View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Illinois Digital Environment for Access to Learning and Scholarship Repository

SURFACE WATER SECTION
FILE COPY

Illinois State Water Survey Division

SURFACE WATER SECTION

SWS Contract Report 458

LAND USE/COVER CLASSIFICATION FOR THE PROPOSED SUPERCONDUCTING SUPER COLLIDER STUDY AREA, NORTHEASTERN ILLINOIS

by Robin B. King, Ming T. Lee, and Krishan P. Singh

Prepared for the
Illinois Department of Energy and Natural Resources

Champaign, Illinois

May 1989

ENR Illinois Department of Energy and Natural Resources

SURFACE WATER SECTION
FILE COPY.

LAND USE/COVER CLASSIFICATION FOR THE PROPOSED SUPERCONDUCTING SUPER COLLIDER STUDY AREA, NORTHEASTERN ILLINOIS

by

Robin B. King
Assistant Supportive Scientist

Ming T. Lee
Professional Scientist

Krishan P. Singh
Principal Scientist

Illinois State Water Survey
2204 Griffith Drive
Champaign, IL 61820-7495

CONTENTS

SURFACE WATER SECTION
FILE COPY

LAND USE/COVER CLASSIFICATION FOR THE PROPOSED SUPERCONDUCTING SUPER COLLIDER STUDY AREA, NORTHEASTERN ILLINOIS

by
Robin B. King, Ming T. Lee, and Krishan P. Singh

INTRODUCTION

On September 2, 1987, in response to a Request for Proposal (RFP) from the U.S. Department of Energy, the state of Illinois submitted a proposal for the construction of a Superconducting Super Collider (SSC) in northeastern Illinois. The SSC is estimated to cost $4.5 billion. It will consist of an elliptical tunnel, 53 miles long and from 10 to 12 feet in diameter, located about 400 feet below the ground. It will accelerate two proton beams to the speed of light, each with 20 trillion electron volts (20-TeV) of energy. Their collision will release 40-TeV of energy. The SSC will ensure the nation's lead in high-energy physics. If it were located in Illinois, it would be a natural extension of the present accelerator at Fermi National Laboratory in Batavia, which is on the eastern edge of the proposed SSC ring. The DOE announced on December 27, 1987 that the Illinois SSC site was on the "best qualified" list of seven sites.

The Illinois Department of Energy and Natural Resources and its divisions conducted a program to develop information pertinent to natural and man-made features within the 36-township region containing the SSC study area (Figure 1). The proposed SSC ring lies within 16 townships (T37N to T40N and R6E to R9E).

This report describes the method used in creating a land use/cover map for the 36-township area from June 1985 satellite imagery, the hardware and software used for image processing, and the results of the land use/cover classification. The land use/cover classification and delineation provide basic spatial information on water bodies, vegetation, and other environmental factors.

The objective of spectral image classification is to translate the raw spectral characteristics of the observed feature (generally a ground scene) into discrete categories that are of interest to the user. Prior to a discussion of image processing strategies, it is necessary to define a few key terms (from Jensen, 1986):

Brightness value: A digital value that represents the amount of reflected or emitted energy that exits the earth's surface. The greater the brightness of the scene, the larger the digital value.

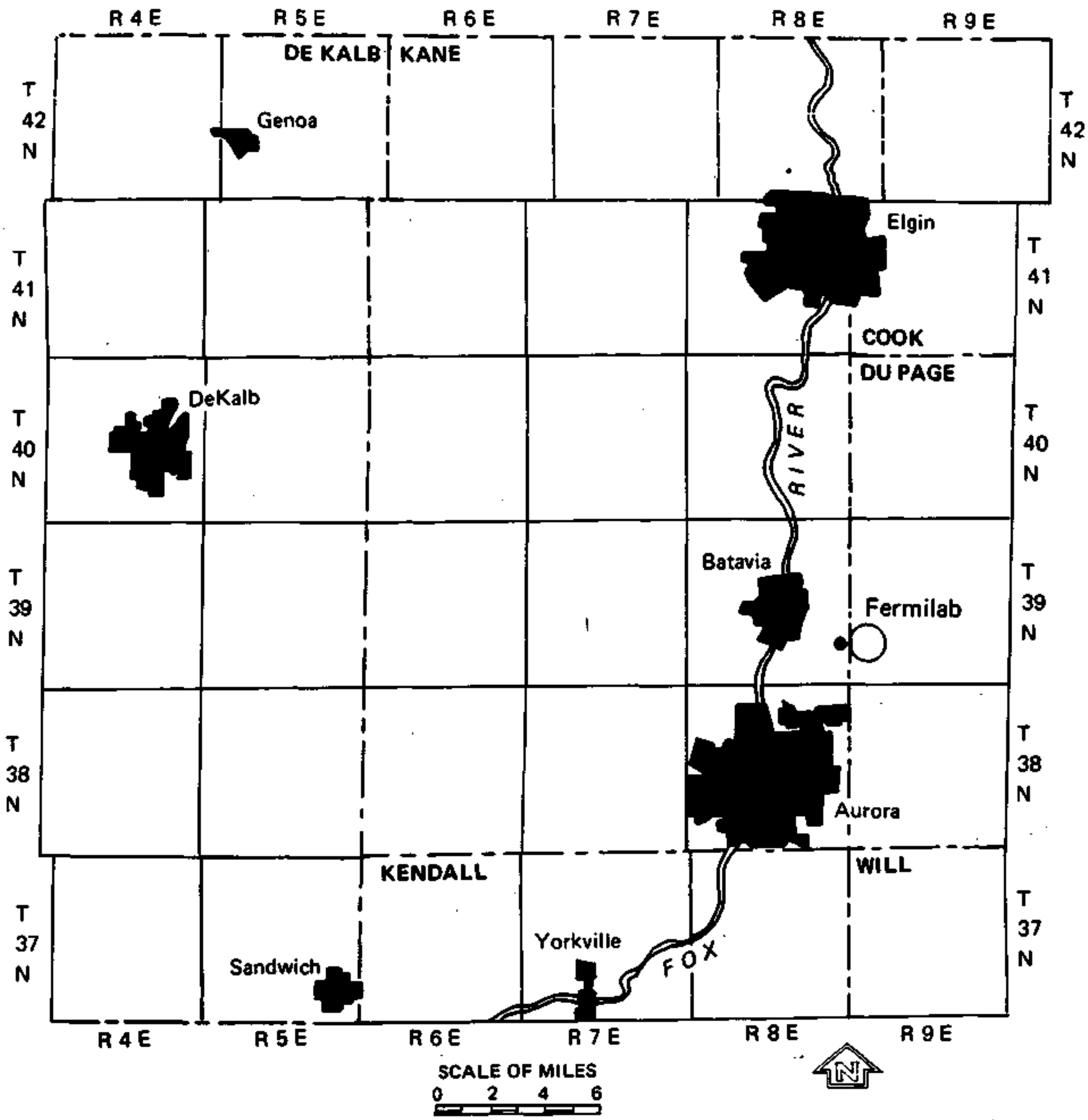


Figure 1. Area of Investigation for Siting the Accelerator Ring

Class: A surface characteristic type such as forest or water that is of interest to the investigator.

