DRAFT

CEO Technical Report 217

June 19, 2000

Final Report

Submitted to
NC Wetlands Restoration Project
NC Division of Water Quality (DWQ)

Land Cover Classification of the Hominy Creek Watershed

Submitted by

Center for Earth Observation
North Carolina State University

U.S. Congressional District 4

CEO Technical Report 217

June 5, 2000

Final Report

Submitted to
NC Wetlands Restoration Project
NC Division of Water Quality (DWQ)

Land Cover Classification of the Hominy Creek Watershed

Submitted by

Siamak Khorram, Principal Investigator, Center for Earth Observation, NCSU
Joe Knight, Project Supervisor, Center for Earth Observation, NCSU
Hui Yuan, Center for Earth Observation, NCSU
Halil Cakir, Center for Earth Observation, NCSU
Zhiyan Mao, Center for Earth Observation, NCSU

Center for Earth Observation
North Carolina State University

U.S. Congressional District 4

Contact:
khorram@ncsu.edu

Box 7106, North Carolina State University
Raleigh, North Carolina 27695-7106
Phone: (919)515-3430
FAX: (919)515-3439

Land Cover Classification of the Hominy Creek Watershed

Summary

The objective of this project was to generate a high spatial detail land cover classification for the Hominy Creek Watershed near Wilson, NC. The target spatial resolution for this classification was one meter. The study are for this project was the Hominy Creek watershed near Wilson, NC. The data type used was digitized NAPP 1:40000 aerial photographs. These photos were digitized, georeferenced, orthorectified, and classified into seven land cover categories: Impervious, Forest, Grassland, Agriculture, Bare Soil, Water, and Shadow (Unknown). The accuracy estimates of the classification are: 1) Classification compared to 450 randomly determined sites on the original aerial photography, 89%; and 2) Classification compared to 61 randomly determined sites where ground observation data were collected, 74%. This method is shown to be useful and appropriate for thematic classifications at this level of detail where cost is a major factor.

Acknowledgments

The results reported here were generated through a contract funded by the North Carolina Wetlands Restoration Project (WRP) in the Division of Water Quality (DWQ). The views expressed in this report are those of the authors and do not necessarily reflect the views of WRP or any of its sub-agencies. The authors would like to thank the WRP in general and Ms. Laura Jones and Mr. Mac Haupt, in particular for their support and for the assistance given on this project. The authors would also like to thank Ms. Linda Babcock of CEO for her indispensable work throughout the duration of this project.

Introduction

The objective of this project was to generate a customized high spatial detail land cover classification for the Hominy Creek Watershed near Wilson, NC. The target spatial resolution for this classification was one meter. A land cover characterization at this level of spatial detail demands high spatial resolution multispectral data. Multispectral data is needed to differentiate between different vegetation cover types. To date, commercial high-resolution (less than 5 meters) multispectral satellite data is not widely available and is very expensive. Alternatives are lower spatial resolution data sources such as the French Systeme pour L'Observation de la Terre (SPOT), NASA's Landsat Thematic Mapper satellite data and USGS Digital Orthorectified Quarter Quads (DOQQ). SPOT multispectral images have 20 meter spatial resolution. Landsat Thematic Mapper data has 30 meter spatial resolution multispectral data. DOQQs have much higher spatial resolution – two meters – however the data is black and white, not multispectral. None of these alternatives is acceptable for use in a one meter land cover classification. This project required the creation of an appropriate dataset.
Given the lack of appropriate satellite datasets and the cost of acquiring airborne multispectral data, we chose to use color infra-red National Aerial Photography Program (NAPP) aerial photographs. These photos are color infra-red (CIR) and have a scale of 1:40,000. Flights were done for a the study area in January-March 1998, with follow-up flights for the eastern Piedmont and other areas in early 1999. CIR NAPP photography is available from the USGS EROS Data Center. Digital CIR NAPP data is still in the development stage and not available for the study area. Therefore, the CIR NAPP photos were digitized to create a CIR DOQQ for the area of interest. This dataset provides both high spatial resolution and multispectral information.

