Developing GIS for Lake Icaria Watershed

Adams County, Iowa

by

Yan Wang

A Thesis Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Department: Civil and Construction Engineering
Major: Civil Engineering (Geometronics)

Iowa State University
Ames, Iowa

1993

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

The objective of this thesis is to develop a Geographic information Systems (GIS) for the Lake Icaria Watershed, Adams County, Iowa. The Lake Icaria Watershed, which covers an area of approximately 70 kilometers, is located in Northeast of Adams County, Iowa (Figure 1-1). Water, forests and farmland are the major land covers for this area.

This project was initially proposed by Dr. Robert Jolly, professor in economics, and Dr. Sunday Tim, professor in Agriculture Engineering both with Iowa State University. The purpose was to design and evaluate changes in agricultural technology, its utility and the local institutions that would foster sustainable development in agricultural-dependent communities. Because this project would involve computer-assessment of various livestock expansion strategies, including economic and environmental consequences, the GIS technique was considered as a potential decision support system for technology design, land use planning, and technical transfer. Therefore, the proposed GIS will be developed to monitor the land cover/land use, hydrography, transportation, soil and other possible parameters within the area.

The GIS developed in this project includes remotely sensed imagery and digitized cartographic products. Hydrography, topography and transportation were digitized from USGS 1:24,000 7.5-minute topographic maps. Tracts, which divided the whole study area into different scale pieces for management purposes, were digitized from photographs taken in 1980 and revised by the Agriculture Soil Conservation Service (ASCS) of Adams County. Soil information was obtained from USDA Soil conservation Service. In order to derive the most current land cover/land use information, SPOT satellite imagery (Figure 1-2.) taken on April 22th, 1990 was purchased. SPOT, an acronym for *Systeme Pour l'Observation de la Terre*, was initially undertaken by

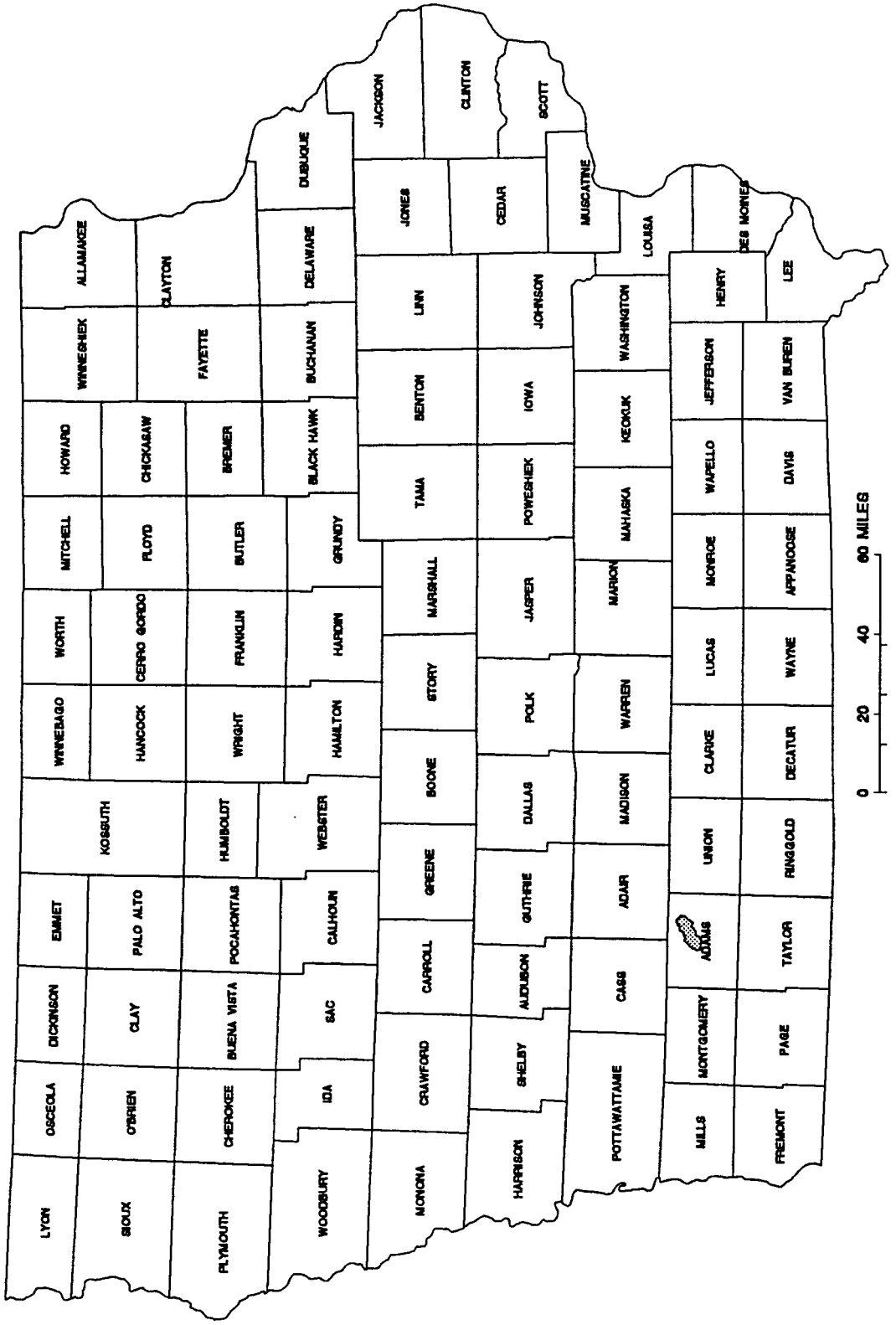


Figure 1-1. The Location of Lake Icaria Watershed

Figure 1-2. Lake Icaria Watershed, SPOT Data

the French government for space development program in early 1978. SPOT data was enhanced, rectified, and classified through digital image processing procedures.The classified land cover/land use from SPOT data may be used to update, identify and monitor patterns and changes in general land cover. This up-to-date land information would be helpful in land management and planning. Color-infrared and black-and-white aerial photographs taken on May 15, 1983 and April 6, 1990 respectively, were available as additional ground truth information for the digital image classification purposes.

In order to rectify the raw SPOT data using the UTM coordinate system, Global Positioning System (GPS) technique was incorporated in the project to obtain relatively accurate coordinates of selected ground control points. These ground control points could be used as registration marks in the image rectification process. To maintain consistency of database projection, the GPS coordinates were transformed from NAD-83 ellipsoid to NAD-27 ellipsoid since all the digital data in this project is based on NAD-27 ellipsoid. The accuracy of map coordinates digitized from USGS 7.5-minute maps were evaluated by comparison with GPS coordinates.

The GIS that was developed was maintained on the DEC workstation network that is part of the Vincent System on the ISU campus. ARC/INFO GIS software (version 6.1.1, released December 23, 1992 by Environment Systems Research Institute, Inc.) was the main GIS system applied. ERDAS Image Processing System (VERSION 7.5, released July 1991 by ERDAS, Inc.) was involved in SPOT data processing. Some other peripheral hardware including GTCO Super L II digitizing tablet, postscript printer, and Calcomp colormaster printer were used from data entry to analysis output.

The thesis consists of three parts. The flow chart depicting the procedures followed in this project can be found in Figure 1-3. In Part I, Chapter 2 introduces some basic concepts of GIS and remote sensing and briefly reviews the two major pack-

ages, ARC/INFO and ERDAS. Chapter 3 describe the basic principles of GPS observation and the comparison between GPS coordinates and digitized coordinates. The accuracy of the digitized coordinates is evaluated. Part II includes digital image processing theories and procedures. Chapter 4 discusses image enhancement techniques used in this project. Image rectification procedures by which the raw satellite imagery is rectified into the UTM coordinate system by least square method will be described in Chapter 5. The Image classification, which yields about 76% classification accuracy, is discussed in Chapter 6. Chapter 7 generally describes the procedures used in post classification which made the classification map more usable. The classified imagery through the post classification processing is considered as the most current land cover/land use information fed back to the GIS database. In Part III, the developed GIS database is applied to solve two practical problems. Chapter 8 discusses application of the developed GIS database to perform surface modeling. The calculated stream network by GIS is compared with the one digitized from existing maps. Chapter 9 uses the GIS database to find the minimum transportation path between any two tracts, which is very important in handling waste removal in this area. Chapter 10 represents the conclusions and recommendations. A more descriptive documentaion of the GIS database is listed in Appendix A. ARCVIEW package is introduced in Appendix B. ARCVIEW can display/query the information stored in GIS database and is considered as a bridge between the GIS database and the end users. Appendix C. contains Arc Macro Language (AML) programs used in this project. Appendix D describes how to use ARC/INFO and ERDAS packages on Iowa State University campus.

This project completed its goal to develop a GIS for the Lake Icaria Watershed. The procedures and approaches applied in this project have been proven to be practical and can be applied to the similar projects in the future. It has been acknowledged

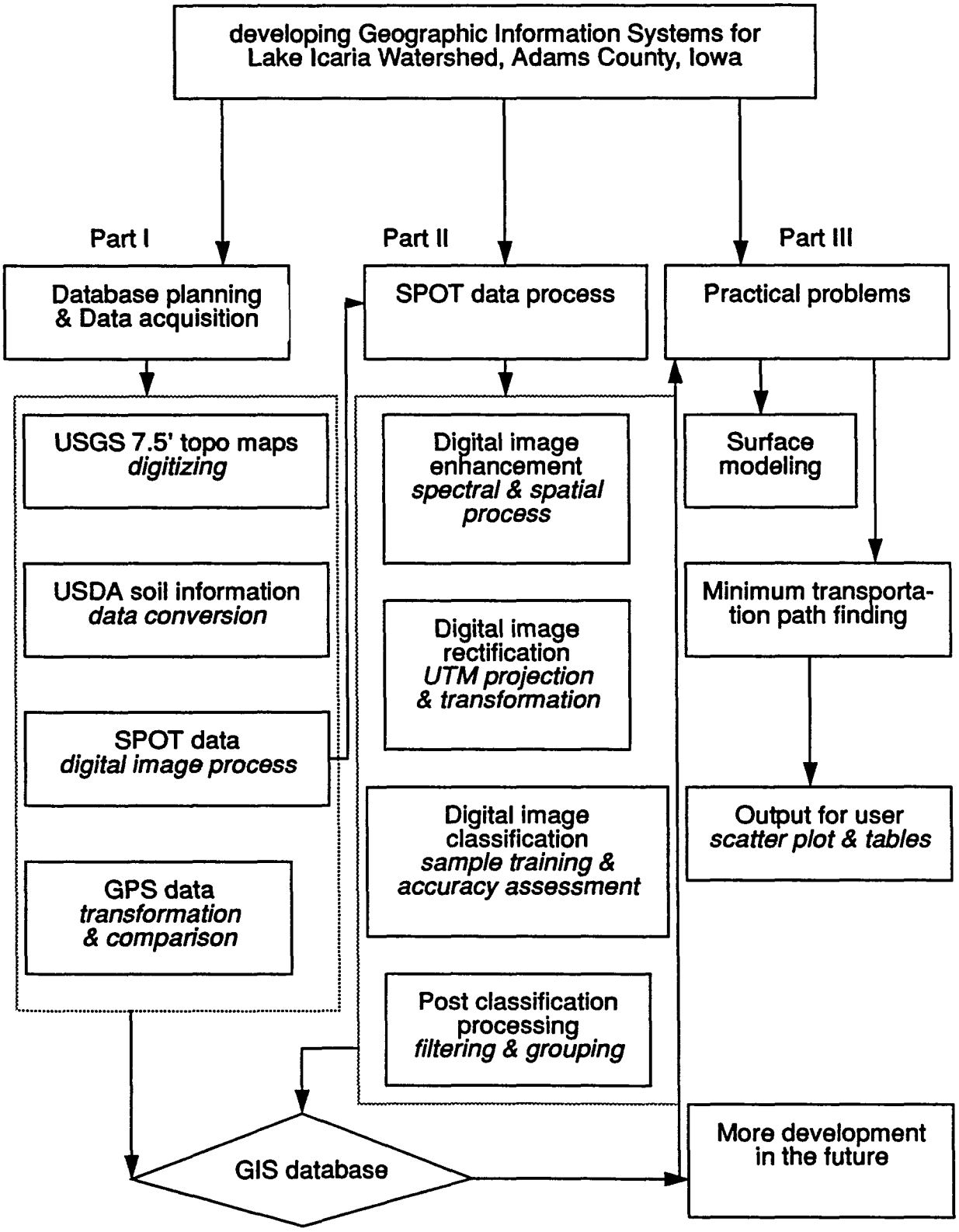


Figure 1-3. Flow Chart Depicting the Procedures Followed in this Project

that GIS and remote sensing can be integrated compatibly. GIS can use the data captured and reduced by remote sensing; therefore, it can enhance the application of remotely-sensed data. Conversely, more utilization of remote sensing data inputs can significantly improve the utility of GIS by supplying more current infomation for use in updating the GIS database. It was concluded that the GIS developed for the Lake Icaria Watershed can also be used as a decision-support system to help economic experts and agricutural engineers to evaluate changes in agricultural technology, and land use planning, etc. More details on the conclusions and recommendations are discussed in Chapter 10.

PART I

GIS, REMOTE SENSING AND GPS OBSERVATION

CHAPTER 2. BASIC CONCEPTS OF GIS AND REMOTE SENSING

Geographic Information Systems

Geographic Information Systems (GIS) are defined as automated systems "for the capture, storage, retrieval, analysis, and display of spatial data" (7). "The first Geographic Information Systems (GIS) were developed in the middle 1960s by governmental agencies as a response to a new awareness and urgency in dealing with complex environmental and natural resource issues" (6). During those days, GIS was very expensive to build as well as to operate. It has only been in the last few years, with the help of high speed computers, "that the true potential and significance of GIS are beginning to be realized" (6). The spatial database developed in a GIS can be used to provide answers to queries of a geographical nature. A good GIS serves as a decision support system for different problem solving environments.

The basic components in a GIS consist of the user, GIS software tools, and database. A schematic account of this relationship is shown on Figure 2-1. The user plays an important role in the GIS especially when complicated analyses, such as spatial overlay analyses or hydrography modeling, must be performed. GIS software and database are required to support the analyses. A GIS database is usually made up of a series of map layers, which represent the many information themes in a real world. Figure 2-1 shows a typical GIS for a sample area.These different map layers can be created and finalized by data acquisition, such as digitizing or scanning, data cleaning, and data management, such as projection of the layers to certain a map projection. The outputs in different format would be the products of GIS analyses. The typical GIS outputs are maps or reports.

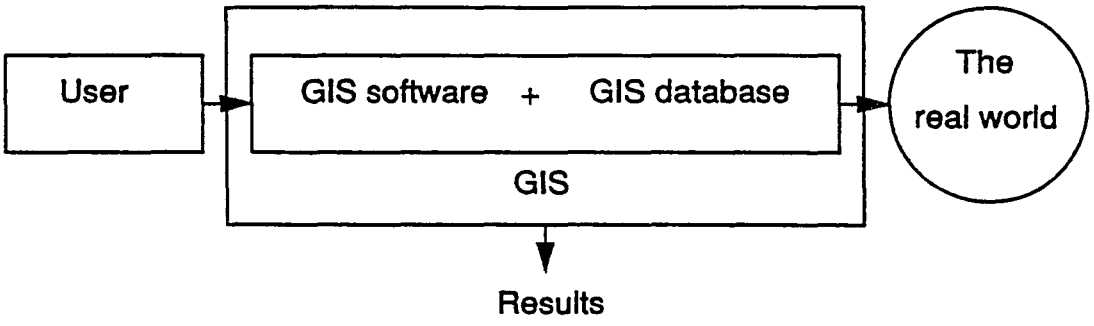


Figure 2-1. Components in a GIS

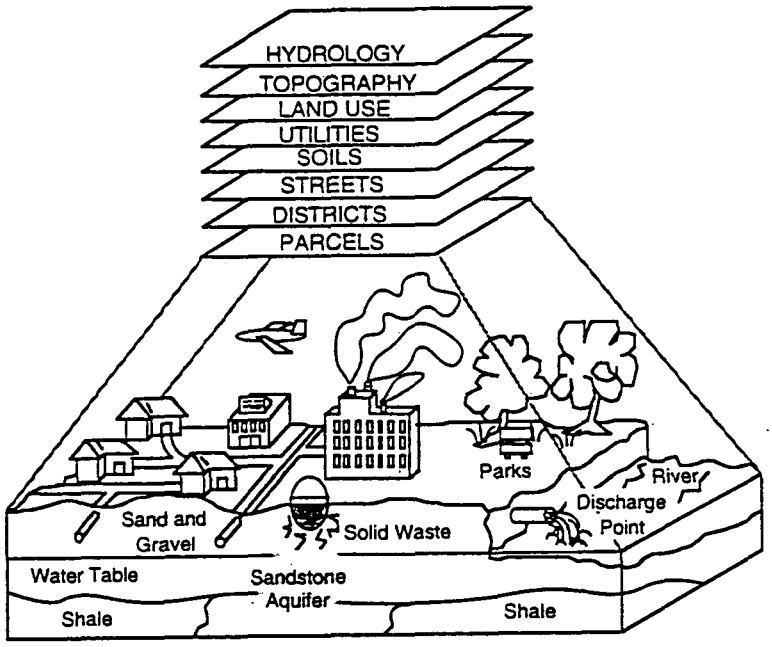


Figure 2-2. Data Layers in Representing Many Geographies (12)

In a GIS database, spatial features are stored in certain spatial structures. There are two major categories of spatial structures: vector and raster. Vector data structures are based on knowing the locations of basic entities such as points, lines, and polygons. The minimum unit in the vector structure is a point. A line might be stored as a starting and an ending point and a series of vertex points which are used

to describe the shape and the direction of the line. A polygon is stored as an area enclosed by lines. In the raster data structure the area of interest is gridded by many picture elements (pixels). A pixel is the minimum unit in raster structure. A pixel has an area and is usually described by its width and height. A line is stored as a series of pixels lined up together from the starting pixel to the ending pixel. A polygon would be stored as many adjacent pixels to form an area.

The selection of the spatial data structure in a GIS should match application needs. In this project, both vector and raster data structures are used. For coverages like roads or streams, a vector structure is more favorable because these spatial features are linear or polygons bounded by small segments of lines. For the satellite remote sensing data, it is more efficient to process it in raster data structure. The ARC/INFO GIS package can process both types of data structures and has the capability to convert from one data structure to another one.

ARC/INFO GIS Package

At present, ARC/INFO is the most widely used GIS package, worldwide. It is a member of the ARC family which are a product series of Environmental Systems Research Institute, Inc. ARC/INFO has multiple functions for dealing with geographic features and their attributes. ARC, ARCEDIT, ARCPLOT, and GRID are the most popularly applied sub-packages which can perform tasks from data acquisition and feature editing to map production and raster data analyses. INFO is a database management system associated with ARC.

The Lake Icaria Watershed GIS was based on ARC/INFO. The GRID sub-package was applied extensively for raster data processing. ARCPLOT was the main supporting system in map production.

There were six basic steps in ARC/INFO to develop the GIS for Lake Icaria Watershed:

1. Getting spatial data into ARC/INFO.
    a. Determine what data to include in the GIS.
    b. Gather the topographic maps. photographs, and satellite data.
    c. Gather the non-graphic attribute data.
    d. Digitize and/or convert data.
2. Making spatial data usable.
    a. Construct topology.
    b. Correct spatial data by using ARCEDIT.
    c. Reconstruct topology.
3. Getting attribute data into ARC/INFO.
    a. Use INFO to create tabular data files.
    b. Link attributes to geographic features.
4. Managing the database.
    a. Transform coordinate system
    b. Join adjacent maps.
    c. Enter new attributes.
5. Performing geographic analysis.
    a. Establish the objectives and criteria for the analysis.
    b. Perform the spatial analysis, such as overlay.
6. Presenting the results of the analysis.
    a. Create maps for display.
    b. Design and generate reports.
    c. Develop a user interface.

Remote Sensing

Remote sensing is a process that acquires data and derivative information about objects or materials located on the earth's surface or in its atmosphere by using sensors mounted on a space platform at a certain distance from the targets to make measurements of interactions between the targets and electromagnetic energy. Figure 2-3 schematically illustrates the generalized processes and elements involved in electromagnetic remote sensing of earth resources. This figure shows the interrelationships among the three major components of the remote sensing system: the earth's surface, the sensor, and the data processing system. The objective of remote sensing is to help in qualitatively or quantitatively interpreting and recognizing the earth features. The sensor system can capture and record electromagnetic energy reflected or/and emitted from objects. The processing system is necessary to make remote sensing raw data more readable and interpretable for further use.

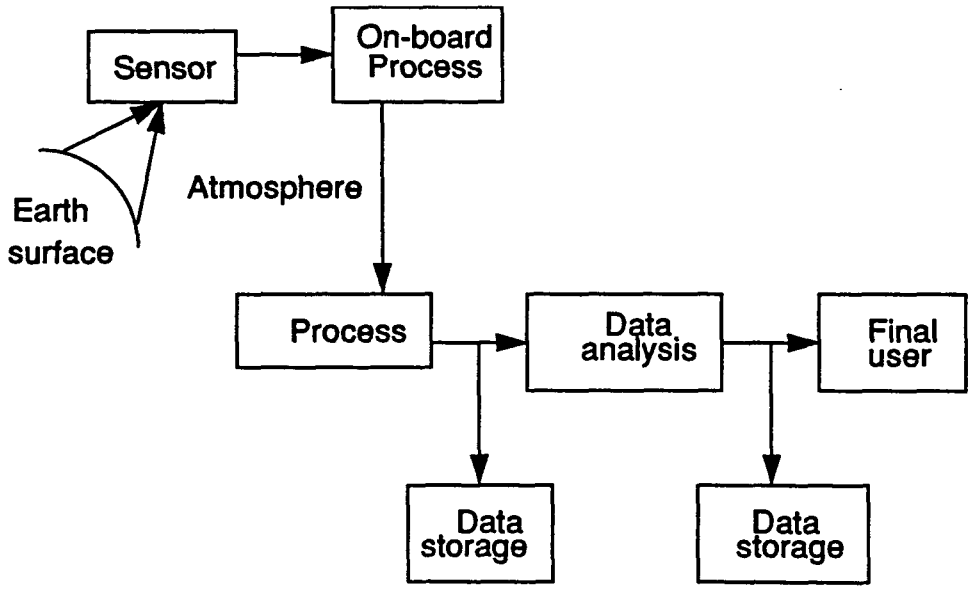


Figure 2-3. A Diagram of Typical Remote Sensing System

Electromagnetic Spectrum

Figure 2-4 shows the entire electromagnetic spectrum which contains all the reflected or emitted information about the earth surface. At the top are shown the different types of energy which are functions of the wavelength. The portion usually used in remote sensing, called optical wavelengths, is displayed in detail on the upper portion of the figure. The human eyes can only sense a small portion of the total electromagnetic spectrum, known as visible light. In traditional photogrammetric measurement, the photographic sensors can detect the photographic wavelengths which cover visible light and a small portion of the infrared region. The mid-infrared and thermal infrared portion of the wavelengths can be used if a more advanced sensor is used. It should be pointed out that only thermal infrared energy is directly related to the sensation of heat; near and mid-infrared energy are both reflected energy. Other portions of the electromagnetic spectrum such as the microwave region are also useful in remote sensing. Microwave remote sensing is not discussed in this project.