Classification: The process of assigning individual pixels of a multispectral image (see definition below) to discrete categories.

Clustering: The statistical analysis of a set of pixels to detect their inherent tendency to form clusters in a multidimensional data space.

Geographic Information System (GIS): A computer hardware and software system designed to collect, manage, analyze, and display spatially referenced data.

Pixel: A picture element having both spatial and spectral properties. The spatial variable defines the apparent size of the resolution cell, and the spectral variable defines the intensity of the spectral response for that cell in a particular band.

Spatial resolution: The ability of an entire remote sensor system to render a sharply defined image. Also, a measure of the smallest angular or linear separation between two objects that can be resolved by the sensor.

Spectral resolution: The dimension and number of specific wave length intervals in the electro-magnetic spectrum to which a sensor is sensitive.

Spectral signature: Term referring to the spectral characteristic of an object in a scene. Implies that each object reflects radiation in a unique and identifiable manner.

Training: The process of informing an image processor which sites to analyze for spectral properties as a prerequisite to a supervised classification (see definition in the "Background Information" section).

Training site: Recognizable area on an image with distinct spectral properties useful for identifying other similar areas.

Acknowledgements

This study was funded by the Illinois Department of Energy and Natural Resources as part of an extensive project. Becky Howard typed the camera-ready copy of the report, and Gail Taylor edited the report.

BACKGROUND INFORMATION

The staffs of the four Divisions of the Illinois Department of Energy and Natural Resources published an environmental screening atlas for siting the SSC in northeastern Illinois (Hines, 1986). The atls includes information on: 1) the physical setting of the study area, the model ring, topography, soils, distribution and thickness of quaternary deposits, sand-and-gravel aquifers, bedrock geology, structural features, seismic risk, and flood hazard areas; 2) conservation and preservation of threatened and endangered species, wetlands and water bodies, natural areas, known paleontological sites, paleontological site probability, known archaeological sites, archaeological site probability, historic site probability, national register sites, and Illinois historic landmarks; 3) cultural features, including land use, population density, land parcel size, public lands, quarries, electric transmission lines, oil and gas pipelines, highways, and railroads; and 4) wells and their depths, borings, samples, and cores.

Environmental evaluations require the preparation of land use information. This study focuses on the development of a land use/cover map from recent satellite imagery. Similar during-construction and post-construction maps allow monitoring of the changes taking place and provide help in decisions regarding whether any remedial measures are needed. To detect the areas of rapid cover changes and the level of such changes, the land use/cover map and data developed here can be compared with those on existing U.S. Geological Survey land use digital analysis (LUDA) maps.

Landsat Digital Data

The earliest origins of remote sensing from space can be traced to the latter part of the nineteenth century, when German-built rocket-propelled camera systems took photographs of the earth's surface. Later, small cameras were mounted aboard captured V-2 rockets and fired from the White Sands Proving Ground in New Mexico. Although the early photographs were crude and primitive, they demonstrated the potential value of remote sensing from space. This value became more apparent during the manned Mercury, Gemini, and Apollo space missions of the 1960s. Numerous high-quality color photographs were taken during these missions, and the value of remote sensing became well recognized.

In 1967, the National Aeronautics and Space Administration, with the cooperation of the U.S. Department of the Interior, began a feasibility study of a sequence of Earth Resources Technology Satellites (ERTS). The study proposed a series of satellite launches designated as ERTS-A, -B, -C, -D, -E, and -F. After successful launch and orbital positioning, the satellites were to be redesignated as ERTS-1, -2, -3, -4, -5, and -6. ERTS-1 was successfully launched on July 23, 1972, and operated until January 6, 1978. All the nations of the world were invited to evaluate the ERTS-1 data. The results from the scientific community pointed to a number of valuable applications of remote sensing in various scientific research missions. On January 22, 1975 (just prior to the launch of ERTS-B), NASA officially renamed the ERTS program the LANDSAT (Land Satellite) program and retroactively changed the name ERTS-1 to LANDSAT-1, or Landsat-1. By 1988, a total of five Landsat satellites had been launched, and the program had evolved into an operational global resource monitoring program. All the Landsat satellites carry 4-band multispectral scanners (MSS) that are sensitive to wavelengths from 0.5 mm to 1.1 mm and have a spatial resolution of from 79 to 82 meters (Table 1). Landsat-1, -2, and -3 also carried a three-channel return beam vidicon (RBV) system that provided an instantaneous camera-like view of the ground scene. The resolution and imaging capability of the onboard sensors was considerably improved on Landsat-4 and -5. The most significant change was the introduction of a 30-meter-resolution Thematic Mapper (TM), which provides relatively high-resolution data in seven different spectral bands.

Table 1. Characteristics of Landsat Multispectral Scanners (MSS) and Return Beam Vidicon (RBV) Systems

Sensor	Landsat mission	Sensitivity (mm)	Resolution (m)
MSS	1-5	0.5 - 0.6	79/82*
		0.6- 0.7	79/82
		0.7 - 0.8	79/82
		0.8- 1.1	79/82
	3	10.4 -12.6**	240
RBV	1-2	0.475-0.575	80
		0.580 - 0.680	80
		0.690 - 0.830	80
	3	0.505 - 0.750	30

* 79 m, Landsat-1 to -3; 82 m, Landsat-4 and -5
** Landsat-3 only, sensor failed shortly after launch

The Landsat series of satellites are in a circular, sun-synchronous, near-polar orbit at a mean altitude of 570 miles (440 miles for Landsat-4 and -5). Each orbit takes 103 minutes (99 minutes for Landsat-4 and -5), so that each satellite completes approximately 14.5 orbits per day, covering the same spot on the earth every 18 days (16 days for Landsat-4 and -5).

