detailed thematic maps are involved. In those situations, it may be difficult to make a very detailed interpretation of the aerial photography. Thus, the individual performance of a photo interpreter may affect the entire accuracy assessment procedure. Congalton and Mead (1983) proposed a technique to quantitatively test the degree of similarity between photo-interpreters and/or photo interpretation variables such as film/filter type, season, and scale. Using the same sets of photos, they examined the accuracy of photo interpretation for different photo interpreters. They showed that it is feasible to use multiple photo interpreters in an accuracy assessment, provided the interpreters are properly trained. Thus, it is crucial, in any project using aerial photo interpretation, that the interpreters be highly trained and that the procedures used be appropriate to the goals of the project.

Study Area and Sampling Scheme

The study area for this project was the U.S. Southeastern States, specifically North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Mississippi, and Alabama. This region contains a wide variety of land cover types, including grassy plains, mountains, saltwater marshes, coastal plains, piedmont, agricultural areas, and urban centers. The land cover data are from the Multi-Resolution Land Characteristics (MRLC) Consortium's National Land Cover Data (NLCD) (Loveland & Shaw, 1996; VanDriel et al, 1999; Vogelmann et al, 1998a,b).

The sampling scheme was a two level design. In the first level, the Primary Sampling Units (PSU) were the size of a USGS National Aerial Photography Program (NAPP) photo. Each PSU was randomly selected from a cluster of 128 photos. These clusters of photos were formed using a geographic unit frame of 30 minutes by 30 minutes. The level stage was a stratified random sample, within the PSUs, of 100 points per land-cover class. The selected points were referred to as Secondary Sampling Units (SSU). The number of points per photo ranged from one to approximately 70. The total number of sample points in the study was 1500 (100 for each of the fifteen land

cover classes in the study area). This approach was chosen over a standard stratified random sampling over the entire study area to reduce the cost of purchasing the NAPP photography (Zhu et al., In Press). The sampling design was based on the following criteria:

- ?? Ensuring the objectivity of sample selection and validity of statistical inferences drawn from the sample data;
- ?? Distributing the points spatially across the region to ensure adequate coverage of the entire region;
- ?? Ensuring a low cost approach in terms of budget and time;
- ?? Ease of implementation and analysis.

Methods

Training.

To establish equal footing for, and consistency among, the interpreters, a comprehensive training program was devised. The program consisted of full-day training sessions and "on the job" training. The formal classroom training sessions were led by two experienced airphoto interpretation and photogrammetry instructors. The training sessions included the following:

- ?? Discussion of color theory and photo interpretation techniques;
- ?? Understanding of the class definitions;
- ?? Interpretation of over 100 sample points of different classes followed by interactive discussions about potential discrepancies;
- ?? Creation of example points for later reference; and
- ?? Repetition of photo interpretation practice after the sessions.

The focus was on likely situations that the interpreters would encounter during the project. Each participant was presented with over 100 pre-selected sites and was asked to provide an interpretation of the land cover for these sites. Their calls were analyzed and subsequently discussed to minimize any misconceptions. During the "on the job" portion of the training, each interpreter was assigned 400-500 training points to examine. Their progress was monitored daily for accuracy and proper methodology. The interpreters kept a log of their calls and the points for which they were uncertain about the land cover classes. On a weekly basis, their questions were addressed by the project supervisors. The problematic sites (approximately 400) were discussed until each team member felt comfortable with the class definitions and their consistency in interpretation.

Photo Interpretation Procedure.

NAPP 1:40000 aerial photos were used as the reference data source for this project. The majority of the photos were color infra-red. However, in some parts of the study area color infra-red photos were not available. In these areas, black and white photos were substituted.

The 1500 sample points were each assigned randomly to one of the three photo interpreters. The interpreters were assigned 500 points plus 75 points drawn randomly from each of the other two interpreters' points. This gave each interpreter a total of 650 points and provided 225 overlapping points (15%) which were interpreted by all three interpreters. The overlapping points were used for consistency analysis and quality assurance.