Study Area

The study are for this project is the Hominy Creek watershed near Wilson, NC. The 14-digit Hydrological Unit Code (HUC) for the watershed is 03020203020040. The Hominy Creek watershed is in Sub-basin 07 of the Neuse River Basin. The study area is estimated to be 11 by 11 miles.

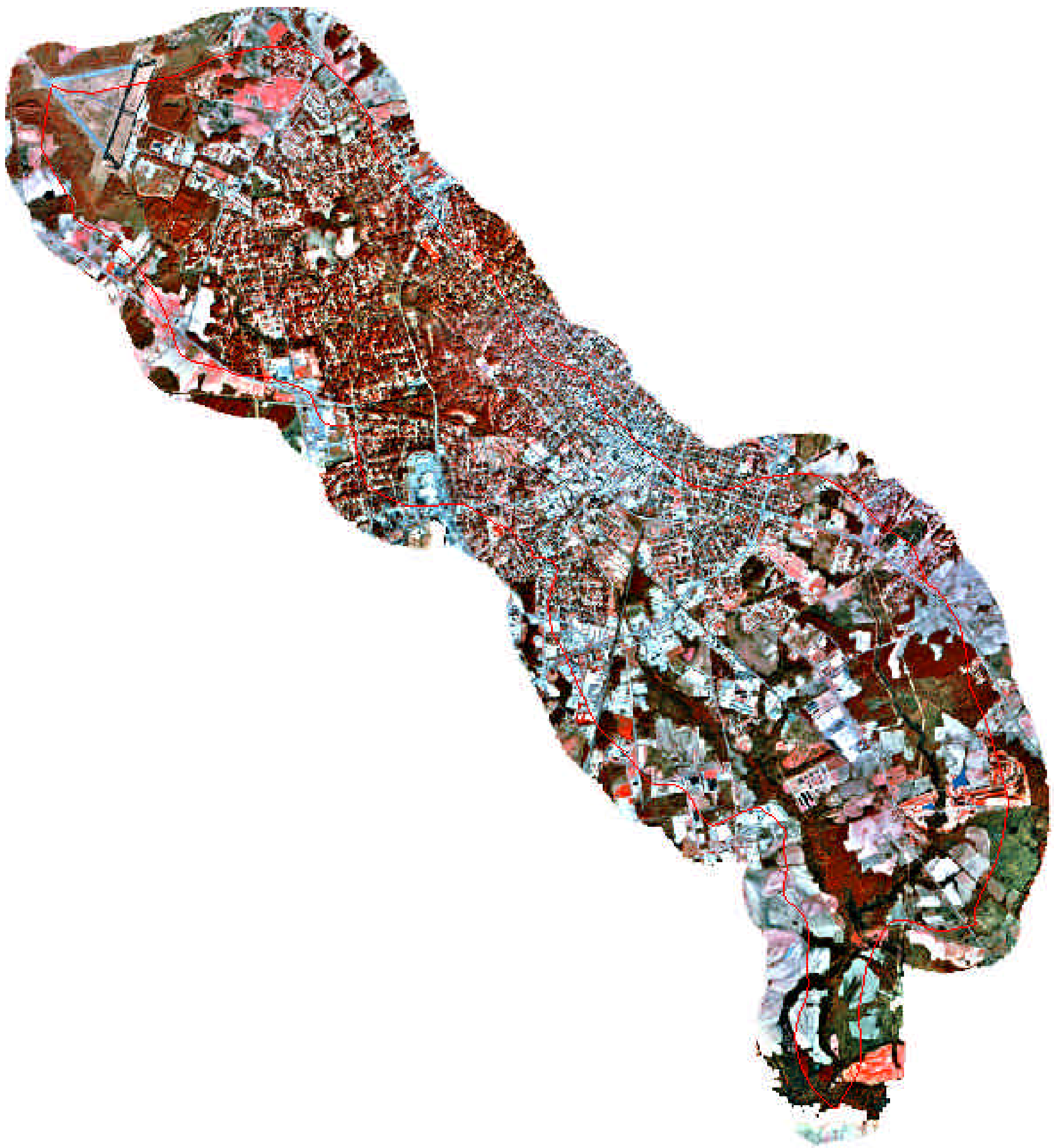

Figure 1. The Hominy Creek watershed. The red line around the edge is the boundary of the watershed.