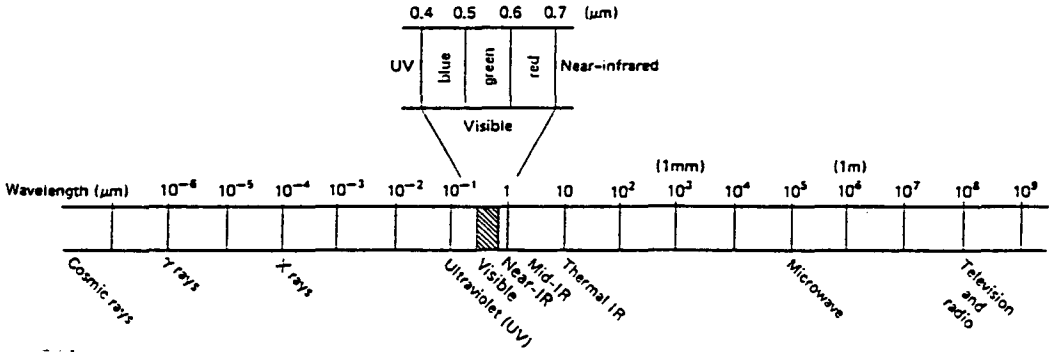


Figure 2-4. The Electromagnetic Spectrum (1)

Energy Interactions in the Atmosphere

As shown in Figure 2-3, any energy detected by the remote sensing sensors passes through the atmosphere before it reaches them. The atmospheric effects change the energy dramatically in both intensity and spectral composition. These effects are caused basically through the atmospheric scattering and absorption.

Atmospheric scattering is defined as the diffusion of energy caused by particles in the atmosphere. When the particles are atmospheric molecules and other tiny particles which are much smaller in diameter than the wavelength of the interacting energy, Raleigh scatter occurs. The effect of Raleigh scatter is inversely proportional to the fourth power of wavelength. Therefore, the short wavelength portions of the energy will be scattered more than the long wavelength portions. In color photography, Raleigh scatter usually causes the imagery to "haze" which diminishes the contrast of the imagery. This "haze" effect can be eliminated or minimized by introducing, in front of the camera lens, a filter which does not transmit short wavelengths. Another type of scatter is called Mie scatter which occurs when the atmospheric particle diameters equal the energy wavelengths. Water vapor and dust are two major causes of Mie scatter. Compared to Raleigh scatter, Mie scatter influences longer wavelengths. Non-selective scatter is the third scatter, which comes about when the diameters of the scattering particles are much bigger than the energy wavelengths being sensed.

Another significant atmospheric effect on remote sensing is called atmospheric absorption. The atmosphere transmits the energy in certain wavelength regions, but blocks the energy in other wavelength regions. The non-blocked spectral regions, called "atmospheric windows" are selected to acquire data in remote sensing. Figure 2-5 shows these atmospheric transmission windows.

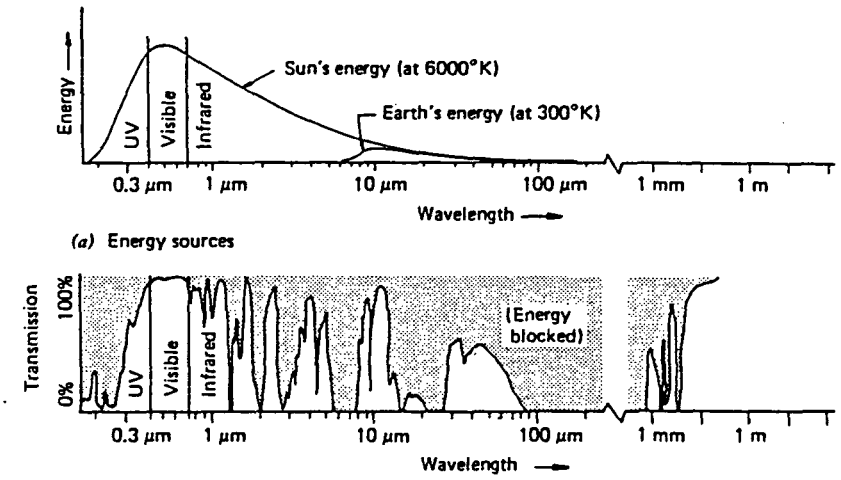


Figure 2-5. Spectral Characteristics of Atmospheric Absorption (1)

Energy Interactions with Earth Surface Features

There are three possible energy interactions with the earth-surface features when electromagnetic energy is incident on the earth's surface. They are reflection, absorption, and/or transmission. This is illustrated in Figure 2-6. The interrelationship between these three energy interactions can be stated as:

$$E_I(\lambda) = E_R(\lambda) + E_A(\lambda) + E_T(\lambda) \qquad \text{Eq. 2-1}$$

where $E_I(\lambda)$ denotes the incident energy, $E_R(\lambda)$, $E_A(\lambda)$, $E_T(\lambda)$ denote the reflected energy, the absorbed energy, and the transmitted energy with all energy components being a function of wavelength λ.

Two important points concerning the relationship between electromagnetic energy and the earth's surface feature should be noted. First, the proportion of energy reflected, absorbed, and transmitted energy varies for different earth surfaces. This makes it possible for us to distinguish different earth features in an scene. Second, the proportion of the reflected, absorbed, and transmitted energy varies at different wavelengths. Two different materials on the earth might be indistinguishable in one spectral region but very clear or distinguishable in another spectral region. Figure 2-7 shows a hypothetical but conceptually accurate graph of the relative reflectance curve as a function of wavelength for three simple ground cover classes: vegetation, soil, and water. We can see that the reflectance for all three features are not distinguishable at wavelength λ_1. But the reflectance value for vegetation at wavelength λ_2 is much higher than both soil and water. This characteristic is widely used in image interpretation to identify different surface features.

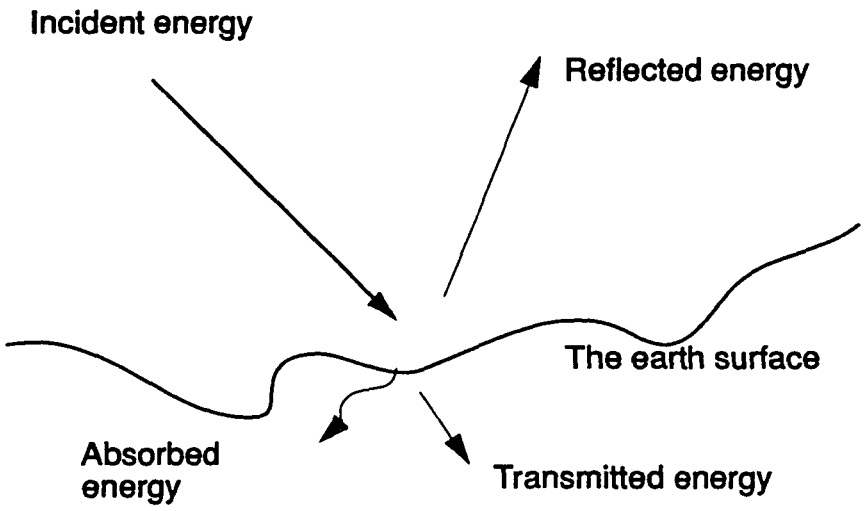


Figure 2-6. Basic Interactions Between Electromagnetic Energy and an Earth Feature

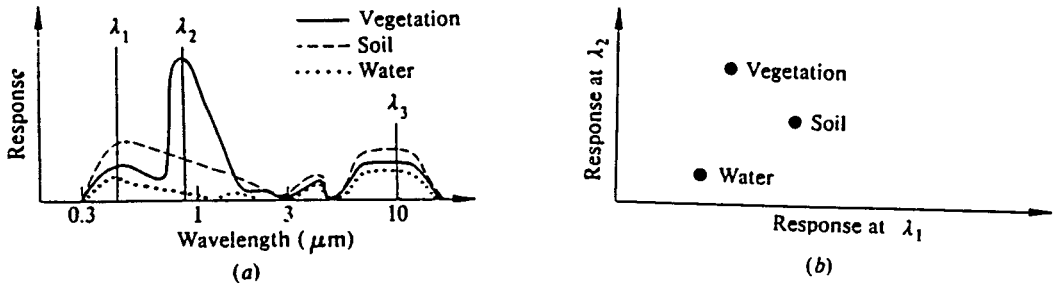


Figure 2-7. Hypothetical Reflectance Curves (3)

Digital Data Format

Traditional photography detects the electromagnetic energy of different surface features by using chemical reactions on the surface of a light sensitive film. This photographic process is relatively simple and inexpensive. Electronic sensors used in remote sensing generate an electrical signal that corresponds to the amount of energy coming from the surface feature. This signal can then be recorded and stored in a digital format. Figure 2-8 illustrates typical digital data. It is actually composed of a two-dimensional array of discrete elements or so called pixels. For each pixel, there is an associated digital number (DN) which represents the brightness value (reflectance) or the amount of energy measured electronically over the ground area corresponding to this pixel through analog-to-digital (A-to-D) signal conversion process.

In this digital format, the image data can be easily analyzed with the aid of a computer. This makes digital image processes such as image enhancement or image classification possible.

54	40	31	20	13	22	32	23	21	24
54	21	21	23	23	32	32	35	31	30
23	13	23	22	13	21	13	23	32	32
23	21	13	32	13	32	43	13	21	43
31	23	12	21	13	34	43	41	56	54

Figure 2-8. A 5 row x 10 column digital image data set (DN values)

ERDAS Image Processing System

The ERDAS Image Processing System was developed by ERDAS, Inc. It incorporates most of the basic operations used in digital remote sensing data analysis. These features include data collecting, displaying/viewing, editing and analyzing of raster data.

ERDAS is capable of displaying up to 256 colors, or gray tones if a single-band image is displayed. Also, three band color composites and graphics overlay are possible. Aside from a variety of digital remote sensing data display options, ERDAS has many fundamental image processing and analysis procedures, including image enhancement, geometric correction, and multispectral classification. Image enhancement includes two-band summing and differencing, and band ratioing techniques. Geometric correction operations in ERDAS provide for image transformation and registration. This step involves ground control point selection on both ground and image, and the calculation of a transformation matrix. Image classification covers both super-

vised and unsupervised classification routines. Parallelepiped, minimum euclidean distance, or maximum-likelihood algorithms are applied in supervised classification, and ISODATA clustering is available in unsupervised classification procedure.

The ERDAS system was initially installed on ISU campus in May 1992. It has been one of the major image process systems in practice. The ERDAS system was applied in this project in several ways.

1. Getting SPOT data into ERDAS format.
    a. Read the entire data set into ERDAS
    b. Clip the area of the Lake Icaria Watershed
2. Performing image enhancement
    a. Initial statistics extraction from raw SPOT data
    b. Spatial enhancement
3. Performing image rectification
    a. Locate the GPS control points on the image
    b. Calculate the transformation matrix
    c. Transform the entire image
4. Performing image classification
    a. Collect ground truth data
    b. Create a signature file by supervised classification procedures
    c. Classify the image by three different classification rules
    d. Compare the outputs by three classification rules
    e. Make the thematic land cover/land use map for the study area
5. Exporting the classified land cover image to ARC/INFO GRID format.

CHAPTER 3. GLOBAL POSITIONING SYSTEM OBSERVATION

Global Positioning Systems (GPS) technology was initially developed by the United States Department of Defense in 1973 as a navigation system. The GPS technique has been widely applied in many non-military fields since then because of its high accuracy and convenience. There are 21 operational satellites at this time. A constellation of up to 24 satellites in three to six orbital planes, 30° to 60° apart, was proposed. It will be assured that four to six satellites about 5° above the horizon at any time basically anywhere in the world will be visible when the planned constellation is completed. Figure 3-1 shows the proposed constellation.

The GPS satellites orbit the earth at an altitude of about 20,183 kilometers with an inclination to the equator. Each satellite completes its orbit around the earth every 12 hours.

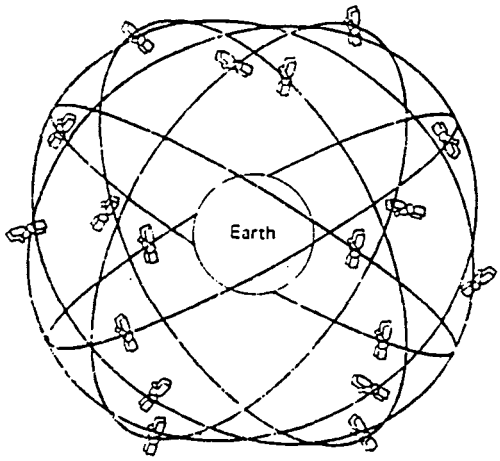


Figure 3-1. Constellation of GPS Satellites (8)

Basic Principles

The basic theory of GPS observation is that a series coded train of pulses emitted simultaneously from four or more satellites determines the position of a receiver by the relative time of their arrival. GPS satellites transmit two types of signals on two L-

band carrier frequencies called L1 and L2. The L1 carrier wave has a center frequency of 1575.42 MHz and the L2 carrier has a center frequency of 1227.6 MHz. The L1 carrier is modulated with two codes carrying information. One is called the clear acquisition or C/A code and has a frequency of 1,023 MHz and a wavelength of approximately 293 meters. The other is called the precision or P-code with a 10.23 MHz frequency and wavelength of about 29.3 meters. The L2 frequency is modulated with only the P-code. Figure 3-2 shows the broadcast signals. GPS measurements utilize one-way ranging as shown in Figure 3-3.

Equation 3-1 can be solved for slope distance d, where is elapsed time, and c is the velocity of electromagnetic energy through the atmosphere.

$$\Delta t = \frac{d}{c}, \text{ or } d = c\Delta t \qquad \text{Eq. 3-1}$$

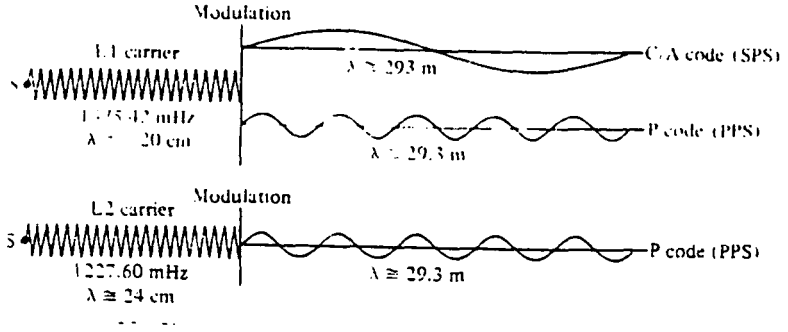


Figure 3-2. GPS Broadcast Signals (9)

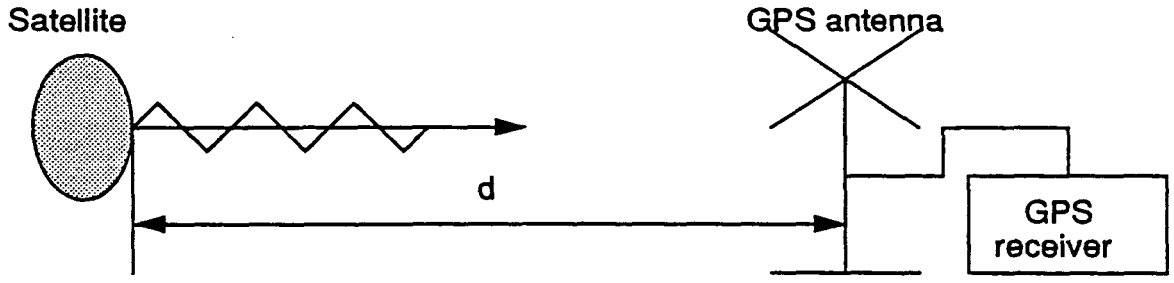


Figure 3-3. GPS One-way Ranging

The above equation is not true without the exact synchronization of receiver and satellite clocks. There is some discrepancy between two satellite clocks. This timing discrepancy, which is generally called the clock bias, will cause serious errors on the range if not corrected. Therefore, these types of ranges are also called pseudo-ranges. Most Global Positioning System receivers can make basically two types of measurements. One is pseudo-ranging and the other is measuring the carrier wave phase.

In order to reduce the effect of the clock bias, four satellites must be used simultaneously. This process can be explained as follows. The position (x, y, z) of the unknown point A in Figure 3-4 can be determined by solving the following equations which incorporate the clock bias.

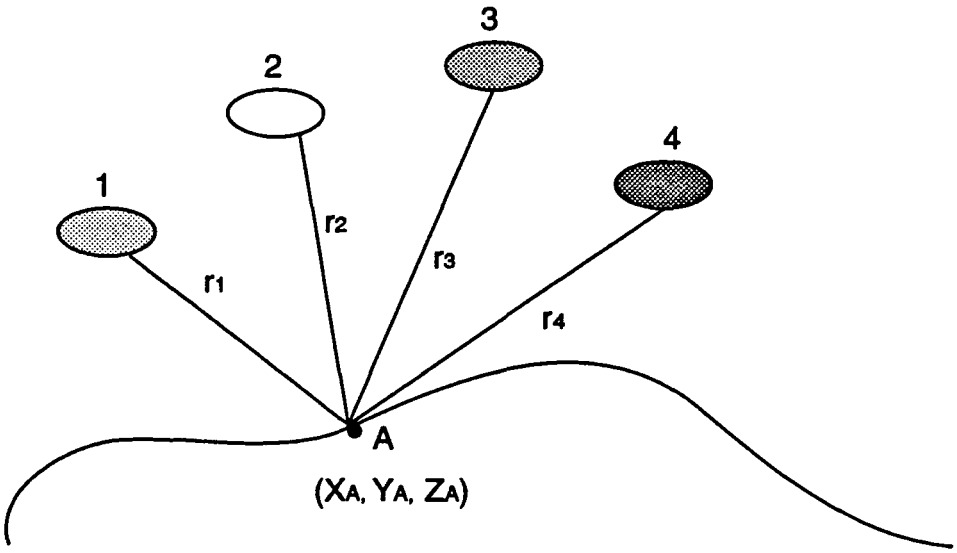


Figure 3-4. Position Determined by Four Satellites

$$
\begin{aligned}
(r_1 + \Delta r)^2 &= (X_1 - X_A)^2 + (Y_1 - Y_A)^2 + (Z_1 - Z_A)^2 \\
(r_2 + \Delta r)^2 &= (X_2 - X_A)^2 + (Y_2 - Y_A)^2 + (Z_2 - Z_A)^2 \\
(r_3 + \Delta r)^2 &= (X_3 - X_A)^2 + (Y_3 - Y_A)^2 + (Z_3 - Z_A)^2 \\
(r_4 + \Delta r)^2 &= (X_4 - X_A)^2 + (Y_4 - Y_A)^2 + (Z_4 - Z_A)^2
\end{aligned}
\qquad \text{Eq. 3-2}
$$

where r_1, r_2, r_3, and r_4 are the distances from each satellite to the GPS receiver. Δr is the error in the range due to clock bias.

The three coordinates of the unknown point X_A, Y_A, Z_A, and Δr can be evaluated by solving these four equations simultaneously. Usually a receiver makes repeated pseudo-range measurements at a station to as many satellites as are visible. If additional equations are available, the least squares solution can be reached to determine the unknown parameters in above equations.

When using two receivers simultaneously, the relative point positioning method is widely used in practice. Using this method, an antenna and a receiver is stationed at a known location and the other set is placed at the unknown location. This procedure can eliminate or greatly reduce several errors such as satellite errors, or errors due to the atmosphere. The relative positioning method produces the components Δx, Δy, and Δz with higher accuracy since errors have been eliminated or reduced. A more accurate location of the unknown point can be reached when these components are applied in the computation.

Four Modes Used in Relative Point Positioning

There are four survey modes which can be operated; static, pseudo-static, kinematic, and pseudo-kinematic. In the static mode, the receivers are set up over the point in a stationary position. The receivers at each station collect data simultaneously for about 45 minutes to 2 hours in every epoch (usually every 10 or 20 seconds). The epoch can be set in advance. The data then is post-processed after the observation. This method yields precise distances between stations by eliminating errors which are associated with satellite information and receiver biases. Similar to static mode, pseudo-static mode is a shorter static observation technique in which the data is collected for only about 15 to 30 minutes at each location. In the kinematic mode, one

receiver generally called the base receiver, is placed at a known point while the second moving receiver (rover) is placed at another known point determined by an antenna swap procedure or a previous survey. The rover will be moved to other unknown locations after a short observation (3-4 minutes) at each unknown point where similar brief observations are made. Because it is fast and time efficient, this procedure is widely used to locate many unknown points, such as the center line of a road. But it requires that at least four satellites be continuously locked by both the base and moving receivers during the whole observation period. The pseudo-kinematic mode is similar to the kinematic mode except that a second known location is required. This technique does not require continuous tracking of four or more satellites. In this mode, a stationary receiver is set up at a known point and another receiver occupies each point for at least two short periods (5 minutes) separated by a longer period of about one hour.

GPS Control Point and Mode Selection

The purpose of GPS observation in this project was to obtain the accurate coordinates of several ground control points (GCPs) which would be used to rectify and georeference, the SPOT data to the UTM coordinate system. With these points, the transformation matrix could be calculated, and the rectified image could be processed further. Generally, at least three nonlinear ground control points are required to rectify an image. In this project, six ground control points scattered around Lake Icaria were selected (Figure 3-5).

The static and pseudo-static modes were selected to perform GPS observation in this project. The static mode was used to determine the location of an unknown point BASE. The pseudo-static mode was used to determine the locations of five other points with the point BASE considered to be known from the static mode.

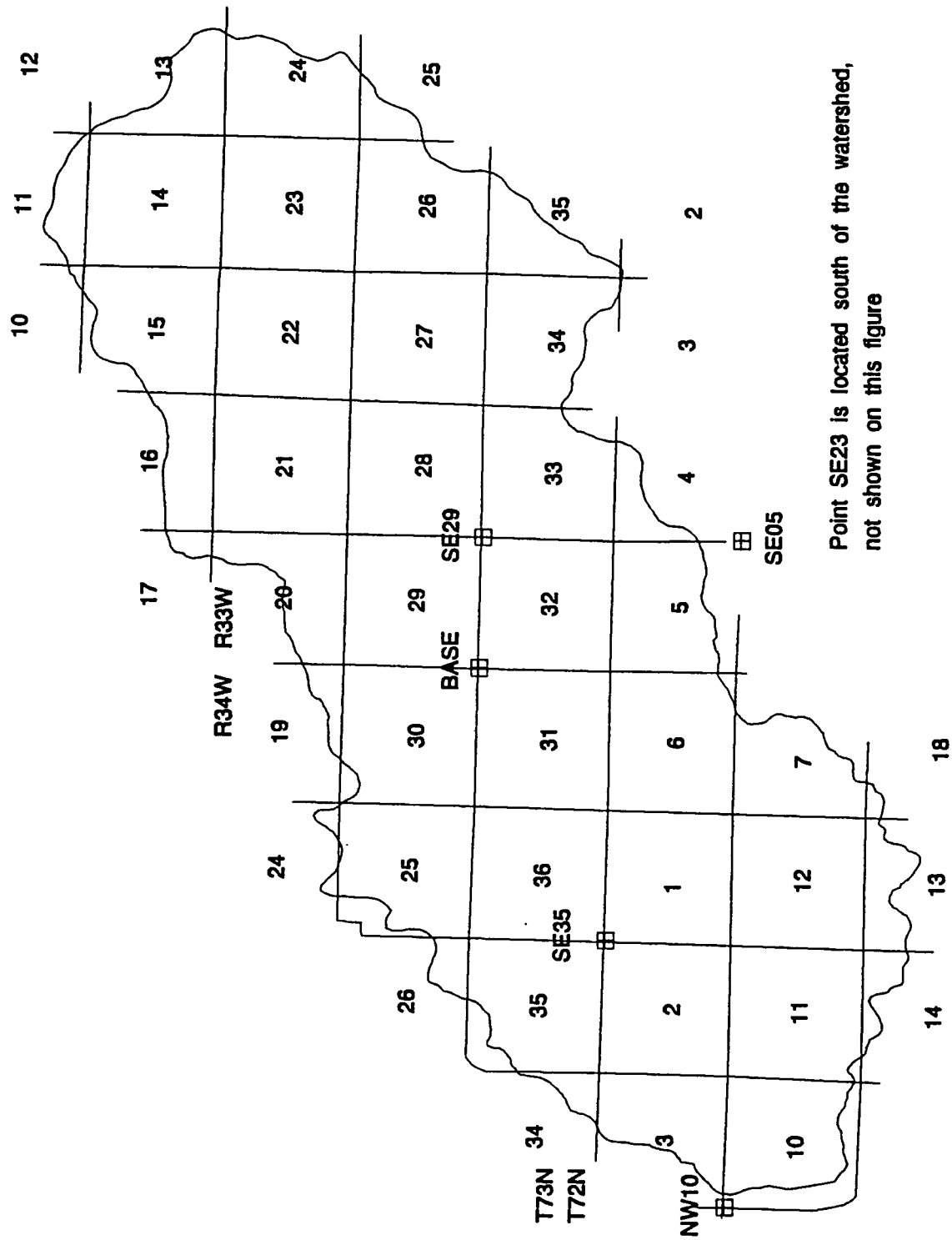


Figure 3-5. The Locations of GPS Points

Using GPS to Observe the Six Control Points

A two hour static GPS observation was completed first to establish the GPS control point BASE. The horizontal control station REICHARDT 2 (Cass County, Iowa, G.C.M., 1953) which was published by the U.S. Department of Commerce Coast and Geodetic Survey was treated as known control point during this static observation. The horizontal control data can be found in Table 3-1.