Landsat Thematic Mapper False Color Composite images of the study area were used to assist in locating the sample points on the photos. The images were displayed using a commercial image processing software package (ERDAS Imagine). The locations of the sample points were then plotted on the images. The interpreters located the sample point on the corresponding NAPP photo based on the context from the TM image. For this study, a "point" was defined as a three-by-three pixel window on the TM imagery (90 meters by 90 meters). This corresponds to a 2.25 mm by 2.25 mm window on the photos.

The interpreters examined the sample points' characteristics on the aerial photo and determined the proper class label for each sample point. This was referred to as the **primary** land cover class for the sample. In situations where, due to heterogeneity of the sample or uncertainty in the correct class, a single land cover class could not be chosen, the interpreters also selected an **alternate** land cover class for the sample. This was very common in residential areas where points often fell in mixed areas of houses, lawns, trees, and other cover types. The primary classes were considered to be the correct classes for the samples and were used in the standard point-for-point accuracy assessment of classification. For the purposes of the additional neighborhood accuracy analysis presented in this paper, the alternate classes were considered to be as correct as the primary classes.

During the photo interpretation it was discovered that photos were not provided for twenty-six of the points. These points were left out of the sample, leaving a total of 1474 points. The 1474 points remaining in the sample

were each assigned a primary land cover class from the fifteen classes present in the study area. In addition, 839 (57%) of the points had an alternate class assigned. As mentioned above, the points with alternate classes were those that were composed of a mix of land cover classes, or where selection of a single class was not reasonable. Also, it should be noted that the Photo Interpreters did not have access to the classification results during the interpretation process.

The MRLC program utilizes a consistent classification scheme for all Federal Regions at approximately an Anderson Level II thematic detail. While there are 21 classes in the MRLC system, only 15 were mapped in the study area. The following classification scheme was applied to the data set:

| MRLC Class Name | Abbreviation |
|---|--------------|
| 1. Water | Wt |
| 2. Low Intensity Residential | R1 |
| 3. High Intensity Residential | R2 |
| 4. Commercial / transportation / industrial | Cm |
| 5. Bare Rock / Sand | Br |
| 6. Quarries / Strip Mines / Gravel | Qu |
| 7. Transitional | Tr |
| 8. Deciduous Forest | DF |
| 9. Evergreen Forest | EF |
| 10. Mixed Forest | MF |
| 11. Pasture/Hay | PH |
| 12. Row Crops | RC |
| 13. Other Grasses. | OG |
| 14. Woody Wetlands | WW |
| 15. Emergent Herbaceous Wetlands | EW |

Table 1. Land cover classes used in the study area and abbreviations

Accuracy Assessment Procedure.

The accuracy assessment was divided into two stages. First, a standard point-for-point accuracy assessment summarized by an error matrix. Second, additional analysis which took into account both the primary and alternate (if present) interpreted classes and a three-by-three pixel neighborhood around the provided sample points. The additional accuracy analysis was broken into six groups:

1. Primary interpreted class matches the classified value for the center pixel only (standard point-for-point method);
2. Primary interpreted class matches the most common classified value in a 3x3 area around the sample point;
3. Primary interpreted class matches any classified value in a 3x3 area around the sample point;
4. Primary or alternate interpreted class matches the classified value for the center pixel only;
5. Primary or alternate interpreted class matches the most common classified value in a 3x3 area around the sample point;
6. Primary or alternate interpreted class matches any classified value in a 3x3 area around the point.

Quality Assurance Procedures.

Strong Quality Assurance and Quality Control (QA/QC) procedures were enforced in this project. Upon completion of the training portion of the project, a test was given to determine how similarly the interpreters would interpret the same sites. The 225 overlap sites were chosen (75 per interpreter) for which all three interpreters would provide a land cover class. These overlap points were then analyzed for consistency. Initially, there was confusion as to the proper interpretation of some of the classes. This necessitated a re-training phase in which the interpreters collaboratively identified a series of training sites. Upon completion of this additional training, a consistency test was re-administered. The average overall agreement between the photo interpreters was 84%.

