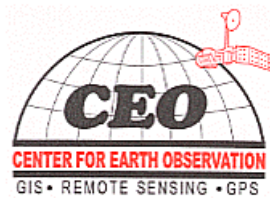


Final Report

Submitted to
Environmental Protection Agency (EPA)
for the MRLC Accuracy Assessment Service Agreement Entitled

**Accuracy Assessment of the EPA Region IV Dataset of the MRLC
Land Cover Mapping Program**

Submitted by



**Center for Earth Observation
North Carolina State University**

U.S. Congressional District 4

Final Report

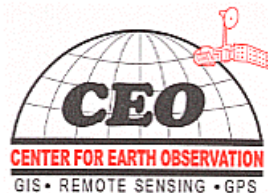
Submitted to
Environmental Protection Agency (EPA)
for the MRLC Accuracy Assessment Service Agreement Entitled

**Accuracy Assessment of the EPA Region IV Dataset of the MRLC
Land Cover Mapping Program**

Submitted by

Siamak Khorram, Center for Earth Observation, NCSU, Principal Investigator
X. Long Dai, Center for Earth Observation, NCSU, Project Supervisor
Joe Knight, Center for Earth Observation, NCSU, Photo Interpretation
Hui Yuan, Center for Earth Observation, NCSU, Photo Interpreter
Halil Cakir, Center for Earth Observation, NCSU, Photo Interpreter
Ziyan Mao, Center for Earth Observation, NCSU, Photo Interpreter

Jim Wickham, Environmental Protection Agency, Technical Monitor
Limin Yang, EROS Data Center, USGS, Project Advisor



Center for Earth Observation
North Carolina State University

U.S. Congressional District 4

Contacts:

khorr@ncsu.edu

xdai@unity.ncsu.edu

Box 7106, North Carolina State University

Raleigh, North Carolina 27695-7106

Phone: (919)515-3430

FAX: (919)515-3439

**Accuracy Assessment of the EPA Region IV Dataset of the MRLC
Land Cover Mapping Program**

Summary

The Center for Earth Observation (CEO) at North Carolina State University (NCSU) was funded by the U.S. Environmental Protection Agency (EPA) under a service agreement to provide labor, technical support, and computer services to EPA for estimation and analysis of the accuracy of the regional MRLC classified land cover products generated from Landsat TM data. The classification of MRLC data was done for EPA Region IV by the USGS EROS Data Center (EDC) in Sioux Falls, S.D. An implementation scheme of the project has been designed by the CEO group to assure the quality of reference data collection. The major components in this scheme include formal training of photo interpretation, on-the-job training of photo interpreters, photo interpretation, weekly group meetings and discussion, PI agreement analysis using overlapping points, and hierarchical QA/QC procedures. Results include the reference database of 1500 sample points collected by photo interpretation, and initial accuracy assessment as compared to the classified data. Recommendations include: 1) Examining the impact of alternate classes in the accuracy assessment; 2) Evaluating and analyzing the effect of positional errors on accuracy assessment; 3) Fuzzy set analysis using fuzzy class assignments for each point, etc. The findings in this project could provide a base for future research and analysis for other regions. The experience from this project will also be valuable for an implementation of similar procedures for other EPA Regions and similar projects in land use/land cover accuracy assessment.

Acknowledgments

The results reported here were generated through a Service Agreement funded by the Environmental Protection Agency (EPA). The views expressed in this report are those of the authors and do not necessarily reflect the views of EPA or any of its sub-agencies. The authors would like to thank EPA, USGS-EROS Data Center (EDC) for their support and for the assistance given on this project. The authors would also like to thank Dr. Heather Cheshire and Ms. Linda Babcock of CEO at NCSU for contributing their expertise in photo interpretation and extended help throughout the duration of this project.

1. Introduction

The U.S. Geological Survey (USGS) EROS Data Center and the U.S. Environmental Protection Agency (EPA) have jointly initialized Multi-Resolution Land Characterization (MRLC) Consortium Land Cover Mapping program using Landsat Thematic Mapper (TM) 30-meter resolution imagery as the baseline data and successfully produced a consistent and conterminous land cover map of the lower 48 states at approximately an Anderson Level II thematic detail. The central goal of the program is to provide a generalized and regionally consistent land cover product for use in a broad range of applications (Lunetta *et al.*, 1998). Each of the ten EPA Federal regions is mapped independently.

Accuracy assessment must be an integral component of any remote sensing-based mapping project. Literally, thematic accuracy assessment is to measure general and categorical qualities of the data (Khorram *et al.*, 1999). A formal accuracy assessment is implemented for each region independently after the land cover mapping program is completed. There are two primary motivations for the accuracy assessment:

- ?? To estimate the overall accuracy and category-specific accuracy
- ?? To evaluate the overall assessment of the reliability of a land cover map from TM data.

Quantitative accuracy assessment of large-area land cover maps, produced from remote sensed data, involves comparing thematic map with reference data often referred to as ground truth (Congalton, 1991). Since there is no suitable existing reference data that can be used consistently for all federal regions, a practical and statistically sound sampling plan is carefully designed to characterize the accuracy of common and rare classes for the map product. The MRLC program has designed a two-stage cluster sampling scheme and selected the National Aerial Photography Program (NAPP) photography to be used as the primary reference data.