Table 3-1. The Coordinates of Control Point REICHARDT 2

Point	Longitude	Latitude	Elevation	G.Height
REICHARDT2	94°44′6″86856	41°10′20″9202	414m	-29.53m

During this observation, the receiver parameters set for these observations were 20 second epoch, 15 degree elevation cut off, and four minimum satellites. After this observation was completed, a pseudo-static GPS observation was carried out. Table 3-2 shows the longitude, latitude and elevation for each point. The point BASE was considered as the known point and the other five points were the unknown points.

Table 3-2. The Longitude and Latitude of GPS Points

Point	Longitude	Latitude	Elevation
Base	94°40′48″76143	41°5′9″9350	398.0394m
SE05	94°39′38″92755	41°3′28″65367	411.0937m
SE23	94°43′9″42966	41°0′54″56319	407.1175m
SE29	94°39′40″08033	41°5′10″09287	425.5349m
SE35	94°43′8″57246	41°4′18″82039	409.0976m
NW10	94°45′25″22159	41°3′29″44982	410.5755m

Comparison of GPS Coordinates and Digitized Coordinates

The GPS observation gave the geographic latitude-longitude of each of the control points based on the NAD-83 datum. But nearly all existing digital maps in our GIS system are based on the NAD-27 datum (Clarke 1866 ellipsoid). This affects not only the altitude values, but also shifts the latitude and longitude by significant amounts. In order to be consistent with the existing database which was initially developed by manual digitizing from USGS 7.5 minutes topographic maps, the geographic longitude and latitude coordinates were first projected to UTM coordinate system in zone 15. Then the UTM coordinates were transformed to NAD-27 ellipsoid from NAD-83 ellipsoid. Table 3-3 shows the UTM coordinates derived from GPS observation on NAD-83 datum, and Table 3-4 lists the UTM coordinates on NAD-27 datum.

Table 3-3. UTM Coordinate Derived from GPS on NAD-83 Datum

Point	X	Y	Z
REICHARDT2	354437.7242	4559143.1614	414.0000m
Base	358869.0674	4549462.3980	398.0394m
SE05	360439.0112	4546307.8874	411.0937m
SE23	355431.5959	4541651.3320	407.1175m
SE29	360471.7037	4549436.5490	425.5349m
SE35	355575.7487	4547949.7507	409.0976m
NW10	352355.9175	4546490.8471	410.5755m

Table 3-4. UTM Coordinate Derived from GPS on NAD-27 Datum

Point	X	Y
Base	358887.5054	4549459.7079
SE05	360457.3920	4546305.1633
SE23	355450.1293	4541648.5581
SE29	360490.0911	4549433.8579
SE35	355594.2886	4547947.0433
NW10	352374.5588	4546488.1269

Also, the UTM coordinates derived from USGS 7.5 minute maps were listed in Table 3-5. The differences of each pair of X, Y between GPS coordinates and digitized coordinates are represented in Table 3-5. From Table 3-5, the maximum X difference was -10.04 meters and Y difference was -15.88 meters, which equals 1/2 and 3/4 pixel width. The mean X difference was 2.6 meters and 4.3 meters, respectively. The Point NW10 contributed the most errors. This was because Point NW10 was located the furthest from the Point Base.

Table 3-5. UTM Coordinates from USGS 7.5 Minute Maps

Point	X	Y	ΔX	ΔY
Base	358886.313	4549458.964	1.20	0.74
SE05	360451.594	4546294.839	6.03	10.32
SE23	355447.875	4541638.965	2.25	9.60
SE29	360483.094	4549424.890	6.99	8.97
SE35	355585.094	4547934.986	9.20	12.06
NW10	352384.594	4546504.004	-10.04	15.88

The relationship between the GPS coordinates and the coordinates from USGS topographic maps could be analyzed by statistical methods. Linear regression models were calculated using Statistical Analysis System (SAS) software. These models were used as a predictive model to describe the relationships between these two types of coordinates. The models were in the following form:

$$X_g = a_{01} + a_{11}X_m + a_{21}Y_m \qquad \text{Eq. 3-3}$$

$$Y_g = a_{02} + a_{12}X_m + a_{22}Y_m \qquad \text{Eq. 3-4}$$

where (X_g,Y_g) are the GPS coordinates, and (X_m,Y_m) are the digitized coordinates.

Table 3-6 lists the ANOVA table and the estimated parameters.

It has been noted that the standard deviation in X and in Y were 6.60 and 9.70, respectively. They were about 3/10 and 1/2 of a single pixel width size. The comparison of GPS coordinates and the coordinates digitized from USGS 7.5 minute maps showed that the accuracy of point coordinate retrieval from USGS 7.5 minute series Quad maps was acceptable for the objectives and precision requirement for image rectification/georeferencing and for the digital map in GIS.

Table 3-6. Linear Regression ANOVA Table

Model One:

Source	DF	SS	MS	F value
Model	2	53207593.61	26603796.80	610361.197
Error	3	130.76092	43.58697	

Model Two:

Source	DF	SS	MS	F value
Model	2	42167383.52	21083691.76	224205.111
Error	3	282.11255	94.03752	

Standard deviation for model one is 6.60 and model two 9.70.

The regression models with estimated parameters:

$$X_g = 644.84 + 1.001493X_m - 0.000259Y_m \qquad \text{Eq. 3-5}$$

$$Y_g = 5968.82 + 0.002537X_m + 0.998489Y_m \qquad \text{Eq. 3-6}$$

PART II

DIGITAL IMAGE PROCESSING

CHAPTER 4. IMAGE ENHANCEMENT

Remotely sensed imagery is generally created by recording the reflected or/and emitted radiant flux from earth surface materials. Different materials generally reflect or emit different amounts of radiant flux throughout the electromagnetic spectrum. Most sensors designed for remote sensing usually detect the visible and near-infrared portion of the electromagnetic spectrum. The infrared and thermal infrared portions of the electromagnetic spectrum are used for detecting emitted heat. As discussed in Chapter 2, remote sensing techniques use the characteristic that different earth surface materials reflect or emit different amounts of energy in most of the electromagnetic spectrum. Sometimes different materials have the same or similar brightness values or Digital Numbers (DNs) because they reflect or emit the same amounts of energy in certain portions of the electromagnetic spectrum. Conversely, the existence of clouds or other atmospheric effects might result in the opposite problem. In other words, the same or the similar surface features might look different. This causes the image to have low contrast and makes it difficult to interpret different earth features from the original data.

The objective of the image enhancement process is to improve the visual interpretability of remotely sensed data by increasing the apparent distinction between the features in the scene.

Enhancement techniques are generally categorized as either point or local operations. Point operations change the DN of each pixel in an image data set independently. This means there is no other surrounding pixels involved in a point operation. The point operation generally refers to spectral enhancement. Local operations modify the DN of each single pixel based on its neighboring DNs. Local operation usually refers to spatial enhancement.

In this project, the spectral enhancement technique called ratioing enhancement was applied to eliminate some physical conditions, such as clouds or shadows, which might affect the quality of the image. Spatial enhancement (spatial filtering enhancement) was also used to obtain the "best" enhanced image for certain criteria which would be selected as a permanent image layer fed to the GIS system for further use and analysis.

Initial Statistics Extraction from SPOT Data

There were three bands in the SPOT data: The blue band (0.5μm to 0.59μm), green band (0.61μm to 0.68μm) and near-infrared band (0.79μm to 0.89μm). This data set had 751 rows and 1001 columns. The total pixel number was 751 X 1001 = 751751. Sample univariate statistics for every band in SPOT data were reported in tabular format, as shown in Table 4-1. In the univariate statistics, band1 exhibits the smallest variance (5.568), the biggest DN (123), the widest range of DNs (from 41 to 123), the highest mean value (62.77). These statistics are of value and could be used to determine the quality of the initial image.

Figure 4-1 shows the histograms of each band. It was noted that the distributions were roughly normal, except for a second small peak in each band which is the result of pixel falling on bodies of water.

Table 4-1. Statistics of SPOT data

Band #	Minimum DN	Maximum DN	Mean	Std. Deviation
Band 1	41	123	62.77	5.568
Band 2	26	114	54.58	7.854
Band 3	14	100	57.23	11.279

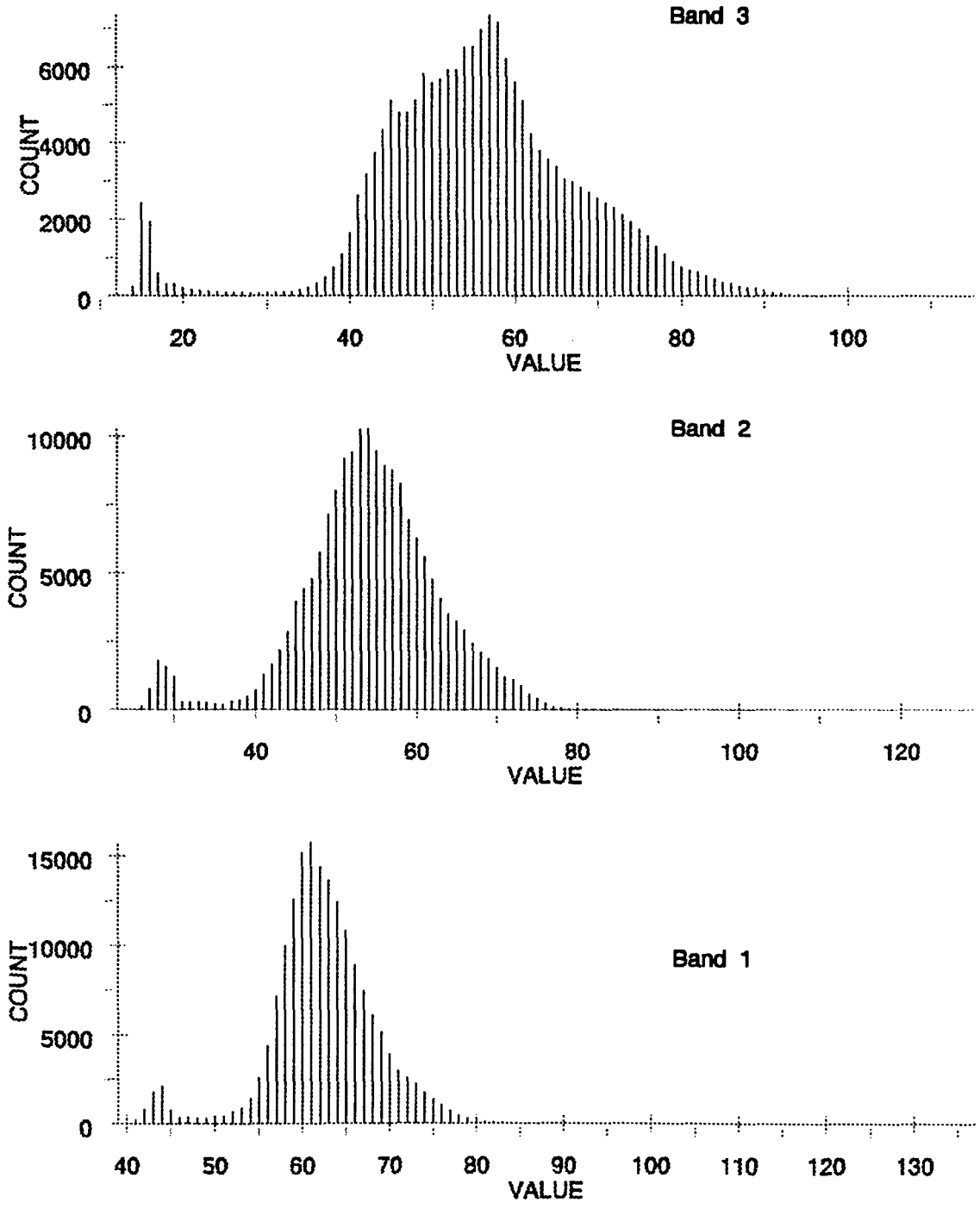
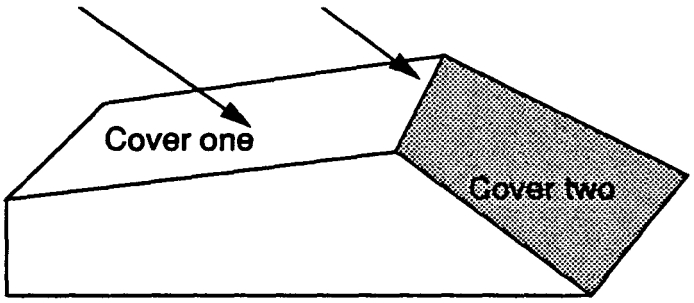


Figure 4-1. Histograms of Each Band

Spectral Enhancement

Sometimes differences in DNs from similar surface materials are caused by topographic conditions, shadows, or seasonal changes in sunlight illumination angle and intensity. This will certainly hamper the ability of an interpreter to correctly identify or classify the surface features from remotely sensed data. Ratio enhancement of remotely sensed data can, in certain instances, be applied to minimize the effects of these environmental conditions. This is so because the ratio values for similar cover features are nearly identical. In addition to minimizing the effects of environmental factors, ratios may also provide unique information not available in any single band. This ratio image results from the division of DNs in one spectral band, by the DNs at the same locations in another band. This concept is illustrated in Figure 4-2.



Land cover/illumination		Band A	Band B	Ratio (band A/band B)
Cover 1:	Sunlit	48	50	0.96
	Shadow	18	19	0.95
Cover 2:	Sunlit	31	45	0.69
	Shadow	11	16	0.69

Figure 4-2. Reduction of Scene Illumination Effects Through Spectral Ratio

In Figure 4-2, there are two different land covers on both sunlit and shadowed sides of an edge line. The DNs for each land cover are lower in the shadowed area than in the sunlit area. It should be noted that the ratio values for each cover type are nearly identical, irrespective of the illumination condition.The mathematical expression of the ratio function is:

$$DN_{i,j,r} = \frac{DN_{i,j,k}}{DN_{i,j,l}} \qquad \text{Eq. 4-1 (2)}$$

where $DN_{i,j,r}$ is the output ratio value for the pixel at row i, column j; $DN_{i,j,k}$ is the DN for the pixel at row i, column j in band k; and $DN_{i,j,l}$ is the DN in band l.

During ratio enhancement processing, many combinations of bands were used to select the "best" ratioed image. The "best" image was defined to yield the sharpest contrast between road intersections and their surrounding fields. Finally, the ratio of band 3 and band 1 was selected as the "best" black and white image for interpretation purpose because it gave the sharpest location of transportation intersections. These intersection locations were determined by the GPS observation (which has been discussed in a former chapter) and used as control points to rectify the raw SPOT data.

Spatial Enhancement

A characteristic of remotely sensed images is a parameter called spatial frequency, defined as "the number of changes in brightness value (DN) per unit distance for any particular part of an image" (2). If there are very few changes in DN over an area, this is called a "low-frequency" area. Conversely, if the DN change rapidly over very short distances, this is a "high-frequency" area.

When conducting spectral enhancement, operations are performed on pixel. Spatial enhancement, on the other hand, modifies the value of a pixel based on the values of its surrounding pixels. A low spatial frequency image consists of a smoothly

varying gray scale which will be processed by applying a low pass filter to emphasize the low frequency cover features. A high spatial frequency image consists of rapidly changing black and white pixels which could be processed by using high pass filter to emphasize the detailed high frequency components of an image.

In practice, a convolution filter which is expressed in a matrix of numbers is usually used to calculate a weighted average for the value of each pixel based on the values of surrounding pixels. These numbers are often called coefficients. To understand how a convolution filter works, imagine this filter is put on the top of the data file values of the image, so that the pixel to be convolved corresponds to the coefficient of the filter, as illustrated in Figure 4-3.

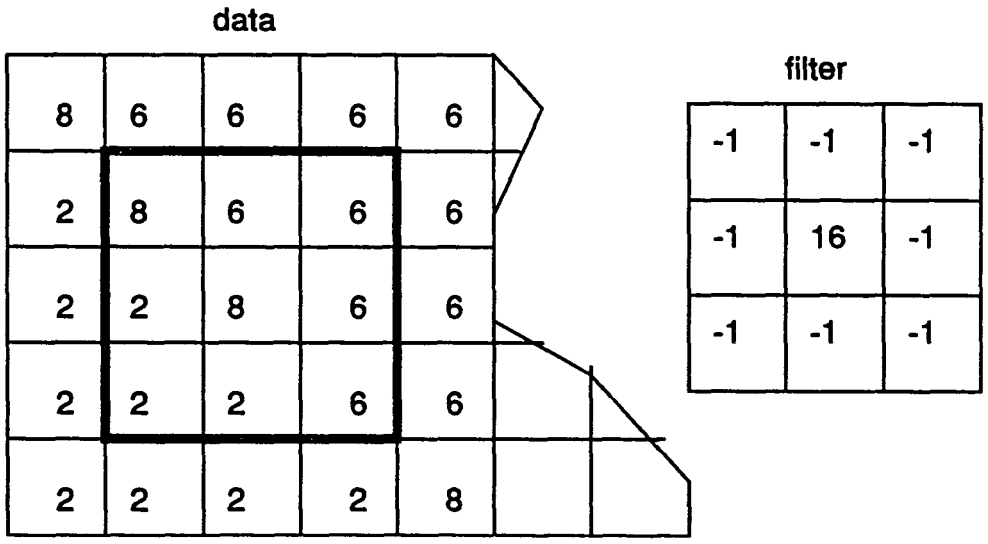


Figure 4-3. Applying a Convolution Filter

The convolution filter used here is a 3x3 matrix. To compute the output value for center pixel during the motion of the filter, the convolution filter is multiplied by the pixel value that corresponds to it. Then the products are summed, and the total is divided by the sum of the values in the filter, as shown in this computation:

int ((-1x8) + (-1 x 6) + (-1 X 6) + (-1 x 2) + (16 x 8) + (-1 X 6) +(-1 x 2) + (-1 x 2) + (-1 x 8) / (-1+-1+-1+-1+16+-1+-1+-1+-1)) = int ((128-40)/(16-8)) = int (88/8) = 11

So, after this 3 x 3 filter is convolved on the image, the output values are as shown in Figure 4-4.

2	8	6	6	6
2	11	5	5	6
2	0	11	6	6
2	1	0	11	6
2	2	2	2	8

Figure 4-4. Convolved DN Values

The filter used in the above example was a high frequency filter. It was noted that the relatively higher values became higher, and the lower values became lower, thus increasing the spatial frequency of the image.

The following formula is used to derive the output data value for the pixel convolved at the center:

$$V = \text{int}\left[\frac{\sum_{i=1}^{q}\left(\sum_{j=1}^{q} f_{ij}d_{ij}\right)}{F}\right] \qquad \text{Eq. 4-2 (12)}$$

where:

f_{ij}= the coefficient of a convolution filter at position i,j,

D_{ij}= the value of the pixel that corresponding to f_{ij},

q = the dimension of the filter, assuming a square filter (if q = 3, the filter is 3 x 3),

F = either the sum of the coefficients of the filter, or 1 if the sum of coefficients is zero,

int = the integer function, which truncates the non-integer number to an integer, and,

V = the output pixel value.

A high-frequency convolution filter, or high pass filter is used in this project to enhance the spatial difference for the SPOT image.The high pass filter used is shown in Figure 4-3.

From the output image (shown in Figure 4-6), we can conclude that this filter serves as an edge enhancer, since it brings out the edges between homogeneous groups of pixels. This enhanced image will be used as a base map to rectify the SPOT data to the UTM coordinate system which will be explained in details in Chapter 5.

Figure 4-6. Convolved Image

CHAPTER 5. IMAGE RECTIFICATION

The remotely sensed data is a representation of the irregular surface of the earth which has some distortion due to systematic and non-systematic geometric errors. Because the satellite is extremely stable, tilt distortion is small and because the flying altitude of the spacecraft is great, topographic displacement which is inversely proportional to the flying height of the satellite, is also very little. Therefore, satellite imagery is nearly orthographic. How to remove the systematic error is beyond the scope of this study. This process is usually completed by the raw data supplier. It is our concern to correct the non-systematic error remaining in the image. This non-systematic error always makes the image non-planimetric.

The objective of rectification is then to make the geometry of an image area planimetric, so it can be represented on a plane surface, or so that it conforms to other images. That means the image rectification process will project the raw data onto a plane and make it conform to a map projection system.

Universal Transverse Mercator Projection System

A map projection is the way in which the spherical surface of the earth is represented on a two-dimensional surface. The cylinder and cone plane are the two most frequently used flat surfaces. The Transverse Mercator projection is made from a normal Mercator cylindrical projection. The projection cylinder is rotated 90 degrees from the North-South polar axis and can then be placed to intersect at a chosen central meridian. The central meridians, the equator, and each line 90 degrees from the central meridian are straight lines. Figure 5-1 shows the two most commonly used projections. Figure 5-2 shows a Transverse Mercator projection.

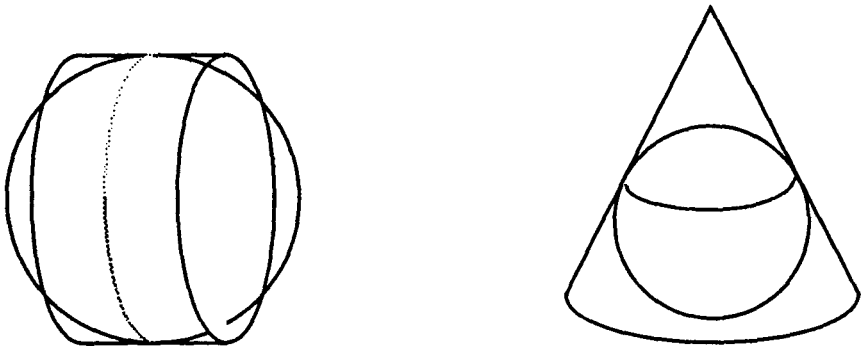


Figure 5-1. Cylindrical and Conic Plane

The Universal Transverse Mercator (UTM) is an international plane (rectangular) coordinate system developed by the U.S. Army which extends around the world from 84 degrees North to 80 degrees South. The world is divided into 60 zones each covering six degrees longitude. Each zone extends three degrees eastward and three degrees westward from its central meridian. Zones are numbered consecutively west to east from the 180 degree meridian (Figure 5-3). From 84 degrees North and 80 degrees South to the respective poles, the Universal Polar Stereographic (UPS) is used. Figure 5-4 illustrates the UTM zones used in the United States. Lake Icaria Watershed lies in UTM zone 15.