In addition to the training procedures and consistency checks, weekly meetings of the project team were held to review project progress and to discuss problems encountered. Apart from these formal meetings, discussions among the research group on an ad hoc basis provided an opportunity to discuss problems that occurred and, in most

cases, provide a solution on the spot. Information gathered from these meetings greatly increased the interpretation quality and consistency.

Results

The results of the accuracy analyses are presented in Tables 2-5. Table 2 is a standard point-by-point error matrix showing the relationship between the **primary** class chosen by the photo interpreters and the classified land-cover class. This table does not take into account any alternate classes or classified neighborhood values. The abbreviations across the top and sides are the classes listed in Table 1. The “Tot” designation is the total number of samples assigned to that class. The next field is the number correct, followed by the percent correct. At the bottom right of the table is the overall accuracy estimate. This method estimates the overall accuracy of the classification to be 43%.

Table 3 lists the accuracy estimates derived from incorporating the different combinations of primary and alternate classes and classified 3x3 neighborhood values.

| Overall Accuracy | | Interpreted Class | | | | | | | | | | | | | | | Tot | Corr | % |
|------------------------|------------|-------------------|------------|-----------|-----------|-----------|-----------|-----------|------------|-----------|------------|------------|------------|------------|-----------|-----------|-------------|------|-----|
| | | Wt | R1 | R2 | Cm | Br | Qu | Tr | DF | EF | MF | PH | RC | OG | WW | EW | | | |
| Land Cover Class | Wt | 87 | | | | | | 1 | | 1 | | | | 1 | 4 | 4 | 98 | 87 | 0.9 |
| | R1 | 1 | 46 | 1 | 3 | 2 | | | 5 | 1 | 24 | 1 | 2 | 11 | 2 | 1 | 100 | 46 | 0.5 |
| | R2 | | 49 | 11 | 21 | 3 | | | | 1 | 3 | 1 | | 11 | | | 100 | 11 | 0.1 |
| | Cm | 3 | 22 | 2 | 32 | 7 | 1 | 1 | 3 | 1 | 7 | 2 | 2 | 12 | 1 | 2 | 98 | 32 | 0.3 |
| | Br | 3 | 1 | | 1 | 33 | 3 | | | | 2 | 15 | 12 | 21 | | 9 | 100 | 33 | 0.3 |
| | Qu | 4 | | | 5 | 19 | 35 | 13 | 9 | | 5 | 4 | | 3 | 2 | 1 | 100 | 35 | 0.4 |
| | Tr | 2 | 2 | | | | | 29 | 7 | 7 | 19 | 12 | 5 | 7 | 8 | 2 | 100 | 29 | 0.3 |
| | DF | | | | | 1 | | 4 | 45 | | 30 | 4 | 3 | 4 | 2 | 1 | 94 | 45 | 0.5 |
| | EF | | 2 | | 1 | 2 | | 2 | 3 | 34 | 43 | | 1 | 1 | 9 | 1 | 99 | 34 | 0.3 |
| | MF | | 1 | | | | | 11 | 6 | 4 | 63 | 4 | 1 | 3 | 5 | | 98 | 63 | 0.6 |
| | PH | | 5 | | | | | 2 | 4 | 1 | 9 | 27 | 37 | 8 | | | 93 | 27 | 0.3 |
| | RC | | 1 | | | 1 | | 4 | 8 | | 4 | 18 | 56 | 4 | 3 | | 99 | 56 | 0.6 |
| | OG | 2 | 25 | 4 | 4 | | | | 1 | 2 | 4 | 9 | 5 | 40 | 1 | | 97 | 40 | 0.4 |
| | WW | 4 | 1 | | | 1 | | 5 | 9 | 6 | 16 | 1 | 2 | 2 | 43 | 8 | 98 | 43 | 0.4 |
| | EW | 3 | | 2 | 1 | 2 | | 1 | | 4 | 6 | 2 | 2 | 4 | 13 | 59 | 99 | 59 | 0.6 |
| | Tot | 109 | 155 | 20 | 68 | 71 | 39 | 73 | 100 | 62 | 235 | 100 | 128 | 132 | 93 | 88 | 1474 | | |
| Corr | 87 | 46 | 11 | 32 | 33 | 35 | 29 | 45 | 34 | 63 | 27 | 56 | 40 | 43 | 59 | | 640 | | |
| % | 0.8 | 0.3 | 0.6 | 0.5 | 0.5 | 0.9 | 0.4 | 0.5 | 0.5 | 0.3 | 0.3 | 0.4 | 0.3 | 0.5 | 0.7 | | 43% | | |