The MSS on board the Landsat satellite is a line-scanning device that continuously scans a 115-mile (185-kilometer) swath along the earth's surface. The image data are separated into individual frames, 115 miles square, during processing. Each image consists of many individual picture elements (pixels) that are obtained in rapid succession by means of an oscillating mirror behind the lens of the MSS. The oscillating mirror scans six scan lines along the 115-mile-long swath perpendicular to the spacecraft. The MSS simultaneously records (in four spectral bands) the amount of light being reflected from each pixel, a 259-foot-square area of the earth's surface.

The MSS video signal is converted to digital data and telemetered to a ground receiving station, either in real time or after being recorded on board the satellite. The final data products include computer-compatible tapes (CCTs), black and white photographs of individual spectral bands (MSS bands 4-7), and color composites comprising several bands, usually MSS bands 4, 5, and 7.

The Landsat digital data can be spatially located on the ground to within one-half of a pixel (Bernstein and Ferneyhough, 1975). One advantage of using the Landsat information is that the data are digital and can be directly entered into a geographic data base. One Landsat scene comprises an area 115 miles square. Also, because Landsat views the same ground point every 16 or 18 days, the accuracy of the land use map can be increased by analyzing data obtained at different times of the year.

Thematic Mapper (TM) Data

TM is an advanced multispectral scanner. This name relates to the intended application of the system's data to spectral pattern recognition techniques that will produce classified images (thematic maps). TM is equipped with a seven-channel scanner designed to maximize vegetative analysis capabilities for agricultural applications. The TM bands and their resolutions are listed in Table 2.

Table 2. Characteristics of Thematic Mapper

Band	Wavelength (mm)	Resolution (m)	Image format
1	0.45-0.52	30	115-mile (185-km)
2	0.52-0.60	30	strip image framed
3	0.63-0.69	30	with 5.4% forward
4	0.76-0.90	30	lap, 7.3% sidelap
5	1.55-1.75	30	at the equator,
6	10.40-12.50	120	increasing toward
7	2.08-2.35	30	poles

Digital Imagery Used for Study Area

The study area is defined as the 36 townships in Cook, DuPage, Will, Kendall, Kane, and DeKalb Counties as shown in Figure 1. The TM scene of path 23, row 31 on June 3, 1985, was purchased from the Earth Observation Satellite Company (EOSAT), a commercial company designated by the U.S. government to sell the Landsat data. The image was selected because of the low cloud coverage and the stage of vegetative coverage on the ground.

Hardware and Software of Image Analysis System

The ERDAS (ERDAS, Inc., 1988) Image Processing System on the IBM-AT was used in this study. The system hardware consists of: 1) IBM-AT system with Enhanced Graphics Adapter and 30 megabyte (MB) fixed disk, 2) Cipher Data Products 9000 series magnetic tape drive, 3) 20 Plus 20 MB Bernoulli removable fixed disk, 4) Tektronix 4696 color dot matrix printer, and 5) Mitsubishi 512 by 512 high-resolution color display. Figure 2 shows the flow diagram of the hardware of this image processing system.

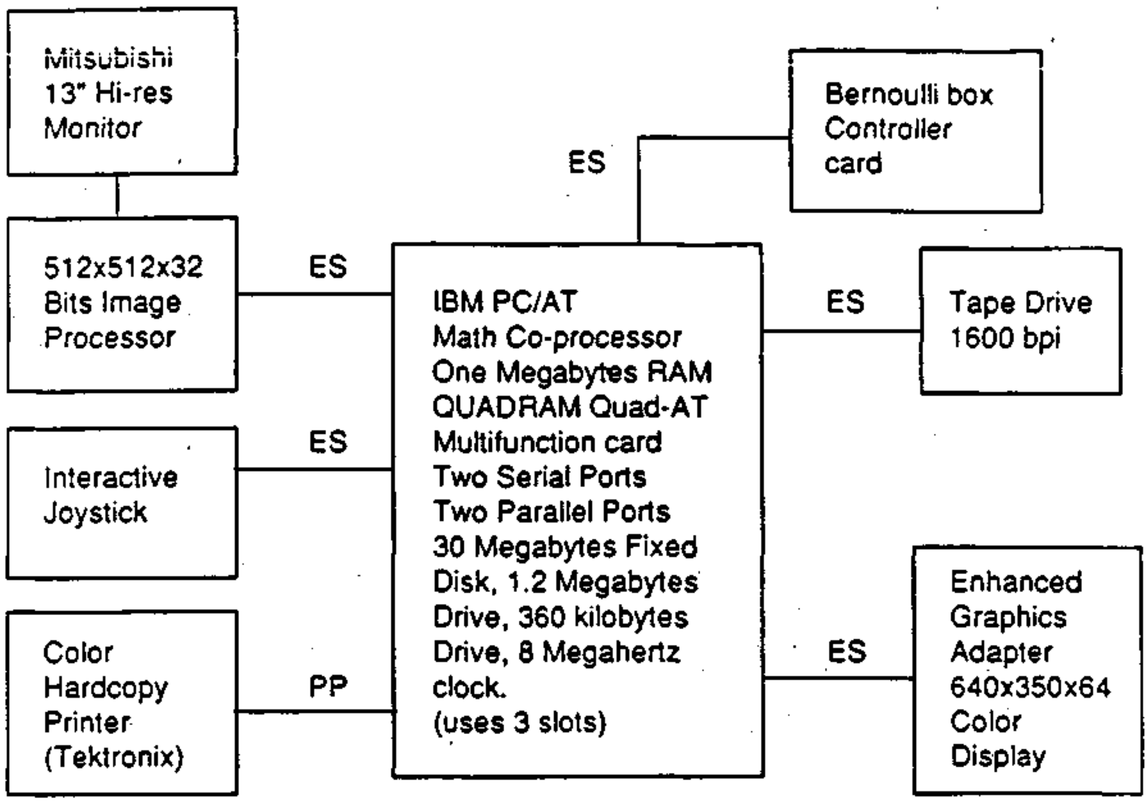
. The ERDAS software for the image analysis system includes the following modules: 1) core module, 2) image processing module, 3) geographic information system (GIS) module, 4) tapes module, 5) color hardcopy module, 6) topographic module, 7) and toolkit module. The software system is driven by the MS-DOS operation system.