To conduct the accuracy assessment for the Region IV (Southeast States), the USGS EROS Data Center (EDC) in Sioux Falls, SD, an MRLC consortium team member, provided the Center for Earth Observation (CEO) of North Carolina State University (NCSU) the following data and information:

- ?? The geographic locations of approximately 1500 reference points in a digital file of X, Y map coordinates;
- ?? NAPP photos containing these reference(sample) points;
- ?? Landsat TM band 3, 4, 5 digital data for the entire region;
- ?? A NAPP flightline coverage and photo index of all photos involved.

Using these data and following the MRLC protocol for reference data collection and evaluation for accuracy assessment, CEO at NCSU provided computer resources and staff to conduct reference data collection by interpreting the preselected 1500 points on NAPP photography and observing the quality assurance (QA) and quality control (QC).

2. Objective

The objective was to collect reference data for accuracy assessment by interpreting NAPP photography and to analyze the accuracy of the regional MRLC classified land cover products generated from Landsat TM data for the EPA Region 4 area by the EROS Data Center in Sioux Falls, SD. Moreover, this project will hopefully serve as a base for future research and provide similar projects with invaluable data, information, and experience.

3. Study Area and Sample Determination

The study area for this project was EPA Region 4, which is composed of 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.

NAPP photographs, 386 in all, were chosen randomly throughout the study area (Figure 1) by the EROS Data Center (EDC). Random point locations were then chosen in each of the selected photos (Figure 2). The number of points per photo ranges from one to approximately 70. The total number of sample points in the study is 1500 (Figure 3). This approach was chosen by the EDC over a standard random sample selection in the entire study area in order to minimize the cost of purchasing the NAPP photography (Zhu et al., In Review).

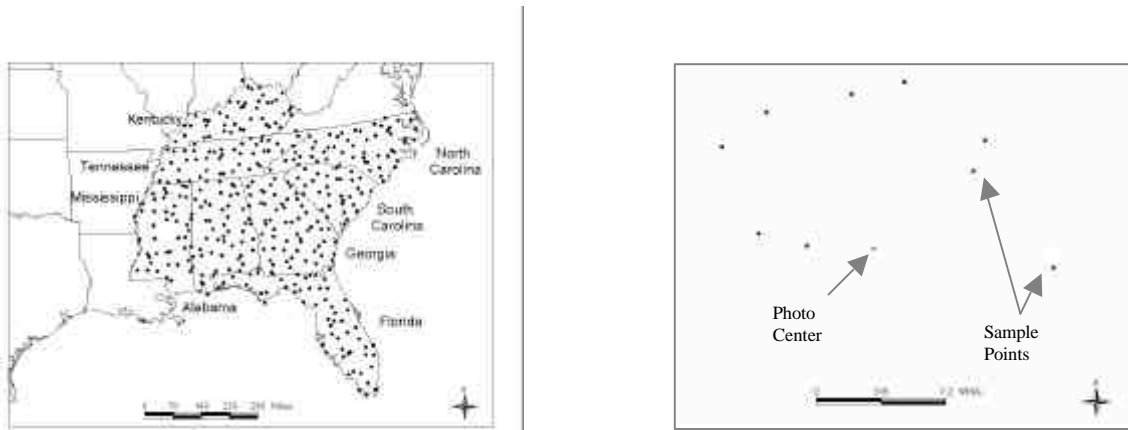


Figure 1. Randomly chosen photograph centers. Figure 2. Sample sites clustered around the photo center.

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;
- ?? Reducing the variance for estimated accuracy parameters;
- ?? Ensuring the low cost approach in terms of budget and time;
- ?? Being easy to implement and analyze.



Figure 3. Distribution of 1500 sample sites.

4. Training

To provide equal footing for and consistency among the interpreters, a comprehensive training program was devised. The program consisted of a full-day training session 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 during the training sessions followed by interactive discussions about potential discrepancies;
- ?? Creation of sample points for later reference; and
- ?? Repetition of PI practice after the sessions.

The focus was on real world situations that interpreters would encounter during the project. Each participant was presented with over 100 pre-selected sites and was asked to provide their 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 approximately 500 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 Photo Interpretation Supervisor. The problem sites (approximately 400) were discussed until each team member felt comfortable with the class definitions and, their consistency in interpretation.

5. Preprocessing of Imagery

The TM imagery for the study area provided by EDC in Sioux Falls was composed of the Landsat TM false color composite images of bands 5 (Infrared), 4 (Red), and 3 (Green). The data

set included 4 subregions and 39 image segments in generic binary format. These image segments are shown in Figure 4 and needed to be properly processed and assembled before they could be used in the project. Data preprocessing included the following six sequential steps:

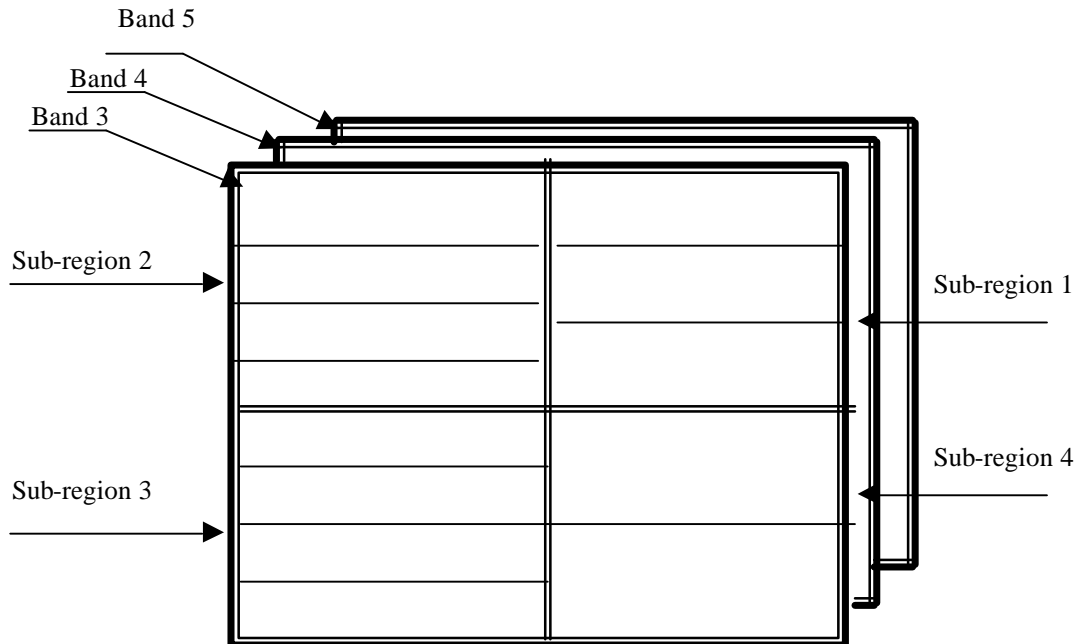


Figure 4. Image preprocessing for EPA region IV.

- ?? Reading generic binary files from the tapes in Unix system.
- ?? Importing these files into Erdas Imagine.
- ?? Re-projection. To have a consistent projection system between data sets, re-projection was performed and the projection system was changed to “Albers Conical Equal Area” coordinate system.
- ?? Mosaicking images. In order to get one image file for each sub-region, image segments were mosaicked.
- ?? Stacking and Subsetting. The corresponding bands were then stacked together. Because total size of the resultant file including bands 3, 4, and 5 were more than two giga-bytes which is over the capacity of the system, each sub-region was subset into two scenes while stacking the layers. Subregion 4 was left as one scene.
- ?? Backup. After all the preprocessing procedures were done, all the subset images were backed up to tapes for data integrity.

6. Photo Interpretation

6.1. Photo Interpretation Protocol

During the photo interpretation, there was a common photo interpretation protocol to follow for all photo interpreters.

- ?? Photo interpretation should be based on spectral imagery and NAPP photographs to maintain the objectivity of the PI.
- ?? The corresponding location on NAPP is found by consulting with spatial patterns apparent on TM color composite image.
- ?? When the PI is interpreting points, land cover class and related information such as homogeneity information should be recorded for later analysis.
- ?? When there are differences in acquisition times between image and photo, the PI results are based on image while recording the change.
- ?? In homogeneous area, the PI result is preferred to be based on only one pixel where the point is located.
- ?? In heterogeneous area, 3x3 pixels are considered to interpret the point.
- ?? For mixed pixels, both primary and secondary land cover classes should be recorded. Either of the two should be thought to be correct.
- ?? Confidence of photo interpretation will be given for each point by each PI.

6.2. Sample Point Preparation and PI Team

The bounding coordinates of the four TM images provided by EDC were used to separate photos into four subregions. The photo center coordinates were overlaid on the image and photos were assigned into one of the four subregions. In case of overlap, the photo centers were arbitrarily assigned to one subregion so as to keep the number of photos per subregion as equal as possible.

Upon the receipt of sample point coordinates, a similar procedure was used to determine on which photo a sample point was located. A bounding rectangle was established which represented the dimensions of a NAPP photo. A Visual Basic program was written to use these dimensions to separate the sample points.

The photos were then assigned randomly to one of the three photo interpreters. Each interpreter completed approximately 500 points plus 15% overlap, for a total 650 points. The overlapping 15% of the points were interpreted twice by all three interpreters, first independently and then in conference, and were used for quality assurance purposes. The 15% overlap was distributed equally between the four subregions of the study area. The PI Team Organization is shown in Table 1.

PI Organization

Photo Interpreters	PI #1 (500 pts + 75 pts from PI #2 and 75 pts from PI #3)	PI #2 (500 pts + 75 pts from PI #1 and 75 pts from PI #3)	PI #3 (500 pts + 75 pts from PI #1 and 75 pts from PI #2)
PI supervisor	Random checking for consistency, checking 225 overlapped points, points with question from three PIs.		
Project supervisor	Checking points with question from PI Supervisor, Random checking of overall points, Overall QA/QC.		
Project director	Procedure establishment, Discussions on issues, Random checking, Overall QA/QC.		

Table 1. Photo-interpretation team organization.