Data Used

Based on all the requirements above, the following data were used:

?? GIS layers of the study area boundaries

1) From GIS coverages, a polygon file for the study area was derived.

2) The study area is within Wilson county. The boundary of the coverage is as follows: Xmin: 701965 Ymin: 211615 (Upper left corner)
   Xmax: 711778 Ymax: 224859 (Bottom right corner)
   Projection information:
   Georeferenced to : State Plane (SP)
   Datum : NAD83
   Zone number : 4901
   Map units : Meters

3) The study area boundary was used to crop all image data to the study area.

?? Black and white (B/W) USGS DOQQ

The B/W DOQQs were imported and mosaicked to form a continuous coverage of georeferenced data for the study area.

?? Color infrared (CIR) NAPP photos

Eleven CIR NAPP photos were ordered from the USGS EROS Data Center. Five of these photos were used to make the CIR DOQQ. The acquisition date of these photos is January 26, 1998. They are listed as follows:

Roll Number	Frame Number	Time	Lens Serial No.
10760	10	Jan. 26, 1998	13081
10760	12	Jan. 26, 1998	13081
10760	108	Jan. 26, 1998	13081
10760	110	Jan. 26, 1998	13081
10760	131	Jan. 26, 1998	13081

?? DEM data

Three 7.5-Minute Digital Elevation Model (DEM) files for the Wilson, Saratoga and Winstead Crossroads USGS topographic quads were downloaded from the USGS. Each 7.5-minute DEM is based on 30-by 30-meter data spacing with the

Universal Transverse Mercator (UTM) projection. The DEMs were converted to the State Plane projection to match the scanned photos and other image data.

?? Calibration Information

Calibration information was acquired from USGS for lens serial number 13081. The main information, which was used in orthorectification, includes the calibrated principal point coordinates, fiducial coordinates, and calibrated focal length for the lens.

Methods

The project was divided into three stages:

?? Pre-processing – Digitizing, georeferencing, and mosaicking the CIR photos
?? Classification – Performing the land cover classification of the study area
?? Accuracy assessment – Developing an estimate of the accuracy of the classification

Pre-processing

In this stage, the CIR DOQQ was created from the scanned NAPP CIR photographs.

Digital Orthorectified Quarter Quads (DOQQs) require several types of inputs to produce an orthogonally rectified image from the original perspective image captured by the sensor. These inputs are the following:

1) The unrectified raster image scanned from the diapositive or directly acquired from a digital sensor (the NAPP CIR photos);
2) The image and ground coordinates of photo identifiable ground control points (from B/W USGS DOQQs);
3) A digital elevation model of the study area; and
4) Calibration information about the sensor.

These four inputs are used to register the image file to the scanner and to the sensor platform, to determine the orientation and location of the sensor platform with respect to the ground, and to remove the relief displacement from the image data.

Digitization

The first step in producing the CIR DOQQ was digitizing the photography. The steps involved in this process are outlined below:

?? Since the eleven provided photos overlap each other, the extent of the study area was used to determine which photos should be scanned. By comparing the photo coordinates and examining the B/W DOQQ for the study area, six photos were selected for digitizing.

?? The scale of the NAPP photography is approximately 1:40,000. To produce a spatial resolution of one meter, a scanning aperture of 25 micrometers must be used, which means that the scanning must be done at 1016 dots per inch (DPI). The scanning processing is an analog-to-digital (A/D) conversion and, like all quantization procedures, will introduce errors. To minimize these errors, the scan settings were consistent from photo to photo. In addition, all pre-processing besides the scanning was done using the ERDAS Imagine™ image processing software package. This software allows for more control over the spatial fidelity of the data than other commercial image manipulation packages.

?? The scanned images were saved in Tagged Image File Format (TIFF) to eliminate distortion caused by using a lossy compression format (such as JPEG).

?? The TIFF images were then imported into Imagine to be orthorectified, georeferenced, and mosaicked.