Land Use/Cover Classification; Unsupervised and Supervised Classifications

In a general sense, the overall objective of image classification is to categorize all the individual elements of a digital image into separate groups or classes that are in some way unique. This objective can be achieved through various methods, and in fact a whole family of image classification strategies exists. Spectrally oriented classification strategies are the dominant type in use. Two spectral classification strategies, the supervised analysis and the unsupervised analysis, form the backbone of image classification activities (Lillesand and Kiefer, 1987).

An unsupervised classification identifies spectrally similar groups or clusters within the multispectral data. Remotely sensed images are usually composed of spectral clusters that are reasonably uniform with respect to brightness in several spectral bands. Although the existence of the clusters (or classes) may not be intuitively obvious, unsupervised classification is defined as the definition, identification, labeling, and mapping of these natural or spectral classes.

A supervised classification entails identifying a *training area* consisting of sample pixels that belong to a known informational class. The spectral characteristics of the training area are computed statistically. The classification algorithm is *trained* to categorize sample pixels on the basis of their statistical similarity to the training area characteristics. The pixels



ES - Takes an expansion slot
SP - Serial port
PP - Parallel port

Figure 2. ERDAS Image Processing System

of the image arc assigned to the informational class that most closely resembles the spectral signature of the sample pixels.

Unsupervised classification has several advantages and disadvantages relative to supervised classification. The advantages are:

a) No extensive *a priori* or prior knowledge of the ground scene is required. However, knowledge of the scene is required to interpret and verify the final results of the classification process.
b) The opportunity for human error is minimized. If inaccurate preconceptions exist regarding the ground scene, they have very little influence on the classification.
c) Identified classes are very uniform with respect to spectral composition.
d) Unique classes are recognized as distinct units. Unsupervised classification will ensure recognition of a class very small in areal extent and will usually prevent it from accidental incorporation into another class.

Unsupervised classification has certain disadvantages and limitations that arise primarily from the reliance upon natural clustering. The disadvantages arc:

a) The procedure identifies spectrally homogeneous classes within the data. These classes do not necessarily correspond to the informational categories that are of interest to the user.
b) There is either limited control or no control over the resultant menu of classes and their identities. The user has very little control over the natural spectral clustering results.
c) Spectral properties of informational classes change over time. Thus relationships between spectral classes and informational classes are not constant, and relationships defined for certain informational classes may not be the same as for other images for the same area.

Conversely, supervised classification holds certain advantages and disadvantages relative to unsupervised classification. The advantages are:

a) The user has complete control over a selected menu of informational classes. These classes can be tailored to a specific application and geographic region.
b) The resultant informational classes are tied to specific areas of known identity. This obviates the problem of matching spectral classes in the final map with useful informational classes.
c) Serious classification errors are easily detected if the training data are examined to determine if they have been correctly classified. Although this alone will not ensure correct classification of the other (non-training) data, it is a reliable assessment of how accurate the training site data are.

The disadvantages are:

a) A classification structure is imposed on the data. The user-defined classes may not match the natural classes within the data that emerge in multidimensional data space.

b) Often training data are defined on the basis of a known informational category and only secondarily on the spectral uniqueness of that category. Some informational categories may share many spectral properties with one another, leading to obvious classification errors.

c) Selecting an acceptable training area can be time-consuming, tedious, and expensive. Even if adequate training data are available, the supervised classification may not recognize special or unique categories because the classifier may not be able to identify them sufficiently.

Thus the obvious question emerges: What classification strategy should the analyst use: supervised or unsupervised? In a general sense, this depends upon the extent of prior knowledge of the ground scene, the amount of correlation between natural spectral classes and desired informational classes, and the analyst's prior experience.

Inherent Difficulties and Uncertainties

Digital image analyses for land use/cover classifications are subject to numerous errors. The image processes using the statistical method are based upon distinctions in spectral reflectance. The land use/cover classifications are artificial and arbitrary because there are no exact signature analogs. For example, forests and parks are considered different land uses, but they may be similiar in spectral reflectance in satellite images. Secondly, the classifications are less complete than human land use classification, which is based on the context, shape, texture, and color of the objects.

Because the signature overlapping of the training sites is unavoidable in the real-world image data, the possibility of misclassification always exists as far as each individual pixel is concerned. The regional statistics in terms of total areas of each class will be better than those derived from pixel-by-pixel comparison. Therefore the final product is best for use in regional analysis rather than in site-specific land use/cover analysis. This study used digital image analysis as a time-saving, cost-effective tool to produce a regional land use/cover map as a supplement to the existing LUDA data.

METHODOLOGY USED

The study area for the Superconducting Super Collider in northeast Illinois is a region of varied land use and land cover. Although the area is dominated by agricultural activity, the eastern part of the study area is presently undergoing rapid urbanization as the sprawling suburban area of Chicago continually pushes westward. A map of the study area that categorizes major land cover features has been produced and is included in this report as Figure 4. The map was created at classification level I of the U.S. Geological Survey's land use and land cover classification system (Table 3). A supervised classification strategy was used.

There was considerable prior knowledge of land use and land cover patterns in the study area. Much of this knowledge is based on relatively old LUDA data created in the early to middle 1970s (U.S. Geological Survey, 1973, 1976); nonetheless, the LUDA data were quite useful in examining the overall patterns and areal extent of the land use and land cover categories.

Level I land use and land cover consists of nine categories, but information from the LUDA maps indicated that only six of these categories (urban, agricultural, forest, water, wetland, and barren) occur in the SSC study area. The general procedure used in producing the land cover map is listed below, and each step of this procedure is then examined in greater detail.

- Load the relevant multi-channel TM (Thematic Mapper) data into the PC and rectify the image with a ground-true map coordinate system. These TM data have a resolution of 30 meters.
- Determine the categories (classes) to be mapped and prepare for a supervised analysis by selecting training sites that give unique spectral definition to the desired classes.
- Perform the actual supervised analysis by using a maximum-likelihood classification algorithm, and produce a GIS-based land cover map.
- Assess the accuracy of the map for site-specific locations.
- Produce a final land cover map in hard copy form.

Rectification of Image

A full Landsat TM scene comes as a set of twelve CCTs (computer-compatible tapes). The full scene covers a ground area of approximately 115 by 115 miles (185 by 185 kilometers). Each scene is divided into four quarters, each of which requires three CCTs. TM bands 1 and 2 are stored on the first tape, bands 3 and 4 on the second tape, and bands 5 through 7 on the third tape.