6.3. Classification Scheme and Class Definitions

The MRLC program utilizes a consistent classification scheme for all EPA Regions at approximately an Anderson Level II thematic detail. The following classification scheme was applied to EPA Region IV data set:

- (1) Water
- (2) Low Intensity Residential
- (3) High Intensity Residential
- (4) High Intensity residential
- (5) Bare Rock / Sand
- (6) Quarries / Strip Mines / Gravel
- (7) Transitional
- (8) Deciduous Forest
- (9) Evergreen Forest
- (10) Mixed Forest
- (11) Pasture/Hay
- (12) Row Crops
- (13) Other Grasses.
- (14) Woody Wetlands
- (15) Emergent Herbaceous Wetlands

The class definitions are as follows:

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

1.1 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).

2.1 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.

- 2.2 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.
 - 2.3 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.
- 3.1 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.
 - 3.2 Quarries/Strip Mines/Gravel Pits – Areas of extractive mining activities with significant surface expression.
 - 3.3 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.
- 4.1 Deciduous Forest – Areas dominated by trees where 75% or more of the tree species shed foliage simultaneously in response to an unfavorable season.
 - 4.2 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.
 - 4.3 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.
- 8.1 Pasture/Hay – Grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops.
 - 8.2 Row Crops – All areas used for the production of crops, such as corn, soybeans, vegetables, tobacco, and cotton.
 - 8.5 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.
- 9.1 Woody wetlands – Areas of forested or shrubland vegetation where soil or substrate is periodically saturated with or covered with water.

9.2 Emergent Herbaceous wetlands – Non-woody vascular perennial vegetation where the soil or substrate is periodically saturated with or covered with water.

6.4. Interpretation Procedures

The Landsat TM images were displayed using Erdas Imagine. By inputting the coordinates of the point, the photo interpreters precisely located each point on the image. Then based on the coordinates and the context from the Landsat TM False Color Composite (FCC) image, PI located the point precisely on the photo. PIs examined the point's characteristics using aerial photo and FCC image and determined the Land Use/Land Cover label for each sample point according to the land cover classification scheme. Each PI was responsible to enter all information into the database sheet.

A computer database was designed and developed to store and organize the data collected in this project. This database was created using Microsoft Access and consists of one table to store data and one two-page form to enter, edit, and display the data. The data sheet consisted of two parts; the first part is for the regular accuracy assessment and the second part is for the fuzzy analysis.

The first part of the data sheet allowed the PI to put additional information such as secondary class for uncertainty of primary class, whether the site was homogeneous or heterogeneous, changes occurred through time, and general site description, etc. It was designed to enter following information: Site (Sample Point) ID number, Coordinates, Photo Acquisition Date, Photo ID, Image ID, Primary/ Dominant Class, Alternative Class (if any), General Site Description, Unusual Observations, General Comments, and the information that whether the site is changed or not. The second part of the data sheet, fuzzy part, allowed the PI to interpret the confidence of each point for different classes. Confidence intervals include: Absolutely Correct: there is no doubt that this is the correct class for the site; Probably Correct: it is likely that this is the correct class; Acceptable: maybe not the best possible answer but it is reasonable and acceptable; Probably Incorrect: not a good answer and there is clearly a better class; Absolutely Incorrect: totally unacceptable. These additional information could be helpful in the future research on this subject.

The database was designed to have a built-in error-checking and data-validation rules. For example, coordinates entered that are not within the study area would be rejected. Also, typographical errors are minimized by having the interpreters select items from menus whenever possible. In addition, comprehensive training was provided on database use and maintenance.

Photo interpretation was based on the TM imagery and NAPP photography. The definition of each site was consistent with that in the sample formation. Each site was defined as one pixel in TM images. A window of the corresponding size on the photography was generated and used to interpret each site. Then, the dominant class within the window was used as labeling protocol. Based on these guidelines, PI determined the land use/land cover class for each sample point, and then entered the information into Microsoft Access data

Each PI interpreted 15% (75 points) of other two PIs' sample points in order to create an overlap to allow us to evaluate the performance of PIs and the agreement between them. Selection of those 75 points was done through random sampling. Total of 225 points (3 PI times 75) were selected. Each PI has then independently interpreted these overlapping points from other two PIs, total of 150 points. For example, PI-1 has interpreted 75 points from PI-2 and 75 points from PI-

3. This scheme gave us 225 points that were interpreted by all three PIs. These 225 points were used for agreement analysis between the three PIs.

6.5. Problems and Discussions

The problems that we have identified and/or encountered over the period of this project have stimulated many fruitful discussions both in the meetings and over the interpretation tables. The recommendations and solutions to these problems were formulated and used in this project. It is further hoped that the information and experience from this project will benefit similar types of projects in the future. These problems, recommendations, and solutions are reported here.

?? Mixed pixel problem.

Interpretation is sometimes problematic due to heterogeneous nature of some areas. The question was to choose the size of reference point (area) on the photo. Should it be an exact point, a pixel size, or three by three pixel size? In the case of low intensity residential area, if you choose 3 by 3 pixel size, the final decision would mostly be the low intensity residential area. But if one chooses a point, it would be hard to locate the exact point and the PI would end up deciding between two classes such as between low intensity residential and other grassland in most cases. Since we are interpreting the satellite images one pixel size is recommended be chosen from reference area. In the case of mixed pixel or heterogeneous areas, further analyzing the neighboring pixels on the satellite imagery is recommended. Usually this method is helpful for interpretation.