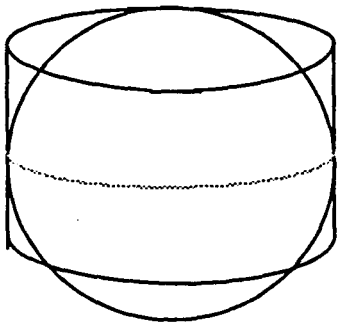


Figure 5-2. Transverse Mercator

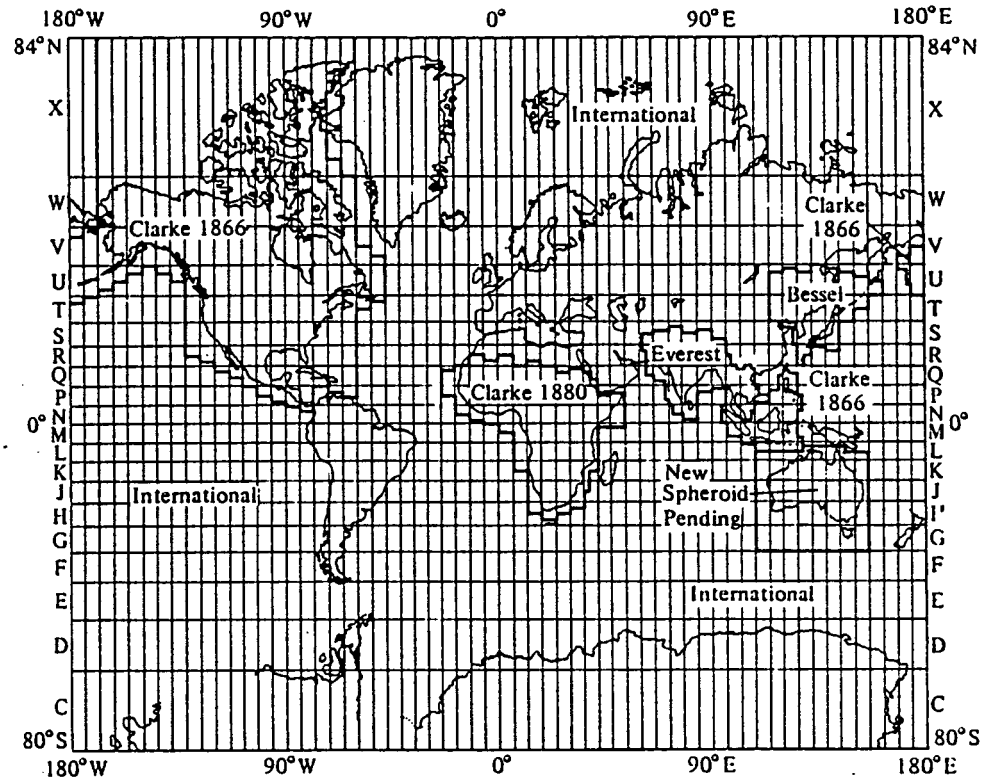


Figure 5-3. Universal Transverse Mercator (UTM) (12)

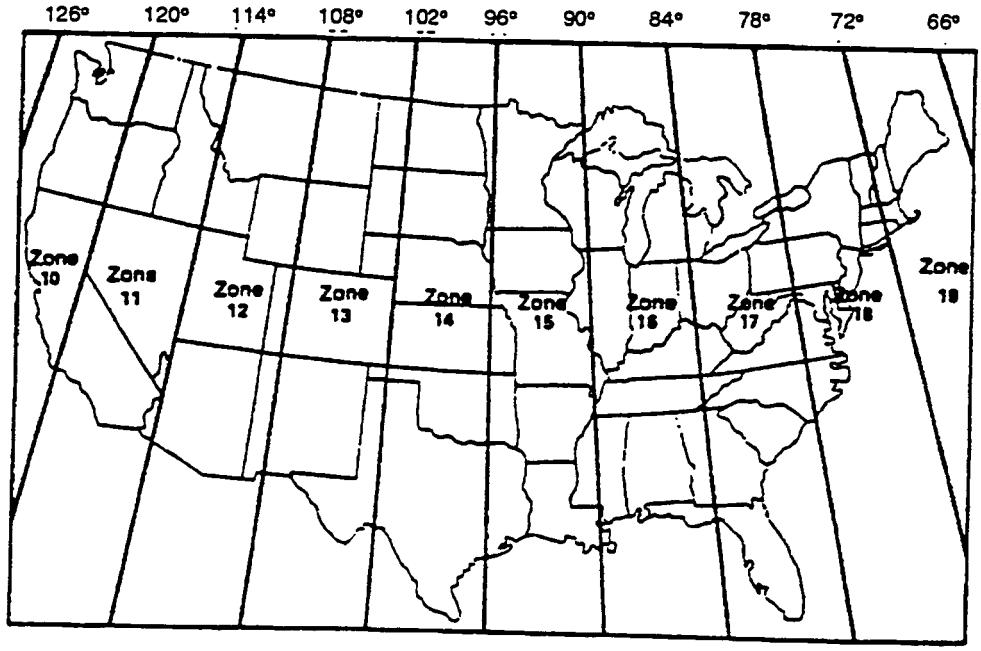


Figure 5-4. UTM zones in the United States (12)

Spatial Interpolation using Ground Control Points (GCPs)

Accurate ground control points (GCPs) are essential to complete a high quality rectification. Rectification process will automatically interpolate all other non-control points relating to the control points. Ideally, as many GCPs should be located throughout the study area as possible. The more dispersed the GCPs are, the more reliable the rectification.

A spatial interpolation is a process that establish the relationship between input pixel (row and column) location and the associated map coordinate of this same point (x, y). Now suppose that the two coordinate systems can be related via a pair of functions f and g so that

$$x' = f(x, y) \qquad \text{Eq. 5-1}$$

$$y' = g(x, y) \qquad \text{Eq. 5-2}$$

where: (x', y') is the position in the source coordinate system (input), and (x, y) is the corresponding position in the output-rectified coordinate system.

For moderate distortions in a relatively small area of an image, a six-parameter affine transformation is sufficient to rectify the imagery to a geographic frame of reference. This six-parameter affine transformation can model six kinds of distortion in the remote sensing data, including: translation in x and y, scale changes in x and y, skew, and rotation (2). A first order transformation or so called linear transformation was apply in this project. All six parameters can be expressed in the following equations:

$$x' = a_1 x + a_2 y + a_0 \qquad \text{Eq. 5-3}$$

$$y' = b_1 x + b_2 + b_0 \qquad \text{Eq. 5-4}$$

or

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_0 \\ b_1 & b_2 & b_0 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad \text{Eq. 5-5}$$

In order to solve the six unknowns in above equations, at least three ground control points are needed. If more than three ground control points are used, the unknowns are determined by least square regression method. This will cause the Root Mean Square (RMS) error not equal to zero.

The RMS error is to measure the image distortion not corrected for by the six-parameter coordinate transformation. A transformation matrix is generally created in the digital image process. The inverse of this transformation matrix is also calculated to retransform the reference coordinates (map coordinates) of the GCPs back to the source coordinate system (rows and columns). When first order transformation was used in this project, some discrepancies between the source coordinates and the retransformed coordinates from reference coordinates were introduced.

RMS error is defined as the distance between the input location of a GCP, and the retransformed location for the same GCP in the same source coordinate system. RMS error is calculated with a distance for each GCP using the Equation 5-6:

$$RMS = \sqrt{(x' - x_o)^2 + (y' - y_o)^2} \qquad \text{Eq. 5-6}$$

where: (x', y') are the input source coordinates, and (x_o, y_o) are the retransformed coordinates.

By computing all GCPs, it is possible to see which GCPs exhibit the greatest error and to calculate the sum of all the RMS errors. RMS error is expressed as a distance in pixel width in the source coordinate system. For example, a RMS error of 2 means that the reference pixel is 2 pixels away from the retransformed pixel.

Normally, a tolerance of RMS error is specified by the user. The tolerance value of RMS error specified can be thought of as a window around each source coordinate, inside which a retransformed coordinate is considered to be correct (that is, close enough to use). A RMS error tolerance of 2 specifies a distance of 2 pixels from the source pixel into which a retransformed location is considered within an acceptable range of accuracy.

The RMS error tolerance can be shown in Figure 5-5. The circle whose radius is 2 pixel width represents the range of the accuracy.

In this project, six GCPs are selected and the RMS error tolerance is set to be 0.5, which approximately is 10 meters on the ground.

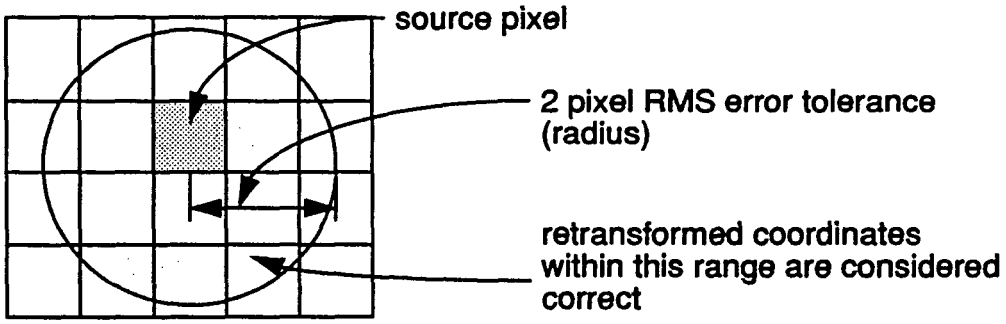


Figure 5-5. RMS Error Tolerance

Intensity Interpolation (Resampling)

Since the grid of pixels in the source image cannot perfectly match the grid for the reference image, the pixels are resampled so that new data DN values for the output data can be calculated (Figure 5-6).

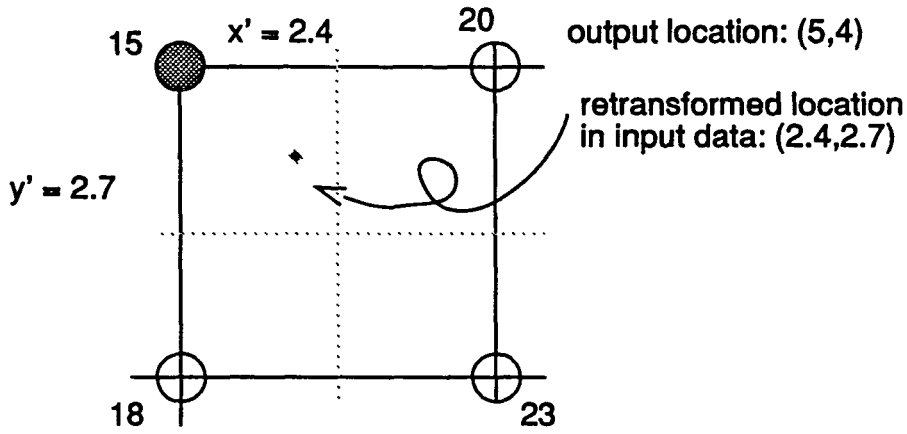


Figure 5-6. Nearest Neighbor Interpolation

This process involves the extraction of a DN from the x′, y′ location in the original (distorted) input image and its relocation to the corresponding x, y location in the output rectified image. Generally the x′ and y′ coordinates to be retransformed are not integers but real numbers. This process can be illustrated in Figure 5-6. The DN value at pixel 5,4 (x, y) in the output image is to be "filled" with the value from the coordinates 2.4, 2.7 (x′, y′) in the input image. The DN value for coordinates at (2.4, 2.7) can be interpolated in several ways. Nearest-neighbor interpolation uses the DN value of the pixel closest to the x′, y′ coordinates to assign to the output value at the corresponding coordinates x, y. For example, in Figure 5-6, the output value at pixel 5,4 (x, y) needs the brightness value (DN) in the input image at location 2.4, 2.7 (x′, y′). There is no value at this position; however, a nearest-neighbor interpolation will assign

the output pixel (x, y) the value of 15, which is the value at the nearest input pixel.

Another widely used interpolation method is bilinear interpolation which assigns the output pixel values by interpolating the DN values of the surrounding four pixels based on the weighted distances to these pixels in a 2 by 2 window. In this project, nearest-neighbor interpolation is applied to resample the output rectified image.

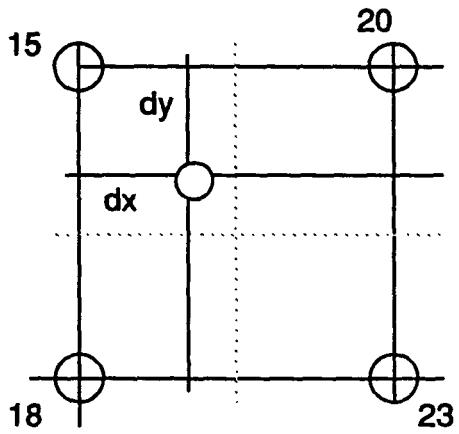


Figure 5-7. Bilinear Interpolation

Rectification Procedure

In this project, the Corning North, Iowa, USGS 7.5-minute quad map (photorevised 1979) was selected as the appropriate base map with which to rectify the SPOT data. Five ground control points were located on the map and the UTM easting and northing of each point were retrieved from the digitized quadrangle map which were developed by Iowa State University GIS Research and Support Facility (Table 5-1). The coordinate accuracy of these points has been discussed in Chapter 3. These ground control points were selected based on the rule that the GCPs should normally

be located throughout the region to be rectified. They were also identified as easy-accessed after field survey. The same five GCPs were then identified in the SPOT data according to their row and column coordinates (also in Table 5-2).

Table 5-1. UTM Coordinates of GCPs and Image Coordinates

Point	Easting	Northing	x pixel	y pixel
BASE	358886.313	4549459.000	1439	2443
SE05	360451.594	4546295.000	1548	2581
SE29	360483.094	4549425.000	1517	2429
SE35	355585.094	4547935.000	1293	2552
NW10	352384.594	4546504.000	1151	2656

Table 5-2. The Residuals when five GCPs Used

Point	Image X	X Residual	Image Y	Y Residual
BASE	1438.77	-0.2290E+00	2443.47	0.4702E+00
SE05	1547.95	-0.5005E-01	2581.12	0.1248E+00
SE29	1517.26	0.2570E+00	2428.35	-0.6479E+00
SE35	1292.94	-0.5896E-01	2552.37	0.3701E+00
NW10	1151.08	0.8105E-01	2655.68	-0.3171E+00

X RMS error= 0.16188 and Y RMS error= 0.42286

Total RMS error= 0.45279

Table 5-3. The Error Contribution by Each Point

Point	Error	Error Contribution by Point
BASE	0.5230	1.1551
SE05	0.1344	0.2969
SE29	0.6970	1.5394
SE35	0.3748	0.8277
NW10	0.3273	0.7229

The final transformation matrix is:

$$\begin{bmatrix} 19.549951 & -4.133923 & 340859.44 \\ -4.2162692 & -19.6401659 & 4603515.4 \end{bmatrix} \qquad \text{Eq. 5-7}$$

The five GCPs were input to the least-squares regression procedure previously discussed to identify (1) the coefficients of the coordinate transformation matrix, and (2) the individual and total RMS errors associated with the GCPs. The output is summarized in Table 5-3. The threshold of 0.5 was satisfied. The six coefficient transformation matrix is found in Equation 5-7. Once the coefficients of the transformation matrix were established, the type of resampling to be performed was specified as nearest-neighbor interpolation. The output pixel size is set to be 20m X 20m which is equal to the SPOT data pixel size.

Because the data is rectified to the USGS 7.5-minute quad maps, it is possible to have the remotely sensed data precisely overlay other digital maps (e.g., roads; streams) in the same area which were digitized from the same quad map.

CHAPTER 6. IMAGE CLASSIFICATION

The objective of digital image classification is to automatically categorize all pixels in an image into different classes or themes by using computer software. Generally, multispectral data are used during the classification, and the spectral pattern presented in each band within the image for each pixel is used as the numerical basis for categorization. That is, different feature types manifest different combinations of digital number values based on their inherent spectral reflectance and emittance properties. Depending on the type of information to be extracted from the original data, classes can be associated with known features on the ground or may simply represent areas that "look different" to the computer.

Supervised and Unsupervised Classification

The classification process usually proceeds in two steps: a training process and a classifying process. First, the computer system must be trained to recognize patterns in the data. The result of training is a set of signatures, which are statistical criteria for a set of proposed classes. Then the pixels in the image data are sorted into classes based on the signatures, by use of a classification decision rule which is a mathematical algorithm that uses particular statistics, contained in the signature, to sort the pixels.

Two image classification procedures are usually used for quantitative analysis of remotely sensed image data. One is called supervised classification and the other is unsupervised classification. In supervised classification the pixel categorization process is supervised by the image analyst. Certain pixels, called training samples, are first selected that present patterns recognized or identified with help from other sources, such as aerial photos, ground truth data or maps. Ground truth data usually

refers to the knowledge about the study area from field survey, analysis of aerial photography, or personal experience, etc. Ground truth data are considered the most accurate (true) data available about the area of study. Ground truth data should be collected at the same time as the remotely sensed data was taken, so that the data correspond as much as possible. The computer system is first trained to recognize pixels with similar characteristics in the supervised classification process. Each pixel in the data is compared numerically to each class in the training areas and labeled with the name of the class it "looks most like". The unsupervised procedure is also applied by two separate steps but does not utilize training data as the basis for classification. The image data are first classified by aggregating them into the natural spectral groupings or clusters presented in the scene. The basic assumption is that values within a given cover type should be close together in the measurement space, whereas data in different classes should be comparatively well separated. After classification, these spectral groups are identified and labeled by comparing the classified image data to ground truth data.

Training Process in Supervised Classification

The objective of the training process in supervised classification is to generate a set of statistics that describes the spectral response pattern for every land cover type to be classified in an image. It is during this process that the location, size, shape, and orientation of the pixels for each land cover class are determined. Very importantly, the quality of the training process will determine the success of the classification and, therefore, the value of the information derived from the entire classification effort. Generally, the training samples are created by outlining training area boundaries on a computer screen. This process can be done with the help of some reference data as background, such as rectified photograph, to reach the required accuracy.

A four section area was selected to create the signature file by supervised classification method in this project. The ground truth data about these four sections were supplied by Adams County ACSC. There are 234 X 193 = 45162 pixels in this sample area. The color infrared photography at a scale of 1:2000 was also used as reference. These four sections are located in T73N R33W section number 31 and 32 and T72N R33W section 5 and 6. Figure 6-1 shows the four sections and their locations in the study area.

The training procedure for supervised classification was carried out with help of ERDAS image processing tools. There were seven signatures in the final signature file. They were water, forest, cedar, pasture, corn, soybean, and an unknown crop. These seven land cover types basically represented major land covers in this area.

Evaluation of Signature File

The signature file was analyzed by spectral distance evaluation methods. The spectral distance between the means of each pair of signatures was computed by ERDAS system. The output is listed in Table 6-1. From Table 6-1, the minimum spectral distance was 8.00 which occurred between soybean and unknown crop. This implied that the reflectance of soybean in April is similar with that of the unknown crop.

Classification Decision Rules

A signature associated with a classification process is a set of statistical information which defines a training sample or cluster. This statistical information is required by the classification decision rule. Specifically, the signature includes the name of each signature, the number of bands used in the training process, the minimum, maximum and mean data file values in each band for each sample or cluster, and the covariance matrix for each sample or cluster.

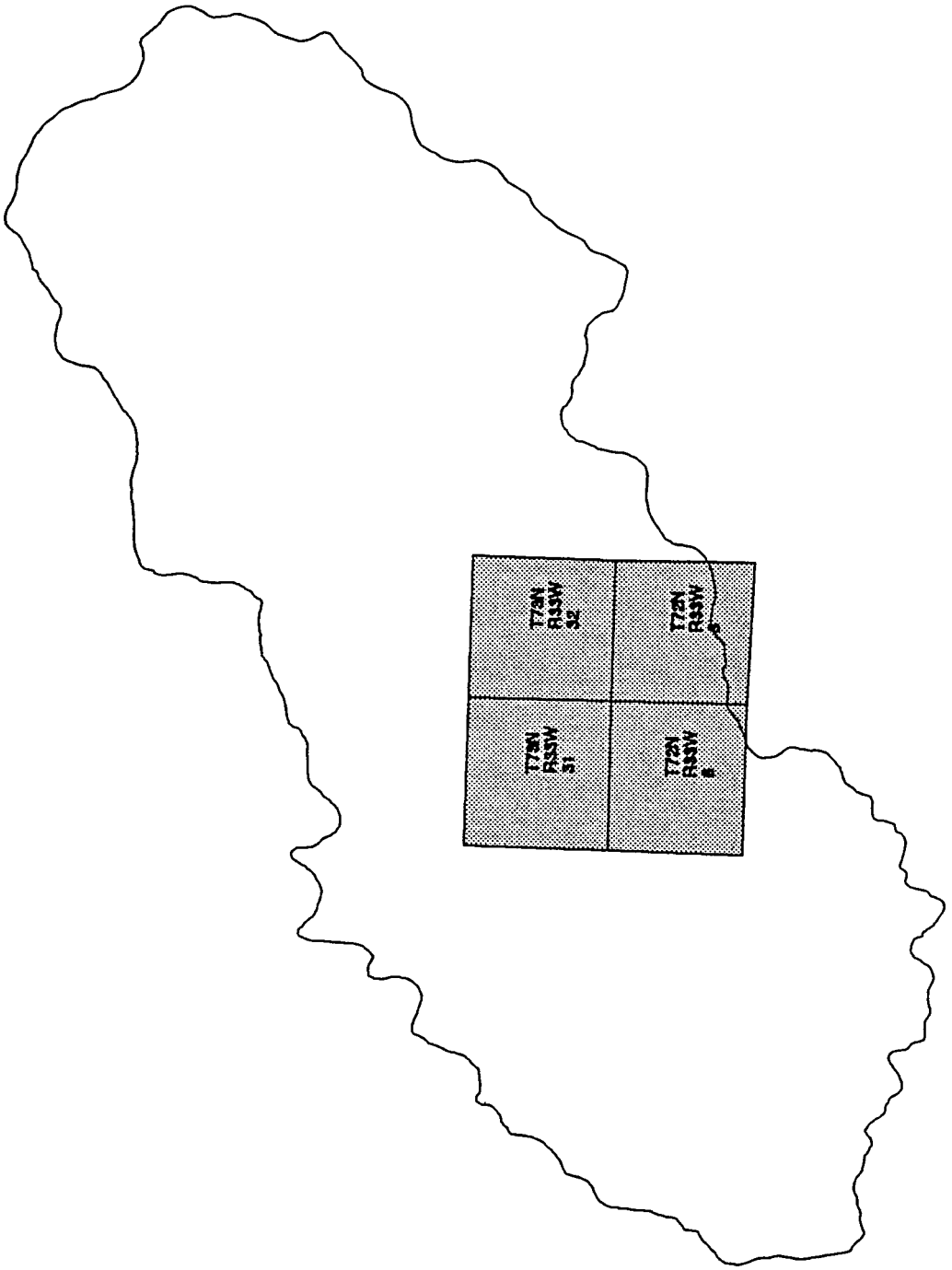


Figure 6-1. Four Sections used as Ground Truth Data

Table 6-1. Euclidean Distance Table

	Water	Forest	Cedar	Pasture	Corn	Soy	Unknown
Water	00.00	30.99	44.02	68.22	39.26	49.52	56.48
Forest	30.99	00.00	19.05	39.12	11.61	20.59	28.38
Cedar	44.02	19.05	00.00	25.08	25.43	28.43	35.24
Pasture	68.22	39.12	25.08	00.00	38.37	33.76	35.94
Corn	39.26	11.61	25.43	38.37	00.00	10.29	17.49
Soy	49.52	20.59	28.43	33.76	10.29	00.00	08.00
Unknown	56.48	28.38	35.24	35.94	17.49	08.00	00.00

One of the most important parameters of signature data is the spectral distance between the mean vectors of each pair of signatures which has been discussed in the former section. The signatures will not be distinct enough if the spectral distance between two classes or clusters is not significant. The spectral distance is also the basis of the minimum distance classification (as explained in the following section). The classification process is performed after the signature file is evaluated. During this process, each pixel is analyzed independently and is assigned to a certain class if this pixel passes the criteria that are established by the decision rule. The most commonly-used decision rules are: parallelepiped, minimum distance, and maximum likelihood decision rules.