Table 3. U.S. Geological Survey Land Use/Land Cover Classification System for Use with Remote Sensor Data

Level I	Level II
1 Urban or built-up land	11 Residential
	12 Commercial and services
	13 Industrial
	14 Transportation, communications, and services
	15 Industrial and commercial complexes
	16 Mixed urban or built-up land
	17 Other urban or built-up land
2 Agricultural land	21 Cropland and pasture
	22 Orchards, groves, vineyards, nurseries, and ornamental horticultural areas
	23 Confined feeding operations
	24 Other agricultural land
3 Rangeland	31 Herbaceous rangeland
	32 Shrub and brush rangeland
	33 Mixed rangeland
4 Forest land	41 Deciduous forest land
	42 Evergreen forest land
	43 Mixed forest land
5 Water	51 Streams and canals
	52 Lakes
	53 Reservoirs
	54 Bays and estuaries
6 Wetland	61 Forested wetland
	62 Nonforested wetland
7 Barren land	71 Dry salt flats
	72 Beaches
	73 Sandy areas other than beaches
	74 Bare exposed rocks
	75 Strip mines, quarries, and gravel pits
	76 Transitional areas
	77 Mixed barren land
8 Tundra	81 Shrub and brush tundra
	82 Herbaceous tundra
	83 Bare ground
	84 Mixed tundra
9 Perennial snow and ice	91 Perennial snow fields
	92 Glaciers

Source: Anderson et al. (1976)

The repetitive nature of the Landsat orbits and the continuous stream of data from the sensors led to the development of an easy-to-use coordinate system, which is referred to as the Worldwide Reference System (WRS). The 233 ground tracks of Landsat 4 and 5 define the paths in the WRS, and each path contains 248 rows that correspond to latitude lines. The nominal center of the scene was determined by using the WRS, and the approximate coordinates of the SSC study area were computed by counting pixels (picture element, one pixel = 30 x 30 meters) from the scene center until the study area was framed. The TM data were loaded into the PC from the CCTs in a Band Sequential (BSQ) format (Band 1, Band 2, ..., Band 7). The bands were stacked or layered one band at a time, creating an image file in Band Interleaved by Line (BIL) format.

Landsat TM data require considerable computer memory space (one full scene alone consumes over 330 megabytes). Because of memory size limitations on the image processing system, it was necessary to divide the SSC study area into three subscenes (fractions of the whole scene area) and to process each one independently. Subscene sizes ranged from 10 to 14 megabytes and were easily managed by the PC. Thus the entire classification process (from raw digital data to the finished product) was performed separately for each subscene, and the final land cover map is an aggregate of these three subscenes. Each subscene contained six bands of TM data. TM band 6 (thermal infrared) was omitted from the classification process because it has 120-meter pixel resolution.

After a subscene was loaded into the PC, the image was geometrically rectified to the Lambert coordinate system. Rectification is necessary because satellite image data have a certain amount of inherent distortion due to the characteristics of satellite remote sensing. Distortion is generally classified as either systematic or nonsystematic and is summarized in Table 4 (Bernstein and Ferneyhough, 1975; Bernstein, 1983). TM data acquired from the Landsat satellite generally have had the systematic distortion removed at the NASA Goddard Space Flight Center (Billingsley, 1983). However, nonsystematic distortion remains in the image and must be corrected if the image is to be planimetric with respect to the earth's surface. It is almost always desirable to have the final product in a planimetric form because it is considerably simpler for the user of such a map to locate and visualize useful information. Further, such a map is much more compatible with other maps and data bases.

To rectify the original digital image, it is necessary to resample the raw data and produce a "new" image that is geometrically correct. This is done by selecting evenly distributed ground control points (GCPs) throughout the image. A GCP is a feature of known location that can be accurately identified on the original image. GCPs used in the study area were typically either a road intersection, road-stream intersection, or an edge of a water body. The Lambert conformal coordinates corresponding to the GCPs were

Table 4. Sources of Image Geometry Errors in Landsat MSS and TM Data

Systematic distortions

Scan skew: Caused by the forward motion of the platform during the time required for each mirror sweep. The ground swath is not normal to the ground track but is slightly skewed, producing cross-scan geometric distortion.

Mirror scan velocity: The MSS mirror scanning rate is usually not constant across a given scan, producing along-scan geometric distortion.

Panoramic distortion: The ground area imaged is proportional to the tangent of the scan angle rather than to the angle itself. Because data are sampled at regular intervals, this produces along-scan distortion.

Platform velocity: If the speed of the platform changes, the ground track covered by successive mirror scans changes, producing along-track scale distortion.

Earth rotation: The earth rotates as the MSS scans the terrain. This results in a shift of the ground swath being scanned, causing along-scan distortion.

Perspective: For some applications it is desirable to have the MSS images represent the projection of points on the earth upon a plane tangent to the earth with all projection lines normal to the plane. This introduces along-scan distortion.

Nonsystematic distortions

Altitude: If the MSS platform departs from its normal altitude, this produces changes in scale.

Attitude: One sensor system axis is usually maintained normal to the earth's surface and the other parallel to the spacecraft's direction of travel. If the sensor surface departs from this attitude, geometric distortion results.

Source: Bernstein and Ferneyhough (1975); Bernstein (1983) determined. A transformation matrix is computed by ERDAS that will transform each of the pixels in the original scene into a geometrically correct location in an output image. The transformation matrix can be expressed in polynomial form as:

$$x' = a_1 + a_2x + a_3y$$
$$y' = b_1 + b_2x + b_3y$$

where x', y' = original image coordinates
x, y = map coordinates
a_i, b_j (i, j = 1,2,3) = coefficients

After the above coefficients are determined, the accuracy of the transformation matrix is examined by computing the root mean square (RMS) error of the GCPs. Selection of a maximum allowable RMS error is somewhat arbitrary in image processing (Bernstein, 1983); however, the maximum total RMS error was limited to 0.5 pixel in our transformation matrix.

At this point the image is ready to be resampled into the Lambert conformal coordinate system. Resampling was done with a bilinear interpolation technique that takes a distance-weighted average of the brightness values of the four pixels nearest the output pixel. The process is essentially the two-dimensional equivalent of linear interpolation. Because a new brightness value is computed on the basis of the weighted distances of the original spectral values, bilinear interpolation acts as a spatial moving filter and tends to smooth the extremes in brightness values throughout the image. However, this "loss" of spectral data is limited to a very small number of pixels and is relatively insignificant for the type of application used (Level I land cover mapping), and the benefit obtained is a georeferenced output image whose features are in high spatial correlation with ground truth.