?? Difference in Acquisition Dates

There was a need for establishing criteria to answer the question whether the site was changed or not between the two acquisition dates (one for Landsat TM and the other for NAPP photos). Most of the questions arose because of the time differences between photo and image acquisition dates. For example, when the PI had to interpret early forest growth areas (based on the class definitions, those sites were defined as transitional areas), PI had to decide whether the site was a transitional or a forested area. If the photo was acquired before the image (sometimes 6 years earlier), it is obvious that those early forest growth sites would show up as forest cover on the satellite image. The question was “should we stick with photos or should we use the best knowledge and say “forested area” as in this example?”

In this case, the recommendation was “use your best knowledge and decide according to satellite imagery (what was recorded by the satellite at that time?).” For these situations, the date of image acquisition was provided by EROS data center to help the PI. Before that, since the subregions consist of many TM scenes, acquisition date of imagery corresponding to a particular reference point was impossible to determine.

?? Heterogeneity Problem

The nature of heterogeneity of many areas caused interpreters much confusion to assign the point to an exact class label. Since the spatial resolution of the Landsat TM data is 30 meters by

meters, in many cases one pixel could consist of more than one land cover class. For example, a pixel where the point is located on the image is often composed of some trees, some grassland and several houses, so its reflectance of the pixel actually combines different reflectance from several classes. In this case, we interpret the point using the 3*3 window and the majority of surrounding land cover.

?? Points without Photos or with Wrong Photos

The following sample points, as shown in Table 2, have been found having problems either having no photographs or with wrong photographs. These points have been reported to EDC for further checking.

Point #	Point ID	Photo ID	Problems
1	1190	NP0NAPP000031209	wrong photo
2	1199	NP0NAPP000036199	wrong photo
3	1217	NP0NAPP000036199	wrong photo
4	1218	NP0NAPP000036199	wrong photo
5	1297	NP0NAPP000036199	wrong photo
6	1332	NP0NAPP000036199	wrong photo
7	1165	NP0NAPP000047116	wrong photo
8	1211	NP0NAPP000047116	wrong photo
9	1215	NP0NAPP000047116	wrong photo
10	1220	NP0NAPP000047116	wrong photo
11	1095	NP0NAPP000052194	wrong photo
12	1156	N10NAPPW04700202	wrong photo
13	1396	N10NAPPW04700203	wrong photo
14	1034	NP0NAPP000697142	no photo
15	1048	NP0NAPP000697142	no photo
16	1071	NP0NAPP000697142	no photo
17	647	NP0NAPP000651076	wrong photo
18	1476	NP0NAPP000651076	wrong photo
19	1054	NP0NAPP000031249	wrong photo
20	208	NP0NAPP000037206	no photo
21	326	NP0NAPP000037206	no photo
22	496	NP0NAPP000037206	no photo
23	133	NP0NAPP000043166	no photo
24	219	NP0NAPP000043167	no photo
25	306	NP0NAPP000043168	no photo
26	566	NP0NAPP000043169	no photo

Table 2. Points without photo or with wrong photos.

7. Quality Assurance/Quality Control (QA/QC)

During the entire PI process, QA/QC procedures were enforced as designed in the PI Organization Chart, as shown in Table 1. In addition to the interactions between the CEO group and the scientists from EPA and EROS Data Center scientists, a weekly meeting, hosted by the

Project Director and Project Supervisor, among the CEO group was held during the entire project to review project progress and discuss problems encountered. Besides, discussions among the CEO group on an individual basis have provided an opportunity to discuss problems that occurred and solve the problems on the spot. Information gathered from these meetings helped the PI quality.

The relationship between QA/QC and PI can be described in Figure 5. Upon completion of the training, a test was given to determine how similarly the interpreters would call the same sites. 225 sites were chosen (75 per interpreter) for which all three interpreters would provide a land cover class. These overlap points were then compared using standard error matrices. In addition, the 225 overlap points were classified by the Photo Interpretation Supervisor to ensure that the interpreters were making calls consistent with their training and previous experience.

The initial results of the overlap analysis revealed that some misunderstandings about class definitions had crept through the training process. As a result, the interpreters re-evaluated the 225 overlap points as a group to "calibrate" themselves. This helped to ensure that their calls were more consistent between interpreters. The consensus results for these 225 overlap points were also used for agreement analysis.

Upon satisfactory completion of the overlap points, the interpreters were assigned to complete the 1500 sample points throughout Region 4. Each interpreter was assigned approximately 500 points. Remaining questions regarding the correct call for a particular point were discussed with the Photo Interpretation Supervisor until any problems were resolved. Based on the final PI results, an agreement analysis was conducted, as shown in Table 3. On average, the agreement between three PIs reached 84%.

