NCEA-C-1698
April 2006

Classification of High Spatial Resolution, Hyperspectral Remote Sensing Imagery of the Little Miami River Watershed in Southwest Ohio, USA

National Center for Environmental Assessment
Office of Research and Development
U.S. Environmental Protection Agency
Cincinnati, OH 45268

NOTICE

The U.S. Environmental Protection Agency through its Office of Research and Development funded and managed the research described here under contract no. 68-C-02-060 to Forest One, Inc. and 3C-R337-NTSA to Hyperspectral Data International, Inc. It has been subjected to the Agency's peer and administrative review and has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

This report summarizes a collaborative project led by the U.S. Environmental Protection Agency to create a high spatial resolution land use/land cover (LULC) dataset for the entire Little Miami River watershed in southeastern Ohio, USA from remotely sensed imagery. The LULC classification was derived from 82 flight lines of *Compact Airborne Spectrographic Imager* (CASI) hyperspectral imagery acquired from July 24 through August 9, 2002 via fixed wing aircraft. Categories within this classification include: water (both lentic and lotic), forest, corn, soybean, wheat, dry herbaceous vegetation, grass, urban barren, rural barren, urban/built, and unclassified. A hierarchical classification approach was used involving object image segmentation in eCognition (Definiens Imaging GmbH., 2003), and spectral angle mapper (SAM) in the ENvironment for Visualizing Images (ENVI) (Research Systems Inc., 2003). A final classification was completed after an extensive Quality Assurance and Quality Control (QA/QC) phase which included manual editing. The final product includes classification results at the original data spatial resolution of 4m x 4m.

Preferred citation:
Troyer, M.E., J. Heo and H. Ripley. 2006. Classification of High Spatial Resolution, Hyperspectral Remote Sensing Imagery of the Little Miami River Watershed in Southwest Ohio, USA. U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, Cincinnati, OH.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACORN	Atmospheric CORection Now
BIP	Band Interleaved-by-Pixel
CART	Classification and Regression Tree
CASI	Compact Airborne Spectrographic Imager
DN	Digital Number
ENVI	ENvironment for Visualizing Images
FGDC	Federal Geographic Data Committee
GMT	Greenwich Mean Time
GPS	Global Positioning System
HDI	Hyperspectral Data International
IFOV	Instantaneous Field of View
LULC	Land Use/Land Cover
MLC	Maximum Likelihood Classification
NLCD	National Land Cover Dataset
PPS	Pulse-Per-Second
QA/QC	Quality Assurance and Quality Control
RMS	Root-Mean-Square
SAM	Spectral Angle Mapper (classification in ENVI)

AUTHORS, CONTRIBUTORS AND REVIEWERS

AUTHORS

Michael E. Troyer, Ph.D.
National Center for Environmental Assessment
Office of Research and Development
U.S. Environmental Protection Agency
Cincinnati, OH 45268

Joon Heo, Ph.D.
Forest One, Inc.
Itasca, IL

Herbert Ripley, FRSPSoc
Hyperspectral Data International, Inc.
Halifax, Nova Scotia

CONTRIBUTORS

Christopher Bolton
Earth Satellite Corp.
Rockville, MD

Michael Diller
Earth Satellite Corp.
Rockville, MD

Christopher Jengo
Earth Satellite Corp.
Rockville, MD

William Jones
Hyperspectral Data International, Inc.
Halifax, Nova Scotia

Sitansu Pattnaik
Forest One, Inc.
Itasca, IL

Laura Roy
Hyperspectral Data International, Inc.
Halifax, Nova Scotia

Francois Smith
Earth Satellite Corp.
Rockville, MD

Michelle Warr
Hyperspectral Data International, Inc.
Halifax, Nova Scotia

INTERNAL PEER REVIEWERS

F. Bernard Daniel, Ph.D.
Senior Environmental Scientist
National Exposure Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency

AUTHORS, CONTRIBUTORS AND REVIEWERS cont.

Noel W. Kohl, M.P.H.
EPA GIS Working Group
Office of Information Services
U.S. Environmental Protection Agency
Region 5
Chicago, IL

Cheryl G. Itkin
EPA GIS Working Group
Information Management
National Center for Environmental Assessment
Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC

Tara A. Maddock, Ph.D.
Geographer
National Center for Environmental Assessment
Office of Research and Development
U.S. Environmental Protection Agency

Mary L. White, Ph.D.
EPA GIS Working Group
Water Quality Branch, Water Division
U.S. Environmental Protection Agency
Region 5
Chicago, IL

EXTERNAL PEER REVIEWERS

Nina Lam, Ph.D.
R.J. Russell Professor of Geography
Louisiana State University
Baton Rouge, LA

Victor Mesev, Ph.D.
Associate Professor of Geography
Florida State University
Tallahassee, FL

R. Douglas Ramsey, Ph.D.
Associate Professor and Director of the Remote Sensing and GIS Laboratory
Department of Forest, Range, and Wildlife Sciences
Utah State University
Logan, UT

ACKNOWLEDGMENTS

Many people provided their time and talents to ensure the success of this project. Essential fieldwork and ground-truth data were provided by the following scientists from the U.S. Environmental Protection Agency’s research facilities in Cincinnati, Ohio: Drs. F. Bernard Daniel, Michael B. Griffith, James M. Lazorchak, Tara Maddock, Matthew A. Morrison, Bruce W. Peirano, Joseph P. Schubauer-Berigan, Christopher Schultz and William Shuster. Peter Mueller and Drew Comley (Dr. Troyer’s summer interns) also assisted in these field efforts as participants of the “Research Apprenticeship Program” sponsored by the University of Cincinnati and the U.S. EPA. Additional aerial imagery used in this project was generously provided by Dr. Lawrence Spencer from the Ohio State University - Center for Mapping. Early discussions on remote sensing, classification and choosing the proper spectral bands and widths for this project were substantively informed by conversations with Dr. Prasad S. Thenkabail of the Center for Earth Observation at Yale University, and David J. Williams of the U.S. EPA’s Environmental Services Division in Reston, Virginia. Appreciation also goes to Bette Zwayer (U.S. EPA National Center for Environmental Assessment), Lana Wood (Intellitech Systems, Inc.) and Elizabeth Fryer (ECflex) who helped with preparing tables, and formatting and editing the document into its final form.

PREFACE

Geographic products such as the land use/land cover (LULC) dataset presented herein enable a wide array of studies important for sustaining both society and nature. For example, opportunities and risks associated with various human-ecological interactions can be identified, as well as a greater elucidation of other spatial relationships and processes affecting the Earth's surface and our environment.

In 2001, an interdisciplinary group of scientists based at the U.S. Environmental Protection Agency's research facilities in Cincinnati, Ohio formed a collaborative effort to study the Little Miami River (LMR) and its watershed. This work was performed in support of the U.S. EPA's watershed (or geographic) approach for protecting designated uses of America's rivers and streams. In particular, development of this LULC dataset provides a basis for studying: (1) social drivers of land use change, (2) how land use change affects the hydrology, sediments and nutrients of streams, and (3) aquatic biological responses as a result of changes in the landscape. This work also assisted in exploring and suggesting refinements to spatially-explicit criteria intended to prevent habitat alteration, excess nutrients, suspended and bedded sediments, pathogens, toxic chemicals, and other stressors affecting the Nation's waters.

A primary goal of this project is to help enhance the use of geographic and spatial analytic tools in risk assessments at U.S. EPA, and to improve the scientific basis for risk management decisions. This is important because environmental problems are inherently spatial. For example, many pollutants originate from *multiple non-point sources* in the landscape and *spread* to other areas within a particular watershed, "airshed," or across ecological and political boundaries. From a technological point of view, applications of this product will represent an improvement from more readily available data with coarser spatial and spectral resolution. The hierarchical classification approach used incorporating both object-based pattern recognition and spectral techniques may be beneficial and transferable to other settings. This dataset may also be useful to others outside the Agency; particularly, those interested in studying anthropogenic and natural processes occurring at watershed or smaller spatial scales.

1. PROJECT DESCRIPTION

The goal of this project was to create a higher spatial resolution land use/land cover (LULC) dataset for the entire Little Miami River watershed in southeast Ohio for a number of studies being conducted by the U.S. Environmental Protection Agency (U.S. EPA). The Little Miami River has a drainage area of 175.5 square miles and stretches in a southwesterly direction for 105.5 miles originating from near South Charleston, Ohio to its confluence with the Ohio River east of Cincinnati, Ohio (Figure 1). It is one of the oldest river groups in the state and became Ohio's first State and National Scenic River (Sanders, 2002). Prior to this project, existing LULC datasets based on Landsat imagery had spatial resolutions of 30 to 60 meters. Hyperspectral Data International (HDI) collected 4 meter spatial resolution, hyperspectral imagery of the watershed from July 24 through August 9, 2002. Forest One Incorporated (Earth Satellite Corporation) subsequently classified the resulting 82 flight-lines of *Compact Airborne Spectrographic Imager* (CASI) data into the 11 classes of land cover type presented here.

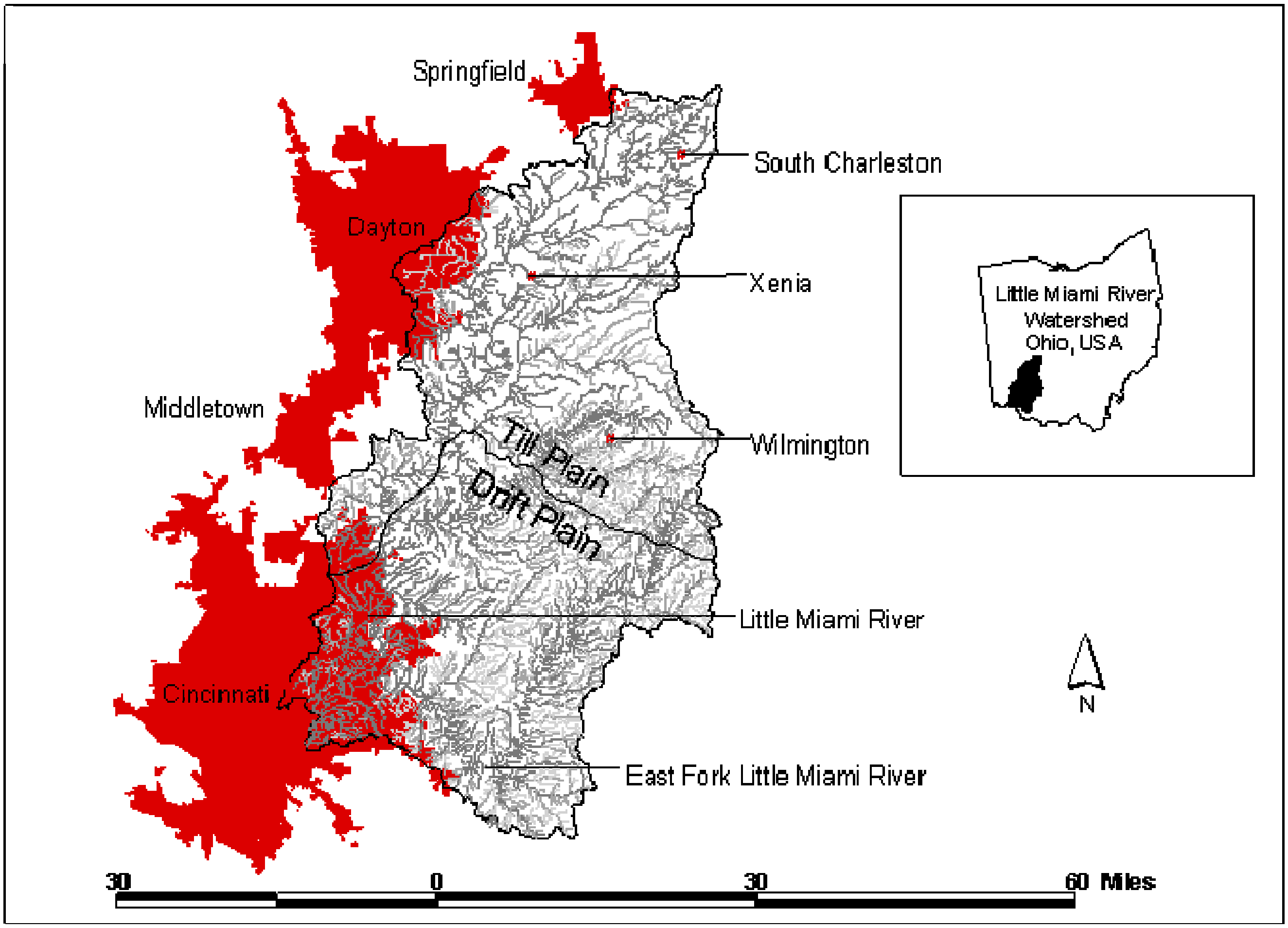


FIGURE 1

Orientation Map of the Little Miami River Watershed
in Southwestern Ohio, USA

2. THE REMOTE SENSING EFFORT

2.1. INTRODUCTION

The remote sensing effort consisted of planning and implementing all the logistics and details necessary for the collection of CASI data from a fixed-wing aircraft. The CASI instrument used in this project was a rack-mounted, fully programmable, high resolution (12 bit resampled to 16 bit) pushbroom imaging spectrometer system with the capability to collect data in 19 channels between 400-950 nm at intervals ranging from as small as 2 nm in width to tens of nanometers. This system was fully georeferenced using data from an aircraft mounted gyroscope, an aircraft mounted Global Positioning System (GPS), base station GPS data, and Digital Elevation Model (DEM) data (see Figure 2). As a result, all recorded data for this project was corrected for variations in aircraft altitude and attitude during flight. Spectral wavelengths and widths chosen for this project evolved from conversations with Herb Ripley of HDI, Inc. (based upon his many years of experience with this particular instrument), Dr. Prasad Thenkabail from the Center for Earth Observation at Yale University (also see Thenkabail et al., 1994, 2000), and David Williams from the U.S. EPA’s Environmental Services Division in Reston, Virginia. The resulting 19 spectral bands chosen were selected in the hope of best discerning both the urban and agricultural landscapes expected in this particular watershed (Table 1). Flight line acquisition dates for this mission are shown in Figure 3. A detailed flight log is also provided in Appendix A.

The remote sensing effort also involved a series of pre-processing corrections (radiometric, geometric, and atmospheric) required to make the CASI data suitable for input and analysis within a geographic information system (GIS) as well as the next step, classification of this data into a land use/land cover product (Section 3).

2.2. RADIOMETRIC CORRECTION

Two files derived from data collected during the over-flights were used as a basis for all image processing steps in this project. One was an ASCII file which contained aircraft CASI pitch, roll, and GPS data. The other was a raw image file in band interleaved-by-pixel (BIP) format. Remote sensing instruments on the aircraft (i.e., the imager, gyroscope, GPS, computer and hard drive storage, etc.) were linked by a Pulse-Per-Second (PPS) signal and time stamp based on Greenwich Mean Time (GMT).