The parallelepiped classification algorithm is based on simple Boolean "and/or" logic. The values of any candidate pixel are compared to upper and lower limits which could be specified in different ways. The limits might be identified by the minimum and maximum data file values of each band in the signature, or by the mean of each band plus and minus a number or standard deviation. When the data value of the candidate pixel falls into the limits for every band in a signature cluster then this pixel is assigned

to that cluster. It is possible that a candidate pixel might satisfy the criteria of more than one class if it falls into overlapping parallelepipeds. In this case, the candidate pixel will be assigned to the first class for which it meets all criteria. For the pixels falling outside any parallelepipeds, no classification occurs (Figure 6-2). Table 6-2 shows the classification output for the four-section sample area by using parallelepiped decision rule.

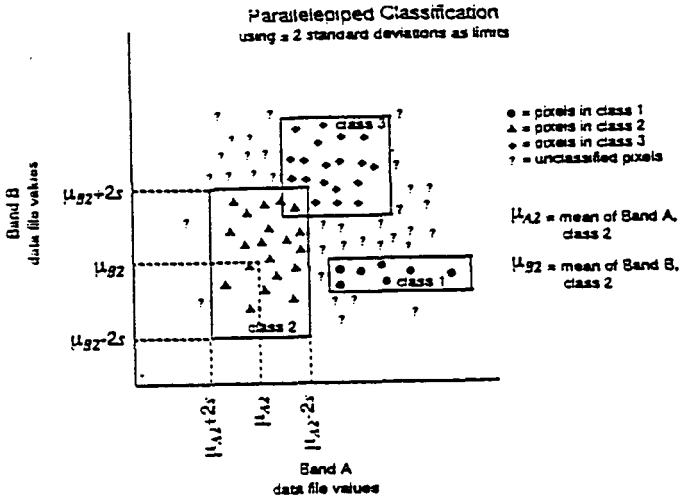


Figure 6-2. Parallelepiped Decision Rule (12)

Table 6-2. Parallelepiped Classification Output

Class	Classified Point #	Percent of total
Not classified	18659.0 points	41.316%
Water	749.0 points	1.658%
Forest	456.0 points	1.010%
Cedar	125.0 points	0.277%
Pasture	6220.0 points	13.773%
Corn	8959.0 points	19.837%
Soybean	7761.0 points	17.185%
Unknown	2233.0 points	4.944%

Another decision rule is called minimum distance decision rule. The spectral distances between the candidate pixel and the mean of each class is calculated (Figure 6-3).

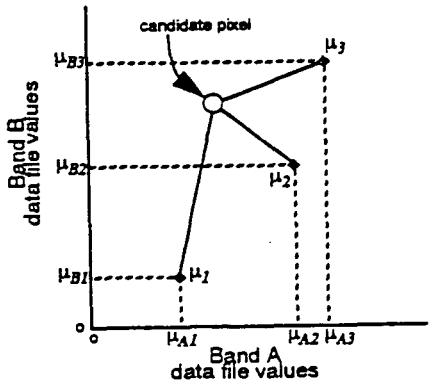


Figure 6-3. Minimum Spectral Distance Decision Rule (12)

After the spectral distance is computed for all possible values of all possible classes, the candidate pixel is assigned to the class for which the distance is the shortest. Since any pixel will find one closest class to it, there will be no unclassified pixel. Table 6-3 shows the classification output for the sample image by using minimum spectral distance decision rule.

Table 6-3. Minimum Spectral Distance Classification Output

Class	Classified point #	Percent of total
Water	1173 points	2.597%
Forest	1599 points	3.541%
Cedar	1664 points	3.685%
Pasture	12478 points	27.629%
Corn	7923 points	17.544%
Soybean	13276 points	29.396%
Unknown	7049 points	15.608%

The maximum likelihood algorithm is commonly used if the training sample data statistics for each class in each band are normally distributed. This decision rule is based on the probability that a pixel belongs to a particular class. For an unknown pixel, the maximum likelihood decision rule computes its probability values for each class. Finally, it assigns the pixel to the class that has the largest (or maximum) probability. The maximum likelihood decision rule is the most accurate classifier among these three major algorithms because it takes most variables into consideration. Obviously, the maximum likelihood decision rule requires more computations per pixel than either parallelepiped or minimum distance classification algorithms.

Figure 6-4 shows the equiprobability contours for a sample area which has some typical land covers, such as urban, sand, corn, hay, water and forest based on multispectral data band 4 and band 3. These equiprobability contours can be considered as the sensitivity outlines of the likelihood classifier to any unclassified pixel. Based on this theory, pixel 1 would be categorized corn because it has the maximum probability value when it falls into corn category. Table 6-4 shows the maximum likelihood classification output.

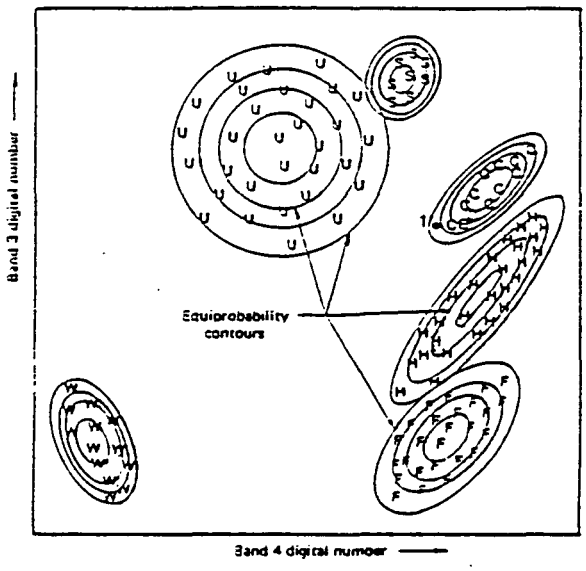


Figure 6-4. Equiprobability Contours in Maximum Likelihood Decision Rule (1)

Table 6-4. Maximum Likelihood Classification Output

Class	Classified point #	Percent of total
Water	1038 points	2.298%
Forest	763 points	1.689%
Cedar	659 points	1.459%
Pasture	16017 points	35.466%
Corn	8019 points	17.756%
Soybean	13225 points	29.283%
Unknown	5441 points	12.048%

The sample image was also classified using the maximum likelihood decision rule. The classification statistics is list in Table 6-4.

Classification Accuracy Assessment

The overlay method was first used to evaluate the classification. The original data were displayed as a color infrared image on the screen, and then the classified image was overlaid. The ground truth data and color infrared photographs were also used to assess the classifications. It was concluded that the image classified by using maximum likelihood classification was closer to the ground truth data than that by parallelepiped and minimum distance decision rules.

A complete accuracy assessment of a classification map would require verification of the class of every pixel. Obviously this is impossible and indeed defeats the purpose of image classification. Therefore, representative test areas must be used to estimate the map accuracy with as little error as possible. It is acceptable that the accuracy of a classification map is estimated by dividing the number of correctly classified test pixels in the class by the total number of test pixels in that class. The correct

classes of the test pixels are determined from independent information, which usually are ground truth data and aerial photography.

Obviously, the pixels used to create training sample in supervised classification can not be used as checking pixels, because this estimation is strongly biased and invariably yields overly optimistic accuracy figures. An approach called random pixel sample testing was used because it is the less biased method. Forty five reference pixels (about 1/1000th of total sample image pixels) were selected randomly by ERDAS so that possibility of bias is lessened or eliminated. The true GIS values for each of these 45 pixels were entered according to the ground truth data. The classification-assigned GIS values for each pixel was compared. The accuracy report shows the percentage of accuracy as in Table 6-5.

Table 6-5. Accuracy Assessment of Classification

Class	Reference	Classified #	Correctly Classified #
Background	0	0	0
Water	2	1	1
Forest	2	1	1
Cedar	0	0	0
Pasture	20	16	15
Corn	9	8	6
Soy	8	13	7
Unknown	4	6	4
Totals	45	45	34

Overall Classification Accuracy = 75.56%

The supervised maximum likelihood classification used in this project yielded about 76% classification accuracy. The same signature file was applied to classify the image covering the whole watershed. Figure 6-5 illustrates the classified land cover.

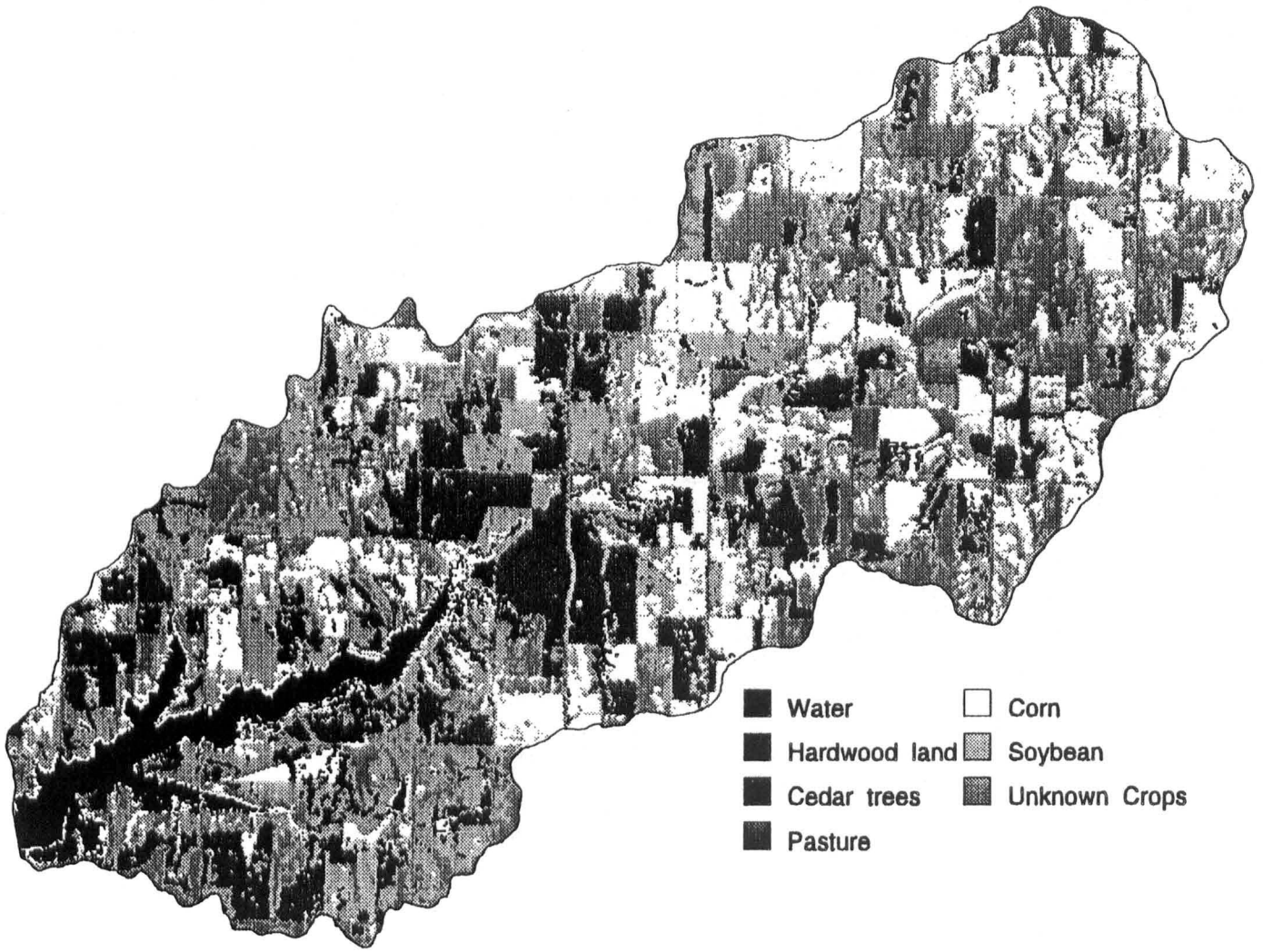


Figure 6-5. Classified Land Cover

CHAPTER 7. POSTCLASSIFICATION PROCESS

The image was converted to ARC/INFO grid after it was classified by ERDAS. This classified image (land cover) was clipped to the area we were interested in, so we could save storage space and computer time when further processing was needed.

Because of the errors that occur either in the classification process or from spurious data provided by the satellite, a 'noisy' appearance often exist. That is, there are many isolated pixels or small groups of pixels whose classification is different from that of their neighbors. Many ARC/INFO GRID functions were applied to clean up noisy or heterogeneous data. Thus it was possible to filter out small regions, smooth borders, expand, shrink, or thin selected zones, and fill holes in data.

Filtering

The classified image was first filtered by majority rule which means the value of each pixel is reassigned based on the majority of its contiguous neighbors, to get rid of the small regions. A spatial window is passed through the classification map, and at each pixel the majority class within the window is determined. If the majority class is different than the class of the center pixel in the window, the center pixel's classification is changed to the majority class. If there is no majority class in the neighborhood, the center pixel is left unchanged. It should be noted that at each new position of the window, only the original pixel classification value is used by the algorithm, not the reassigned value from the previous window position. Figure 7-1 illustrates the majority algorithm. Figure 7-2 shows the filtered land cover data by majority algorithm.

(a) CENTRAL PIXEL CHANGED

original classification			class	# pixels	final classification		
A	A	A	A	6	A	A	A
A	C	B	B	1	A	A	B
C	A	A	C	2	C	A	A

(b) CENTRAL PIXEL UNCHANGED

original classification			class	# pixels	final classification		
A	A	B	A	3	A	A	B
C	C	B	B	3	C	C	B
C	B	A	C	3	C	B	A

Figure 7-1. Majority Algorithm (5)

Boundary Simplification

Another technique usually used in practice is called boundary cleaning. The boundaries between zones in a classified image might be simplified or smoothed by removing the pits and bumps at the boundary between zones. There are two steps involved in boundary cleaning. The first step is to expand the selected zones, followed by the second step to shrink the same zones. This expand-then-shrink process can be illustrated in Figure 7-3. The first step is to expand the higher priority zone, into the zone of lower priority, by one cell in all eight directions. If a no-data area exists it will have the lowest priority. So all zones can invade all no-data areas. The second step is to shrink any cell in the expanded zone back to its original value if the cell is not completely surrounded by cells of the same value. In other words, the cell value will change if a cell is not the center of a 3-by-3 window of cells. A cell in the original input which has no neighbors of the same value (a single cell region) will not recover its original value after the shrinkage. This works just like the majority filter function we discussed

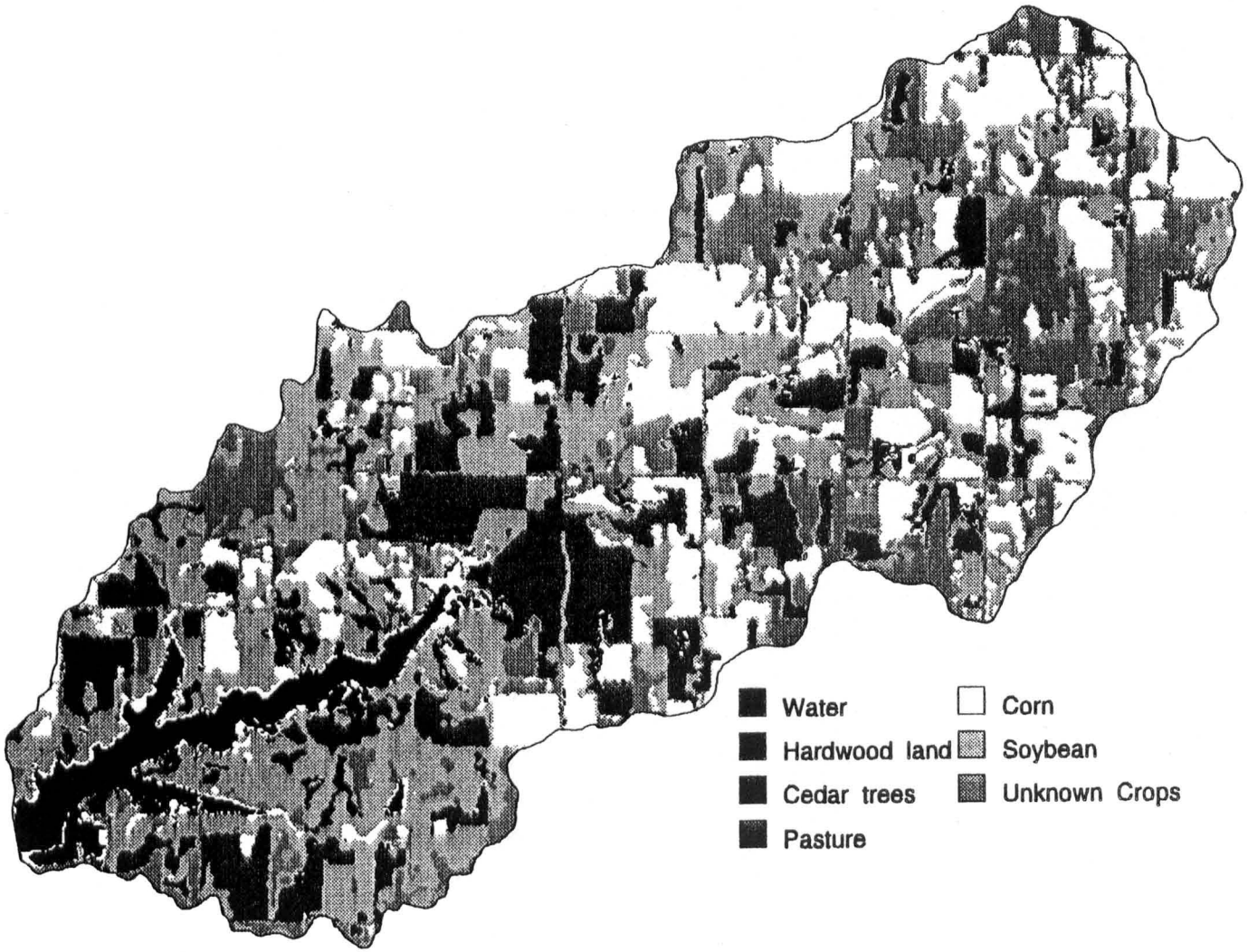


Figure 7-2. Filter Data

in the former section. It should be noted that the thin islands inside a zone, which could be considered as sharing boundaries with this zone, might also be changed. From the discussion above we can conclude that the smallest size region that can be kept is a 3-by-3 block of cells. A thin region which is 2 cells wide and 5 cells long will be removed, since it cannot recover after shrinking.

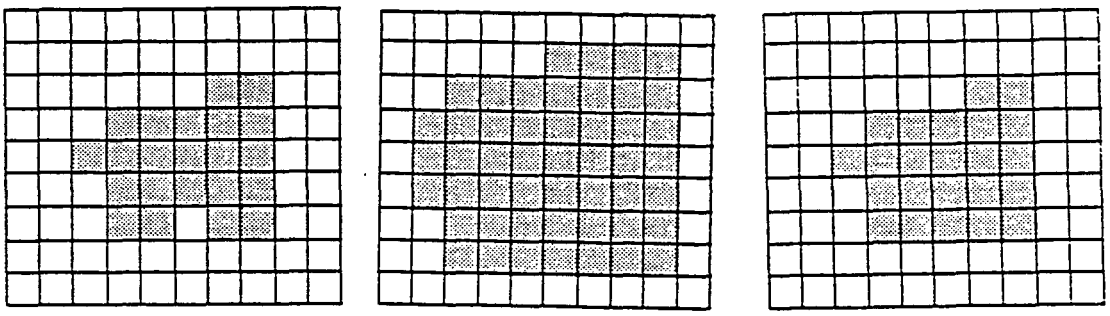


Figure 7-3. Boundary Simplification (13)

Generally, it is hard to determine the priority of each zone. In this project we assumed the zones with smaller total areas have a high priority to expand into zones with larger total areas. Figure 7-4 shows the land cover map after boundary simplification.

Nibble Mask

The classified image has been cleaned by majority filtering and boundary cleaning methods. However, there are still some small regions which could be considered as noisy or misclassified zones that might be caused by spurious data or that are irrelevant to future processing. So we assigned each region a unique value. A function called region grouping was applied here to eliminate small regions. This is illustrated in Figure 7-5. In this processing, any pixel will be assigned one unique value which is

Figure 7-4. Boundary Simplified Data

determined as the identity of its connected region within the analysis window. After each region has a unique identifier, a minimal acceptable area for a region was determined. A new image which meets the requirement that all the regions have an area greater than the minimal threshold was also created as a mask image. The nibble mask function was then applied to the image to eliminate those small regions whose areas are equal to or below the determined minimal area. The nibble mask function assigns to the cell locations that belong to a region below the minimal area threshold, the value of their closest neighbor. That means its closest neighbor will swallow this cell location. In this project, the minimal area was determined as 4 cells, which equals a 40-by-40 meter square. Figure 7-6 is the nibbled land cover map.