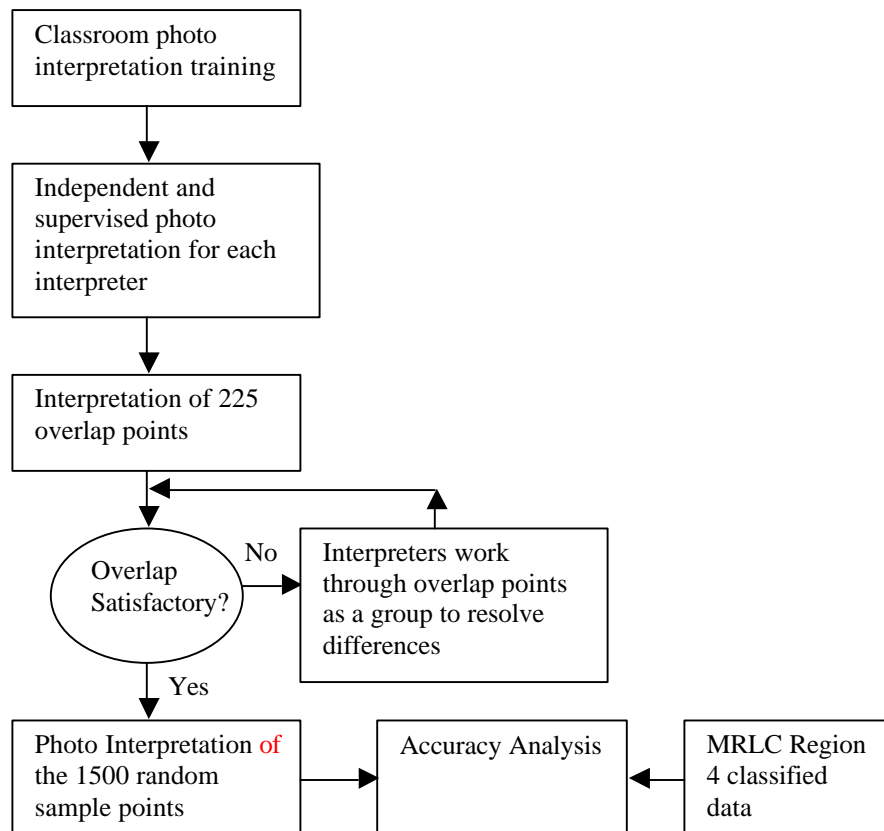


Figure 5. Training, PI, and QA/QC Procedures.

		Overlap Analysis															Tot	%	Corr	
		Overlap Consensus																		
Interpreted Results		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15				
	1	18	1													1		20	0.9	18
	2		21												1			22	1	21
	3			3	1													4	0.8	3
	4				9													9	1	9
	5					4								2				6	0.7	4
	6	1					6											7	0.9	6
	7							16	1							1		18	0.9	16
	8								2	14		1	1	1				19	0.7	14
	9								2		7							9	0.8	7
	10							3	2	1	26						4	36	0.7	26
	11											10	1					11	0.9	10
	12											3	10		1			14	0.7	10
	13								1						16		2	19	0.8	16
	14										2					13	1	16	0.8	13
	15																15	15	1	15
Tot		19	22	3	10	4	6	23	18	8	29	14	14	18	19	18	225			
%		0.9	1	1	0.9	1	1	0.7	0.8	0.9	0.9	0.7	0.7	0.9	0.7	0.8		0.84		
Corr		18	21	3	9	4	6	16	14	7	26	10	10	16	13	15			188	

Table 3. Agreement analysis between PIs: interpreter call vs. overlap consensus for the 225 overlap points.

8. Results

8.1 Initial Analysis

The PI results of 1474 sample points are stored in Access format and ready to be delivered.

The following are two error matrices from our initial analysis that were generated during the training, QA, and interpretation processes. Each is accompanied by an explanation of what it represents. Unless noted otherwise, the numbers across the top and sides of the matrices are the 15 MRLC landcover classes.

Overall Accuracy		Classified															Tot	%	Corr
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
Interpreted	1	87			3	3	5	2					2	4	2	108	0.8	87	
	2		47	49	22	1		2		2	1	5	1	24	1	155	0.3	47	
	3			1	10	2								4		19	0.5	10	
	4			3	22	32	1	5		1				4		69	0.5	32	
	5			2	3	6	33	18		1	2			1	1	69	0.5	33	
	6					1	3	34								38	0.9	34	
	7	1					13	33	4	2	12	3	4		5	78	0.4	33	
	8		6		3		8	6	46	3	7	3	7	1	9	99	0.5	46	
	9	1	1	1	1			7		34	4			2	6	61	0.6	34	
	10		24	3	7	2	6	16	29	42	62	9	4	4	16	228	0.3	62	
	11			2	1	2	15	4	11	4		4	28	18	11	103	0.3	28	
	12			1		3	11	1	6	3	1	1	37	57	3	128	0.4	57	
	13	1	10	11	13	20	3	7	4	1	3	8	4	41	2	131	0.3	41	
	14	4	2		1	1	2	8	2	10	4		3	1	43	96	0.4	43	
	15	4	1		2	10	1	2	1	1					9	91	0.7	60	
Tot	98	100	100	98	100	100	100	94	99	98	93	99	97	99	98	1474			
%	0.9	0.5	0.1	0.3	0.3	0.3	0.3	0.5	0.3	0.6	0.3	0.6	0.4	0.4	0.6		0.44		
Corr	87	47	10	32	33	34	33	46	34	62	28	57	41	43	60			647	

Table 4. This matrix shows the interpreted results for the 1474 points that were finished versus the MRLC's classification for those points.