Training Site Selection

Upon completion of image rectification, the next step was selecting suitable training sites and defining acceptable training statistics. Training sites were selected that represented each of the map categories desired in the final map. The relevant statistics computed for each training site are the mean, minimum, maximum, standard deviation, and covariance of the pixel brightness values for each band of imagery. It is extremely important to select training site samples that have normally distributed signatures and do not exhibit excessive variance. If a training sample is bimodal and/or has a large amount of variance, it becomes increasingly difficult for the algorithm to distinguish it from neighboring classes during the classification process. After several training sites were selected for each map class, training site statistics were examined for separability. Separability refers to the extent to which the signatures (training sites) are unique within their spectral space. Signatures with high separability have significant spectral separation. A graphical representation of a training signature is referred to as a feature space plot. Figure 3 shows the feature space for a simple two-dimensional, two-signature case. The brightness values in one band of a given feature are plotted against brightness values of the same feature in a different band. In feature space plot a, considerable confusion exists between signature 1 and signature 2. A classification algorithm would have a difficult time discriminating between these two signatures. Sample pixels that occur in the shaded spectral region could quite easily be erroneously categorized. The likely errors would be errors of commission (a pixel assigned to a class to which it does not belong) and errors of omission (a pixel not assigned to its appropriate class). On the other hand, feature space plot b has excellent signature separation. It is very unlikely that either a commission or omission error would result here during the classification process.

Several training sites for each desired map class were selected and rigorously examined for spectral separability. Two-dimensional feature space plots were created for each of the possible band combinations from every training area. Training signatures that lacked spectral uniqueness in most or all of the band pairs were discarded from the analysis. Signatures were further tested by an ERDAS routine that compares the training signature to the original training areas. The routine operates by actually classifying individual pixels within each training area into their most likely classes. The result is a contingency table that shows how closely the individual training site pixels correlate to the overall training signature. Ideally this correlation should be 100%, but signatures were rejected only if their correlation fell below 98%.

Supervised Classification

After an acceptable set of signatures has been assembled, the georectified image is ready for classification. ERDAS software provides three supervised classification techniques: maximum likelihood, minimum distance, and Mahalanobis distance. The minimum distance classification computes the Euclidean (linear) distance from the sample pixel to the class mean. The sample pixel will be assigned to the class that has the smallest distance to mean. The Mahalanobis classification behaves much like the minimum distance classification except tiiat class covariance is considered in the minimum distance computation.

The maximum likelihood classification, the technique used in our methodology, is a powerful Bayesian classifier tiiat computes the likelihood of the sample pixel belonging to each class on the assumption tiiat the class signatures are normally distributed. The class

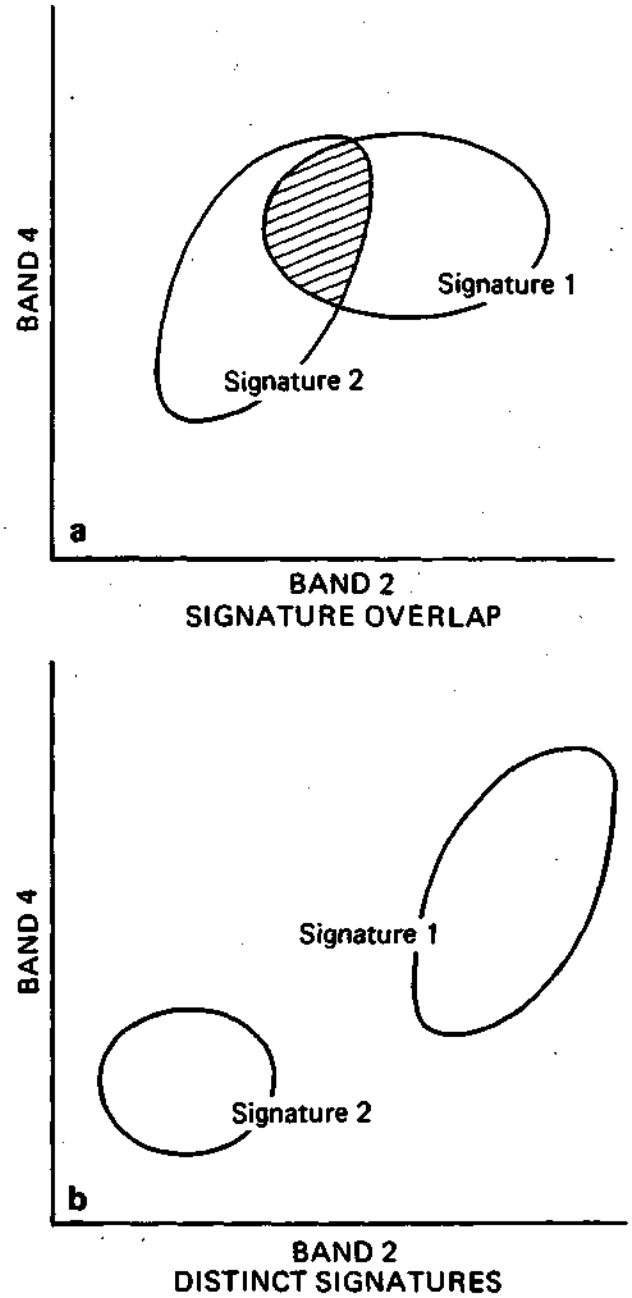


Figure 3. Signature Overlap and Separation
(after ERDAS, Inc., 1988)

with the maximum likelihood is chosen as the output class. The function that governs the technique is given by:

$$p(X/W_i) = \frac{1}{2\pi^{N/2} \quad C_i^{1/2}} \exp[-1/2(X-M_i)^T C_{i-1}(X-M_i)]$$

where X is the sample pixel vector

W_i is class i (vector)

C_i is the covariance matrix for class i

M_i is the mean vector for class i

N is the number of bands

p is the conditional probability

$(X-M_i)^T$ is the transpose of $(X-M_i)$

As previously mentioned, the prior knowledge of the study area was relatively high. Thus the reliability of the maximum likelihood classification was further enhanced by the use of *a priori* probability factors. The *a priori* probability acts as a weighting factor for each class and can increase or decrease the output value of the maximum likelihood function. Usually this factor is left at a value of 1, but it can be assigned any value if there is reason to believe that one class is more (or less) likely to occur than another class. It was decided to assign relative class weights on the basis of the probability of occurrence of each class in the original LUDA data set. The breakdown was as follows:

Class	Percent of LUDA data set	A priori probability
Urban	11.8	0.118
Agricultural	82.5	0.825
Forest	3.2	0.032
Water	0.4	0.004
Wetland	0.2	0.002
Barren	1.9	0.019

These weights resulted in a classification strategy that, for example, made the agricultural class 7 times more likely to occur than the urban class and the forest class 16 times more likely to occur than the wetlands class.