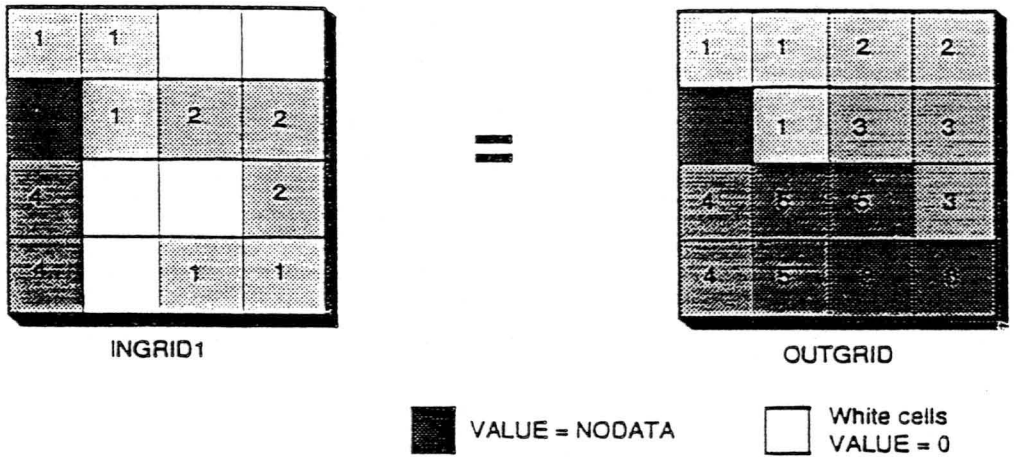


Figure 7-5. Region Grouping (13)

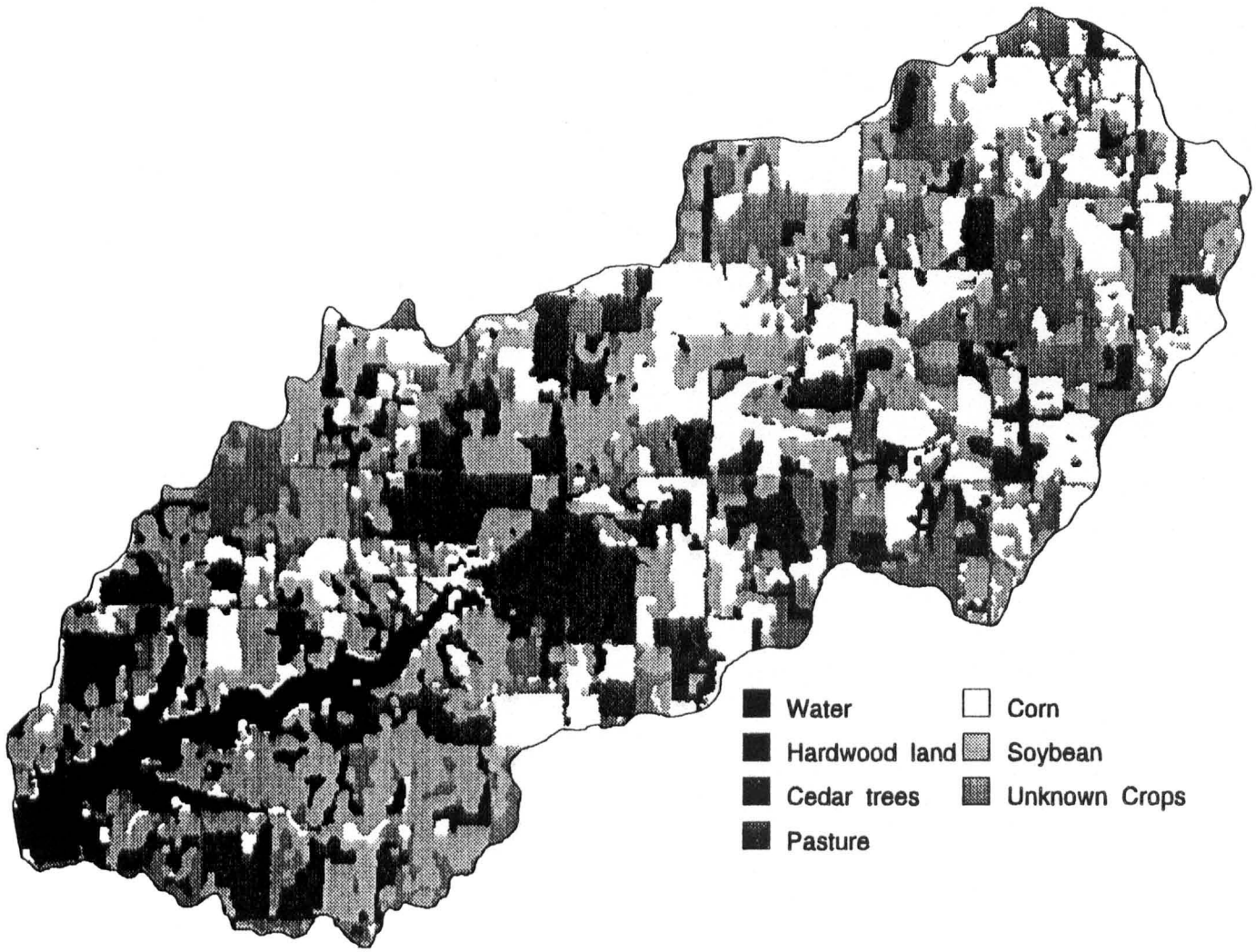


Figure 7-6. Nibble Masked Data (Final)

PART III

PRACTICAL APPLICATIONS

CHAPTER 8. SURFACE MODELING

In this chapter, digital elevation data obtained by manual digitizing from USGS 7.5-minute topographic map is used to simulate a three dimensional surface model for the Lake Icaria Watershed. This 3-D surface model helps to determine, for any location in the model, the upslope area contributing to that point and the downslope path water would follow. A stream network will be calculated from digital elevation data. This stream network can be compared with the stream network digitized from 7.5-minute quad map.

Slope and aspect are two basic concepts about the surface model which are usually used to describe the surface characteristics. The slope of a surface refers to the maximum rate of change in z values across a region of the surface. Slope is generally expressed as a percentage. A region with a slope less than 1% is considered to be flat. The aspect of a surface is the direction of the maximum rate of change in elevation in the downward direction. Aspect is also referred to the slope direction. Aspect is expressed in positive degree from 0 to 359, measured clockwise from the north.

Digital Elevation Data

Digital Elevation Data often refers to Digital Elevation Models (DEM) which represent the Earth's continuous surface in cell-based format. The DEM data consists of an array of elevations referenced horizontally in one coordinate system such as the UTM coordinate system. Figure 8-1 shows a 3x3 array of DEM. Each digital value represents the elevation for each single cell.

78	72	69
74	67	56
69	53	44

Figure 8-1. A Sample of Digital Elevation Model

Triangulated Irregular Network (TIN)

Triangulated Irregular Network (TIN) is a set of adjacent, non-overlapping triangles computed from irregularly spaced points with x, y coordinates and elevation (z) values. A tin model will store the topological relationships between triangles and their adjacent neighbors. This special data structure will make the efficient generation of surface models for the analysis and display of terrain or other types of physical surface. Each triangle in TIN describes the behavior of a portion of the TIN's surface. Very important information about the facet, such as aspect, slope, surface area, and surface length could be derived from the x, y, z coordinate values of a triangle's three nodes.

The surface modeling in this project was performed with the help of TIN package in ARC/INFO. Since it is not possible to get a z value for every location on the surface, TIN uses interpolation to calculate z values at surface locations where no samples have been taken. A point lying between two contour lines is a typical example here. Two interpolators are generally used on tin surface models: linear and quintic interpolations. Both methods use the z values of each node of the triangle to compute the z value for a specified point within this triangle.

The linear interpolation simply assumes the surface to be a continuous faceted surface formed by triangles. The normal to a surface will be constant throughout the extent of each triangle facet. In this interpolation, the surface value to be interpolated is computed based only on the z values for the nodes of the triangle within which the point lies. The equation for linear interpolation in a triangle facet is:

$$Ax + By + Cz + D = 0 \qquad \text{Eq. 8-1 (14)}$$

Where A, B, and C are constants determined by the coordinates of the triangle's three nodes.

Linear interpolation is simple and quick. It produces fixed first derivatives (slope of the surface) within a triangle which means the slope for any location in this triangle will be the same. This method is acceptable if the study area is big enough. But a serious problem is that the normal to a triangle changes abruptly to that of the next triangle if a point crosses over an edge which separates two adjacent triangles. Figure 8-2 illustrates the linear interpolation process.

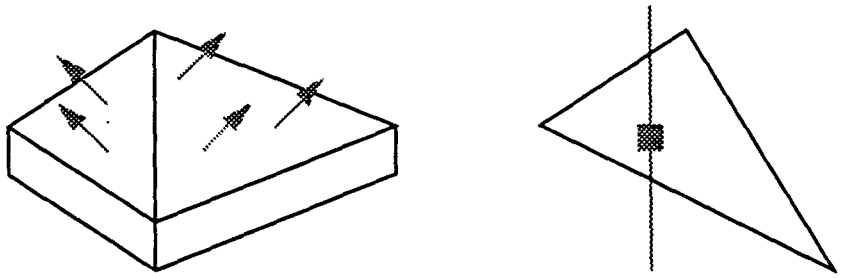


Figure 8-2. Linear Interpolation

Quintic interpolation considers the surface model not only to be continuous but also to be smooth. In another words, the normal to the surface changes continuously within each triangle. Also, there are no abrupt changes in the normal as it crosses an edge between triangles. This smooth surface model is accomplished by considering

the geometry of the neighboring triangles when the z value in a triangle is interpolated. The interpolation algorithm used in quintic interpolation is a bivariate fifth-degree polynomial in x and y. The generalized equation for quintic interpolation of the z value of a point on a surface

$$z = \sum_{j=0}^{5} \sum_{k=0}^{5-j} q_{jk} x^j y^k \qquad \text{Eq. 8-2 (14)}$$

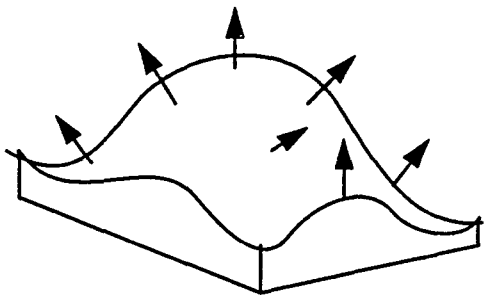


Figure 8-3. Quintic Interpolation

There are 21 coefficients in the above equation. The determination of these coefficients depends on the values of three triangle nodes and the characteristics of the model surface. The values of the function itself and its first-order and second-order partial derivatives are given at each triangle node, yielding 18 coefficients. The three remaining coefficients can be determined by considering the surface to be both continuous and smooth in the direction perpendicular to the three triangle edges.

The quintic interpolation method is more realistic than linear interpolation because it produces continuous first derivatives (slope) within a triangle and even at triangle edges. But it is computationally more intensive than the linear interpolation.

Here is an example to compare these two types of interpolation. As shown in Figure 8-4, A and B are two randomly selected points whose elevation values are unknown. Both linear and quintic interpolation methods have been applied.

Table 8-1. Comparison of Linear and Quintic Interpolations

Point	X value	Y value	Zl	Zq
Point A	360385.8	4550243	359.675	359.675
Point B	356262.6	4545327	363.757	363.629

The coordinates shown in Table 8-1 are in UTM coordinate system in zone 15. The unit is meter. The Zl is the interpolated elevation by linear method, and the Zq by quintic method. For point B the Zl and Zq values are different. This was because that point B lay in the middle of two contours which are 359.664 and 365.760 (Figure 8-4). The interpolated elevations for point A by two different interpolations was the same because the point A happened to lie in the lake area. During the creation of this 3-D surface it was assumed that the lake surface was a plane.

Surface Display

Once the surface model has been created, it could be viewed in different ways. Figure 8-5 is the shaded relief display of the created surface with the lake superimposed. The sun illumination angle was zero, which meant the sun was directly overhead. This picture was very useful for enhancing the visualization of the surface.

Surface Analysis

After the surface was determined, the direction of flow across a surface could be determined by the aspect at each location and the energy of flow could be determined by the slope of the surface. As shown in Figure 8-6, a steeper slope causes greater energy. A stream increases its ability to transport more and larger particles when its energy increases.

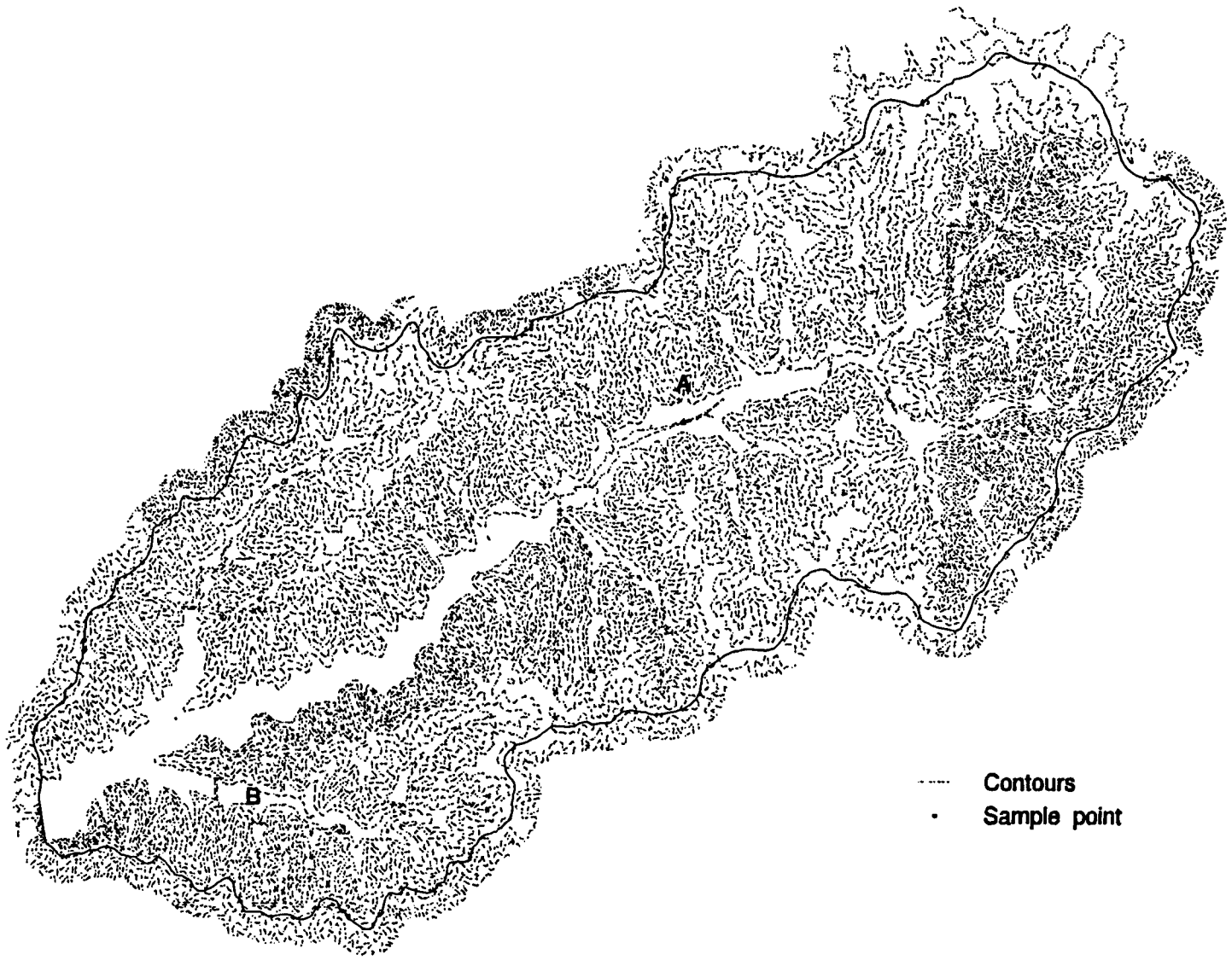


Figure 8-4. The Locations of Two Randomly Selected Points

Figure 8-5. Shaded Relief Display

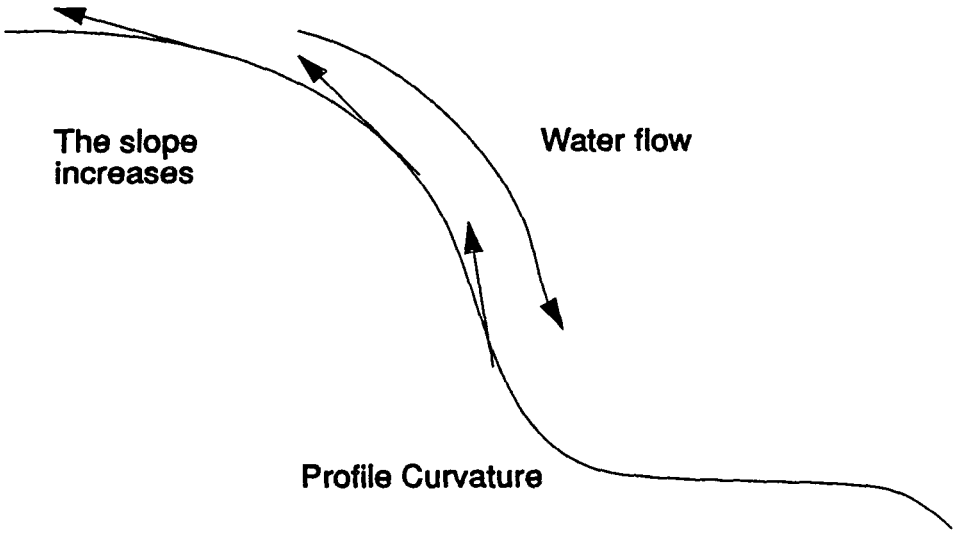


Figure 8-6. The Direction of Flow

The flow across a surface will always go in the steepest downslope direction. Once the direction of flow out of each cell is determined, which and how many cells flow will enter can also be determined. This procedure can be used to determine stream network. The output stream network will be used to delineate the water flow when a certain amount of precipitation occurs in this area.

The direction of flow was determined by calculating the direction of the steepest descent, or maximum drop, from each cell. The formula used in calculation was:

maxim drop = change in elevation (z) / spatial distance

The spatial distance stands for the distance between cell centers. If the cell size is 1, the distance between two orthogonal cells will be 1, and the distance between two diagonal cells will be the square root of 2, which is 1.414216. In Figure 8-7, the direction of flow at center cell will be from the center cell to the lower-right cell, because there is the biggest maximum drop value (16.26) in this direction

78	72	69
74	67	56
69	53	44

maximum drop = (67 - 44) / 1.414216 = 16.26

Figure 8-7. The Determination of Flow Direction for Center Cell

If the elevation for the processing cell is higher than any of its neighbors, this processing cell is called a peak. The direction of flow from a peak is from this cell to any other surrounding cells; but the direction of flow is undefined. Similarly, If the elevation of the processing cell is lower than any of its surroundings' elevation, this cell is called a sink, and the direction of flow is also undefined.

Figure 8-8 shows the calculated stream network. This stream network can be used to determine/predict the downslope path along which water follows. Compared with the stream channels digitized from existing USGS 7.5 minute topographic maps, it is noticed that certain overlap occurred. This confirms the accuracy of the calculated stream network. Because this stream network was determined by the physical surface of the earth, it can be used to determine/predict the downslope path that water would follow.

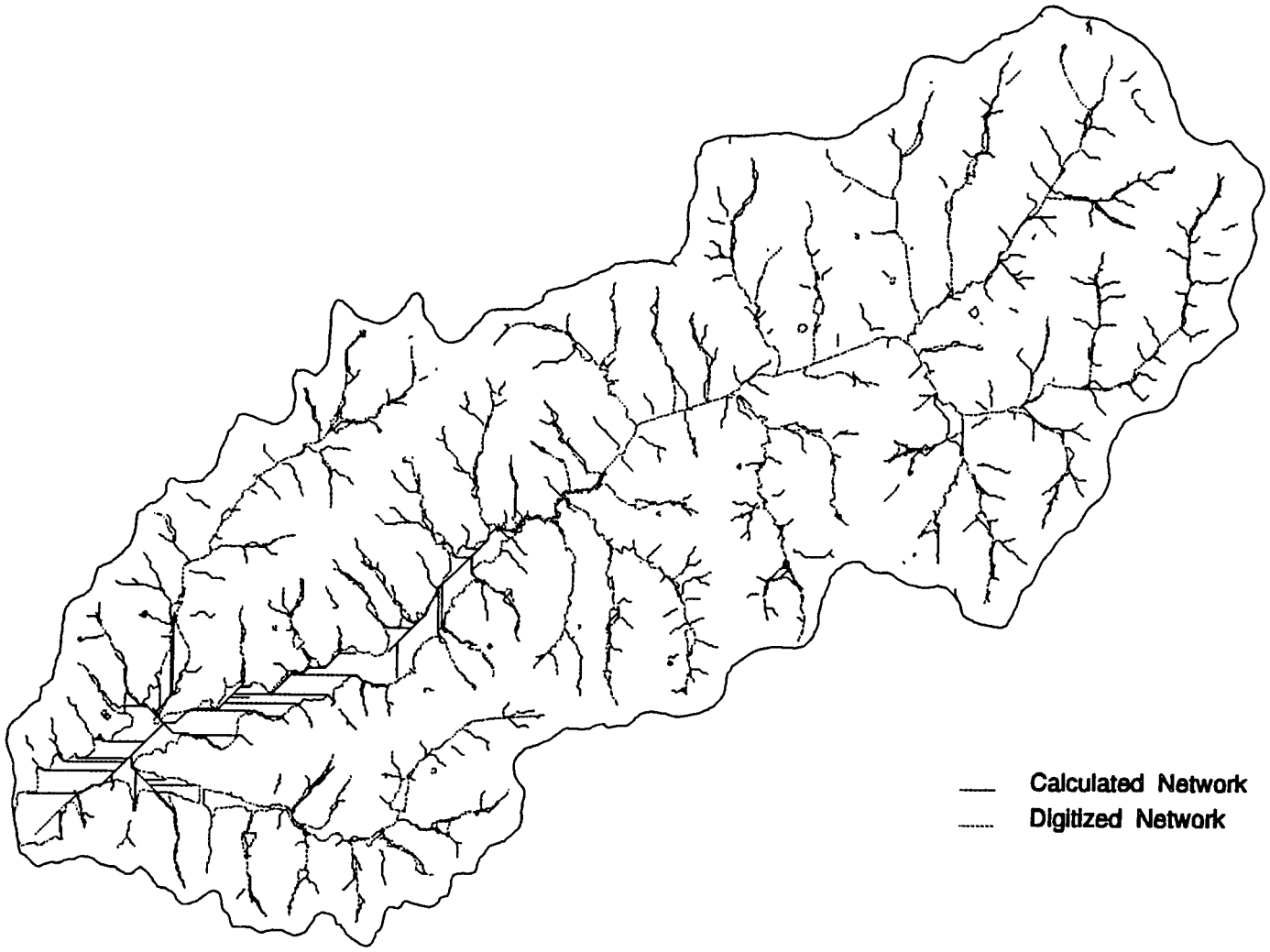


Figure 8-8. Calculated against Digitized Stream Network

CHAPTER 9. MINIMUM TRANSPORTATION PATH FINDING

There are numerous agricultural activities around Lake Icaria watershed. Livestock is the most important among these activities. The solution for finding the best transportation path to remove the waste from one livestock site to another one will help the land use planners to make the right decisions. In the Lake Icaria Watershed GIS, minimum transportation distances can be calculated by two different methodologies, which are raster and vector analysis. This chapter discusses the algorithms applied and the procedures performed in these calculations. Also the output for both approaches will be compared and evaluated. The advantages and disadvantages of each approach will be summarized.

The objective of this task was to find the minimum transportation path/distance between each single pair of tracts in Lake Icaria Watershed. There is a total of 211 tracts in this area, which would yield 22,155 pairs. Therefore, the computational time became a very important factor in the calculation.

ARC/INFO GRID and NETWORK were the two major packages applied in this analysis. The transportation coverage used in the calculation was created by digitizing from USGS 7.5 minute topographic maps. The tract coverage was digitized from aerial photos which were taken in 1980 and revised by Agriculture Soil Conservation Service (ASCS).

The Concept of Minimum Transportation Distance

The concept used to define the minimum transportation distances in this application is illustrated in the Figure 9-1. T1 the starting tract and T2 the destination tract are represented by their centroids (C1 or C2). The goal of the calculation is to find the minimum transportation path when we move from Tract 1 to Tract 2. This transporta-

tion distance will involve three quantities. They are, the starting tract centroid (C1) to the closest point (A) on the road (L1), the destination tract centroid (C2) to the closest point (B) on a road (L2), and the road distance between point A and B. Therefore, the minimum transportation distance between Tract 1 and Tract 2 is calculated by the following model (Eq. 9-1).

Total min. transportation distance from T1 to T2 = L1 + AB + L2 Eq. 9-1

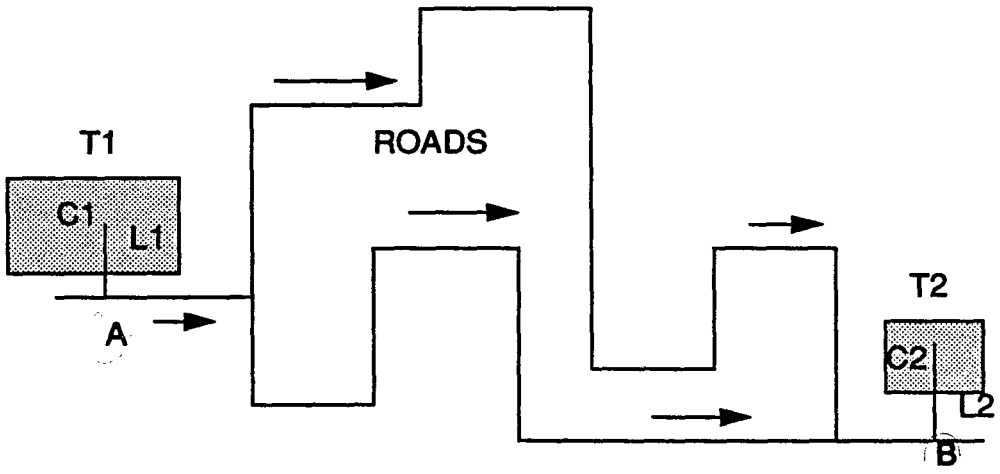


Figure 9-1. Transportation Path From T1 to T2

Apply the Calculation Model in ARC/INFO GRID

The first method used in this task was to apply the above model (Eq. 9-1) in the ARC/INFO GRID cost distance calculation function. The cost distance here represents the cost of moving from one cell through certain cells to another. The cost may be travel time, dollars, preference, and so forth. The ARC/INFO cost distance function will calculate the cost distance from the source grid to each cell in the study area. The source grid contains single or multiple zones, which may or may not be connected by cells. The cost grid assigns an impedance in some uniform-unit measurement system

that depicts the cost involved in moving through any particular cell. The value of each cell in the cost grid is assumed to represent the cost-per-unit distance of passing through the cell, where a unit distance corresponds to the cell width. The source grid in this task was created by converting the tract center point coverage to point grid with 20m x 20m cell size. The cost grid was derived by converting the road coverage to a road grid also with cell size 20m x 20m. The values for on-road cells and for off-road cells were assigned as 1 and 10, respectively. This means that travel on the off-road field will be ten times more expensive than travel on the road. In other word, this would effectively force the calculated least cost path from the source site to the destination site along the road.