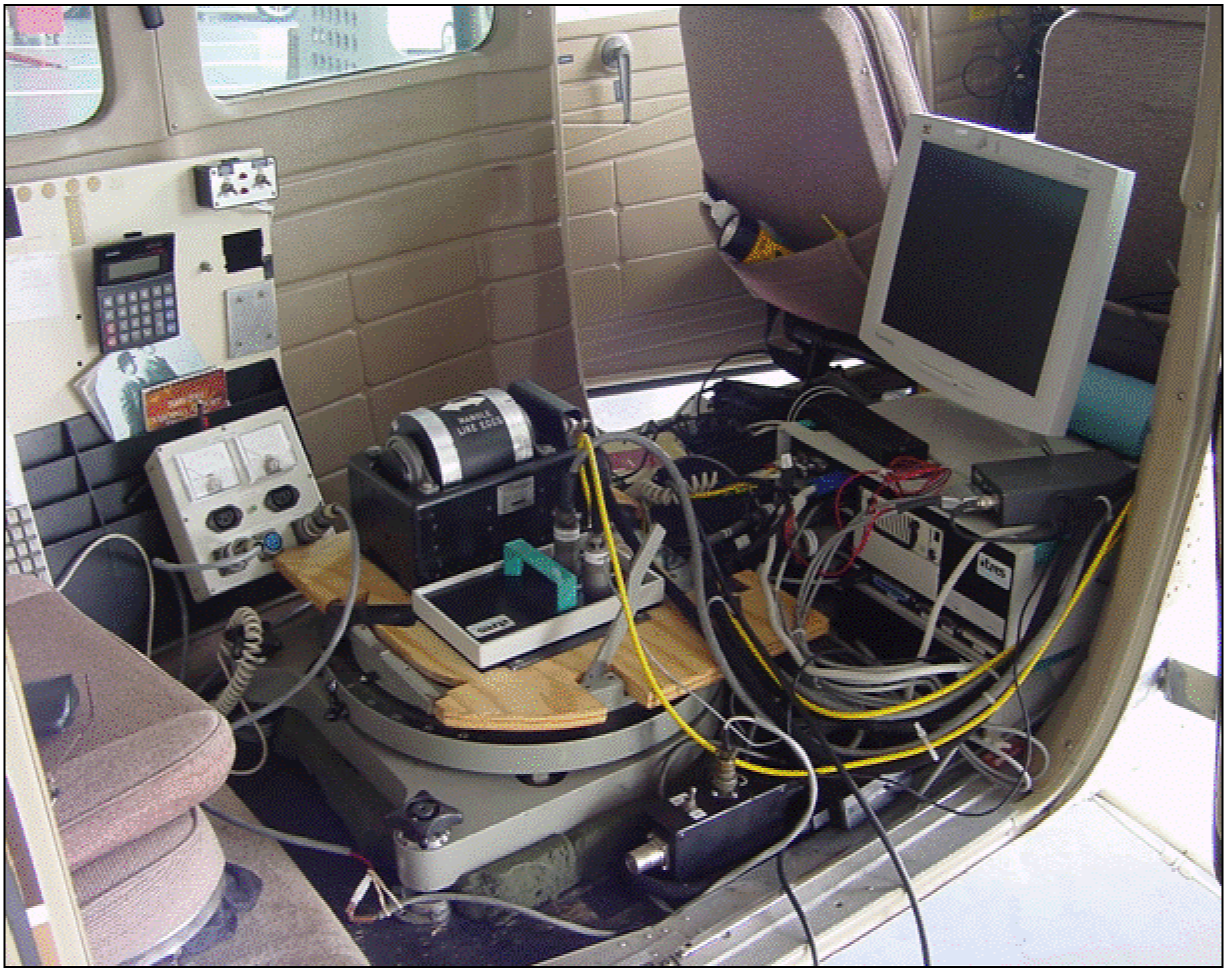


FIGURE 2

Remote Sensing Instrumentation Used in this Project

TABLE 1

Spectral Bands Remotely Sensed in the Little Miami River Project

Band	Spectral Region	Band Center Nm	Band Width (+/-) Nm	Band Range Nm	Comments
1	Blue	449.6	15.0	30.0	
2	Blue	490.4	15.0	30.0	Crop to soil reflectance ratio minima.
3	Green	520.2	9.5	19.0	1st order derivative. Positive change in reflectance per unit change in wavelength is maximized. Pigment content.
4	Green	550.2	9.5	19.0	Green band peak. Related to total chlorophyll.
5	Green	574.6	7.7	15.4	1st order derivative. Negative change in reflectance per unit change in wavelength is maximized. Pigment content.
6	Green-Red	600	8.6	17.2	
7	Red	619.8	7.7	15.4	
8	Red	659.6	7.7	15.4	Chlorophyll absorption pre-maxima.
9	Red	674.8	7.8	15.6	Chlorophyll absorption maxima. Greatest crop-soil contrast. Related to chlorophyll a and b.
10	Red	691	4.9	9.8	
11	Red-edge	700.5	4.9	9.8	Chlorophyll absorption post-maxima.

TABLE 1 cont.

Band	Spectral Region	Band Center Nm	Band Width (+/-) Nm	Band Range Nm	Comments
12	Red-edge	719.6	6.8	13.6	1st order derivative. Maximum change in slope of reflectance spectra per unit change in wavelength. Vegetative stress. Nitrogen status of plants.
13	Red-NIR	750.1	10.7	21.4	Red edge/vegetative stress.
14	NIR	799.9	10.7	21.4	
15	NIR	820.1	9.7	19.4	Atmospheric water absorption/correction
16	NIR Shoulder	845.1	9.8	19.6	Center of near-infrared (NIR) shoulder. Broad or narrow band may provide same result. Related to total chlorophyll.
17	NIR Peak1	899.9	10.7	21.4	Crop growth, stress or senescense. Useful for crop moisture sensitive index.
18	NIR Peak2	920.2	9.8	19.6	Crop growth, stress or senescense.
19	NIR-Moisture Sensitive	937.5	7.9	15.8	Atmospheric water absorption/correction.

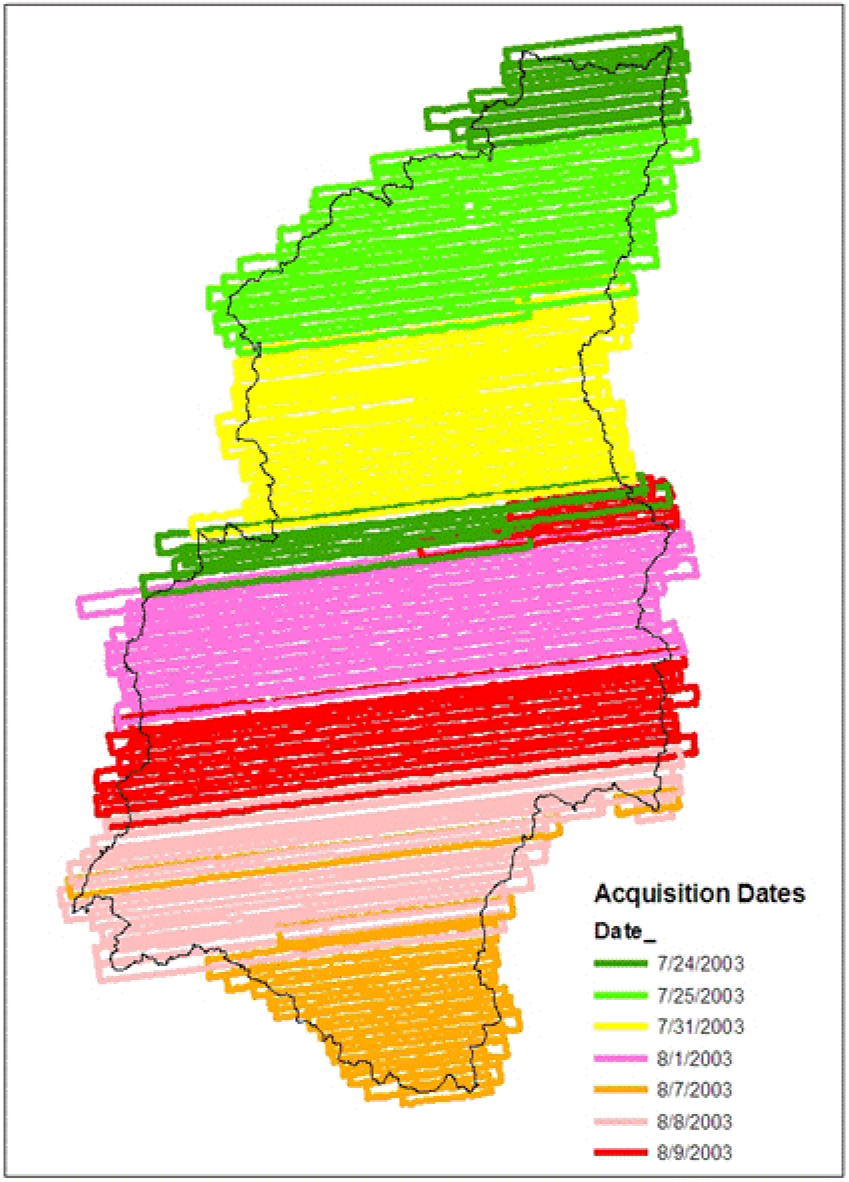


FIGURE 3

Flight Line Acquisition Dates of CASI Data
in the Little Miami River Watershed

Radiometric correction was the first and most important step in the image processing stream. In short, this correction calibrates pixel digital numbers to radiance units. This is necessary because imagery from any given sensor may be influenced by factors such as changes in scene illumination, atmospheric conditions, viewing geometry variations, and instrument response characteristics (Lillesand and Kiefer, 1994). The CASI instrument used in this project normally undergoes at least one and sometimes more system calibrations per year. The system manufacturer conducts this calibration in the laboratory using an U.S. National Institute of Standards and Technology (NIST) radiance standard. Full frame uniformity and absolute data at each aperture stop are obtained to determine Radiant Sensitivity Coefficients (RSCs) for all pixels in the CCD array. The RSC files are used to convert scene data into units of spectral radiance. The calibration procedures compensate for variations in optical transmission and CCD responsivity. Signal contributions arising from electronic offset, dark current, frame shift smear and scattered light are also removed. CASI image scan lines were then correlated and interpolated with the internal navigation data records, i.e., position (X, Y) and CASI gyro-based attitude (pitch and roll) measurements, to create a single file suitable for geocorrection.

2.3. GEOMETRIC CORRECTION

Each flightline was geometrically corrected and geographically registered using the CASI manufacturer's software (ITRES Research Limited, 2006). This project generated CASI data with a nominal accuracy of 3 pixels RMS or 4 meter RMS accuracies which fell within the published accuracies of the CASI instrument. PCI Geomatica geocorrection software was employed subsequent to the ITRES procedures to further refine the geographic accuracy. Geometric correction also involved the use of road vectors (Wessex Inc., 1997) to ensure accurate map registration.

2.4. ATMOSPHERIC CORRECTION

Each geocorrected flight line was then atmospherically corrected using ACORN (Atmospheric CORection Now) software (Analytical Imaging and Geophysics LLC., 2002). In brief, ACORN assesses, models, and compensates for the atmosphere to convert input radiance spectra to apparent surface reflectance. After atmospheric correction, the spectral absorption features inherent to surface materials are revealed. This software uses the MODTRAN4 algorithm for atmospheric correction of calibrated hyperspectal and multispectral data in the 350 to 2500 nm spectral range.

3. IMAGE CLASSIFICATION

3.1. INTRODUCTION

The end product objective for this project was to develop a 4-meter spatial resolution classification of the entire Little Miami River watershed, and to do so with an accuracy of 80% or better for each class. Particular emphasis was placed on discerning landscapes thought to potentially contribute nutrients to the Little Miami River and its tributaries (i.e., various chemical species of nitrogen or phosphorus). A variety of classification methods were considered during the preparation of a quality assurance project plan for this project. The final classification approach chosen (a hierarchical, two-step scheme incorporating the strengths of both object-image segmentation and pixel-based approaches) is detailed in the next section.

3.2. THE CLASSIFICATION PROCESS

Key steps for deriving the Little Miami River Watershed Land Use/Land Cover Classification presented in this report included:

- Reviewing the quality of the CASI data for purposes of classification (Section 3.2.1);
- Additional ground-truth work through the collection and interpretation of aerial images from 2002 and 2003 (Section 3.2.2);
- Generating classification results (Section 3.2.3); and
- Ensuring consistency of the classification across flightlines (Section 3.2.4).

3.2.1. Quality Review of the CASI Data for Purposes of Classification. Considering the climate conditions during the period of collection, the quality of the CASI data was deemed sufficient for image classification purposes. However, two artifacts present in the data did require some additional work. The first type of artifact was the presence of clouds and corresponding cloud shadow in some of the imagery. Overall, this type of artifact was minor except for a 342 acre gap between flight lines 13a and 13b which remained unclassified (Figure 4).

The second type of artifact was cross-track illumination effects. Cross-track illumination artifacts routinely occur in airborne hyperspectral data and are a result of sun-sensor-target-geometry, atmospheric conditions, differential path length (across the instantaneous field of view or IFOV) and spectral band selection. The CASI dataset of the Little Miami River watershed is variably affected by cross-track illumination artifacts

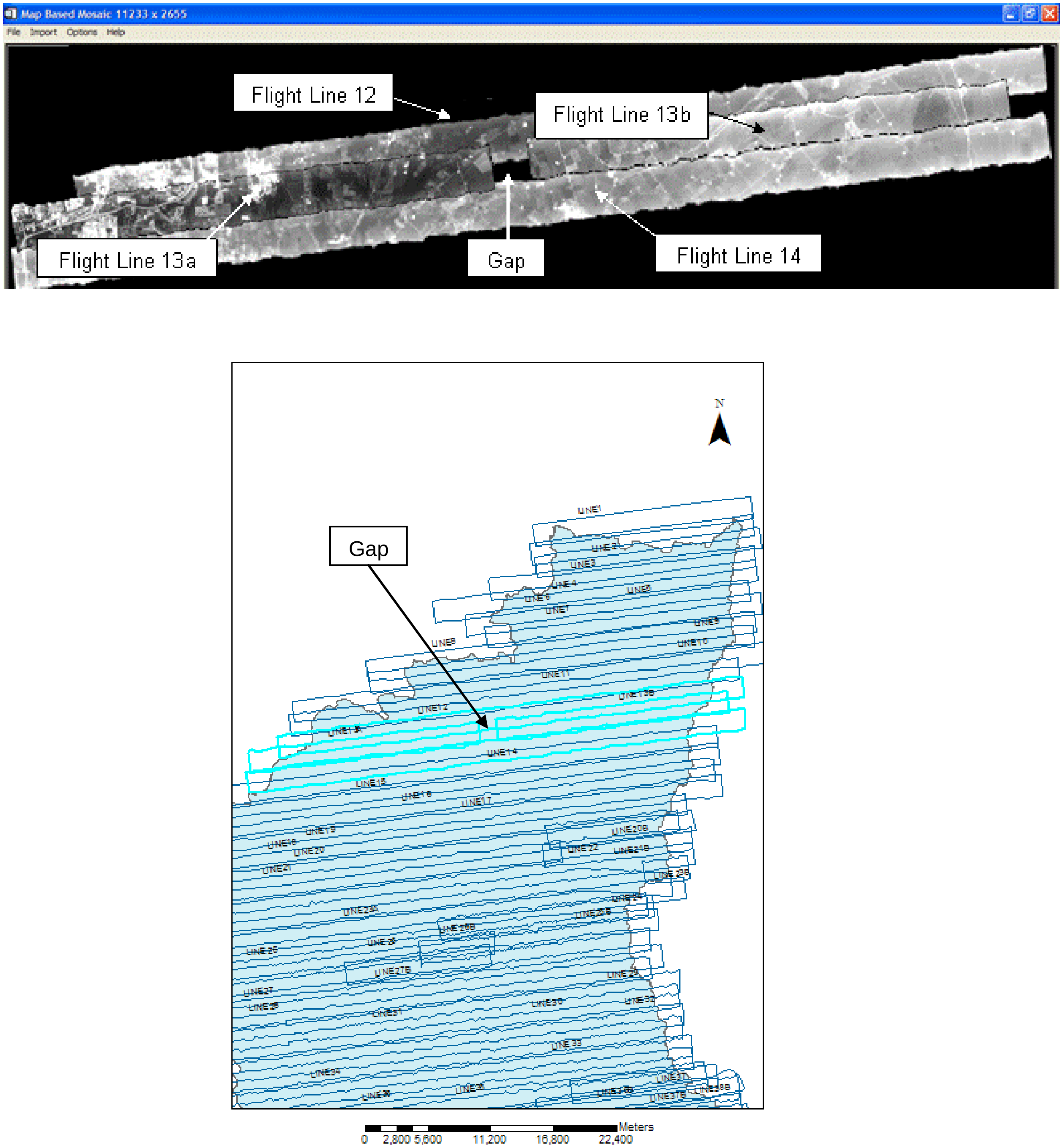


FIGURE 4

An Unclassified Gap Between Flight Lines 13a and 13b, July 25, 2002 with some flight lines more affected than others. The procedure used to evaluate the data was to examine each flight line both visually and statistically. Figures 5 and 6 provide a graphical representation of cross-track illumination. These figures show a section of uncorrected CASI data for two different flight lines acquired on July 24 and July 25 respectively. A distinct brightening on the right side of the images is apparent and visible in the cross sections on the right side of the graphics. These cross sections were generated by averaging the Digital Numbers (DNs) under a graphic mask in each band in the cross-track direction of the flightline. A polynomial curve has been fitted to the averaged profile to demonstrate the increased DN values at the edges of the flight line relative to the center. There is variability in the cover type under the mask as can be seen in the spectral profiles, however, the polynomial curve represents an averaged measure and is generally representative of overall trends. The elevated DNs on the left of the flight line are not overly apparent in the image on the left.