Georeferencing and orthorectification

After scanning the images are just pictures without any coordinate system. Furthermore, variations due to aircraft tilt, feature geometry, and lens distortion are still present. To make the images useable, they must be georeferenced and orthorectified. Georeferencing is the process of assigning a coordinate system to an image. This allows for overlaying other datasets on the image. Orthorectification takes a raw digital image, applies an elevation model (DEM), and incorporates calibration information to create an orthoimage (digital orthophoto). An orthoimage has all variation due to tilt, geometry, and lens distortion removed. The process is performed by inputting the camera and lens calibration parameters into a orthorectification model in Imagine. Then, Ground Control Points (GCPs) are selected on another image that has a coordinate system (the B/W DOQQ). In this way, the image is both orthorectified and georeferenced in one step. This was done for each of the five scanned photos. The resulting five images had an average root mean square (RMS) error of approximately five meters.

Mosaicking

Following orthorectification and georeferencing, the five images were mosaicked to form one large CIR DOQQ for the study area. This was done using the "Mosaic" model in Imagine. The input images were histogram matched and the image boundaries were feathered at the joins to make the DOQQ consistent from photo to photo. The resulting image is a CIR DOQQ for the Hominy Creek watershed.

Classification

The mosaicked CIR DOQQ was classified into seven land cover classes. The following is a list of the classes with examples of each:

1. Impervious cover – Urban areas and roads and highways
2. Grassland – Lawns and golf courses
3. Forest – Deciduous, coniferous, or mixed forest
4. Agriculture – Row crops and pasture
5. Bare Soil – Construction sites, bare agricultural land
6. Water – Ponds and lakes
7. Shadow/Unknown – Areas that were in shadow on the photos

During initial classification attempts, it became apparent that differentiating certain bright rooftops and bare soil was problematic. The reflectance characteristics of these cover types are very similar. To remedy this problem, agricultural areas were manually clipped out of the image. This resulted in two images, one with only the agricultural areas and one with only the non-agricultural areas. These images were then classified separately. Another problem encountered during test classifications was that shadowed areas on the photos were not classified correctly. Due to the use of aerial photos to create the CIR DOQQ, this problem could not be avoided. This issue is addressed in the Recommendations section of this report.

The final classification was performed with a supervised classification using a maximum likelihood decision rule.

Accuracy Assessment

To provide an estimate of the accuracy of the classification, both ground and photo reference points were compared to the classified values. 450 points were interpreted from the original aerial photos, and 61 points were visited on the ground. Both of these samples were constructed using a standard random sampling scheme. In the ground sampling, particular effort was made to ensure that the sample covered the entire study area.

Results

Classification

The land cover classification of the Hominy Creek watershed produced the following areal estimates of the land cover classes of interest:

Land Cover Class	Area (acres)
Urban	3348
Forest	4785
Grassland	2567
Agriculture	2648
Bare Soil	65
Shadow	1360
Total	**14846**

Note that the Shadow class should be considered to be “unknown” for the purposes of this classification.

Accuracy Assessment

The following tables give error matrices for the photo and ground accuracy assessment samples. Table 1 shows the results of the accuracy assessment procedure using the 450 photo reference points. Table 2 gives the results of the ground visitation sample. Table 3 shows the ground sample data again, but with the bare soil and agriculture classes combined. This was done to compensate for the fact that the classification was done using photos that were taken in the winter (when much of the agricultural land was bare), but the ground reference data was collected in the spring (when much of the agricultural land has been planted).

These tables are in the form of a standard error matrix. They show the classified results on the left as compared with the reference data (ground truth) for the corresponding points along the top. For example, examine the Forest class in Table 1. A total of 142 of our 450 test points were classified as forest. Of those, 128 were determined to actually be forest, while two were impervious, eleven were grassland, and one was agriculture. The overall accuracy figure of 88.7% was calculated by adding up all of the diagonal (correctly classified) elements in the table (128, 83, 47, 29, 32, 27) and dividing that by 450.