The calculation computed the distance between the first selected tract and any other tracts. Suppose we are looking for the distance between T1 and T2. According to the computation model we set up in Equation 9-1, three quantities L1, AB, and L2 need to be determined first. With the help of GRID Euclidean function, L1 and L2 could be determined. Then the cell number from the tract center to the closest road can be calculated by:

$$\text{Cell number} = \text{Euclidean distance (L1)} / \text{Cell size} \qquad \text{Eq. 9-2}$$

here, the cell size is 20 meters. Call the calculated cell number from T1 and T2, N1 and N2 respectively. From understanding the algorithm of ARC/INFO cost distance function, the minimum cost distance from T1 to T2 is calculated by:

$$\text{Minimum cost distance} = N1 \times 10 \times 20 + AB + N2 \times 10 \times 20 \qquad \text{Eq. 9-3}$$

$$\text{or } AB = \text{Minimum cost distance} - N1 \times 10 \times 20 - N2 \times 10 \times 20 \qquad \text{Eq. 9-4}$$

where, 10 is the off road value for cost grid, and 20 is the cell size.

Then the transportation distance between T1 and T2 can be determined by Eq. 9-1, which can be written as:

$$\text{Transportation distance} = N1 \times 20 + N2 \times 20 + AB \qquad \text{Eq. 9-5}$$

Under certain circumstances, the model in Eq. 9-1 does not apply. This is illustrated in Figure 9-2. In Figure 9-2, the minimum cost path between T1 and T2 will go from C1 to C2 directly instead of going on the road. In this case, the transportation distance is calculated as:

Transportation distance = C1 to C2 Euclidean distance / 20 Eq. 9-6

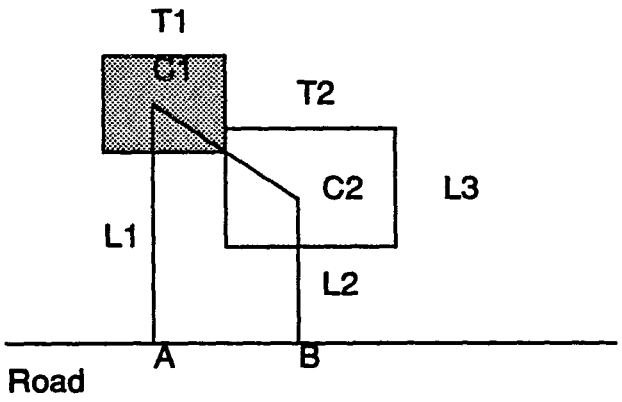


Figure 9-2. The Case in Which Model Cannot Apply

Pathfinding Algorithm in ARC/INFO NETWORK Analysis

There are many algorithms that find least-cost paths through a network. ARC/INFO NETWORK makes use of Dijkstra's algorithm. Figure 9-3 illustrates how this algorithm works to find the least-cost path between nodes A and G. In this example, the least-cost can be replaced by least-travel-time. First of all, the scanning process will start with node 'a' to search for the node with the lowest cumulative cost from node 'a'. Obviously, there are only two possible paths available. The travel time between node 'a' and node 'f' (8 seconds) is the least. So 'af' is selected as the least cost path at this point. Then the node 'f' will be treated as origin node and the scanning process will start again from here. Node 'a' is not qualified as a potential destination node this

time. After comparison, path 'fe' is selected as least-cost (2 seconds) path. Then the scanning repeats until the destination node 'g' is encountered. In Figure 9-3, the calculated least-cost path is from node a to f to e to d to g. The total travel time is 19 minutes.

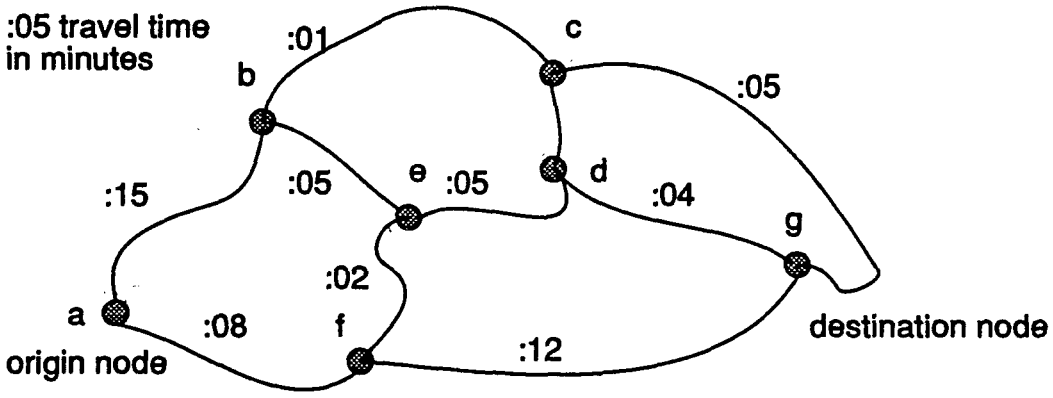


Figure 9-3. Network Illustration

Results and Conclusions

The calculation was performed on 8 tracts within the study area, which yielded 28 pairs of distances. These 8 tracts were selected to meet the criteria that the area must be bigger than 1,000,000 square meters. Table 9-1 lists all 28 pairs of distances calculated by two approaches. Figure 9-4 shows the locations of these 8 tracts with the reference of roads. The scatterplot shown in Figure 9-5 represents the distances derived from GRID versus the distances from NETWORK and shows that both distances are very close.

The initial objective of this task was achieved by both GRID and NETWORK approaches. Considering the efficiency and the accuracy of calculation, we would recommend use of the distances obtained from the NETWORK method because the GRID method was time consuming. Also GRID method was based to a certain extent on approximation in computation.

Table 9-1. Comparison of Calculated Distances by Two Approaches

From Tract	To Tract	GRID Distance	NETWORK Distance
5	34	2873.137	3029.169
5	36	1760.000	2311.018
5	69	5295.141	5378.721
5	97	5629.705	5814.207
5	188	14102.904	14053.998
5	221	21187.115	21025.549
5	227	15078.248	15138.569
34	36	1456.568	1653.531
34	69	2402.005	2524.566
34	97	4221.421	4416.827
34	188	11209.752	11199.843
34	221	18293.949	18171.395
34	227	12185.094	12284.414
36	69	3878.573	4003.083
36	97	4213.137	4438.570
36	188	12686.320	12678.360
36	221	19770.531	19649.912
36	227	13661.664	13762.931
69	97	3210.289	3280.701
69	188	10198.619	10063.718
69	221	17282.831	17035.270
69	227	11173.964	11148.289
97	188	9208.331	9198.101
97	221	16292.527	16169.653

From Tract	To Tract	GRID Distance	NETWORK Distance
97	227	10183.672	10282.672
188	221	8968.058	8829.685
188	227	2739.216	2818.971
221	227	7067.421	7049.173

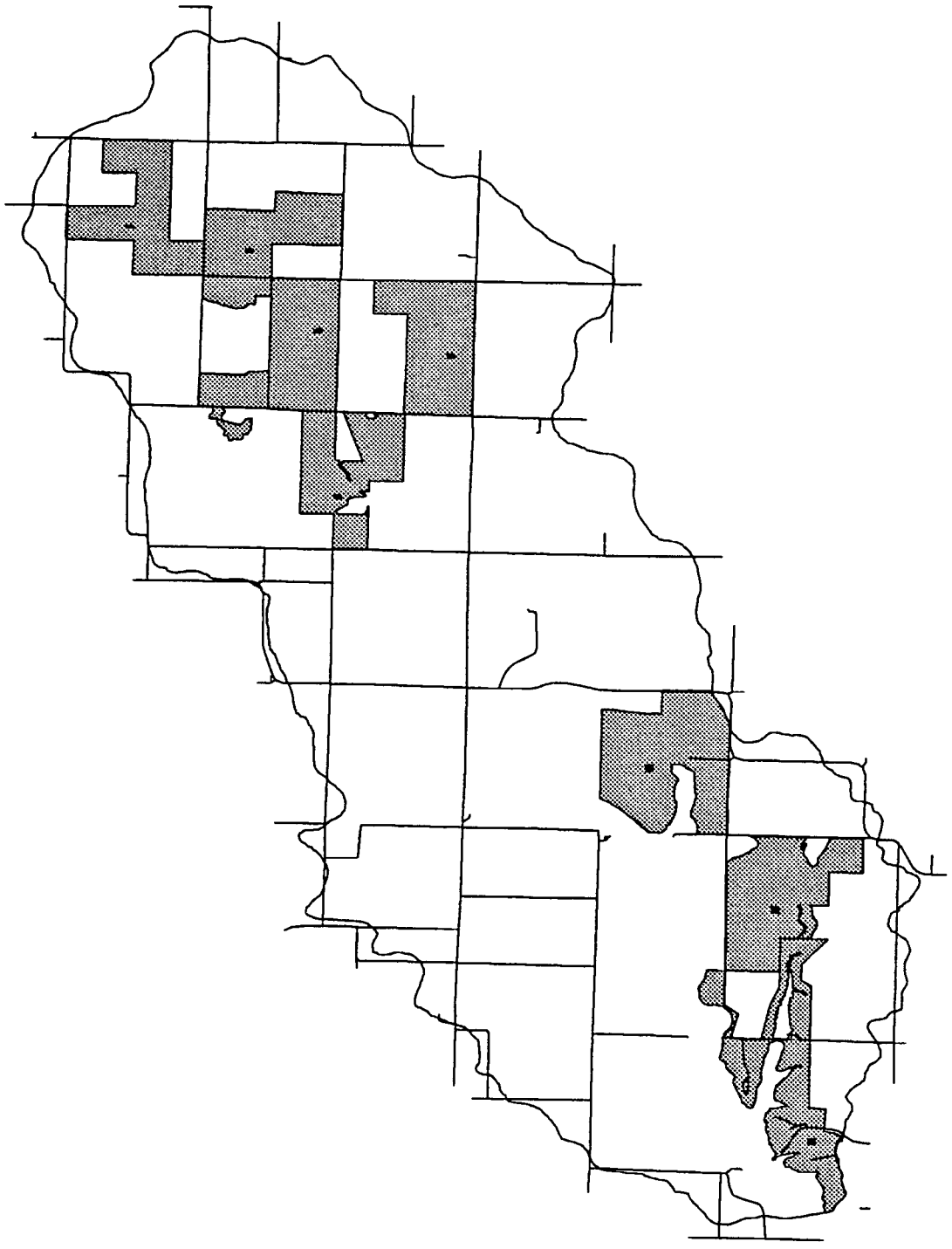


Figure 9-4. The Locations of 8 Selected Tracts

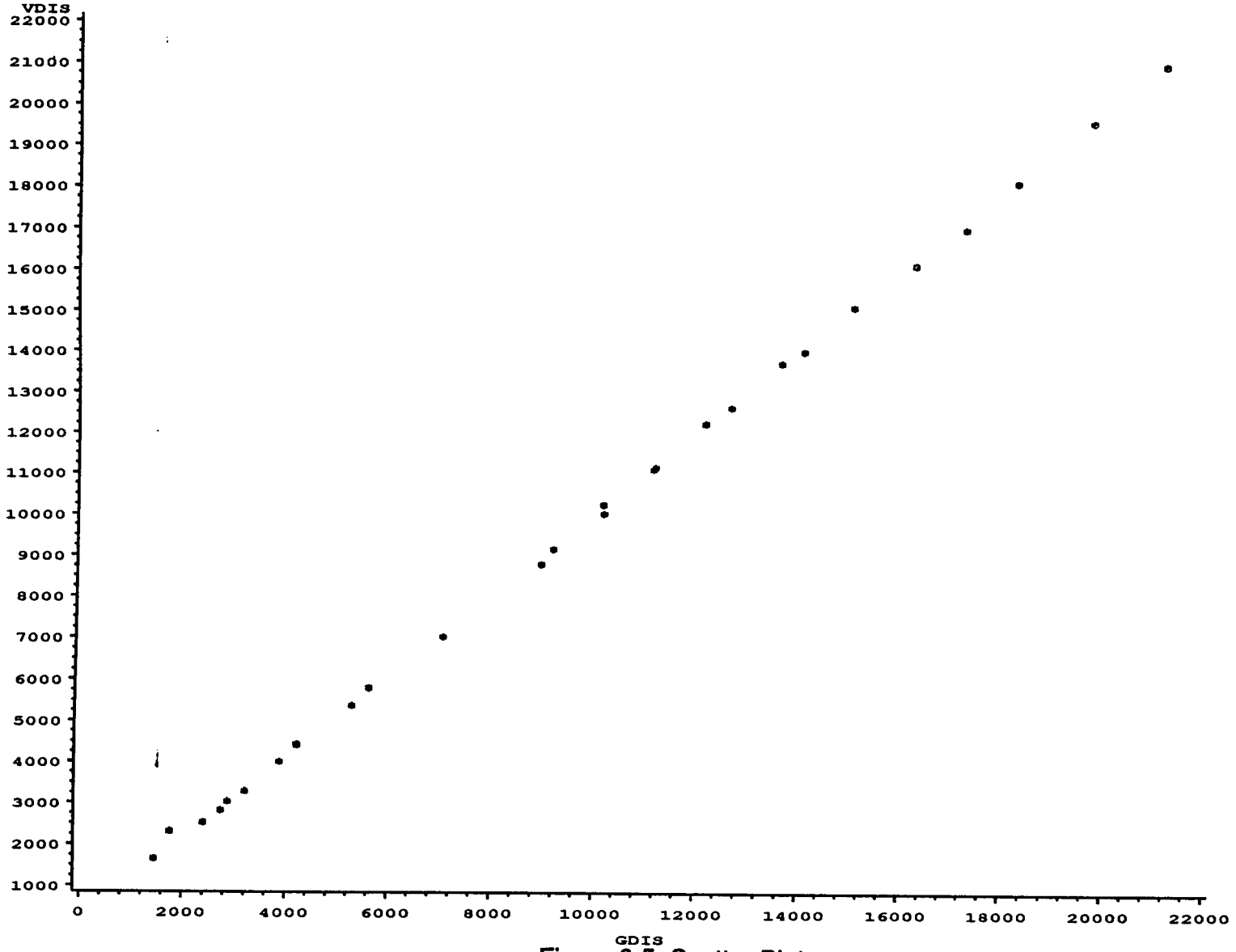


Figure 9-5. Scatter Plot
GRID Distance vs. NETWORK Distance

CHAPTER 10. CONCLUSIONS AND RECOMMENDATIONS

Conclusions

GIS was successfully developed for the Lake Icaria Watershed in Adams County, Iowa. The system accomplishes its goal of providing the user with information necessary to make decisions for management and analysis. The approaches and procedures used in this project have proven to be practical and can be applied to the similar scale projects in the future.

Many methods can be applied to develop GIS database. Traditional approach to creating digital data is to digitize from existing paper maps. Remote sensing data were effectively input to the GIS in this project. It has been proven in this project that GIS can import, store, display, analyze, and export the remotely sensed data easily. Conversely, remote sensing can improve the utility of GIS by supplying more current information.

SPOT satellite data was processed by enhancement, rectification, and classification procedures. Image enhancement process made the raw data more readable and interpretable by increasing the distinction between the features in the scene. An image rectification process geo-references the raw data to the UTM coordinate system. This rectification procedure made the satellite data consistent with other digital data in GIS database in terms of the coordinate system, so some analytical processes such as overlay can be performed. Instead of interpreting satellite data visually (qualitatively) as in image enhancement process, image classification processes analyze the remotely sensed data quantitatively. This procedure assigns each pixel to one feature category. The final output of classification is usually a thematic map to represent land cover/land use. The digital image processing yielded the most current land cover/land use information.

GPS was used to establish the initial six control points for image rectification purposes in this research project. The Ashetech XII GPS receiver we used in observation has centimeter accuracy and works efficiently. The GPS coordinates were converted from geographic latitude-longitude to the UTM coordinates on NAD-27 datum (Clarke 1866 ellipsoid). In order to compare the accuracy of coordinates derived from a USGS 7.5 minute topographic map by manual digitizing, linear regression was performed. The output of the analysis of variance shows the accuracy of coordinates retrieval from USGS 7.5 minute quad map by manual digitizing was acceptable for the objectives and precision requirement for image rectification/geo-referencing purposes.

The stream network created by elevation data in the study area can help the user to predict the flow direction following a rainfall event. This stream network will supply user with more information than the steams coverage developed by digitizing from USGS 7.5 minute topographic maps, because the USGS maps describe only those permanent or intermittent water features. The output of minimum transportation path between each pair of tracts assists the user to make the best decision when many options are available. This technique can also be applied in many similar fields such as city emergency response system, etc.

Recommendations

As discussed in the previous chapters, most geographic data in this GIS database, such as streams, are developed by digitizing currently available USGS 7.5 minute topographic maps. This certainly supplies baseline information for the study area, but accuracy is questionable. The information is only as up-to-date as the aerial photographs from which it was mapped. However, some information might not change or change very slowly, such as contour or soil type. But for some other information as

land use/land cover, the change is rapid. Thus, the information shown for many areas may be now incorrect.

It has been shown that to develop and to update GIS database by using satellite imagery are both efficient and convenient. In this project, a land cover/land use thematic map was derived from SPOT satellite data. This produced the most current land cover/land use information for the study area.

To update existing GIS database, it is recommended that more effort be put on research into digital image processing. It is also recommended to use remote sensing techniques for creating shorelines/streams, transportation, and even soil coverages by using remote sensing data. In fact, remote sensing has been applied in practice to detect shoreline changes using aerial photography. Therefore, shoreline change analysis with the help of satellite imagery has great potential. However, it should be noted that it is difficult to map some point and linear features, digitally, due to the fact that they are not always recognizable at the spatial resolution of the data, nor are they represented at their "true" location due to sensor errors in satellite data collection. On the other hand, Global Positioning System (GPS) data is static and could provide accurate positional information for those point and linear features. Additional research must be done into the integration of GPS technology and digital image processing to develop and update GIS database.

BIBLIOGRAPHY

1. Thomas M. Lillesand and Ralph W. Kiefer. Remote Sensing and Image Interpretation. New York: John Wiley & Sons, 1987.

2. John R. Jensen. Introductory Digital Image Processing: A Remote Sensing Perspective. New Jersey: Prentice-Hall, 1986.

3. Philip H. Swain and Shirley M. Davis. Remote Sensing: The Quantitative Approach. New York: McGraw-Hill, 1978.

4. Paul M. Mather. Computer Processing of Remotely-Sensed Images: An Introduction. New York: John Wiley & Sons, 1987.

5. Robert A. Schowengerdt. Techniques For Image Processing and Classification in Remote Sensing. New York: Academic Press, 1983.

6. Donna J. Peuquet and Duane F. Marble. Introductory Readings in Geographic Information Systems. New York: Taylor and Francis Ltd, 1990.

7. K.C. Clarke. Advances in geographic information systems, Computers, Environment and Urban Systems, 10, pp. 175-184, 1986.

8. Ashtech XII Receiver Operating Manual. Sunnyvale: Ashtech Inc., 1989.

9. P. R. Wolf and R. C. Brinker. Elementary Surveying. New York: Harper and Row, 1989.

10. F. H. Moffitt and H. Bouchard. Surveying. New York: Harper and Row, 1987.

11. D. Wells, N. Beck, D. Delikaraoglou, A. Kleusberg, E. J. Krakiwsky, G. Lachapelle, R. B. Langley, M. Nakiboglu, K. P. Schwarz, J. M. Tranquilla, and P. Vanicek. Guide to GPS Positioning. Fredericton, New Brunswick, Canada: University of New Brunswick Graphic Services, 1987.

12. Environmental Systems Research Institute, Inc. Understanding GIS: The ARC/INFO Method. Redlands: ESRI, Inc., 1992.

13. ERDAS, Inc. Field Book. Atlanta: ERDAS,Inc., 1991.

14. Environmental Systems Research Institute, Inc. ARC/INFO USER'S GUIDE: Cell-based Modeling with GRID. Redlands: ESRI, Inc., 1991.

15. Environmental Systems Research Institute, Inc. ARC/INFO USER'S GUIDE: Surface Modeling with TIN. Redlands: ESRI, Inc., 1991.

16. Environmental Systems Research Institute, Inc. ARC/INFO USER'S GUIDE: Network Analysis. Redlands: ESRI, Inc., 1991.

ACKNOWLEDGMENTS

I would like to thank my major professor, Dr. Kandiah Jeyapalan, for his assistance in the completion of this project as well as my other faculty committee members, Dr. Steve Jungst and Dr. Da Yin Lee. I also wish to express my appreciation for the time Dr. Jungst has spent to revise my thesis. A special thanks to Mr. James Majure, Director of GIS Support and Research Facility at Iowa State University, for his input and assistance in various aspects throughout the project. I also wish to acknowledge my fellow graduate students, Mark Stein, Robert Buffour, and Jeff Wang, who helped me in GPS observations.

I finally wish to express my gratitude to my wife who encouraged and supported me whenever I needed it.

APPENDIX A: GIS DATABASE FOR LAKE ICARIA WATERSHED

The Lake Icaria Watershed geographic information system (GIS) developed in this project is part of the Adams County Project, which is being carried out by Dr. Robert Jolly, principle investigator, and Dr. Sunday Tim at Iowa State University. This document contains the explanation of this GIS database. The GIS is maintained on the network of DEC workstations that are part of the Vincent system on the Iowa State University campus. The ARC/INFO GIS software is the major GIS system being used. Satellite imagery can be processed by ERDAS image processing software. Attribute data are maintained using INFO which is database management system associated with ARC.

Types of Data

Data that was developed for the Lake Icaria Watershed includes:

1. LAND COVER: The updated land cover derived from SPOT data
2. HYDROGRAPHY: The streams and water bodies. Digitize from USGS 1:24,000 topographic maps
3. TRANSPORTATION: Roads. Digitized from 1:24,000 quads
4. TOPOGRAPHY: Contours. Digitize from 1:24,000 quads
5. SOILS: Lines and attributes from SCS detailed soil surveys in cooperation with the Land Use Analysis Lab
6. TRACTS: Ownership/operation status, farm program status, and other attribute information for each field. Digitize from revised photographs.

Directory Structure Supporting The GIS Organization

The data is kept in the directory, **/home/gis5/rocky/thesis**. Within this directory there are three subdirectories: covers, spot, and writing. These directories respectively contains the digital coverages, SPOT data and its associated files, and the documentation.

Data Dictionary Elements

A data dictionary is maintained for each data set in database which includes:

1. data set name
2. location of the data set in the file system of the computer
3. a brief explanation of the data contained in the data set
4. the software used to support the data set
5. date the data set was digitized
6. the source agency responsible for the creation of the data set
7. the source from which the data set was digitized, such as a published map
8. the date the source was created
9. a contact knowledgable about the source
10. the scale of the source
11. the projection in which the data set is maintained
12. geographic extent of the data set
13. annotation contained in the data set
14. description of all attribute files associated with the data set
15. the lineage of the data set, including any operations that changed the character of the digital data, such as generalization, etc.
16. appropriate comments

Data Dictionary

Data set name: LAND COVER

Location: /home/gis5/rocky/thesis/spot

Explanation: This data set was originated from SPOT data. The raw data was processed by image enhancement, rectification, classification and post classification procedures.