The results of image classification using data with cross-track illumination effects can be variable and is dependent on the spectral properties of the individual classes relative to the scene. In general, the ability to extract any particular class from a dataset is based on the presence of a unique statistical signature that represents a subset of the full scene variance. In the case of multiple classes, the individual classes must also be unique relative to each other. Class confusion occurs when there is statistical overlap between two distinct classes. In images with cross-track illumination effects, the spectral signature of ground cover types is different at the edges of the image than at the center. Consequently, training areas selected at the edges of the image will have a different spectral signature, or statistical representation, than at the center of the image.

Two approaches were considered to mitigate the effects of these cross-track illumination artifacts: (1) correction of the individual flight lines to remove the cross-track illumination effects, or (2) adapt to the cross-track illumination effects during the classification phase of the project. It was determined that the first approach was beyond the scope and available budget of the project since it would require substantial reprocessing of the data including performing corrections on all 82 flightlines and a reapplication of geometric and atmospheric corrections to each flight line. As such the second approach of adapting to the cross-track illumination effects in the classification procedure presented a more economical approach for this project and may in fact be superior to the first approach given the paucity and nature of correction algorithm tools available in most image processing software. Moreover, the second approach can effectively deal with misclassification due to cloud and shadows. The chosen approach

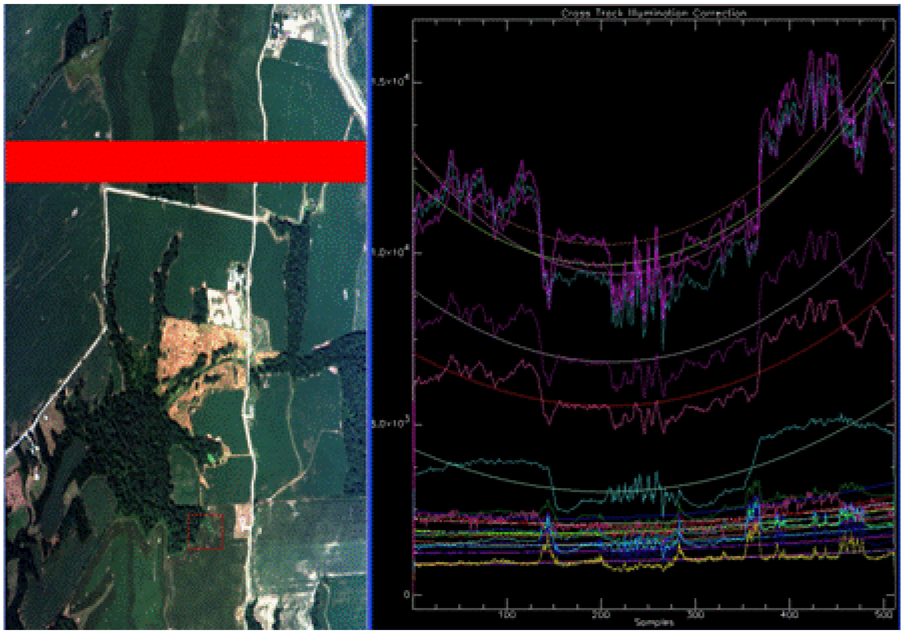


FIGURE 5

Cross-Track Illumination Artifacts from Flight Line 38 Acquired on July 24, 2002

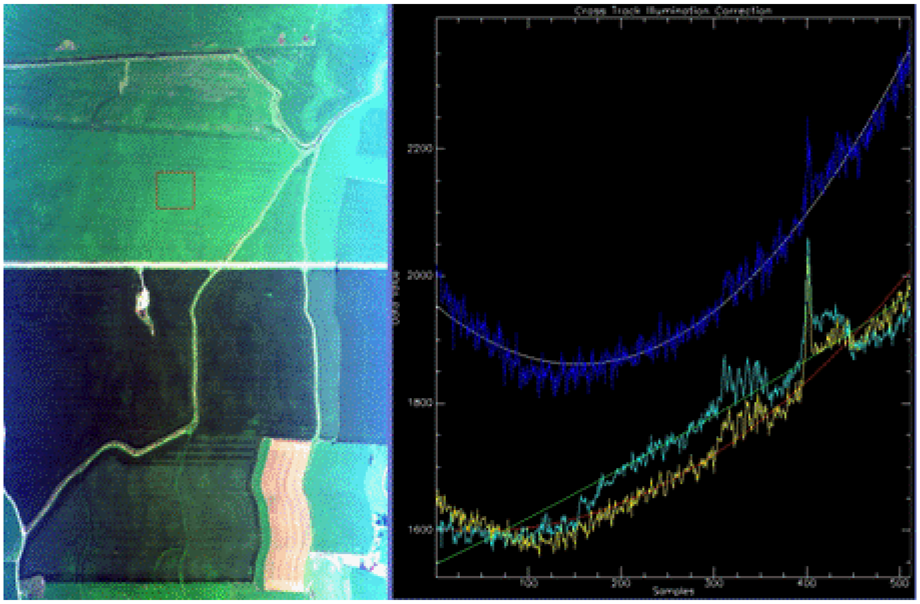


FIGURE 6

Cross-Track Illumination Artifacts from Flight Line 10 Acquired on July 25, 2002 basically amounts to increasing the number and strategic distribution of training areas in the affected flight lines such that an accurate representation of each class is acquired.

3.2.2. Ground Truth. In addition to 390 ground truth points collected by U.S. EPA personnel during the image acquisition in 2002, supplementary ground training samples were collected using two different sources of high resolution aerial images. The first was the 2003 metropolitan Cincinnati digital orthophoto quads (Aerials Express Inc., 2003) used to sample urban classes. The second dataset consisted of un-rectified 2002 aerial images from the Center for Mapping at the Ohio State University used to increase the number of sample locations in agricultural fields. Overall, 3613 and 351 reference polygons were obtained from each data set, respectively. Reference polygons from aerial images were selected from homogeneous areas of 40m x 40m (or 0.4 acre) minimum. Some of this data was used for training the two supervised classifications: image object segmentation and spectral angle mapper (SAM). Other portions of it were used for accuracy assessment following classification (Section 4.3).

Obtaining adequate ground truth data was a key component of this project and a critical aspect to acquiring ground truth data from secondary sources was in setting business rules or guidelines which precisely specified the process to be followed. The key guidelines followed included:

1. Location of sample site

   It is not uncommon to collect information on the wrong location because of inadequate procedures. This project used orthophotos of high spatial accuracy (<2m) as base layers for the ground truthing process.

2. Data collection and entry error

   Data collection errors occur when measurements are done incorrectly and variables of the classification scheme are misidentified (i.e., crop type). Ground truth data given by EPA, particularly on specific variables such as crop type, was the reference for the ground truthing process. Errors were monitored and removed from the data set when found.

3. Data collection consistency

   Training and the development of objective data collection procedures ensure data collection consistency. It is important to ensure that everyone identifying ground truth points follow the same process. Appendix B contains the protocol used by EPA to collect ground truth data. Appendix C contains the supplemental ground truth data collection protocol for the project.

4. Date of reference data

   If change in land use/land class occurs between the date of imagery capture and the date of ground truth data, the results may be impacted. Therefore, ground truth data were collected as close as possible to the date of the CASI image acquisition. For example, EPA's ground truth data from the field coincided with the CASI flyover. Supplemental aerial images used as secondary sources of agricultural ground truth came also from 2002. The 2002 LMR CASI dataset and the Aerials Express Inc. (2003) dataset (used as a secondary source of urban ground truth) are only a year or less apart, and seasonal difference within the CASI dataset itself was also minimal (ranging only from late July through early August 2002).

5. Sampling design

   The choice and distribution of ground truth samples is an important aspect of accuracy assessment. The concept of randomness is a central issue to ensure that the samples are chosen without bias and eventually the accuracy of the map is statistically sound. For this project, the 82 flight lines were used as an independent spatial framework to guarantee that the required minimum number of ground truth data is collected for each class, and that they are statistically random samples.

6. Number of ground truth data

   In the remote sensing community, a general guideline or good "rule of thumb" is to collect a minimum of 50 samples for each class. The objective was to collect samples for each class for accuracy assessment and training set samples for each class on every flight line. Overall, 4354 ground truth samples were obtained; 902 of these were used for the accuracy assessment.

3.2.3. Classification Method. A number of classification algorithms including Maximum Likelihood Classification (MLC), Spectral Angle Mapper (SAM), and Classification And Regression Tree (CART) were tested on the CASI imagery collected from the Little Miami River watershed. However, no single algorithm alone proved to be capable of successfully classifying the hyperspectral data. Some of the problems encountered included the following:

- Sub-classes of forest (e.g., deciduous versus conifer) were not separable, especially because the areas are difficult to find ground-truth for.
- Other vegetation, such as corn, grass, soybean, etc. gets classified into forest, and vice versa.
- Pasture, soybeans, corn and grass get confounded, particularly in urban areas.
- Roads are not discernible from urban/built and barren land covers.
- Rock and stone quarries were not discernible either.

Nevertheless, a solution was found to resolve the above problems and to obtain as many classes as possible under the study constraint of 80% accuracy per individual class. The solution was to use a hierarchical approach using two different classification algorithms: “Image Object Segmentation” and “Spectral Angle Mapper.”

Image object segmentation is an innovative method which uses homogenous image objects as processing units at a given scale for classification, instead of pixels. One motivation for the image object approach is to make use of powerful generalizations of the image to generate level-1 strata of different contexts. With the creation of these strata (urban, rural and water), further independent classification (e.g., pixel-based or other algorithms) of each stratum can be applied and then the results combined later to improve the accuracy of the final classified dataset.

Figure 7 shows an example result of image object segmentation versus pixel-based classification of the same forest area in the Little Miami watershed. Forest areas typically contain areas of dark shadows, bright canopy tops, and medium bright illumination under the tree canopy, which pixel-based approaches without level-1 strata classify into all different classes, for example: forest (dark green), corn (gold), and soybean (light green). Unreasonable classes, such as corn and soybean crops in the midst of forest land will be removed if a hierarchical approach is used. Also, image object segmentation could also be used without any negative effect on delineating riparian features such as a thin forest strip along a river.

On the other hand, the Spectral Angle Mapper (SAM) is a whole pixel, spectral method for supervised classification developed specifically for hyperspectral data. It is based on entirely different principles than common multispectral classification algorithms such as the probability-based Maximum Likelihood Classifier. SAM is thought of as a similarity classifier and can be used on multispectral data too, often with improved results in comparing the spectral properties of materials. In brief, it is a physically-based spectral classification that uses an n-dimensional angle to match the spectra of imaged (remotely sensed) pixels to reference spectra. The algorithm determines the spectral similarity between image and reference spectra by calculating the spectral angle between them, treating them as unit vectors in spectral space with dimensionality equal to the number of bands (in this particular case, 19 bands or dimensions). Since only the angle between the vectors matters (the smaller the angle, the better the match) and not the vector’s length, brightness variations such as topographic and cross-track illumination effects are well accounted for. In other words SAM’s advantage over common or more “traditional” classifiers is its relative

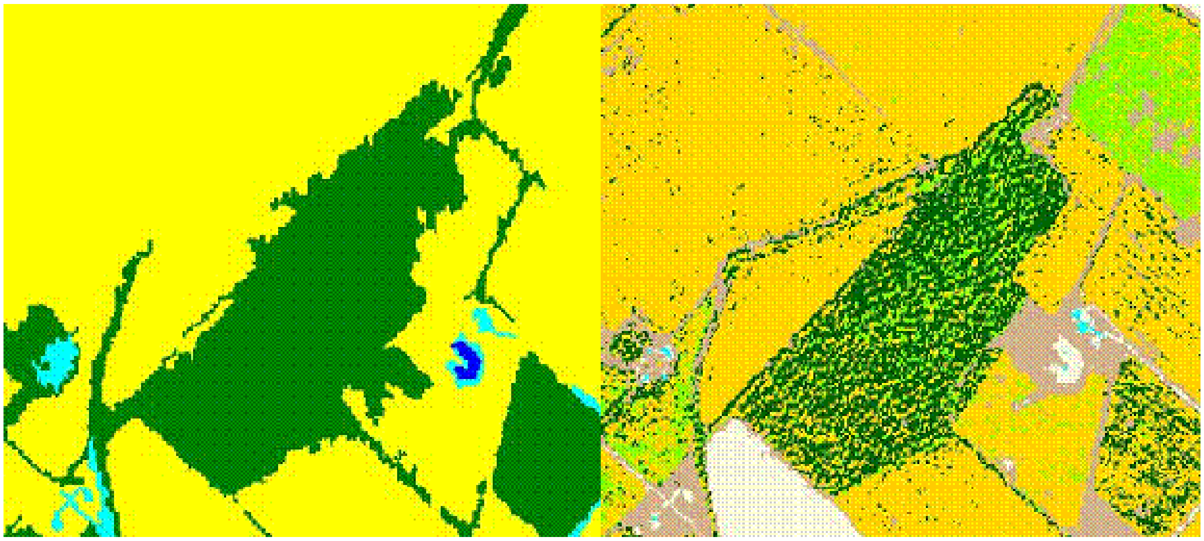


FIGURE 7

Image Object Segmentation versus Pixel-Based Classification
of a Forest Area in the Little Miami River Watershed

insensitivity to illumination and albedo effects inherent with remotely sensed imagery (Research Systems Inc., 2002; Kruse et al., 1993).

In order to combine the power of both approaches, object image segmentation in eCognition (Definiens Imaging GmbH., 2003) was applied as a “Level 1” classification of water, urban, and rural features which required consideration of large-scale factors, as well as area-based parameters such as adjacency, texture and shape. Next, Spectral Angle Mapper in ENVI (Research Systems Inc., 2003) was applied as a “Level-2” classifier to discern the urban and rural areas into more specific classes (barren, built, grass, corn, soypbeans, wheat, etc.). Wire diagrams of the methodology used are shown in Figures 8 and 9. Throughout the process, image object segmentation created a spatially exclusive mask of urban and rural regions and then each region was filled with the classification result from SAM. As a consequence, the final classification result remained a uni-scale product with a spatial resolution of 4 meters throughout the entire watershed.

Training samples were chosen along all flight lines and selected in order to capture intra-class variation. The same classification rules were applied to each flightline. After the first round of classification, results were fine-tuned by adding training sets to accurately define inter-class boundaries. The classification results for each individual flight line were assessed for accuracy and accepted if they did not show any overall discrepancies with respect to the aerial images. Classification results were also re-examined after joining the classified flight lines together in a mosaic of the watershed.