Table 2. This matrix shows the interpreted results for the 1474 completed points versus the land cover classification for those points.

| | |
|---|------------------|
| Primary interpreted class matches classified pixel: | 640/1474 = 43 % |
| Primary interpreted class is most common in classified 3x3 area: | 674/1474 = 46 % |
| Primary interpreted class matches any pixel in classified 3x3 area: | 859/1474 = 58 % |
| Primary or alternate interpreted class matches classified pixel: | 824/1474 = 56 % |
| Primary or alternate interpreted class is most common in 3x3 area: | 985/1474 = 67 % |
| Primary or alternate interpreted class matches any pixel in 3x3: | 1170/1474 = 79 % |

Table 3: Summary of extended analysis

| Class | Num | Primary PI Matches Classif. | Primary PI is Mode of 3x3 | Primary PI Matches any 3x3 | Prim or Alt PI Matches Classif. | Prim or Alt PI is Mode of 3x3 | Prim or Alt PI Matches any 3x3 |
|---------------|-------------|--|--|---|--|--|---|
| Wt | 109 | 87 | 84 | 92 | 95 | 94 | 100 |
| R1 | 155 | 46 | 60 | 81 | 69 | 124 | 135 |
| R2 | 20 | 11 | 8 | 11 | 11 | 15 | 16 |
| Cm | 68 | 32 | 35 | 41 | 39 | 44 | 49 |
| Br | 71 | 33 | 27 | 30 | 35 | 34 | 42 |
| Qu | 39 | 35 | 34 | 36 | 36 | 35 | 37 |
| Tr | 73 | 29 | 33 | 42 | 44 | 40 | 52 |
| DF | 100 | 45 | 60 | 68 | 55 | 79 | 83 |
| EF | 62 | 34 | 44 | 48 | 39 | 52 | 54 |
| MF | 235 | 63 | 68 | 110 | 98 | 148 | 187 |
| PH | 100 | 27 | 27 | 38 | 39 | 46 | 64 |
| RC | 128 | 56 | 56 | 83 | 82 | 83 | 102 |
| OG | 132 | 40 | 33 | 53 | 61 | 56 | 91 |
| WW | 93 | 43 | 47 | 59 | 53 | 68 | 84 |
| EW | 88 | 59 | 58 | 67 | 68 | 67 | 74 |
| Totals | 1474 | 640 | 674 | 859 | 824 | 985 | 1170 |

Table 4. Summary of extended accuracy analysis by class represented by the number of points in each category.