RESULTS

Several training sites were selected for each informational class. This was necessary because often a desired informational class, such as an agricultural field, will have greatly differing spectral characteristics at different scene locations. When training sites are selected that represent the range of spectral characteristics exhibited by an informational class, the opportunity for classification errors to occur is gready diminished if not entirely eliminated. The result of the maximum likelihood classifier is then a set of three subscenes, each containing informational classes corresponding to the original training sites. The final land use and land cover map was created by recoding each informational class into its appropriate land use and land cover category. For example, eight agricultural training sites from the northern sub-area were selected. The result of the classification process was a classified subscene in which eight different sets of pixels represent the agricultural land informational class. These eight classes were recoded into a single informational class. This procedure was repeated for the entire study area, resulting in three GIS-based classified subscenes, each containing a menu of six informational classes. The subscenes were edge-matched according to their respective Lambert coordinates and merged together. The land use and land cover map categories are summarized in Table 5. The final land use and land cover map is seen in Figure 4.

Table 5. Acreages, Percentages, and Totals of Land Use Categories in the SSC Study Area

Class number	Class description	Acres	Percent
1	Urban or built-up	112,169	13.95
2	Agricultural	652,216	81.14
3	Forest	32,123	3.99
4	Water (streams and lakes)	4,637	0.58
5	Wetland	730	0.09
6	Barren (quarries and pits)	1,985	0.25
	Total	803,860	100.00

The results of the maximum likelihood method from three sub-areas were verified individually. The classification corresponding to each training site was cheeked on the basis of LUDA data. The numbers of classes were merged to six USGS Level-I land use classes. The numbers of training sites for each sub-area are as follows:

Sub-area	Number of training sites
North	18
Middle	70
South	42

The purpose of selecting training sites for the sub-areas was to produce less overlapping of the histogram of each spectral band among the specified land use classes. In each sub-area, for a known land use, the spectral variation was still quite large. To solve this problem, multiple training sites were selected. The number of training sites depends upon the complexity of the land use pattern in the sub-area. The computation time increases as the number of training sites increases; therefore there is a practical limit to the number of training sites in each classification. The middle sub-area was analyzed by using the PRIME-ERDAS system. Because the PRIME-ERDAS system performs computations faster than the IBM-AT ERDAS system, it is possible to use more training sites. That is why the middle sub-area had the largest number of training sites.

The output of the maximum likelihood supervised classification process is a GIS-based land use and land cover map (Figure 4) that is georeferenced to the Lambert conformal map coordinate system. Every individual pixel in the map now represents one of the six possible categories of land use and land cover. In classifying original image pixels into informational classes, the algorithm considers each pixel to be a discrete element For example, it decides that an image pixel is either agricultural land or forest, and will not classify an image pixel as part agricultural or part forest If it did, the final map would be of much less utility to the ultimate users and consumers of the data. Opportunities for classification errors are usually present and in practice almost always exist to some extent. In theory, classification errors would be non-existent if the training site signatures of each informational class correlated perfectly with the natural spectral properties of every corresponding informational class in the ground scene. But since this is rarely the case, the objective of image processing is to minimize the classification errors to an acceptable level. Defining and measuring the acceptable level of classification error is known as accuracy assessment.

The overall accuracy level of image classification for earth resources management applications should generally be at least 85% and should be approximately equal for each informational category (Anderson et al., 1976; Milazzo, 1980). The accuracy of the SSC study area land use and land cover maps produced by the supervised classification was examined at several specific locations within the scene. Of particular interest were areas where several classes occurred in significant sizes and were located relatively near each other. For comparison of LUDA data and digital classified maps, the LUDA data must be