Any problem areas noted during this stage were addressed during the QA/QC phase (Section 3.2.4 and Section 4).

Overall, 11 land cover classes were mapped. A twelfth class included pixels that could not be classified, for example, due to cloud cover. The 11 land cover classes consisted of:

Lentic: Open water associated with still water systems, such as lakes, reservoirs, potholes, and stockponds. Such bodies typically do not have a defined channel or associated floodplain.

Lotic: Open water associated with running water systems, such as rivers or streams. Such waterways typically have a defined channel and an associated floodplain.

Forest: Contains either or both deciduous and coniferous trees in any degree of mixture. Single stemmed, woody vegetation with canopy spanning greater than 4 meters and tree canopy accounting for 25-100% of the cover

Corn: Area under cultivation of food and fiber, where corn is the primary crop.

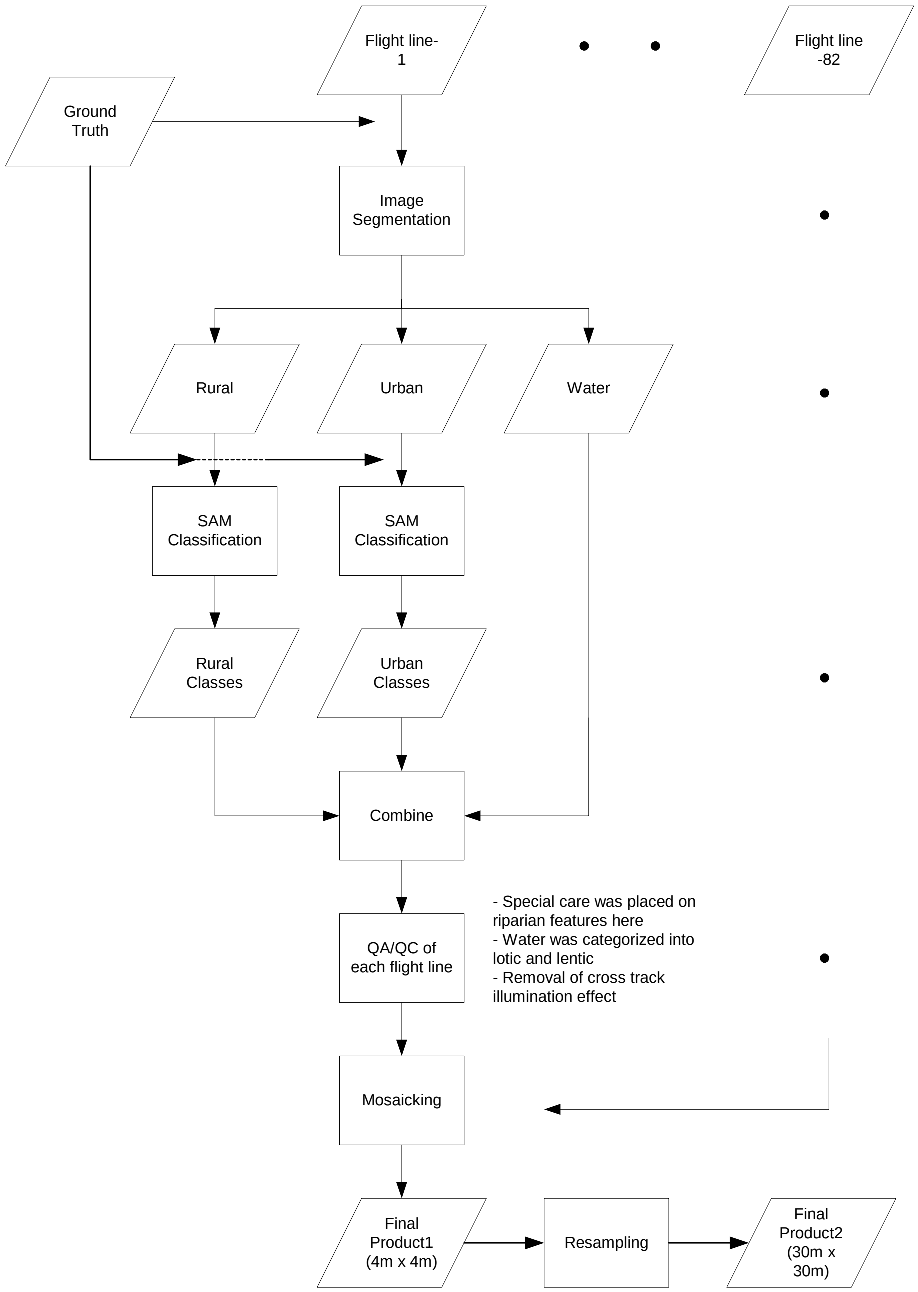


FIGURE 8

Flowchart of the Classification Method

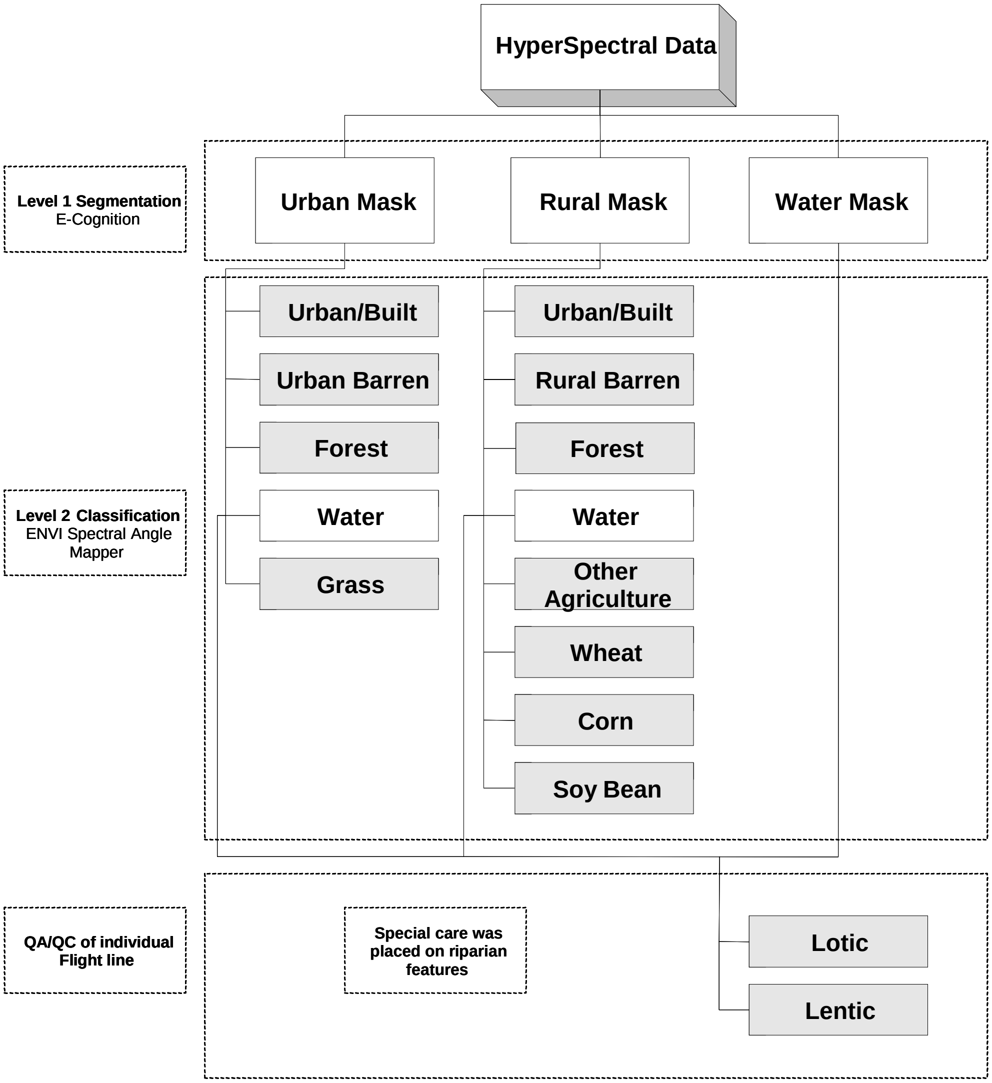


FIGURE 9

Schematic of the Hierarchical Classification and Derived Classes (shadowed)

Soybean: Area under cultivation of food and fiber, where soybean is the primary crop.

Wheat: Area under cultivation of food and fiber, where wheat is the primary crop.

Dry Herbaceous: Dominated by dry and/or less vigorous herbaceous vegetation; herbaceous vegetation accounts for more than 25% of the ground cover. This class mainly includes naturally occurring and unmanaged herbaceous vegetation, and dried out, unhealthy, or stressed croplands. Dry herbaceous vegetation prevailed in croplands, as well as, “Other Agriculture” lands (fallow, hay, pasture, or natural grassland prairies or fields), due to drought in the Summer of 2002. Dry herbaceous vegetation had little chlorophyll content and very similar spectral signatures without regard to vegetative species.

Grass: Dominated by cultivated grasses planted in developed settings for recreation, erosion control, or aesthetic purposes. Examples include parks, lawns, golf courses, airport grasses, and industrial site grasses.

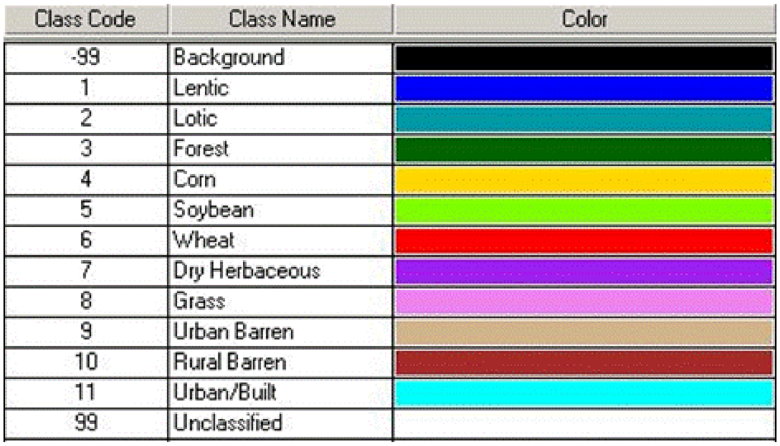
Urban Barren: Composed of bare soil, rock, sand, silt, gravel, or other earthen material with little (less than 25%) or no vegetation within urban areas. Examples include exposed soil in urban areas and construction sites.

Rural Barren: Composed of bare soil, rock, sand, silt, gravel, or other earthen material with little (less than 25%) or no vegetation in rural areas. Typically fallow fields are included in this class too.

Urban/Built: Areas covered by structures and impervious surfaces in urban, suburban, and rural areas. Typically buildings, parking lots, and paved roads.

Unclassified: This class includes areas of image gaps among flight-lines and cloud cover where land cover classification was not feasible.

Figure 10 shows the class code, name and corresponding color schemes for all classification results. Figure 11 provides an overview of the LULC classification.



Class Code	Class Name	Color
-99	Background	⬛
1	Lentic	🟦
2	Lotic	🟦
3	Forest	🟩
4	Corn	🟨
5	Soybean	🟩
6	Wheat	🟥
7	Dry Herbaceous	🟪
8	Grass	🟪
9	Urban Barren	🟫
10	Rural Barren	🟫
11	Urban/Built	🟦
99	Unclassified	⬜

FIGURE 10

Legend for the Final 4m x 4m LULC Product

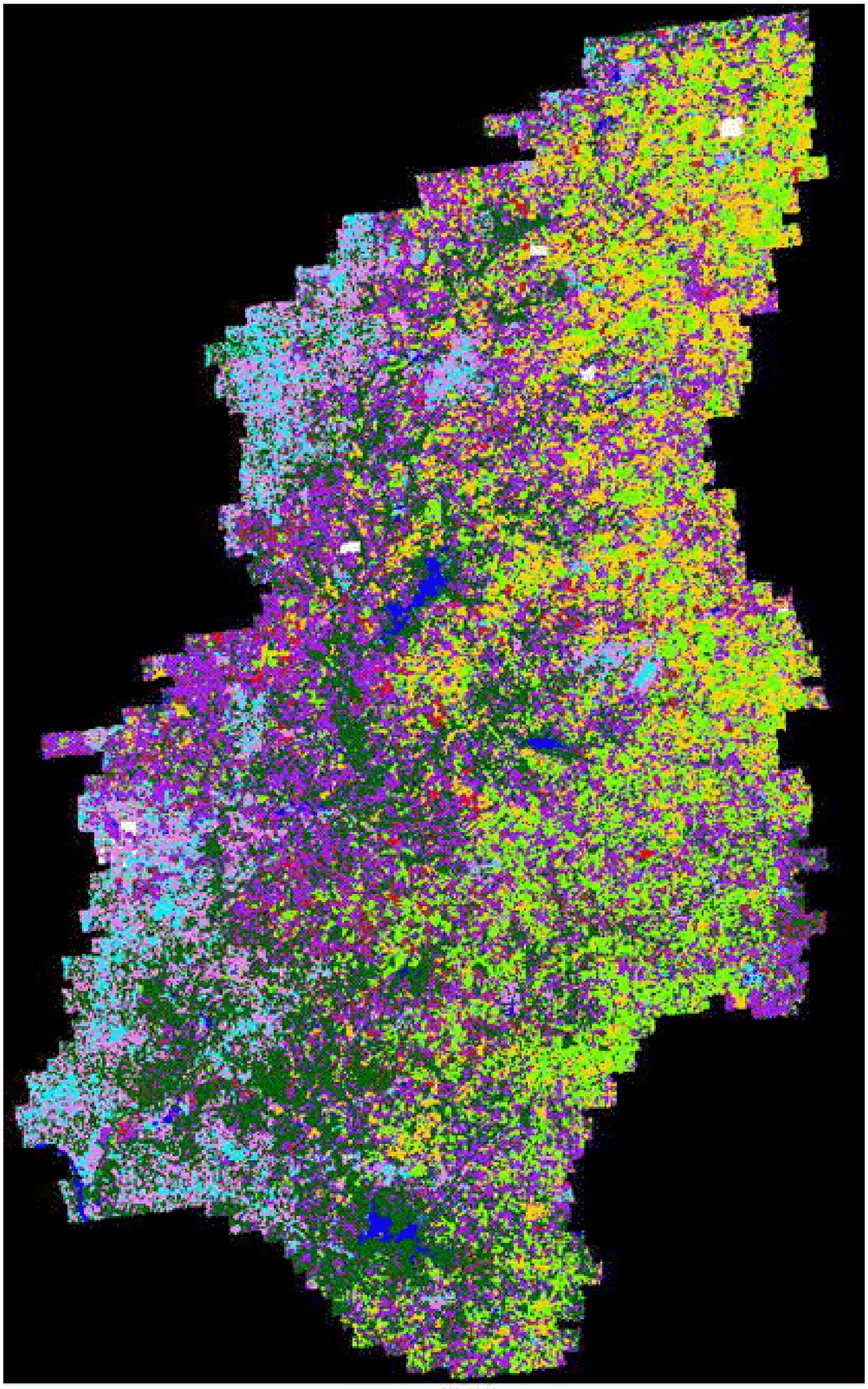


FIGURE 11

Overview of the 4m x 4m LULC Classification for the Little Miami River Watershed

Final products were delivered in the projection and data format provided in Table 2.