| Class | Num | Primary PI Matches Classif. | Primary PI is Mode of 3x3 | Primary PI Matches any 3x3 | Prim or Alt PI Matches Classif | Prim or Alt PI is Mode of 3x3 | Prim or Alt PI Matches any 3x3 |
|---------------|--------------|--|--|---|---|--|---|
| Wt | 100.0 | 80.6 | 77.8 | 85.2 | 88.0 | 87.0 | 92.6 |
| R1 | 100.0 | 30.0 | 38.7 | 52.3 | 44.5 | 80.0 | 87.1 |
| R2 | 100.0 | 57.9 | 42.1 | 57.9 | 57.9 | 78.9 | 84.2 |
| Cm | 100.0 | 46.4 | 50.7 | 59.4 | 56.5 | 63.8 | 71.0 |
| Br | 100.0 | 47.8 | 39.1 | 47.8 | 50.7 | 49.3 | 60.9 |
| Qu | 100.0 | 92.1 | 89.5 | 94.7 | 94.7 | 92.1 | 97.4 |
| Tr | 100.0 | 37.2 | 42.3 | 53.8 | 56.4 | 51.3 | 66.7 |
| DF | 100.0 | 46.5 | 60.6 | 68.7 | 55.6 | 79.8 | 83.8 |
| EF | 100.0 | 55.7 | 72.1 | 78.7 | 63.9 | 85.2 | 88.5 |
| MF | 100.0 | 27.2 | 29.8 | 48.2 | 43.0 | 64.9 | 82.0 |
| PH | 100.0 | 27.2 | 26.2 | 36.9 | 37.9 | 44.7 | 62.1 |
| RC | 100.0 | 44.5 | 43.8 | 64.8 | 64.1 | 64.8 | 79.7 |
| OG | 100.0 | 31.3 | 25.2 | 40.5 | 46.6 | 42.7 | 69.5 |
| WW | 100.0 | 44.8 | 49.0 | 61.5 | 55.2 | 70.8 | 87.5 |
| EW | 100.0 | 66.3 | 63.0 | 72.8 | 73.9 | 72.8 | 80.4 |
| Totals | 100.0 | 43.0 | 45.7 | 58.3 | 55.9 | 66.8 | 79.4 |

Table 5. Summarized percent classification accuracy by class. All values are in percent.

Tables 4 and 5 give the results of the extended analysis broken down by class. Table 4 shows the actual number of samples in each class, while Table 5 presents the same information in percents.

Discussion

The point-for-point method of accuracy assessment (Table 2) produced an overall accuracy estimate for this data set of 43%. Some of the classes were particularly problematic. The error matrix in Table 2 shows that there was much confusion between low intensity residential, high intensity residential, and commercial/transportation. It is likely that the definitions of these categories (see Appendix) contributed the most to this confusion. Categories that cannot be directly determined from remotely sensed data not only compromise the accuracy of classification,

but also the accuracy of reference data collection. In this case, all three categories belong to a general urban category and are distinguished from each other by the amount of vegetation. Generally speaking, it is difficult to identify the vegetation amount within a pixel, i.e. at sub-pixel level. Therefore, the definitions themselves contain ambiguity. The interpreters assigned many high intensity residential areas in the classified images to the low intensity residential and commercial/transportation class. Similarly, the interpreters had difficulty differentiating cropland and pasture/hay. This was attributed to the fact that both of these classes have very similar spatial patterns and commonly co-exist within the same general areas. Confusion also existed between classes of evergreen forest vs. mixed forest; deciduous forest vs. mixed forest; and transitional vs. several classes. The similarity of these classes on the photos made differentiation a very complex task for the interpreters. In addition, there are, no doubt, errors associated with incorrect location of the sample points on the photos. It is difficult to quantify such errors, but the interpreters reported that, for some points, they were not certain that they had accurately located the sample point on the photo. This was frequently due to spatial resolution differences between the image data and the photos and land cover change between the acquisition dates of the imagery and photos.

The main factors which combined to lower the overall estimate of classification accuracy included: ambiguous class definitions, heterogeneity of the study area, similarity between certain land classes, and positional errors in sample point location. To decrease the effects of these factors on the overall accuracy, extended accuracy analysis was performed which examined, not only the primary interpreted class, but also an alternate class (if specified), as well as the classified values in a three-by-three pixel neighborhood around the center point. The addition of an alternate interpreted class helped to compensate for confusion of class definitions, heterogeneity, and class similarity. Inclusion of the three-by-three classified neighborhood values was used to help compensate for positional errors.