Photo Reference

		Fr	Im	Gr	Ag	Br	Wt	Sh	Tot	Pct
	Forest (Fr)	**128**	2	11	1	0	0	0	**142**	**90.1**
	Impervious (Im)	4	**83**	3	0	4	0	1	**95**	**87.4**
Classified	**Grassland (Gr)**	11	2	**47**	1	0	0	0	**61**	**77.0**
Data	**Agriculture (Ag)**	0	0	0	**53**	0	0	0	**53**	**100.0**
	Bare Soil (Br)	0	2	0	0	**29**	0	0	**31**	**93.5**
	Water (Wt)	0	0	0	0	0	**32**	0	**32**	**100.0**
	Shadow (Sh)	9	0	0	0	0	0	**27**	**36**	**75.0**
	Total	**152**	**89**	**61**	**55**	**33**	**32**	**28**	**450**	
	Pct	**84.2**	**93.3**	**77.0**	**96.4**	**87.9**	**100.0**	**96.4**		**88.7**

Table 1. Photo reference sites. Error matrix showing classified results compared to 450 randomly determined sample sites on the aerial photos

Ground Reference

		Fr	Im	Gr	Ag	Br	Wt	Sh	Tot	Pct
	Forest (Fr)	**9**	0	2	0	0	0	0	**11**	**81.8**
	Impervious (Im)	0	**17**	1	0	1	1	0	**20**	**85.0**
Classified	**Grassland (Gr)**	0	0	**12**	0	1	0	0	**13**	**92.3**
Data	**Agriculture (Ag)**	0	0	2	**4**	4	0	0	**10**	**40.0**
	Bare Soil (Br)	0	0	0	0	**2**	0	0	**2**	**100.0**
	Water (Wt)	0	0	0	0	0	**1**	0	**1**	**100.0**
	Shadow (Sh)	3	1	0	0	0	0	**0**	**4**	**0.0**
	Total	**12**	**18**	**17**	**4**	**8**	**2**	**0**	**61**	
	Pct	**75.0**	**94.4**	**70.6**	**100.0**	**25.0**	**50.0**	**---**		**73.8**

Table 2. Ground reference sites. Error matrix showing classified results compared to 61 randomly determined sample sites based on ground observations

Ground Reference

		Fr	Im	Gr	Ag+Br	Wt	Sh	Tot	Pct
	Forest (Fr)	**9**	0	2	0	0	0	**11**	**81.8**
	Impervious (Im)	0	**17**	1	1	1	0	**20**	**85.0**
Classified	**Grassland (Gr)**	0	0	**12**	1	0	0	**13**	**92.3**
Data	**Agriculture + Bare (Ag+Br)**	0	0	2	**10**	0	0	**12**	**85.7**
	Water (Wt)	0	0	0	0	**1**	0	**1**	**100.0**
	Shadow (Sh)	3	1	0	0	0	**0**	**4**	**0.0**
	Total	**12**	**18**	**17**	**12**	**2**	**0**	**61**	
	Pct	**75.0**	**94.4**	**70.6**	**83.3**	**50.0**	**---**		**80.3**

Table 3. Combined ground agriculture and bare soil sites. Error matrix showing classified results compared to 61 randomly determined sample sites based on ground observations, but with seasonal agricultural effects removed.

Recommendations

This method produced a reliable, accurate classification. There are actions which could be taken to improve it. These are:

1) The inclusion of an ancillary data type to enable the classification of areas that are in shadow on the photos. An attractive data type for this purpose is radar data. Radar data is insensitive to time of day, weather conditions, and solar insolation. The Canadian Radarsat can provide eight meter spatial resolution data that would be helpful for resolving areas that are in shadow.

2) The inclusion of an ancillary data type to help discriminate between the tops of buildings and bare soil. Radar data may be suitable for this purpose, but a high resolution multispectral dataset, such as IKONOS, may be more effective.

3) Depending on their future availability, the classification could be done with USGS CIR DOQQs. Not available at the time of this project, the CIR DOQQs are expected to become available sporadically in the next year or so. Using these data would eliminate errors caused by scanning and orthorectifying the aerial photos.

Conclusions

The methods outlined in this paper have produced an accurate high resolution land cover classification for the Hominy Creek watershed. The innovative procedures used have allowed for the inclusion of multispectral information while not sacrificing high spatial resolution. It is expected that this approach could be replicated in other watersheds to produce similarly good results.