TABLE 2

Projection for the Little Miami River Land Use/Land Cover

Projection:	Albers Conic Equal-Area
1st Standard Latitude:	29 degrees, 30 minutes, 00 seconds
2nd Standard Latitude:	45 degrees, 30 minutes, 00 seconds
Latitude of Projection's Origin:	23 degrees, 00 minutes, 00 seconds
Central (Meridian) Longitude Origin:	-96 degrees, 00 minutes, 00 seconds
Datum (Ellipsoid)/Spheroid	NAD83/GRS80
Units:	Meters
Orientation:	North up
Pixel Size:	4 meters
Precision for mosaicked flight lines:	± 3 pixels
Data Format:	Erdas Imagine 32 bit signed integer .img file format
Naming Convention	Original/unsmoothed LULC product: little_miami_river_watershed_4m_before_clumping_signed32bit.img Smoothed LULC product: little_miami_river_watershed_4m_signed32bit.img

3.2.4. Consistency of Classification Across Flightlines. Since there were 82 independent flight lines of imagery, it was important to manage consistency of classification across the flight lines. In order to do this, the classification team classified and conducted an extensive accuracy assessment on 8 selected flight lines (10% of total number of images). As the selected flight lines were classified to an acceptable level of accuracy, they were used to aid in ensuring consistency across all the neighboring images. When an initial classification was completed for any given flight line, it was compared to all of its neighbors and any distinct classification differences along the boundaries or overlap between flight lines were addressed. This process helped to mitigate categorical edge-matching errors when the 82 individual classified flight lines were finally stitched together into the watershed mosaic. Discontinuities along image boundaries were removed using polynomial based rubber-sheeting using ground control points and corresponding cut-lines. The final mosaicked product has three pixel RMS errors on 4 meter spatial resolution data.

4. POST-PROCESSING AND QA/QC

Image post-processing and QA/QC following the Level 1 and 2 classifications included manual editing (Section 4.1) and map generalization to create a second LULC product (Section 4.2). Follow-up steps used to ensure the accuracy and proper interpretation of the end product included:

- The completion of an accuracy assessment including an error matrix and computations of overall, producer, and user accuracy (Section 4.3); and
- Generating metadata for the final product compatible with the Federal Geographic Data Committee (FGDC) standard (Appendix D).

4.1. MANUAL EDITING

Manual editing was used as a final QA/QC step. Final image edits included:

1) Differentiation of water bodies into lotic and lentic. The discrimination of lotic versus lentic waters was completed using the U.S. EPA's National Hydrography Dataset or "NHD."

2) Misclassifications due to clouds, shadows, or haze. Hyperspectral images were systematically examined along with the classification results to identify any misclassified pixels due to the presence of clouds, shadows, or haze. These pixels were then manually included in a mask to convert them to the "Unclassified" class.

3) Removal of boundary effects. Cross-track illumination affected the SAM classification results along the boundary of each flight line. As such, manual editing was necessary to correctly assign the classes. The other task was to remove any inconsistent classifications in overlapping flight lines due to the temporal gap in image acquisition. Atmospheric conditions varied day-by-day, and a 16-day gap in image acquisition throughout the watershed is long enough to expect some changes in vegetation condition. The mosaicking phase was prolonged due to cross-track illumination effects and the temporal gap in image acquisition. All the image boundaries had to be manually inspected and the correct class in the overlap area determined. The percentage of "dry herbaceous" may have increased in the southern region due to the later dates of imagery acquisition. Wheat class was scarce in the southern region as the fields had been harvested, or had senesced. Small portions of some soybean and corn fields had also senesced and, in such cases, were not assigned to the correct class. Such fields were manually interpreted and assigned to the correct class.

4) Other discernible misclassifications. Dark shadow areas and black asphalt pavement were easily confused with water bodies. In such cases, these areas were manually interpreted and assigned to the correct class.

4.2. MAP GENERALIZATION

A “clump-sieve-and-fill” technique was used to eliminate single pixels or groupings of pixels that were smaller than the minimum target mapping unit (e.g., random pixels of “Forest” denoting scattered trees in an otherwise homogeneous 40-acre plot of “Corn”). As a result, a second LULC mosaic product was produced which eliminated the “salt and pepper” effect common in classifications of smaller pixel or “finer” spatial resolution imagery. Although smoothed images are generally more “pleasing” to the eye, U.S. EPA was concerned about any smoothing process which might potentially wipe out important small features, or thin, linear features such as riparian forest. On the other hand, available literature also suggested that whereas unsmoothed or “salt and pepper” classification results may be *more realistic* (e.g., cases where random pixels of “Forest” or scattered trees really do exist within a “Corn” field), smoothed classifications may be *more meaningful* in terms of deriving land use statistics important for interpreting ecological processes across the landscape (e.g., Burnett and Blaschke, 2003; Dorren et al., 2003). As a result, two LULC products were created for the Little Miami River watershed, one smoothed and one not. For the smoothed product, the minimum mapping unit of the classification result is about 0.04 of an acre as represented by 10 pixel clusters (at 4m x 4m spatial resolution), or linear chains of minimally four contiguous pixels in any direction. The original, unsmoothed product remains at a 4m x 4m spatial resolution.

4.3. ACCURACY ASSESSMENT

The accuracy assessment was based on whether the majority of classed pixels in a 3 x 3 pixel window, centered on a ground truth site, agreed or not. Thus, if five or more pixels were classified as corn, and ground truth indicated corn, then the majority criterion was satisfied and “corn class” would be considered correct for that site. A standard error matrix was used in reporting classification accuracies (Table 3). This matrix reports the number of pixels assigned to a particular category in a classification relative to the number of pixels assigned to a particular category in a reference classification. In this case, the classified data, represented by rows in Table 3 and Table 4, are the land cover classifications derived for this project, and the reference data are represented by the columns in the tables. A total of 902 independent ground truth sites were used for the accuracy assessment, including primary data (i.e., data collected by U.S. EPA scientists in the field at the time of the overflights), and secondary data from 2002 and 2003 aerial images of the watershed as explained in Section 3.2.2.

TABLE 3

Classification Error Matrix
(units are the number of reference points)

		Reference (Known Cover Types)											
		Lentic	Lotic	Forest	Corn	Soybean	Wheat	Dry Herbaceous	Grass	Urban Barren	Rural Barren	Urban/ Built	Row Total
Classified	Lentic	**85**	1	0	0	0	0	0	0	0	0	0	**86**
	Lotic	3	**17**	0	0	0	0	0	0	0	0	0	**20**
	Forest	0	2	**95**	0	0	0	3	4	0	2	0	**106**
	Corn	0	0	1	**106**	3	0	1	0	0	0	0	**111**
	Soybean	0	0	0	3	**107**	2	5	1	0	0	0	**118**
	Wheat	0	0	0	0	0	**18**	0	0	0	0	0	**18**
	Dry Herbaceous	2	0	3	9	12	8	**82**	16	0	8	0	**140**
	Grass	1	0	1	0	0	0	8	**75**	8	0	0	**93**
	Urban Barren	0	0	0	0	0	0	0	1	**57**	1	4	**63**
	Rural Barren	6	1	0	2	0	10	1	3	0	**20**	1	**44**
	Urban/Built	1	3	0	0	0	0	0	0	4	0	**95**	**103**
	Column Total	**98**	**24**	**100**	**120**	**122**	**38**	**100**	**100**	**69**	**31**	**100**	**902**

TABLE 4

Summary Statistics of the Accuracy Assessment
(Overall Accuracy of Classification = 83.92%)
(KHAT statistic = 0.82)

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Lentic	98	86	85	86.73%	98.84%
Lotic	24	20	17	70.83%	85.00%
Forest	100	106	95	95.00%	89.62%
Corn	120	111	106	88.33%	95.50%
Soy	122	118	107	87.70%	90.68%
Wheat	38	18	18	47.37%	100.00%
Dry Herb	100	140	82	82.00%	58.57%
Grass	100	93	75	75.00%	80.65%
Urban Barren	69	63	57	82.61%	90.48%
Rural Barren	31	44	20	64.52%	45.45%
Urban/Built	100	103	95	95.00%	92.23%

Three major descriptive measures were used for accuracy assessment, namely: “overall accuracy,” “producer’s accuracy,” and “user’s accuracy.” Overall accuracy is computed by dividing the total number of correctly classified pixels (the sum of the elements along the major diagonal in the error matrix) by the total number of reference points. Producer’s accuracy indicates how well reference pixels of the given cover type are classified. It is computed by dividing the number of correctly classified pixels in each category (on the major diagonal of the error matrix) by the number of reference points used for the class (the column total). User’s accuracy is a measure of commission error and indicates the probability that a pixel classified into a given category actually represents that category on the ground. It is computed by dividing the number of correctly classified pixels in each category (on the major diagonal of the error matrix) by the total number of pixels classified into that category (row total) (Lillesand and Kiefer, 1994).

Tables 3 and 4 present the quantitative summary of the accuracy assessment for the classification result using 902 reference points. The overall classification accuracy was 83.92%, which is above the project’s target of 80% accuracy. However, the producer’s and user’s accuracy of some classes fell short of the target of 80%. The KHAT statistic, a comparative statistic to use in comparing this classification product with others, was 0.82. In other words, the classification result here was 82% better than one resulting from chance. Overall, the strengths of this classification relative to other existing LULC datasets of this watershed, such as the National Land Cover Dataset or NLCD (Vogelmann et al., 2001), and the State of Ohio Land Cover (Ohio DNR, 1994) are:

- Higher spatial resolution (e.g., 4m x 4m rather than 30m x 30m pixel resolution of previous existing classifications of this watershed),
- Higher spectral resolution (the classification is derived from more spectral bands of information than provided by Landsat platforms),
- Ground-truth data used as a basis for this classification includes primary sources of information taken in the field at the same time the remote sensing data was collected,
- The discernment of corn from soybeans (the two major agricultural crops found within this watershed), and
- Better spatial and updated temporal discernment of urban built (existing and new infrastructure) versus urban barren (newly developing areas).

5. RESULTS AND DISCUSSION

A basic understanding of the physical (i.e., geological and anthropogenic) processes at work in the Little Miami River watershed will help the user of this LULC dataset interpret some of the resulting land use patterns shown in Figure 11. For example, one should note that the northern half of the watershed is relatively flat and composed of "Till Plains" with soils that developed from the loamy, limy deposits of the Wisconsinan glaciation roughly 18,000 years ago. These soils normally have better natural drainage and fertility than those of the southern half of the watershed (or "Drift Plain"). The southern half of the watershed has more deeply-leached, acidic, pre-Wisconsinan till and thin loess, as well as, very poorly-drained soils with fragipans (clays). The southern half of the watershed also exhibits relatively modest relief, but with dissected areas and somewhat more complex topography than the north (Omernik, 1987; Woods et al., 1998). As such, the northern and southern parts of the watershed may be expected to have different types and proportions of certain land uses or land covers based on the differing soils and micro-climates found in these two distinct "ecoregions" (the "Till Plain" and the "Drift Plain," refer to Figure 1).

Spatial patterns separating western and eastern portions of the watershed exist too. Perhaps most notable is the western urban/exurban corridor stretching from Cincinnati (in the south) to Dayton and Xenia (in the north) and beyond, encompassing portions of Hamilton, Warren, and Montgomery Counties. These growing urban landscapes run parallel to and already straddle much of the mainstem of the Little Miami River which can be observed as a nearly contiguous linear band of riparian forest running up along the western part of the image. The eastern half of the watershed tends to be more agricultural in character, particularly in the north. But this appears to wane in the east-central part of the image near the city of Wilmington (a crossroads or pole for the primary economic sector in this region, as well as, a major air transportation hub), and in the south as well, particularly along the East Fork of the Little Miami River in Clermont County where urban development and human population continue to rapidly grow.

Other major patterns in the final classification include concentrations of "dry herbaceous" land cover in the western part of the watershed near the urban/rural fringes of Warren, Montgomery and Green Counties, as well as along, or at the source of, many headwater streams to the east (i.e., dry, thin vegetative areas buffering perennial or low-flow streams from adjacent croplands). Recall that "dry herbaceous" was defined in this project as a variety of “Other Agricultural” components including hay, pasture, fallow, dried out crops, and natural herbaceous vegetation. As such, concentrations of dry herbaceous land cover in the west are likely non–irrigated lands, areas practicing water conservation, or perhaps areas left fallow due to failed crops or in anticipation of near-term development. The thin lines of dry herbaceous cover bordering perennial or low-flow streams in the east was likely spectrally distinct, or separable from corn and soybeans based on aerial photography, yet still dry from drought and “managed” only in the sense that it was not mowed or turned over by farmers.

The ratio of corn to soybeans is lower in the southern part of the watershed. This is expected to a certain extent since soybeans are well known to be common and well adapted to spring soil wetness in the southern half of the Little Miami River watershed (Woods et al., 1998). This observation may well represent the reality of crop planting patterns in 2002, and/or failed corn crops susceptible to the early flood then drought extremes experienced that year (and thus, classifications of “dry herbaceous” rather than “corn”). The planting of corn was delayed at least three weeks in many areas of Ohio in 2002 due to heavy spring rains and flooded fields that year. However, subsequent to seeding, drought conditions ensued and lingered in 2002.

The summer drought of 2002 likely affected other classes in this LULC as well. For example, lotic or running waters were rarely detected because many drainage ways, or headwater or perennial streams during the period of late July to early August were already dry or experiencing low flows. Un-watered grasses in otherwise managed grassy areas might also have been classified as “dry herbaceous.”

Nevertheless, the resulting LULC is an important dataset for a variety of environmental and geographic studies within the Little Miami River watershed. Even given the predominance of the “dry herbaceous” class, it remains meaningful in terms of studying several urban and agricultural patterns or gradients, as well as, anthropogenic and natural processes within the watershed. Many hydrological, ecological, and geological applications may be possible too, along with the ability to assist in urban planning, study urban sprawl, and contribute to measurements and evaluations of zoning, congestion, pollution and human health.

6. REFERENCES

Aerials Express Inc. 2003. Cincinnati Metropolitan Aerial Photography. Tempe, AZ.

Analytical Imaging and Geophysics LLC. 2002. ACORN Quick Start Guide. ACORN Version 4.10, March, 2002 Edition. Boulder, CO.

Burnett, C. and T. Blaschke. 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. Ecol. Model. 168:233-249.

Definiens Imaging GmbH. 2003. eCognition 3.0. Munich, Germany.

Dorren, L.K.A. , B. Maier and A.C. Seijmonsbergen. 2003. Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. Forest Ecol. Manag. 183:31-46

ITRES Research Limited. 2006. ITRES: Airborne Hyperspectral Remote Sensing Systems & Solutions. Calgary, Alberta, Canada. Available at http://www.itres.com.

Kruse, F.A., A.B. Lefkoff, J.B. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, and A.F.H. Goetz. 1993. The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging spectrometer Data. Remote Sens. Environ. 44:145-163.

Lillesand, T.M. and R.W. Kiefer. 1994. Remote Sensing and Image Interpretation, 3rd ed. John Wiley and Sons, Inc., New York, NY.

Ohio DNR (Ohio Department of Natural Resources). 1994. State of Ohio Land Cover. Division of Real Estate and Land Management. Columbus, OH.

Omernik, J.M. 1987. Ecoregions of the conterminous United States. Map (1:7,500,000). Ann. Assoc. Am. Geograph. 77(1):118-125.

Research Systems Inc. 2002. Spectral Analysis with ENVI. Boulder, CO: The Professional Services Group, Research Systems, Inc., A Kodak Company.

Research Systems Inc. 2003. ENVI 3.6 (the Environment for Visualizing Images). Boulder, CO. Available at http://www.ResearchSystems.com.

Sanders, R.E., Ed. 2002. A guide to Ohio streams. Columbus, Ohio: Streams Committee, Ohio Chapter of the American Fisheries Society. In partnership with the Ohio Environmental Education Fund, Ohio Department of Natural Resources, and the Ohio Environmental Protection Agency.

Thenkabail, P.S., A.D. Ward, J.G. Lyon and C.J. Merry. 1994. Thematic Mapper vegetation indices for determining soybean and corn crop parameters. Photogramm. Eng. Rem. S. 60:437-442.