		MRLC Categories							Tot	%	Corr
		1	2	3	4	5	6	7			
PI Results Cats	1	87	4	7		2	7	2	109	0.8	87
	2		187	7	4	39	4	2	243	0.8	187
	3			13	90	3	1	3	110	0.8	90
	4	1	45	16	228	33	41	33	397	0.6	228
	5	1	42	55	21	204	13	24	360	0.6	204
	6	8	6	12	18	4	124	10	182	0.7	124
	7	1	1	13	17	6	6	29	73	0.4	29
	Tot	98	298	200	291	289	198	100	1474		
%	0.9	0.6	0.5	0.8	0.7	0.6	0.3		0.64		
Corr	87	187	90	228	204	124	29			949	

Table 5. This matrix shows the interpreted results for the 1474 points that were finished versus the MRLC's classification for those points.

The standard 15 MRLC classes were grouped into the following categories:

- (1) Water
- (2) Urban
- (3) Bare surface
- (4) Agriculture
- (5) Forest
- (6) Wetland
- (7) Transitional

The following additional analysis was completed for the interpreted versus MRLC classified data. General speaking, photo interpretation is not the end of the accuracy assessment and an accurate and well-planned analysis scheme is the key to the understanding and utilization of dataset under assessment (Congalton and Green, 1993). One of the factors that should be considered is positional and correspondence error. To taking this factor into consideration, six cases were made in this project: 1) Primary matches classified pixel; 2) Primary or alternate matches classified pixel; 3) Primary is most common in classified 3x3 area; 4) Primary is most common in classified 3x3 area; 5) Primary is most common in classified 3x3 area; and 6) Primary or alternate matches any pixel in 3x3. "Interpreted" refers to the classes chosen during the airphoto interpretation process, "primary" and "alternate" are the most likely land cover classes for a particular point, and "classified" refers to the MRLC classification result for that point. The analysis results for each land cover class in six cases are presented in Tables 6 and 7. The overall accuracies under various scenarios are as follows:

- ?? Primary interpreted matches classified pixel: 648/1474 = 44.0 %
- ?? Primary or alternate matches classified pixel: 824 = 55.9 %
- ?? Primary is most common in classified 3x3 area: 674 = 45.7 %
- ?? Primary matches any pixel in classified 3x3 area: 985 = 66.8 %
- ?? Primary or alternate is most common in 3x3 area: 859 = 58.3 %
- ?? Primary or alternate matches any pixel in 3x3 : 1170 = 79.4 %

Table 6. Summary of further accuracy analysis by land cover class: number of points.

Class	Num	Primary PI Matches MRLC	Prim or Alt PI Matches MRLC	Primary PI is Mode of 3x3	Primary PI Matches any 3x3	Prim or Alt PI is Mode of 3x3	Prim or Alt PI Matches any 3x3
1	108	87	95	84	92	94	100
2	155	47	69	60	81	124	135
3	19	10	11	8	11	15	16
4	69	32	39	35	41	44	49
5	69	33	35	27	30	34	42
6	38	34	36	34	36	35	37
7	78	33	44	33	42	40	52
8	99	46	55	60	68	79	83
9	61	34	39	44	48	52	54
10	228	62	98	68	110	148	187
11	103	28	39	27	38	46	64
12	128	57	82	56	83	83	102
13	131	41	61	33	53	56	91
14	96	43	53	47	59	68	84
15	92	61	68	58	67	67	74
Totals	1474	648	824	674	859	985	1170

Table 7. Summary of further accuracy analysis by land cover class: percentage of points for each class.

Class	Num	Primary PI Matches MRLC	Prim or Alt PI Matches MRLC	Primary PI is Mode of 3x3	Primary PI Matches any 3x3	Prim or Alt PI is Mode of 3x3	Prim or Alt PI Matches any 3x3
1	100.0	80.6	88.0	77.8	85.2	87.0	92.6
2	100.0	30.3	44.5	38.7	52.3	80.0	87.1
3	100.0	52.6	57.9	42.1	57.9	78.9	84.2
4	100.0	46.4	56.5	50.7	59.4	63.8	71.0
5	100.0	47.8	50.7	39.1	43.5	49.3	60.9
6	100.0	89.5	94.7	89.5	94.7	92.1	97.4
7	100.0	42.3	56.4	42.3	53.8	51.3	66.7
8	100.0	46.5	55.6	60.6	68.7	79.8	83.8
9	100.0	55.7	63.9	72.1	78.7	85.2	88.5
10	100.0	27.2	43.0	29.8	48.2	64.9	82.0
11	100.0	27.2	37.9	26.2	36.9	44.7	62.1
12	100.0	44.5	64.1	43.8	64.8	64.8	79.7
13	100.0	31.3	46.6	25.2	40.5	42.7	69.5
14	100.0	44.8	55.2	49.0	61.5	70.8	87.5
15	100.0	66.3	73.9	63.0	72.8	72.8	80.4
Totals	100.0	44.0	55.9	45.7	58.3	66.8	79.4

8.2. Discussion of Results

In Table 4, we notice that there is confusion between low intensity residential, high intensity residential and commercial/transportation. There are many factors that might cause the confusion. However, we think the definitions of these categories contribute the most to this confusion. In fact, all categories that are not or partially not remotely sensible will compromise the accuracy of classification and the accuracy of reference data collection as well. In this case, all three categories belong to urban and are distinguished from each other by the amount of vegetation. Technically speaking, it is difficult to identify the vegetation amount within a pixel, i.e. at subpixel level. Therefore, the definitions themselves contain ambiguity.