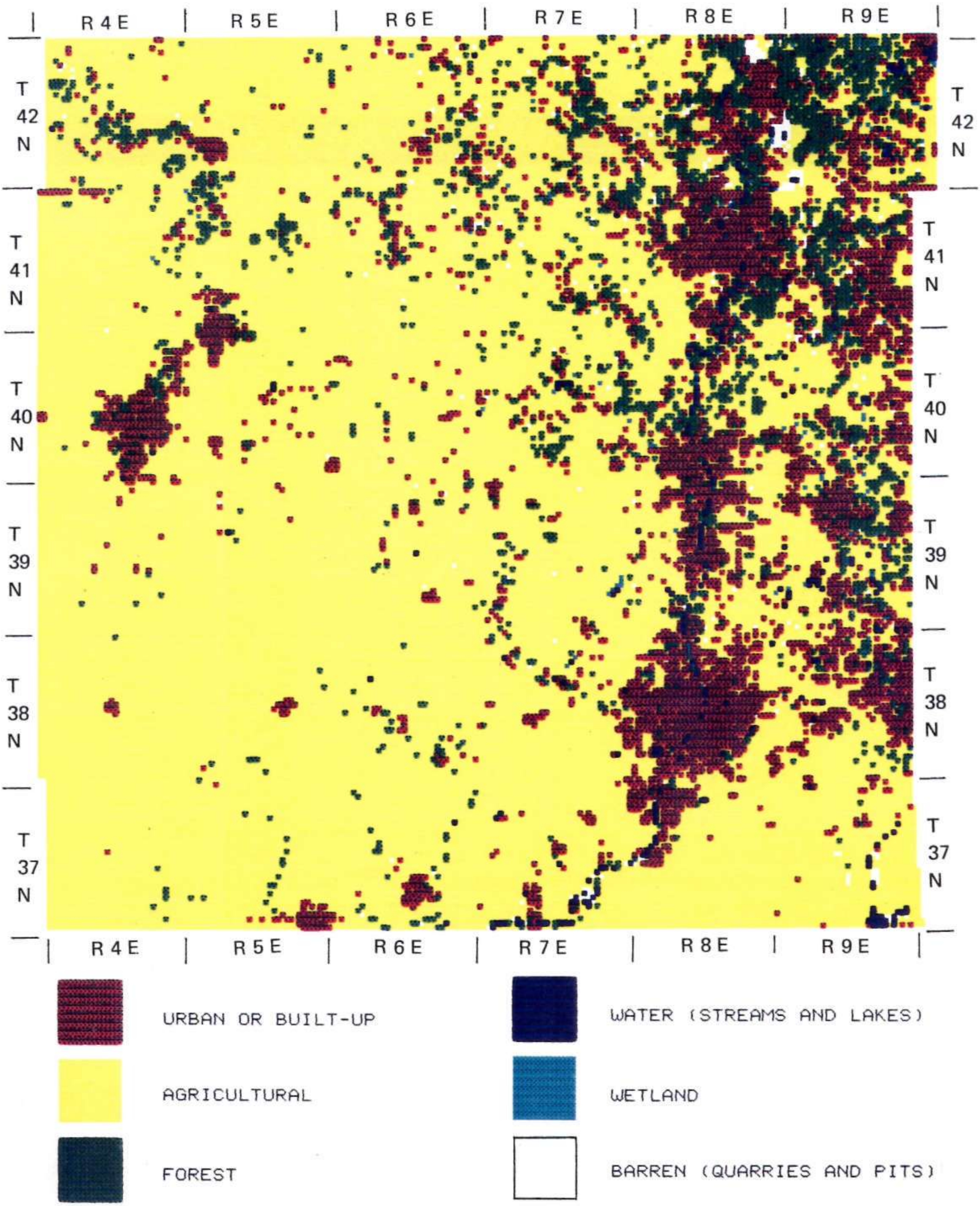


Figure 4. Final Land Use and Land Coverage Map of the Superconducting Super Collider Study Area

imported from the ARC/INFO system to the ERDAS system. Because of the limited time frame, an alternative approach was developed mat involved manual checking of selected areas where there are various combinations of urban land, agricultural land, barren land, forest, and water. All these classes occur in a relatively small area.

A hard-copy print of this area was produced at a scale of 1:200,000 and compared to an existing land use and land cover map (ISGS, 1985) derived from LUDA digital data. It was found that the ERDAS-produced map and the LUDA-based map tended to be highly correlated in terms of classification and geographic representation. Other regions within the SSC study area were similarly checked for classification accuracy. The correlation between the two data sets again was found to be high, ranging from 90 to 99%. However, LUDA map data are generally based on information obtained in the early to middle 1970s. Perhaps for many applications the relatively old age of this data would not be significant; but this is not the case when analyzing land use and land cover patterns in the SSC study area. Because of the growing population (Illinois Bureau of the Budget, 1987) and expanding economy of the region, its land use and land cover patterns have been undergoing rapid changes. Preparation of new LUDA maps from new aerial photographs and their comparison with the existing LUDA maps can provide the change in area under different land use and cover classifications. Urban outgrowth patterns resulting from residential and commercial construction activities are present at many fringe areas of suburban communities in the maps prepared from Landsat images. Many recent small lakes and mining pits emerged on the classified image that were not on the old LUDA-based map. Therefore the digital image analysis is a promising approach for quick updating of the land cover maps. However, it cannot substitute for land use mapping based on aerial photographs.

Conclusions and Suggestions

The six-county region of northeast Illinois that encompasses the proposed site of the Superconducting Super Collider is undergoing rapid change due to economic growth and a growing suburban population. Area land use and land cover data have been catalogued into spatial data bases. However, much of the data is 10,15, or even more than 20 years old. The completed research created an updated land use data set based on recent satellite imagery.

The conversion of Illinois farmland and natural areas into urban or barren land is of great significance in relation to the maintenance and preservation of the state's finite natural resources. Policymakers at all levels of government must have the best information available when making decisions that directly affect the conservation and management of these resources. At present, much of this information is stored in various GIS data bases, but often the GIS data are old. Satellite remote sensing can provide a rapid and efficient method for updating existing databases with recent information. This, in turn, gives public policy decision makers the benefit of improved and more reliable data to use as a tool in their decision-making process.

A detailed and complete comparison of LUDA data and satellite remote sensing classified maps is necessary to ensure the data quality. The classified maps, though of less accuracy than the LUDA maps, are helpful in updating regional land use/cover information at a small cost (10% or less of the cost for the LUDA mapping), identifying areas with rapid change, and quantifying the changes so that LUDA mapping may be undertaken, if necessary, for such areas rather than for the whole region.

REFERENCES

Anderson, J.R., E. Hardy, J. Roach, and R. Witmer. 1976. "A Land Use and Land Cover Classification System for Use with Remote Sensor Data," U.S. Geological Survey Professional Paper 964.

Bernstein, R. 1983. "Image Geometry and Rectification," Manual of Remote Sensing, R.N. Colwell, ed., American Society for Photogrammetry and Remote Sensing, Falls Church, VA, Chapter 21, Volume 1, pp. 875-881.

Bernstein, R., and D.G. Ferneyhough. 1975. "Digital Image Processing," Photogrammetric Engineering and Remote Sensing, Volume 41, pp. 1465-1476.

Billingsley, F.C. 1983. "Data Processing and Reprocessing," Manual of Remote Sensing, R.N. Colwell, Ed., American Society for Photogrammetry and Remote Sensing, Falls Church, VA, Chapter 17, Volume 1, pp. 758-763.

ERDAS, Inc. 1988. "ERDAS User's Guide, Version 7.3," Atlanta, Georgia.

Hines, J.K., editor. 1986. "Siting the Superconducting Super Collider in Northeastern Illinois: Environmental Screening Atlas," Illinois State Geological Survey, Champaign.

Illinois Bureau of the Budget. 1987. "Illinois Population Trends 1980 to 2025," State of Illinois, Bureau of the Budget, 282 p.

Illinois State Geological Survey (ISGS). 1985. "Land Use and Land Cover Map of Northeastern Illinois," Department of Energy and Natural Resources, Illinois State Geological Survey, Champaign.

Jensen, John R. 1986. "Introductory Digital Image Processing," Prentice-Hall, 379 p.

Lillesand, T.M., and R.W. Kiefer. 1987. "Remote Sensing and Image Interpretation," John Wiley and Sons, Inc., 721 p.

Milazzo, V.A. 1980. "A Review and Evaluation of Alternatives for Updating U.S. Geological Survey Land Use and Land Cover Maps," U.S. Geological Survey Circular 826.

U.S. Geological Survey. 1973, 1976. "Digital Land Use and Land Cover for Aurora (1973) and Rockford (1976)," quadrangles, scales 1:250,000.