Thenkabail, P.S., R.B. Smith and E. De Pauw. 2000. Hyperspectral vegetation indices for determining agricultural crop characteristics. Remote Sens. Environ. 71:158-182.

Vogelmann, J.E., S.M. Howard, L. Yang, C.R. Larson, B.K. Wylie and N. Van Driel. 2001. Completion of the 1990s national land cover data set for the conterminous United States from Landsat thematic mapper data and ancillary data sources. Photogramm. Eng. Rem*. S.* 67:650-662.

Wessex, Inc. 1997. “First St.” software. U.S. Streets CDROM, Release 4.0. Winetka, IL. Available at www.wessex.com.

Woods, A.J., J.M. Omernik, C.S. Brockman, T.D. Gerber, W.D. Hosteter and S.H. Azevedo. 1998. Ecoregions of Indiana and Ohio. Interior-Geological Survey, Reston, VA. Accessed May 19, 2005 at ftp://ftp.epa.gov/wed/ecoregions/oh_in/ohin_front.pdf.

APPENDIX A

FLIGHT LOGS OF THE CASI DATA COLLECTION IN THE LITTLE MIAMI RIVER WATERSHED
June 24-August 8, 2002

DATE	LINE	FILE	START TIME	STOP TIME	*f*STOP	INT TIME	ALTITUDE (FEET)	HEADING	VISIBILITY (MILES)	SPEED (MPH)
July 24, 2002		1	12:55:23	12:56:30	5.6	30	10,500	Dark Data		
	1	2	13:10:49	13:15:50	5.6	30	10,500	E	10	124
	2	3	13:18:57	13:24:26	5.6	30	10,500	W	10	120
	3	4	13:27:51	13:33:01	5.6	30	10,500	E	10	127
	4	5	13:36:01	13:42:40	5.6	30	10,500	W	10	121
	5	6	13:46:01	13:53:45	4	30	10,500	E	10	128
	6	7	13:57:05	14:04:00	4	30	10,500	W	10	121
	7	8	14:07:00	14:13:20	4	30	10,500	E	10	128
		9	14:13:49	14:19:30	4	30	10,500	Dark Data		
	25	10	14:24:00	14:34:54	4	30	10,500	W	10	127
	26	11	14:38:15	14:49:20	4	30	10,500	E	10	123
	27	12	14:52:45	15:04:12	4	30	10,500	W	10	127
	36	13	15:15:20	15:29:40	5.6	30	10,500	E	8	124
	37	14	15:32:45	15:47:11	5.6	30	10,500	W	8	127
	38	15	15:50:10	16:00:10	5.6	30	10,500	E	8	127
	39	16	16:01:47	16:13:10	5.6	30	10,500	W	8	125
July 25, 2002		1	12:31:28	12:32:40	4	30	10,500	Dark Data		
	8	2	12:48:10	12:58:02	4	30	10,500	W	5	118

DATE	LINE	FILE	START TIME	STOP TIME	*f*STOP	INT TIME	ALTITUDE (FEET)	HEADING	VISIBILITY (MILES)	SPEED (MPH)
	9	3	13:00:45	13:09:10	4	30	10,500	E	5	130
	10	4	13:13:15	13:24:45	4	30	10,500	W	5	119
	11	5	13:27:20	13:37:10	4	30	10,500	E	5	130
	12	6	13:39:55	13:51:25	4	30	10,500	W	5	118
	13	7	13:54:20	14:00:00	4	30	10,500	E	5	132
	13	9	14:00:10	14:05:00	4	30	10,500	E	5	116
	14	10	14:08:40	14:20:30	4	30	10,500	W	5	132
	15	11	14:24:10	14:35:45	4	30	10,500	E	5	116
	16	12	14:39:05	14:52:35	4	30	10,500	W	5	132
	17	13	14:56:42	15:09:00	4	30	10,500	E	5	116
	18	14	15:12:03	15:25:40	4	30	10,500	W	5	132
	19	15	15:28:30	15:40:00	5.6	30	10,500	E	5	116
	20	16	15:43:00	15:55:50	5.6	30	10,500	W	5	132
	21	17	15:58:20	16:06:34	5.6	30	10,500	E	5	116
		18	16:07:12	16:08:15	5.6	30	10,500	Dark Data		
July 31, 2002		1	12:38:19	12:39:20	4.0					
			12:39:50	12:40:50	5.6					
	22	3	12:51:23	13:02:00	4.0	31	10,500	W	8–10	133
	23a	4	13:05:30	13:19:00	4.0	31	10,500	E	8–10	105

DATE	LINE	FILE	START TIME	STOP TIME	*f*STOP	INT TIME	ALTITUDE (FEET)	HEADING	VISIBILITY (MILES)	SPEED (MPH)
	28	5	13:34:30	13:44:30	4.0	31	10,500	W	8–10	130
	29	6	13:47:01	14:00:05	4.0	31	10,500	E	8–10	105
	30	7	14:03:10	14:12:55	4.0	31	10,500	W		130
	31	8	14:15:10	14:28:15	4.0	31	10,500	E		105
	32	9	14:30:14	14:39:45	4.0	31	10,500	W		133
	33	10	14:41:45	14:54:30	4.0	31	10,500	E		105
	34	11	14:56:25	15:05:30	4.0	31	10,500	W		142
	35	12	15:10:29	15:25:38	5.6	32	10,500	E		105
	24	13	15:31:15	15:41:30	5.6	32	10,500	W		140
	27	14	15:46:48	15:50:38	5.6	32	10,500	E		105
	26	15	15:51:56	15:53:32	5.6	32	10,500	W		140
	25	16	15:54:45	16:00:25	5.6	32	10,500	E		105
	23b	17	16:02:17	16:03:33	5.6	32	10,500	E		105
	21	18	16:05:10	16:08:20	5.6	32	10,500	W		140
	20	19	16:09:35	16:13:00				E		105
August 9, 2002	57	2	12:58:56	13:15:57	4	31	10,500	W	>10	130
	56	3	13:17:48	13:36:24	4	31	10,500	E	10	112
	55	4	13:38:26	13:54:55	4	31	10,500	W	10	127
	54	5	13:56:41	14:14:40	4	31	10,500	E	10	115
	53	6	14:18:34	14:34:59	4	31	10,500	W	10	130

DATE	LINE	FILE	START TIME	STOP TIME	*f*STOP	INT TIME	ALTITUDE (FEET)	HEADING	VISIBILITY (MILES)	SPEED (MPH)
	52	7	14:36:32	14:54:00	4	31	10,500	E	10	120
	51	8	14:56:00	15:12:01	5.6	31	10,500	W	10	135
	50b	9	15:13:35	15:17:45	5.6	31	10,500	E	10	115
	39b	10	15:25:14	15:33:25	5.6	31	10,500	E	10	115
	38b	11	15:34:57	15:39:51	5.6	31	10,500	W	10	135
	37b	12	15:41:07	15:45:48	5.6	31	10,500	E	10	116
	36b	13	15:47:25	15:50:38	5.6	31	10,500	W	10	135
August 1, 2002		1	12:46:47	12:48:00	5.6	32				
		2	12:48:37	12:49:55	40	32				
	40	3	13:10:30	13:26:40	40	32	10,500	W	8–10	136
	41	4	13:20:05	13:48:55	40	32	10,500	E	8–10	114
	42	5	13:51:29	14:06:10	40	32	10,500	W	8–10	138
	43a	6	14:09:20	14:14:50	40	32	10,500	E	8–10	110
	43b	7	14:17:46	14:32:35	40	32	10,500	E	8–10	110
	44	8	14:39:20	14:54:10	40	32	10,500	W	8–10	137
	45	9	14:56:25	15:14:50	5.6	32	10,500	E	8–10	108
	46	10	15:16:45	15:30:55	5.6	32	10,500	W	8	138
	47	11	15:33:25	15:51:20	5.6	32	10,500	E	8	110
	48	12	15:54:07	16:08:30	5.6	32	10,500	W	8	137

DATE	LINE	FILE	START TIME	STOP TIME	*f*STOP	INT TIME	ALTITUDE (FEET)	HEADING	VISIBILITY (MILES)	SPEED (MPH)
	49	13	16:10:35	16:29:20	5.6	32	10,500	E	8	110
		14	16:36:10	16:37:10	5.6	32	5000			
August 3, 2002	7	1	13:19:12	13:23:28	4.0	31	10,500	W	3–4	112
	13	2	13:27:26	13:29:35	4.0	31	10,500	W	3–4	110
	23	3	13:34:56	13:37:23	4.0	31	10,500	E	3–4	130
August 4, 2002	1	1	12:59:12	13:00:28	4	31	10,500	W	<3	117
	2	2	13:01:56	13:04:55	4	31	10,500	W	<3	117
	3	3	13:07:29	13:09:18	4	31	10,500	E	<3	133
	4	14	13:09:21	13:10:29	4	31	10,500	E	<3	133
	5	19	13:11:02	13:12:31	4	31	10,500	E	<3	133
August 6, 2002	50a	1	13:06:02	13:24:15	4	31	10,500	W	6	130
August 7, 2002	68a	1	13:16:21	13:23:34	4	31	10,500	W	>10	120
	74	2	13:27:02	13:33:38	4	31	10,500	E	>10	120
	75	3	13:35:43	13:42:24	4	31	10,500	W	>10	120
	76	4	13:45:11	13:51:42	4	31	10,500	E	>10	120
	77	5	13:53:28	13:59:52	4	31	10,500	W	>10	120
	78	6	14:02:20	14:08:09	4	31	10,500	E	>10	120
	79	7	14:10:46	14:16:18	4	31	10,500	W	>10	120
	80	8	14:17:53	14:22:07	4	31	10,500	E	>10	120

DATE	LINE	FILE	START TIME	STOP TIME	*f*STOP	INT TIME	ALTITUDE (FEET)	HEADING	VISIBILITY (MILES)	SPEED (MPH)
	81	9	14:23:41	14:27:42	4	31	10,500	W	>10	120
	82	10	14:30:21	14:33:08	4	31	10,500	E	>10	120
	73	11	14:37:37	14:44:19	4	31	10,500	W	>10	120
	72	12	14:47:04	14:54:24	5.6	31	10,500	E	>10	120
	71	13	14:56:29	15:04:30	5.6	31	10,500	W	>10	120
	70	14	15:07:05	15:15:44	5.6	31	10,500	E	>10	120
	62-2	15	15:21:57	15:23:00	5.6	31	10,500	E	>10	120
	61-2	16	15:25:31	15:27:26	5.6	31	10,500	W	>10	120
August 8, 2002	60	1	13:09:19	13:27:22	4	31	10,500	W	10	120
	61-1a	2	13:30:20	13:46:49	4	31	10,500	E	10	120
	62-1	3	13:49:53	14:04:36	4	31	10,500	W	10	120
	63	4	14:06:54	14:21:54	4	31	10,500	E	10	120
	64	5	14:23:12	14:37:35	4	31	10,500	W	10	120
	65	6	14:39:09	14:52:30	4	31	10,500	E	10	120
	66	7	14:54:03	15:07:00	5.6	31	10,500	W	10	120
	67	8	15:09:00	15:21:00	5.6	31	10,500	E	10	120
	69	9	15:22:45	15:34:56	5.6	31	10,500	W	10	120
	68b	10	15:37:23	15:44:25	5.6	31	10,500	E	10	120
	61-1b	11	15:48:19	15:55:02	5.6	31	10,500	E	10	125

DATE	LINE	FILE	START TIME	STOP TIME	*f*STOP	INT TIME	ALTITUDE (FEET)	HEADING	VISIBILITY (MILES)	SPEED (MPH)
	59	12	15:58:43	16:16:05	5.6	31	10,500	W	10	125
	58	13	16:18:12	16:35:44	5.6	31	10,500	E	10	122

APPENDIX B
Protocol for Collecting Representative Ground Truth Land Covers for the Little Miami River Watershed 2002

B.1. INTRODUCTION

The purpose of this protocol is to obtain example landcovers within the Little Miami River Watershed to assist in the classification of remotely-sensed imagery during the period of July 1 to July 30, 2002.

B.2. EQUIPMENT NEEDED

1. Field sheets (see Attachment A)
2. Highway maps
3. Map of assigned random lat/long starting points
4. GPS (calibrated by GPS Coordinator beforehand)
5. Compass
6. Digital or 35mm camera
7. Range finder (optional)
8. Adequate disk space on digital camera or several rolls of 35mm film
9. New and/or fully charged batteries for GPS unit and camera
10. Copies of official letter for private landowners
11. Pencils
12. Tape measure

B.3. PROCEDURE

On the day before going out to the field, all GPS units and their respective manuals (if available) need to be temporarily given to the GPS coordinator. The GPS coordinator will standardize all the units with respect to the datum, spheroid, and the lat/long format to be used, and return them to the field crew the following day.

Using highway maps and GPS, proceed to first assigned random starting point (see Attachment B).

If a landcover of interest (see Section B.4 below) exists at the starting point, fill out your first record here.

From this point, ascertain how to travel and look for additional land covers within a mile radius of the starting point. This can be done by traveling in four different directions emanating from this point (e.g., N, S, E, W) or as best you can throughout this area.

While traveling throughout the area (i.e., within a mile radius of each starting point), locate, stop and fill out a record for as many unique landcovers as possible. **Use only one (1) field sheet per landcover found**. Once a particular landcover is recorded around a random starting point do not provide repeated records of this type of landcover unless it differs significantly (e.g., an agricultural crop at different stages of growth, or till or no-till, etc.) or unless you're beginning at a new random starting point within the watershed.

For each type of landcover found, ideally move to a point where you are surrounded by the landcover for 100m on all sides (approximately the length of a football field). If the GPS signal is poor (e.g., while located in a forest with dense canopy or an urban setting with tall buildings) or you are not able to gain access to a private site, attempt to take the coordinates of a point just outside of the landcover area and using a compass and range finder (if available) note on the field sheet where these GPS coordinates are with respect to the landcover being sampled (e.g., northeast corner of corn field). If a GPS reading would ideally be taken on private property, receive permission first.

Once in position:

1) set the GPS unit in a stationary position and allow the GPS to acquire a 3D fix (i.e., on at least 4 satellites) for at least 5 minutes before recording lat/long coordinates,
2) record each satellite # with a solid black bar fix (e.g., on the Garmin units), or otherwise the signal strength of each satellite being used by the GPS unit.
3) take GPS measurements at this single point for a minimum of three minutes, recording all latitude and longitude coordinates displayed by the GPS unit for up to twelve unique sets of coordinates.
4) take a picture or digital image of the landcover sampled, and using a compass, note on the field sheet the view and orientation of the picture or image (e.g., "corn plants 6 inches high taken from sample point in field looking NW," or "Geist reservoir taken from SE shore looking NE").
5) record how to cross-reference the camera images with the field sheets (note: if you run out of camera film or file space for a digital camera, please provide as detailed a description of the landcover and its adjacent surrounding area as possible).
6) fill out the remaining field sheet in as much detail as possible including information on any notable landmarks nearby and any characteristics about the landcover surface which might assist in classifying the remotely-sensed imagery later (e.g., is the terrain flat or hilly? What is the name/type, color, height, or texture of the vegetation or other surface sampled?).

B.4. LAND COVER CLASSES OF INTEREST AND THEIR DEFINITIONS (Modified from the NLCD Land Cover Classification System Key - Rev. July 20, 1999)

B.4.1. Land Cover Classification System Key.