Tables 3, 4, and 5 summarize the results of the extended analysis. The greatest increases in the estimate of overall accuracy occur when the alternate classes are taken into consideration. The estimate increases from 43%, with the point-for-point method, to 56%, with the inclusion of the alternate class values. The impact of the neighborhood values are not as great. After including the "most common" value in the classified three-by-three area, the overall estimate shows small increases over using the alternate classes alone. Only when the interpreted values are matched with "any pixel" in the classified three-by-three area are large increases in the overall accuracy estimate evident. In studies where it is possible to accurately locate the sample points on the reference source, the authors advocate use of the "primary or alternate interpreted class is most common in 3x3 area" and "primary or alternate interpreted class matches classified pixel" methods. These techniques allow the flexibility of using both primary and alternate classes, as well as, in the former, including the neighborhood values, but do not have the weakness of allowing a point to be classified as correct when only one nearby classified value matches the interpreted result. In studies where the reference source is fairly homogeneous, or in which there is some uncertainty in the spatial fidelity of the reference data, the "primary or alternate interpreted class matches any classified pixel" method may be appropriate.

Conclusion

Clearly, the estimate of overall accuracy was dramatically affected by the inclusion of alternate classes and classified neighborhood values. The question then becomes: Is this valid? The answer to this question is not a simple one. The validity of this approach depends upon the purpose of the classification. One would not want to use the technique described in a study where there is little similarity between classes. Inclusion of an alternate class that is assumed to be as correct as the primary in an application where there are great inter-class differences would certainly not be appropriate. Also, assessing the accuracy of highly detailed thematic maps, as opposed to regional or national mapping projects such as MRLC's NLCD, using an approach where neighborhood values are taken into consideration may not be desirable. It is in large-scale studies, where the thematic information of interest contains potentially very similar classes, such as land cover classes, that an approach such as the one described is appropriate and useful.

Acknowledgements

This research was funded by the United States Environmental Protection Agency as a part of the MRLC-NLCD Region IV accuracy assessment. Sample point coordinate locations and NAPP photography were provided by the EROS Data Center in Sioux Falls, S.D. The authors would like to thank Jim Wickham, Limin Yang, and Steve Stehman for their valuable guidance in the completion of this project.

Appendix: MRLC Class Definitions

Water – All areas of open water or permanent ice/snow cover

1. Water - All areas of open water, generally with less than 25% cover of vegetation/land cover.

Developed – Areas characterized by high percentage of construction materials (e.g., asphalt, concrete, building, etc).

2. Low intensity residential – Land includes areas with a mixture of constructed materials and vegetation or other cover. Constructed materials account 30-80 percent of the total area. These areas most commonly include single-family housing areas, especially suburban neighborhoods. Generally, population density values in this class will be lower than in high-intensity residential areas.
3. High Intensity residential – Includes heavily built-up urban centers where people reside. Examples include apartment complexes and row houses. Vegetation occupies less than 25% of the landscape. Constructed materials account for 80-100% of the total area. Typically, population densities will be quite high in these areas.
4. High Intensity Commercial/Industrial/transportation – Includes all highly developed lands not classified as “High Intensity Residential”, most of which is commercial, industrial, and transportation.

Barren – Bare rock, sand, silt, gravel, or other earthen material with little or no vegetation regardless of its inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the vegetated categories.

5. Bare Rock/Sand – Includes areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, beach, and other accumulations of rock and /or sand without vegetative cover.
6. Quarries/Strip Mines/Gravel Pits – Areas of extractive mining activities with significant surface expression.
7. Transitional – Areas dynamically changing from one land cover to another, often because of land use activities. Examples include forestlands cleared for timber, and may include both freshly cleared areas as well as areas in the earliest stages of forest growth.