The PIs had assigned many high intensity residential areas in classified image to the low intensity residential and commercial/transportation. This is considered to be reasonable since the high intensity residential is in the middle of the urban categories and can be easily confused with lower intensity and higher intensity urban development.

The PIs encountered many problems when interpreting cropland and pasture/hay since both of the classes had very similar signatures and patterns and can occur in the same agricultural area. In addition, cropland could have been changed to pasture/hay during the interval of two acquisition dates or vice versa. This also causes problems for interpreting these two classes.

Confusion also existed within classes of evergreen forest and mixed forest, deciduous forest and mixed forest, barren ground and other grassland, low intensity residential and mixed forest, transitional and all other classes. The difference between image classification and photo interpretation is that the automatic classification is mostly based on the reflectance values of the

pixels whereas the photo interpreters not only reads the special signature of the objects, but also considers the context of the background and many other characteristics intuitively.

In terms of the photo interpretation, the factors contributed to the accuracy of land use/land cover included accurate land use definition, heterogeneity and familiarity of the study area, spectral similarity between certain land classes, and the gap in the data acquisition dates. To decrease the effects of these factors on the overall accuracy we make the following two suggestions. First, we can create the error matrix based only on those samples with high confidence since many errors occur on the edge of the heterogeneous and mixed areas. Second, a field data collection of the area where rare or confusion classes exist would be helpful to characterize spectral signature of these rare classes.

9. Recommendations for Further Analysis

?? Examine the impact of alternate classes in the accuracy assessment.

During the classification of the sample sites, the interpreters provided both primary and alternate classes for each site. The primary site is the most likely land cover class for the area, while the alternate class was seen to be nearly as likely by the interpreters. For example, in a situation where an agricultural area did not have obvious rows of planted crops, the primary class may be Row Crops while the alternate may be Pasture. We feel that if we include these alternate classes in the accuracy assessment will improve the reported accuracy of the data, but questions regarding the validity of this approach must be carefully considered.

?? Evaluate and analyze the effect of positional errors on accuracy assessment.

Positional errors may be caused by many factors, such as image/data misregistration (Dai and Khorram, 1998) and feature correspondence errors or inaccuracies during photo interpretation. The option of matching any of the 3x3 pixels is trying to recover from this type of errors. However, more systematic and in-depth research is needed to develop analysis techniques to minimize this bias. These techniques involve estimation, modeling, and simulation.

?? Conduct fuzzy set analysis using the fuzzy class assignments for each point made by the interpreters.

In addition to the primary and alternate classes, the interpreters provided fuzzy class values for each of the land cover classes. The values given were "Absolutely Correct", "Probably Correct", "Acceptable", "Probably Incorrect", and "Absolutely Incorrect". For example, for a site which appeared to be agricultural, Row Crops might be called "Probably Correct", Pasture "Acceptable", and Water "Absolutely Incorrect". Given the large number of points in this database, significant work is required to analyze these fuzzy data to evaluate different methods of accuracy assessment (Gopal and Woodcock, 1994).

?? Rotate the photo interpreters' fuzzy class assignments for additional 225 or similar number of overlap points.

An analysis of how different people assign fuzzy class tolerances could be conducted using additional 225 overlap points. This would help to describe the validity of using interpreter assigned fuzzy classes derived from airphotos.

?? Analyze error matrices by integrating fuzzy data sets and other techniques.

?? Conduct research on alternate methods of accuracy assessment to handle the outliers.

?? Collect field data for the 225 or an adequate number of overlapping sample sites.

?? Analyze the satellite data with higher temporal resolution to get a better handle on the change detection and analysis between the acquisition of Landsat TM data and NAPP photography. These data types include NOAA-AVHRR and MODIS data.

References

- R. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sensing of the Environment*, vol. 37, pp. 35-46, 1991.
- R. Congalton and K. Green, "A practical look at the sources of confusion in error matrix generation," *Photogrammetric Engineering and Remote Sensing*, vol. 59, No. 5, pp. 641-644, 1993.
- X. L. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol.36, no.5, pp. 1566-1577, 1998.
- S. Gopal and C. Woodcock, "Theory and methods for accuracy assessment of thematic maps using fuzzy sets," *Photogrammetric Engineering and Remote Sensing*, vol. 60, No. 2, pp. 181-188, 1994.
- S. Khorram, G. S. Biging, N. R. Chrisman, D. R. Colby, R. G. Congalton, J. E. Dobson, R. L. Ferguson, M. F. Goodchild, J. R. Jensen, and T. H. Mace, "Accuracy assessment of remote sensing-derived change detection," *Monograph, American Society of Photogrammetry and Remote Sensing (ASPRS)*: Bethesda: Maryland, 64p. 1999.
- R. S. Lunetta, J. G. Lyon, B. Guidon, and C. D. Elvidge, "North American Landscape Characterization Dataset Development and Data Fusion Issues," *Photogrammetric Engineering & Remote Sensing*, vol. 64, no. 8, pp. 821-829, August 1998.

Z. Zhu, L. Yang, S. V. Stehman, and R. L. Czaplewski, "Accuracy Assessment for the U.S. Geological Survey Regional Land Cover Mapping Program: New York and New Jersey Region," Submitted to *Photogrammetric Engineering & Remote Sensing*. In Review.