- Water
- Developed
    - Concrete
    - Asphalt
    - Other construction material
    - Low Intensity Residential
    - High Intensity Residential
    - Commercial/Industrial/Transportation
- Barren
    - Bare Rock/Sand/Clay
    - Quarries/Strip Mines/Gravel Pits
    - Transitional
- Forested Upland
    - Deciduous Forest
    - Evergreen Forest
    - Mixed Forest
- Shrubland
- Non-natural Woody
    - Orchards/Vineyards/Other
- Herbaceous Upland
    - Grasslands/Herbaceous
- Herbaceous Planted/Cultivated
    - Pasture/Hay
    - Row Crops
        - Corn
        - Soybeans
    - Small Grains
        - Wheat
    - Fallow
    - Urban/Recreational Grasses
- Wetlands
    - Woody Wetlands
    - Emergent Herbaceous Wetlands

B.4.2. Land Cover Class Definitions.

Water. All areas of open water; typically 25% or greater cover of water (per pixel).

Developed. Generally defined as areas of intensive human use with a high concentration (30% or higher) of constructed materials (e.g., asphalt, concrete, buildings, glass and metal structures, etc.). If possible, look for and take coordinates in the center of large areas (100m by 100m) which are uniformly asphalt or concrete, or composed of other constructed materials. In addition, record points in the center of the following general areas too:

> *Low Intensity Residential* - Includes areas with a mixture of constructed materials and vegetation. Construction materials account for 30-80% of area. Vegetation may account for 20-70% of area. These areas most commonly include single-family housing units. Population densities will be lower than in high intensity residential areas.
>
> *High Intensity Residential* - Includes highly developed areas where people reside in high numbers. Examples include apartment complexes and row houses. Vegetation accounts for less than 20% of the cover. Constructed materials account for 80-100% of the cover.
>
> *Commercial/Industrial/Transportation* - Includes infrastructure (e.g., roads, railroads, etc.) and all highly developed areas (e.g., central business district, shopping or strip mall, warehouses, etc.) not classified as High Intensity Residential.

Barren area. Areas characterized by bare rock, gravel, sand, silt, clay, or other earthen material with little or no "green" vegetation present regardless of its inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the "green" vegetated categories; lichen cover may be extensive.

> *Bare Rock/Sand/Clay* - perennially barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, beaches, and other accumulations of earthen material.
>
> *Quarries/Strip Mines/Gravel Pits* - Areas of extractive mining activities with significant surface expression.
>
> *Transitional* - Areas of sparse vegetative cover (less than 25% of cover) that are dynamically changing from one land cover to another, often because of land use activities. Examples include forest clearcuts, a transition phase between forest and agricultural land, temporary clearing of vegetation, new construction sites, and changes due to natural causes (e.g., fire, flood, etc.).

Forested Upland. Areas characterized by tree cover (natural or semi-natural woody vegetation, generally greater than 6 meters tall); tree canopy accounts for 25-100% of the cover.

Deciduous Forest - Areas dominated by trees where 75% of more of the tree species shed foliage simultaneously in response to seasonal change.

Evergreen Forest - Areas dominated by trees where 75% of more of the tree species maintain their leaves year-round. Canopy is never without green foliage.

Mixed Forest - Areas dominated by trees where neither deciduous or evergreen species represent more than 75% of the cover present.

Shrubland. Areas characterized by natural or semi-natural woody vegetation with aerial stems, generally less than 6 meters tall, with individuals or clumps not touching to interlocking. Both evergreen and deciduous species of true shrubs, young trees, and trees or shrubs that are small or stunted because of environmental conditions are included.

Shrubland - Areas dominated by shrubs; shrub canopy accounts for 25-100% of the cover. Shrub cover is generally greater than 25% when tree cover is less than 25%. Shrub cover may be less than 25% in cases when the cover of other life forms (e.g., herbaceous or tree) is less than 25% and shrubs cover exceeds the cover of the other life forms.

Non-natural Woody. Areas dominated by non-natural woody vegetation; non-natural woody vegetative canopy accounts for 25-100% of the cover. The non-natural woody classification is subject to the availability of sufficient ancillary data to differentiate non-natural woody vegetation from natural woody vegetation.

Orchards/Vineyards/Other - Orchards, vineyards, groves, nurseries, ornamental horticultural areas or other areas planted or maintained for the production of fruits, nuts, berries, or ornamentals.

Herbaceous Upland. Upland areas characterized by natural or semi-natural herbaceous vegetation; herbaceous vegetation accounts for 75-100% of the cover.

Grasslands/Herbaceous - Areas dominated by upland grasses and forbs. In rare cases, herbaceous cover is less than 25%, but exceeds the combined cover of woody species present. These areas are *not subject to intensive management*, but they are often utilized for grazing.

Planted/Cultivated. Areas characterized by herbaceous vegetation that has been planted or is intensively managed for the production of food, feed, or iber; or is maintained in developed settings for specific purposes. Herbaceous vegetation accounts for 75-100% of the cover.

Pasture/Hay - Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops. Visual clues to these areas include grazing areas used by cows or horses, and fields of alfalfa hay in different stages of cutting.

Row Crops - Areas used for the production of crops, such as corn, soybeans, vegetables, and tobacco, or cotton. In the LMR, we are most likely to find corn and soybeans with some vegetables and tobacco.

Small Grains - Areas used for the production of graminoid crops such as wheat, barley, oats, or rice. In the LMR, we are most likely to find wheat. Oats may be a possibility, but this crop is more common in northerly to northeastern counties of Ohio.

Fallow - Areas used for the production of crops that are temporarily barren or with sparse vegetative cover as a result of being tilled in a management practice that incorporates prescribed alternation between cropping and tillage.

Urban/Recreational Grasses - Vegetation (primarily grasses) planted in developed settings for recreation, erosion control, or aesthetic purposes. Examples include parks, lawns, golf courses, airport grasses, and industrial site grasses.

Wetlands. Areas where the soil and substrate is periodically saturated with or covered with water as defined by Cowardin, et al.

Woody Wetlands - Areas where forest or shrubland vegetation accounts for 25-100% of the cover and the soil or substrate is periodically saturated with or covered with water.

Emergent Herbaceous Wetlands - Areas where perennial herbaceous vegetation accounts for 75-100% of the cover and the soil or substrate is periodically saturated with or covered with water.

Attachment A: Remote Sensing Classification Field Data Sheet
Little Miami River Watershed 2002

Today's Date:_______________ **Investigator(s)**:_______________________________

Random Starting Point _________, **Landcover#** __________, **Picture#** __________

3D fix? Yes / No **Satellite #s**:____

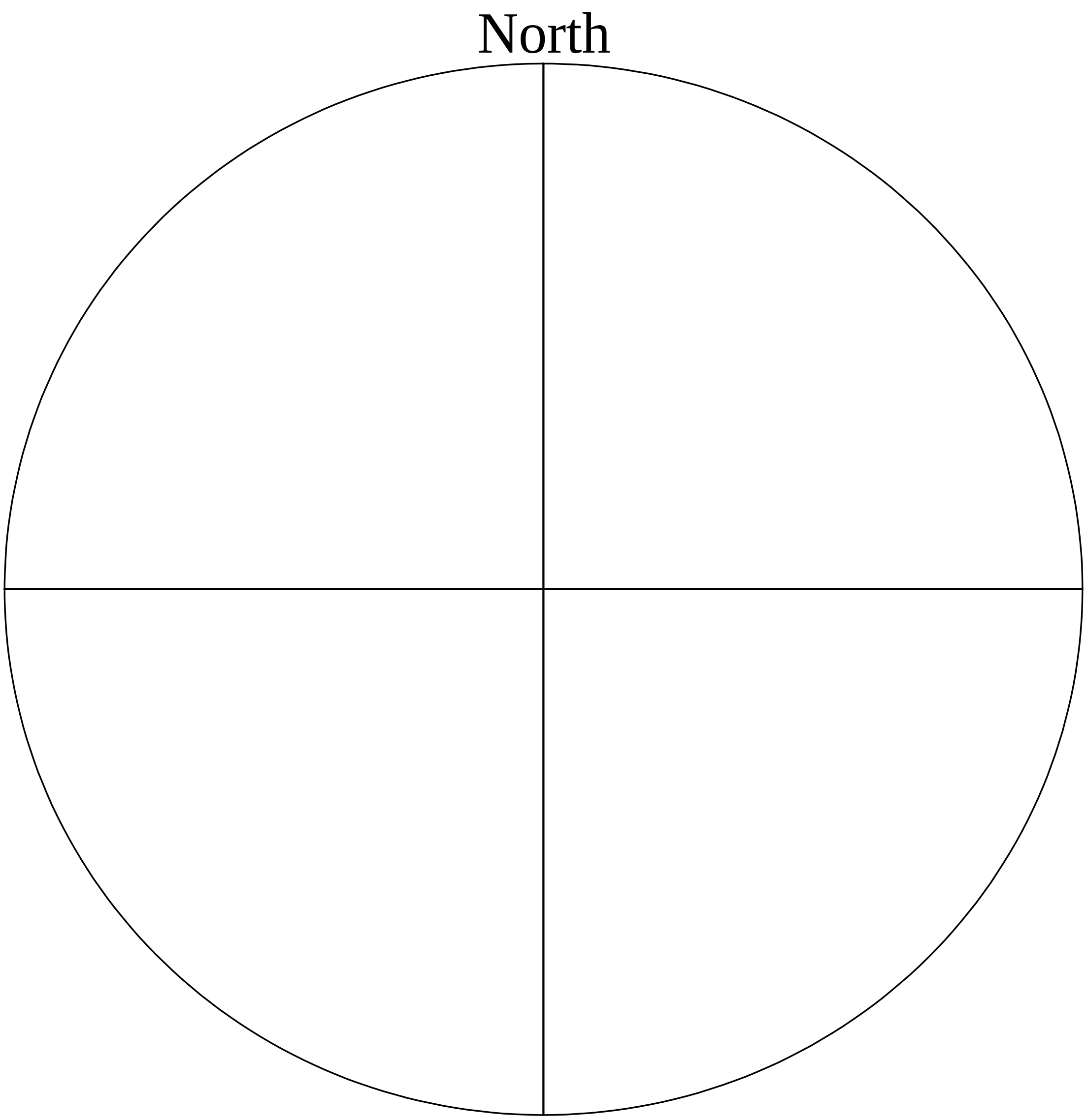
Displayed Accuracy? __________ **Begin/End Time** : ____________ ____________

Latitude:	**Longitude:**
_________________________	____________________________
_________________________	____________________________
_________________________	____________________________
_________________________	____________________________
_________________________	____________________________
_________________________	____________________________
_________________________	____________________________
_________________________	____________________________
_________________________	____________________________
_________________________	____________________________
_________________________	____________________________

Landcover Type (circle one):

Cement
Asphalt
Other construction material _________________
Low-density residential
High-density residential
Commercial/Industrial/Transportation
Other urban or built-up land: ________________
Corn
Soybeans
Wheat
Pasture/Hay (horse? cow? alfalfa hay? _________)
Fallow (i.e., croplands temporarily out of production)
Grass (residential, golf course, park, airport, erosion control, industrial, developed, or managed site)
Grasslands (primarily upland, natural or semi-natural, or not intensively managed)
Orchard, grove, vineyard, nursery, or ornamental horticultural area
Other crop or agricultural land: ________________

Open Water
Woody Wetland
Emergent Herbaceous Wetland
Barren areas:
BareRock/Sand/Clay
Quarry/Strip Mine/Gravel Pit
Transitional (clearcuts, clearings, burnt or flood area)
Deciduous Forest
Coniferous Forest
Mixed Forest
Shrubland



Radius from center (GPS location): ~100 meters

Detailed description of landcover type. For example, predominate species present and growth stage or height of vegetation or other structures? Dry or wet conditions? Dust or puddles present? Flat or hilly topography? Adjacent landcovers, roads or landmarks?

APPENDIX C
GENERAL RULES USED FOR OBTAINING TRAINING SET AND SUPPLEMENTAL GROUND TRUTH DATA

The following guide sets forth the rules to follow while creating polygons of land areas that are samples of a particular land cover/land use class. These sample areas should be as homogenous in pattern and color as possible; in short, the “best” or “cleanest” examples that can be found of that particular land cover/land use class.

General Rules:

1. Polygons should be no smaller than 5x5 pixels or 20m x 20 m, preferably larger. Exceptions would be narrow linear features such as roads and riparian forest. See the Key below for “Roads”.

2. Scale should be approximately 1:5000. Urban areas will likely require a larger scale.

3. Do not create a polygon that intersects or crosses a flight-line.

4. There should be at minimum a 10m border between the edge of the polygon and the outside edge of the land use being sampled. In the case of roads, the border should be at least 5 meters. The purpose of the border is to leave ample room for error.

5. Choose only areas that are uniform in color and pattern for a particular class.

APPENDIX D
METADATA

Metadata ("data about the data") created for this project was passed through the USGS's metadata-parser software to insure that the metadata files were error free. Metadata included in this appendix includes information about the image sources, dates, datums, projections, resampling algorithms, processing steps, file records, accuracy assessment, and other pertinent information associated with this geographic product. This metadata complies with Federal Executive Order 12906.