Natural Forested Upland (non-wet) – A class of vegetation dominated by trees generally forming > 25% canopy cover.

8. Deciduous Forest – Areas dominated by trees where 75% or more of the tree species shed foliage simultaneously in response to an unfavorable season.
9. Evergreen Forest – Areas dominated by trees where 75% or more of the tree species maintain their leaves all year. Canopy is never without green foliage.
10. Mixed Forest – Areas dominated by trees where neither deciduous nor evergreen species represent more than 75% of the cover present.

Herbaceous Planted/Cultivated – Areas dominated with vegetation which has been planted in its current location by humans, and/or is treated with annual tillage, a modified conservation tillage, or other intensive management or manipulation. The majority of vegetation in these areas is planted an/or maintained for the production of food, fiber, feed, or seed.

11. Pasture/Hay – Grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops.
12. Row Crops – All areas used for the production of crops, such as corn, soybeans, vegetables, tobacco, and cotton.
13. Other Grasses – Vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes. Examples include parks, lawns, and golf courses.

Wetlands – Non-woody or woody vegetation where the substrate is periodically saturated with or covered with water.

14. Woody wetlands – Areas of forested or shrubland vegetation where soil or substrate is periodically saturated with or covered with water.
15. Emergent Herbaceous wetlands – Non-woody vascular perennial vegetation where the soil or substrate is periodically saturated with or covered with water.

References

- Biging, G. S. and R. G. Congalton. (1991). A Comparison of Photointerpretation and Ground measurements of Forest Structure. Proceedings, 1991 ACSM-ASPRS Annual Convention, Baltimore, MD, Vol. 3, pp. 6-15.
- Congalton, R. G. and R. A. Mead. (1983). A Quantitative Method to Test for Consistency and Correctness in Photointerpretation. Photogrammetric Engineering and Remote Sensing, Vol. 49, No. 1, pp. 69-74.
- Congalton, R. G. (1988). A Comparison of Sampling Schemes Used in Generating Error Matrices for Assessing the Accuracy of Maps Generated from remotely Sensed Data. Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 5, pp. 593-600.
- Congalton, R. G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote Sensing of Environment, Vol. 37, pp. 35-46.
- Congalton, R.G., K. Green. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. CRC Press, Inc., Boca Raton, FL., pp 29-30.
- George, T. H., 1986. Aerial Verification of Polygonal Resource Maps; A Low-Cost Approach to Accuracy Assessment. Photogrammetric Engineering and Remote Sensing, Vol. 52, No. 6, pp. 839-846.
- Loveland, T.R., and D.M. Shaw. (1996). Multiresolution land characterization: building collaborative partnerships, Gap Analysis: A Landscape Approach to Biodiversity Planning (J.M. Scott, T. Tear, and F. Davis, editors), Proceedings of the ASPRS/GAP Symposium, Charlotte, North Carolina, National Biological Service, Moscow, Idaho, pp 83-89.
- Van Driel, N., T. Loveland, and D. Lauer. (1999). The MultiResolution Land Characteristics Consortium, a government success story. The 50th International Astronautical Congress, October, 1999, Amsterdam, The Netherlands.
- Vogelmann, J.E., T.L., Sohl, P.V. Campbell, and D.M. Shaw. (1998a). Regional land cover characterization using Landsat Thematic Mapper data and ancillary data sources, Environmental Monitoring and Assessment 51: 415-428.
- Vogelmann, J.E., T. Sohl, and S.M. Howard. (1998b). Regional characterization of land cover using multiple sources of data, Photogrammetric Engineering and Remote Sensing 64: 45-57.
- Zhu, Z, L. Yang, S.V. Stehman, and R.L. Czaplewski. In Press. Accuracy assessment for the U.S. Geological Survey regional land cover mapping program: New York and New Jersey Region. Photogrammetric Engineering and Remote Sensing.