Sample Metadata of the 4-meter Classification of the Little Miami River Watershed

```
Identification_Information:
  Citation:
    Citation_Information:
      Originator: U.S. Environmental Protection Agency (EPA)
      Publication_Date: 2006
      Title: Little Miami River Watershed Hyperspectral Classification Dataset
      Geospatial_Data_Presentation_Form: Map
      Publication_Information:
        Publication_Place: EPA, Cincinnati, OH
        Publisher: EPA, Cincinnati, OH
      Online_Linkage: http://www.epa.gov
      Larger_Work_Citation:
        Citation_Information:
          Originator:
            This is a land cover classification based upon Compact Airborne
            Spectrographic Imager (CASI) data for the Little Miami River
            watershed, an EPA project.
          Publication_Date: 2006
          Title:
            Classification of High Spatial Resolution, Hyperspectral Remote
            Sensing Imagery of the Little Miami River Watershed in Southwest
            Ohio, USA
          Publication_Information:
            Publication_Place: EPA, Cincinnati, OH
            Publisher: EPA
          Other_Citation_Details:
            This classification is based on 82 CASI flight lines acquired on
            7/24/2002, 7/25/2002, 7/31/2002, 8/1/2002, 8/7/2002, 8/8/2002/,
            8/9/2002.

            Troyer, M.E., J. Heo and H. Ripley.  2006.  Classification of High Spatial
            Resolution, Hyperspectral Remote Sensing Imagery of the Little Miami
            River Watershed in Southwest Ohio, USA.  U.S. Environmental
            Protection Agency, Office of Research and Development, National Center
            for Environmental Assessment, Cincinnati, OH.
Description:
    Abstract:
      This is a final classification. This data set is the classification of the
      Little Miami River Watershed. This data set consists of 82 mosaicked
      flight lines that were analyzed according to the protocols as set forth
      in a Quality Assurance Project Plan (QAPP) to determine land use/land
      cover. It was created using CASI hyperspectral data by applying the object
      segmentation approach in eCognition (version 3.0) and the Spectral Angle
      Mapper (SAM) approach in ENVI (version 4.0). Then data were manually
      edited and mosaicked. This layer is also available in selected
      sub-watersheds.
    Purpose:
      To classify existing hyperspectral imagery and produce several land
```

```
      use/land cover products suitable for generating landscape metrics and
      analyses by the EPA.
  Time_Period_of_Content:
    Time_Period_Information:
      Range_of_Dates/Times:
        Beginning_Date: 20020724
        Ending_Date: 20020809
    Currentness_Reference: Dates of the CASI collection
  Status:
    Progress: Complete
    Maintenance_and_Update_Frequency: None
  Spatial_Domain:
    Bounding_Coordinates:
      West_Bounding_Coordinate: -84.507357
      East_Bounding_Coordinate: -83.547827
      North_Bounding_Coordinate: 39.991621
      South_Bounding_Coordinate: 38.874112
   Keywords:
    Theme:
      Theme_Keyword_Thesaurus:
      Theme_Keyword: Land Cover Analysis
      Theme_Keyword: Hyperspectral
    Place:
      Place_Keyword_Thesaurus:
      Place_Keyword: Little Miami River Watershed
      Place_Keyword: Ohio
  Access_Constraints: None
  Use_Constraints:
    Data set is not for use in litigation. While efforts have been
    made to ensure that these data are accurate and reliable within
    the state of the art, EPA, cannot assume liability for any
    damages, or misrepresentations, caused by any inaccuracies in the
    data, or as a result of the data to be used on a particular
    system. EPA makes no warranty, expressed or implied, nor does
    the fact of distribution constitute such a warranty.
  Native_Data_Set_Environment: Erdas Imagine signed 32bit integer (.img)
Data_Quality_Information:
  Attribute_Accuracy:
    Attribute_Accuracy_Report:
      According to accuracy assessment performed by Forest One,
      the overall accuracy is 83.92% and 81.99% Kappa.
      Each class accuracy is as follows: (Producers Accuracy/Users Accuracy)

      1: Lotic = 86.73%/98.84%
      2: Lentic = 70.83%/85.00%
      3: Forest = 95.00%/89.62%
      4: Corn = 88.33%/95.50%
      5: Soybean = 87.70%/90.68%
      6: Wheat = 47.37%/100.00%
      7: Dry Herbaceous = 82.00%/58.57%
      8: Grass = 75.00%/80.65%
      9: Urban Barren = 82.61%/90.48%
      10: Rural Barren = 64.52%/45.45%
      11: Urban/Built = 95.00%/92.23%

      The validation points were both assembled from EPA provided ground truth
      and interpreted in the lab using color aerial photographs.
      There were 902 points used for accuracy assessment total.
      Field collected validation points were collected concurrently
      (or nearly so) with the imagery in July and August of 2002. The field data
      and 2002 unrectified color aerial images were used for ground truthing of
      agricultural classes. 2003 high-resolution color aerial orthophotos were
      used for ground truthing of the rest of the classes: lotic, lentic,
      forest, grass, urban barren, rural barren, and urban/built.
  Logical_Consistency_Report:
    Tests for logical consistency indicate that all row and column
    positions in the selected latitude/longitude window contain data.
    Conversion and integration with vector files indicates that all
    positions are consistent with earth coordinates covering the same
    area. Attribute files are logically consistent.
```

```
Completeness_Report:
  Data exists for all classes.
  All pixels (other than background) have been classified.
Lineage:
  Source_Information:
    Source_Citation:
      Citation_Information:
        Originator: Hyperspectral Data International, Inc.
        Publication_Date: 20030428
        Title:
          High Resolution Remote Sensing of the Little Miami River
          Watershed in Southwest Ohio, USA
        Publication_Information:
          Publication_Place: NA
          Publisher: NA
        Online_Linkage: NA
    Type_of_Source_Media: DVD+R and CD-ROM
    Source_Time_Period_of_Content:
      Time_Period_Information:
        Range_of_Dates/Times:
          Beginning_Date: 20020724
          Ending_Date: 20020809
      Source_Currentness_Reference: Unknown
    Source_Citation_Abbreviation: NA
    Source_Contribution: NA
  Process_Step:
    Process_Description:
      This dataset was created by Forest One (Joon Heo - Principal
      Investigator, and Sitansu Pattnaik), Earth Satellite Corp.
      (François Smith, Christopher Jengo, Christopher Bolton, and
      Michael Diller),and Hyperspectral Data International, Inc.
      (Herbert Ripley, William Jones, Laura Roy, and Michelle Warr) under
      EPA purchase orders 3c-R337-NTSA, and 1C-R328-NALX. Project
      Officer: Dr. Michael E. Troyer. The study area is the Little
      Miami River Watershed in Southwest Ohio.

      Summary:
      This section outlines the classification procedure for the
      Little Miami River Watershed. Atmospheric correction and
      rectification were applied to the raw hyperspectral imagery.
      EPA field data, rectified aerial photography (Orthophotos),
      and unrectified aerial images were used for guidance
      in training point selection. The hyperspectral data were then
      segmented into three Level 1 classes (Urban, Rural, Water) in
      eCognition.  The training sets were used as inputs into Level
      2 classifications within the Urban and Rural Level 1 classes.
      Level 2 classification was performed with the Spectral
      Angle Mapper (SAM) method within ENVI. The resulting clusters
      of data were combined into categories related to the
      classification scheme. QA/QC was performed often resulting
      in a re-combination of clusters to better represent the final
      classes. Some categories were lumped to correspond to the final
      classification scheme. Data were classified by flight line.
      Then the data were manually edited and mosaicked. Then a final
      QA/QC was completed.

      Pre-Processing steps:
      The data received for this project were in a state ready
      for classification. Hyperspectral Data International, Inc
      performed atmospheric correction via Atmosphere Correction
      Now (ACORN) software at the time of data acquisition.
      Initial data examination by Forest One and Earth Satellite
      Corp. determined that cross-track illumination variations
      on the boundary could be handled by manual editing after
      classifications. Classification was performed on data
      geo-rectified using nearest neighbor resampling.

      Field-Collected Data:
      The EPA made available field data collected by EPA
      personnel during the time of image data acquisition in 2002.
```

The 390 points were used as a guide in selecting training sets for classification.

Classification:
Classification for this project was performed in a hierarchical manner. First, a Level 1 classification was performed. Training sets guided by EPA groundtruthing and high-resolution color aerial photography were selected for the Level 1 classes (Urban, Rural, Water). eCognition was used to perform the segmentation. QC work on the segmented image was performed to reduce the occurrence of obviously misclassified pixels. For instance, dark rooftops are often misclassified as water. The next step was to perform Level 2 classification for the Urban and Rural Level 1 classes. Training sets were chosen for the Level 2 classification. For Urban, these included Water, Grass, Urban Barren, Urban/built, and Forest. For Rural, these included Water, Corn, Soybean, Dry Herbaceous, Rural Barren, Urban/Built, and Forest.
Water and Forest were included in Level 2 classification in case any pixels remained improperly classified from Level 1. The Level 2 classification was performed using the Spectral Angle Mapper algorithm in ENVI. The Level 1 and Level 2 classes were then combined, and QA/QC was performed to locate any misclassified pixels. Also, water was differentiated to lotic and lentic at this point.
The final classes for this project are:
-99:Background
1: Lotic
2: Lentic
3: Forest
4: Corn
5: Soybean
6: Wheat
7: Dry Herbaceous
8: Grass
9: Urban Barren
10: Rural Barren
11: Urban/Built
99: Unclassified

Ancillary Datasets:
Non-CASI datasets used are 2003 high-resolution color aerial orthophotos from Aerials Express, acquired in July, 2003, and 2002 unrectified color aerial images from the Center for Mapping at the Ohio State University, The acquisition dates of 2002 images were between Aug. 30, 2002 and Sept. 13, 2002. EPA provided ground truth, 35 shape files of sub-basins as well as transportation and hydrographic vectors for the Little Miami River Watershed.

Post-Processing Steps:
The 82 flight lines were classified separately and then mosaicked to produce the land use/land cover for the entire watershed at 4m spatial resolution. The 4m dataset was aggregated to produce an additional 30m spatial resolution land use/land cover product too. Data products at 4m and 30m spatial resolution were also created for 35 selected sub-basins. Data were later subset by sub-watershed.
Process_Date: 20040215
Process_Contact:
Contact_Information:
Contact_Person_Primary:
Contact_Person: Dr. Michael E. Troyer
Contact_Organization: U.S. Environmental Protection Agency
Contact_Address:
Address_Type: mailing and physical address
Address: 26 West Martin Luther King Drive
City: Cincinnati
State_or_Province: OH
Postal_Code: 45268
Country: USA
Contact_Voice_Telephone: 513-569-7399
Contact_Electronic_Mail_Address: troyer.michael@epa.gov

```
Spatial_Reference_Information:
  Horizontal_Coordinate_System_Definition:
    Planar:
      Map_Projection:
        Map_Projection_Name: Albers Conical Equal Area
        Albers_Conical_Equal_Area:
          Standard_Parallel: 29.5
          Standard_Parallel: 45.5
          Longitude_of_Central_Meridian: 96 West
          Latitude_of_Projection_Origin: 23 North
          False_Easting: 0.00000
          False_Northing: 0.00000
      Planar_Coordinate_Information:
        Planar_Coordinate_Encoding_Method: Row and column
        Coordinate_Representation:
          Abscissa_Resolution: 4 meters
          Ordinate_Resolution: 4 meters
        Planar_Distance_Units: Meters
    Geodetic_Model:
      Horizontal_Datum_Name: North American Datum 1983
      Ellipsoid_Name: GRS80
      Semi-major_Axis: 6378137.0
      Denominator_of_Flattening_Ratio: 298.257
Entity_and_Attribute_Information:
  Detailed_Description:
    Entity_Type:
      Entity_Type_Label: Little Miami River Watershed, Southwest Ohio, USA
      Entity_Type_Definition:
        Little Miami River Watershed as delineated by EPA
        and extent of imagery
      Entity_Type_Definition_Source: EPA
    Attribute:
      Attribute_Label: Land Cover Classification
      Attribute_Definition: Land Cover Classification as determined by EPA
      Attribute_Definition_Source: EPA
      Attribute_Domain_Values:

        Enumerated_Domain:
          Enumerated_Domain_Value: -99 Background
          Enumerated_Domain_Value_Definition:
            This class contains no data due to data voids.
          Enumerated_Domain_Value_Definition_Source: EPA

        Enumerated_Domain:
          Enumerated_Domain_Value: 1 Water - Lotic
          Enumerated_Domain_Value_Definition:
            Open water associated with running water system, such as a river
            or stream.  Such waterways typically have a defined channel
            and an associated floodplain.
          Enumerated_Domain_Value_Definition_Source: EPA

        Enumerated_Domain:
          Enumerated_Domain_Value: 2 Water - Lentic
          Enumerated_Domain_Value_Definition:
            Open water associated with still water system, such as lakes,
            reservoirs, potholes, and stock ponds.  Such bodies typically
            do not have a defined channel or associated floodplain.
          Enumerated_Domain_Value_Definition_Source: EPA
        Enumerated_Domain:
          Enumerated_Domain_Value: 3 Forest
          Enumerated_Domain_Value_Definition:
            Contains either or both deciduous and coniferous trees in any degree
            of mixture, single stemmed, woody vegetation with canopy spanning
            greater than 4m and tree canopy accounting for 25-100 percent of
            the cover.
          Enumerated_Domain_Value_Definition_Source: EPA

        Enumerated_Domain:
          Enumerated_Domain_Value: 4 Corn
          Enumerated_Domain_Value_Definition:
```

```
    Area under cultivation of food and fiber,
    where corn is the primary crop.
  Enumerated_Domain_Value_Definition_Source: EPA

Enumerated_Domain:
  Enumerated_Domain_Value: 5 Soybean
  Enumerated_Domain_Value_Definition:
    Area under cultivation of food and fiber where soybeans are the
    primary crop.
  Enumerated_Domain_Value_Definition_Source: EPA

Enumerated_Domain:
  Enumerated_Domain_Value: 6 Wheat
  Enumerated_Domain_Value_Definition:
    Area under cultivation of food and fiber where wheat is the primary
    crop.
  Enumerated_Domain_Value_Definition_Source: EPA

Enumerated_Domain:
  Enumerated_Domain_Value: 7 Dry Herbaceous
  Enumerated_Domain_Value_Definition:
    Dominated by dry and/or less vigorous herbaceous types of
    vegetation; herbaceous vegetation accounts for no less than 25%
    of the cover. This class mainly includes naturally occurring and
    unmanaged herbaceous vegetation, and dried out, unhealthy, or
    stressed crop.  Dry Herbaceous vegetation prevailed in crop fields
    as well as natural fields, due to a high degree of drought in the
    summer of 2002.  These dry herbaceous types of vegetations had little
    chlorophyll content and very similar spectral signatures.  Not
    enough variation was present in the spectral signatures to further
    classify this class into different vegetative species.
  Enumerated_Domain_Value_Definition_Source: EPA

Enumerated_Domain:
  Enumerated_Domain_Value: 8 Grass
  Enumerated_Domain_Value_Definition:
    Dominated by cultivated grasses planted in developed settings for
    recreation, erosion control, or aesthetic purposes. Examples include
    well-watered parks, lawns, golf courses, airport grasses, and
    industrial site grasses.
  Enumerated_Domain_Value_Definition_Source: EPA

Enumerated_Domain:
  Enumerated_Domain_Value: 9 Urban Barren
  Enumerated_Domain_Value_Definition:
    Composed of bare soil, rock, sand, silt, gravel, or other earthen
    material with little (less than 25%) or no vegetation within urban
    areas.  Examples in this class include exposed soil in urban areas
    and construction sites.
  Enumerated_Domain_Value_Definition_Source: EPA

Enumerated_Domain:
  Enumerated_Domain_Value: 10 Rural Barren
  Enumerated_Domain_Value_Definition:
    Composed of bare soil, rock, sand, silt, gravel, or other earthen
    material with little (less than 25%) or no vegetation in rural areas.
  Enumerated_Domain_Value_Definition_Source: EPA

Enumerated_Domain:
  Enumerated_Domain_Value: 11 Urban/Built
  Enumerated_Domain_Value_Definition:
    Areas covered by structures and impervious surfaces in urban,
    suburban, and rural areas.  Buildings, parking lots, and
    paved roads typically fall into this class.

  Enumerated_Domain_Value_Definition_Source: EPA

Enumerated_Domain:
  Enumerated_Domain_Value: 99 Unclassified
```

```
          Enumerated_Domain_Value_Definition:
            This class includes areas of image gaps among flight lines and
            cloud cover where land cover classification is impossible.
          Enumerated_Domain_Value_Definition_Source: EPA
Distribution_Information:
  Distributor:
    Contact_Information:
      Contact_Organization_Primary:
        Contact_Organization:
          National Center for Environmental Assessment, Office of Research
          and Development, U.S. Environmental Protection Agency
        Contact_Person: Dr. Michael E. Troyer
      Contact_Address:
        Address_Type: mailing and physical address
        Address: 26 West Martin Luther King Drive.
        City: Cincinnati
        State_or_Province: OH
        Postal_Code: 45268
        Country: USA
      Contact_Voice_Telephone: 513-569-7399
      Contact_Electronic_Mail_Address: troyer.michael@epa.gov
  Resource_Description: Little Miami River Watershed Hyperspectral Classification
  Distribution_Liability: NA
  Standard_Order_Process:
    Digital_Form:
      Digital_Transfer_Information:
        Format_Name: Erdas Imagine signed 32bit integer (.img)
      Digital_Transfer_Option:
        Offline_Option:
          Offline_Media: DVD or CD-ROM
          Recording_Format: ISO 9660
          Compatibility_Information:
            ISO 9660 format allows the media to be read
            by most computer operating systems.
    Fees: NA
Metadata_Reference_Information:
  Metadata_Date: 20040217
  Metadata_Review_Date: 20060330
  Metadata_Contact:
    Contact_Information:
      Contact_Organization_Primary:
        Contact_Organization:
          National Center for Environmental Assessment, Office of Research
          and Development, U.S. Environmental Protection Agency
        Contact_Person: Dr. Michael E. Troyer
      Contact_Address:
        Address_Type: mailing and physical address
        Address: 26 West Martin Luther King Drive
        City: Cincinnati
        State_or_Province: OH
        Postal_Code: 45268
        Country: USA
      Contact_Voice_Telephone: 513-569-7399
      Contact_Electronic_Mail_Address: troyer.michael@epa.gov
  Metadata_Standard_Name:
    FGDC (Federal Geographic Data Committee)
    CSDGM (Content Standard for Digital Geospatial Metadata)
  Metadata_Standard_Version: FGDC-STD-001-1998
```