Software: ERDAS/ARC/INFO

Date digitized:

Source agency: SPOT Inc.

Source: SPOT data.

Date source created: April 22nd, 1990.

Source contact:

Source scale:

Projection information:

Projection: UTM

Zone: 15

Units: meters

Geographic extent: Lake Icaria Watershed.

Annotation:

Lineage:

Notes:

Attribute files:

PAT file

Items:

AREA: The area of the polygon

PERIMETER: The perimeter of the polygon

LANDCOVER#: The internal polygon number

LANDCOVER-ID: The polygon user id

GRID-CODE: Indicating the different land cover number

LUT (Look up table):

Items:

GRID-CODE: Indicating the different land cover number

1: Water

2: Hardwood land

3: Cedar trees

4: Pasture

5: Corn

6: Soybean

7: Unknown Crops

SYMBOL: Symbol numbers assigned to each class

DESCRIPTION: Descriptive explanation of each class

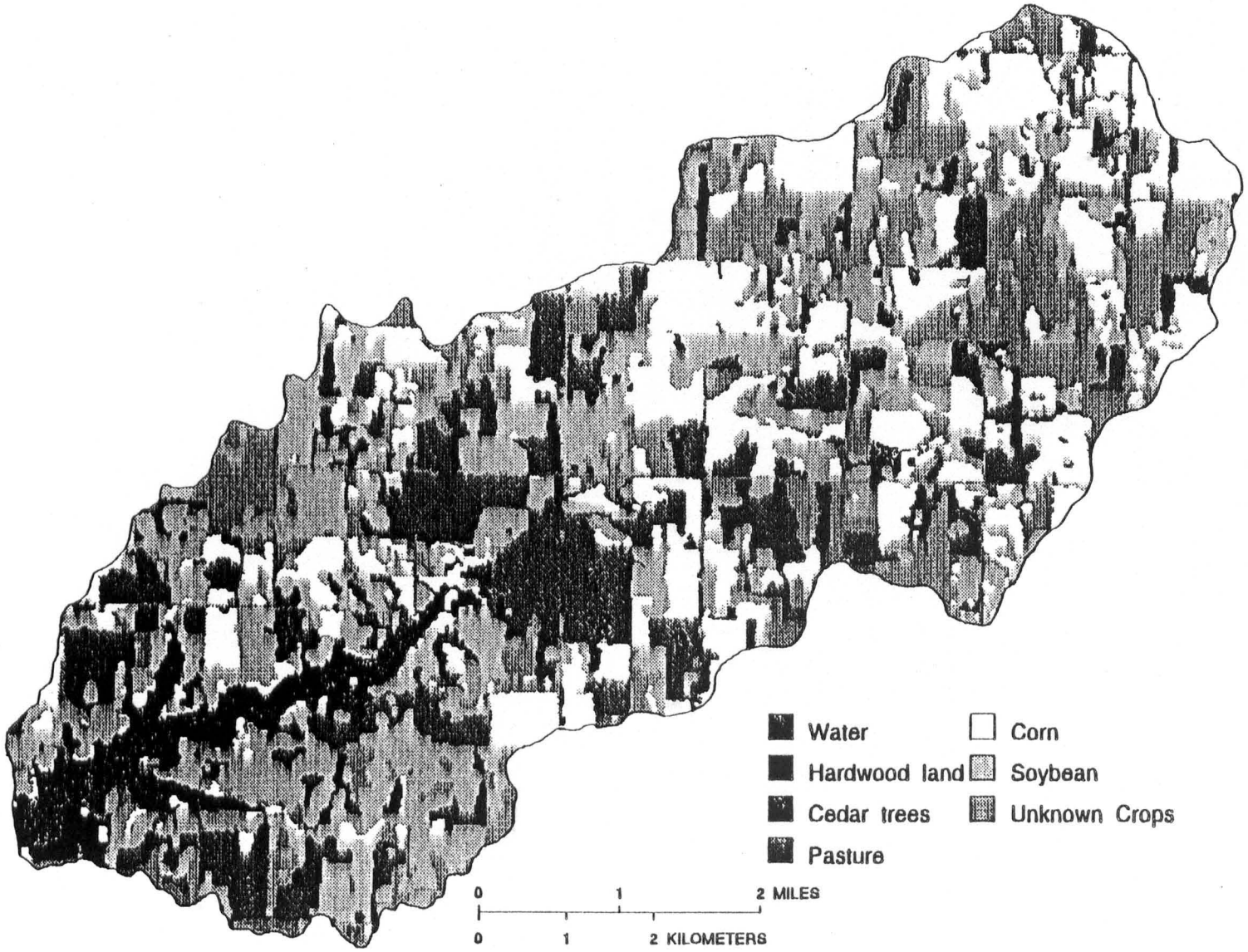


Figure A-1. Lake Icaria Watershed Database, land cover

Data set name: HYDROGRAPHY

Location: /home/gis5/rocky/thesis/covers

Explanation: This data set is a line coverage of the stream network in the icaria watershed area.

Software: ARC/INFO

Date digitized: July, 1991

Source agency: USGS

Source: USGS topographic maps Bridgewater, Carbon, Corning, Nevinville, and Prescott

Date source created: 1980

Source contact:

Source scale: 1:24,000

Projection information:

Projection: UTM

Units: meters

Zone: 15

Geographic extent: North central Adams county, Iowa.

Lineage: Generated by manual digitizing into AutoCad, imported into Arc/Info and attributed.

Notes:

Attribute files:

AAT file

Items:

FNODE#: The arc from node

TNODE#: The arc to node

LPOLY#: The left polygon number

RPOLY#: The right polygon number

LENGTH: The length of the arc

STREAMS#: The internal arc number

STREAMS-ID: The arc id number

TYPE: The stream type

I: Intermittent

P: Perennial

L: Lake or pond

U: Unknown

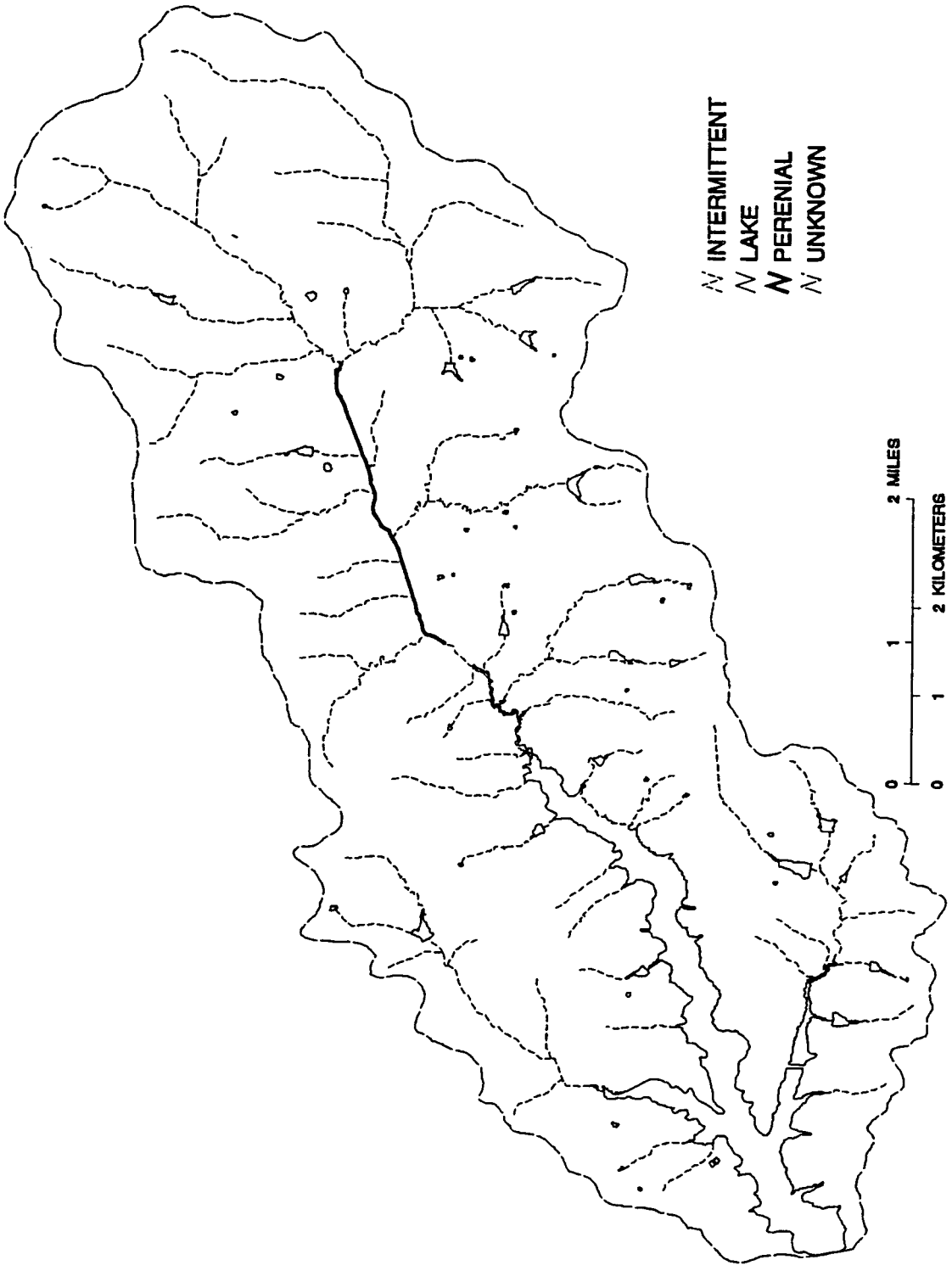


Figure A-2. Lake Icaria Watershed Database, hydrography

Data set name: TRANSPORTATION

Location: /home/gis5/rocky/thesis/covers

Explanation: This data set is a line coverage that contains transportation network digitized at 1:100,000 (figure 2).

Software: ARC/INFO

Date digitized:

Source agency: US Census Bureau

Source: TIGER data base

Date source created:

Source contact:

Source scale: 1:100,000

Projection information:

Projection: utm

Units: meters

Zone: 15

Geographic extent: Adams county, Iowa

Annotation:

Lineage:

Attribute files:

AAT file:

Items:

FNODE#: The arc from node

TNODE#: The arc to node

LPOLY#: The left polygon number

RPOLY#: The right polygon number

LENGTH: The length of the arc

ROADS#: The internal arc number

ROADS-ID: The arc id number

CLASS: The classification of roads.

LUT (Look up table) file:

Items:

CLASS: The classification of roads

L: LIGHT-DUTY ROAD, HARD OR IMPROVED SURFACE

P: PRIMARY HIGHWAY, HARD SURFACE

S: SECONDARY HIGHWAY, HARD SURFACE

U: UNIMPROVED ROAD

SYMBOL: Symbol numbers assigned to each class

DESCRIPTION: Descriptive explanation of each class.

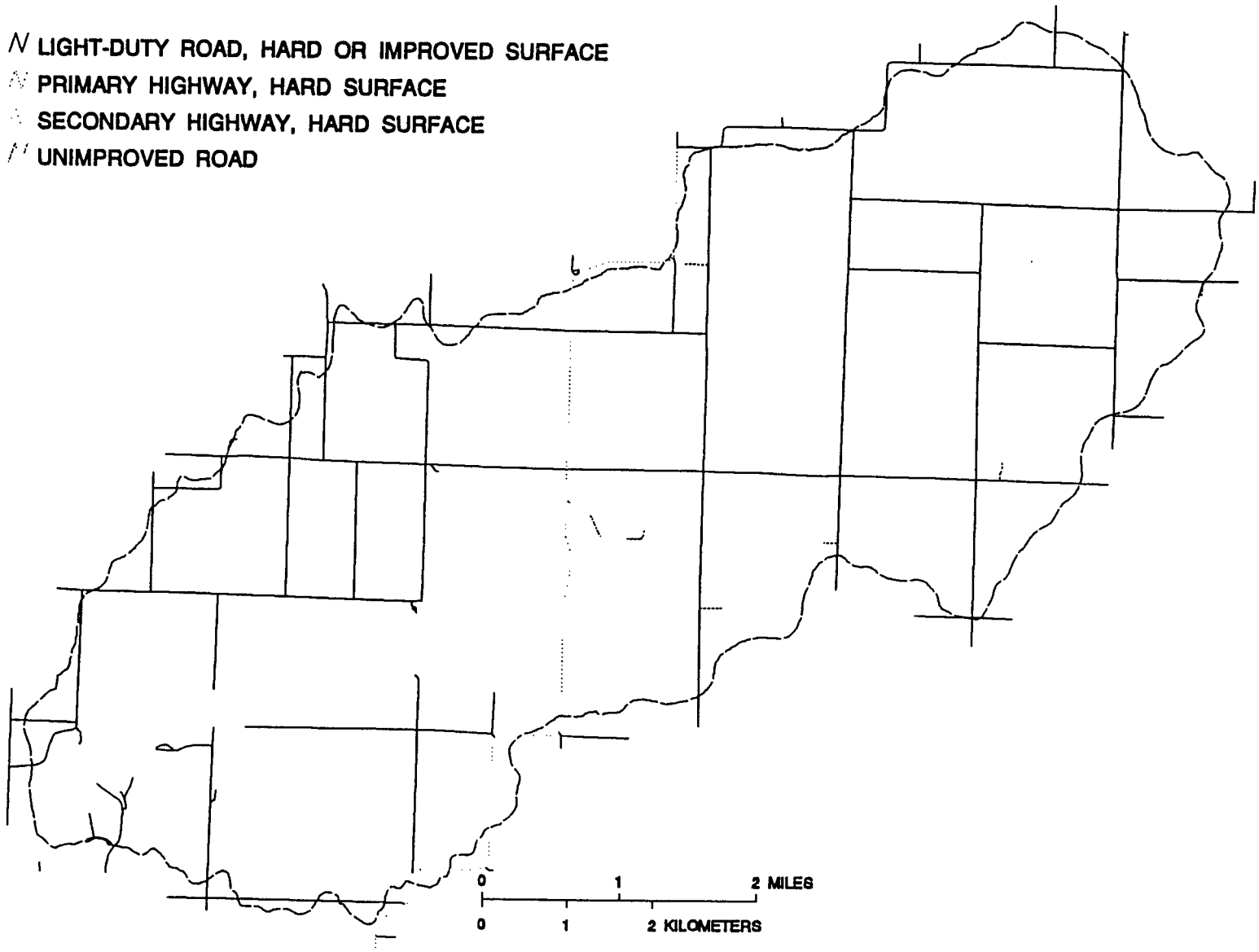


Figure A-3. Lake Icaria Watershed Database, transportation

Data set name: TOPOGRAPHY

Location: /home/gis5/rocky/thesis/covers

Explanation: This coverage delineates a portion of the Adams County contour lines.

Software: ARC/INFO

Date digitized: spring 92

Source agency: USGS, DBA systems inc.

Source: USGS topographic maps

Date source created: 1980

Source contact:

Source scale: 1: 24,000

Projection information:

Projection: UTM

Units: meters

Zone: 15

Geographic extent: Carbon, Corning, Prescott quad sheets in Adams County.

Lineage: Generated from scanned in mylar separates of the quad sheets.

Note: DBA systems did a poor job in working on this with us.

Attribute files:

AAT file

Items:

FNODE#: The arc from node

TNODE#: The arc to node

LPOLY#: The left polygon number

RPOLY#: The right polygon number

LENGTH: The length of the arc

CONTOURS#: Arc Number

CONTOURS-ID: Arc user id

ALTITUDE: Elevation in feet

SYMBOL: Symbol number with the lowest elevation number starting at 1 and going up from there.

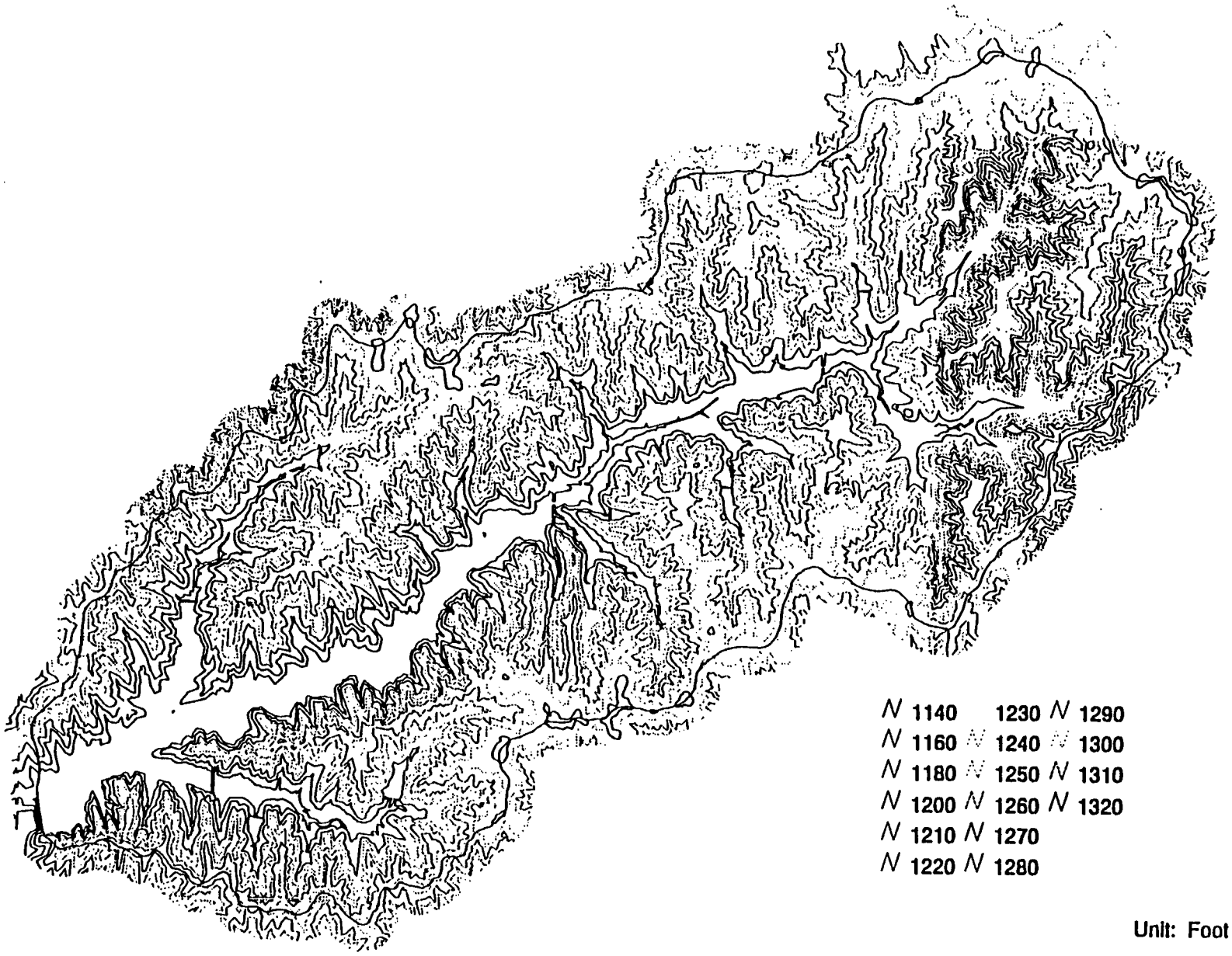


Figure A-4. Lake Icaria Watershed Database, topography

Data set name: SOILS

Location: /home/gis5/rocky/thesis/covers

Explanation: This data set is a polygon coverage that soils polygons (figure 7).

Software: ARC/INFO

Date digitized:

Source agency: USDA Soil Conservation Service

Source: Adams County, Iowa, Soil Survey

Date source created: May, 1963

Source contact:

Source scale: 1:15,840

Projection information:

Projection: utm

Units: meters

Zone: 15

Geographic extent: North central Adams county, Iowa.

Lineage: The data were originally scanned and processed for use in the ISOIL program, Iowa Cooperative Soil Survey Program. The following steps were used in processing the data:

- Scanned at 300 dpi
- Processed to resolution of 50 dpi for use on a CGA monitor
- Converted to ARC/INFO and referenced to section corners that were digitized from USGS 7.5 minute quads

Notes:

Attribute files:

PAT file:

Items:

AREA: The area of the polygon

PERIMETER: The perimeter of the polygon

SOIL#: Internal polygon number

SOIL-ID: Polygon id number

LABEL: The soil mapping unit label

Data file name: ADAMS.ISP

Location: /home/gis5/adams/icaria24k

Explanation: The ADAMS.ISP file is the Iowa Soil Properties and Interpretations Data Base. The following documentation contains a brief explanation of the items in the file. For a more detailed description, see the ISOIL User's Manual.

Items:

SMU: Soil Map Unit (SMU)

SCSSOILS: SCS-Soils-5 Number

KINDOFMU: Kind of Map Unit

KINDCOMP: Kind of Component

SLOPERNGL: Slope Range (low) in percent

SLOPERNGH: Slope Range (high) end in percent

EROSIONC: Erosion Class

LCC: Land Capability Class/Subclass

CSR: Corn Suitability Rating

OMM: Organic Matter Midpoint in percent

OMR: Organic Matter Range (?) in percent

PERM: Permeability

TILTHRTG: Tilth Rating

CORNYLD: Corn Yield in bu/ac

SOYBNYLD: Soybean Yield in bu/ac

OATYLD: Oat Yield in bu/ac

WHEATYLD: Wheat Yield in bu/ac

ALFBRMYLD: Alfalfa-Bromegrass Yiel in T/ac

TIGRSYLD: Tall Introduced Grasses Yield in aum/ac

KYBGYLD: Kentucky Bluegrass Yield in aum/ac

PRIMELND: USDA Prime Farmland

MSA: Major Soil Area (MSA)

SOILNAME: Soil Name

DRNCLASS: Drainage Class (natural)

DRNCLSCD: Drainage Class Code (natural)

NATIVEVEG: Native Vegetation

PARENTMAT: Parent Material

SUBSOILGRP: Subsoil Group

LNDSCPPOS: Landscape Position

TEXTSURHOR: Texture (surface horizon)

SANDSIZESH: Sand Size (surface horizon)

SANDCONTSL: Sand Content (surface) (low) in percent

SANDCONTSH: Sand Content (surface) (high) in percent

SURBDL: Surface Bulk Density (low) in g/cm3

SURBDH: Surface Bulk Density (high) in g/cm3

SUBSLBDL: Subsoil Bulk Density (low) in g/cm3

SUBSLBDH: Subsoil Bulk Density (high) in g/cm3

DEPTHCNTRL: Depth to Strongly Contrasting Particle-Size Class (0-40")

TONSRES: Tons of Residue/Acre in T/ac

LEAGFMLND: LEAG Farmland Units

ORDER: Taxonomic Classification

SUBORDER: Taxonomic Classification

GREATGROUP: Taxonomic Classification

SUBGROUP: Taxonomic Classification

FAMILY: Taxonomic Classification

MOLCOLMD: Thickness of Mollic Colors Midpoint in inches

MOLCOLRNG: Thickness of Mollic Colors Range (?) in inches

SURCOLVL: Surface Layer Color Value

SURCOLCHR: Surface Layer Color Chroma

MLRA: Major Land Resource Area

ADDMLRA: No. of Additional MLRAs

PWRINDX: Power Index

SMS: Soil Map Symbol

OMRLO: Organic Matter Range (low) in percent

OMRHI: Organic Matter Range (high) in percent

DTEXCMPDSL: Depth to Textural or Compositional Discontinuity (0-60") (low) in inches

DTEXDMPDSH: Depth to Textural or Compositional Discontinuity (0-60") (high) in inches

TEXCMPDSCD: Textural or Compositional Discontinuity Code

HYDSOILCD: Hydric Soil Code

HEL: Highly Erodible Land Code

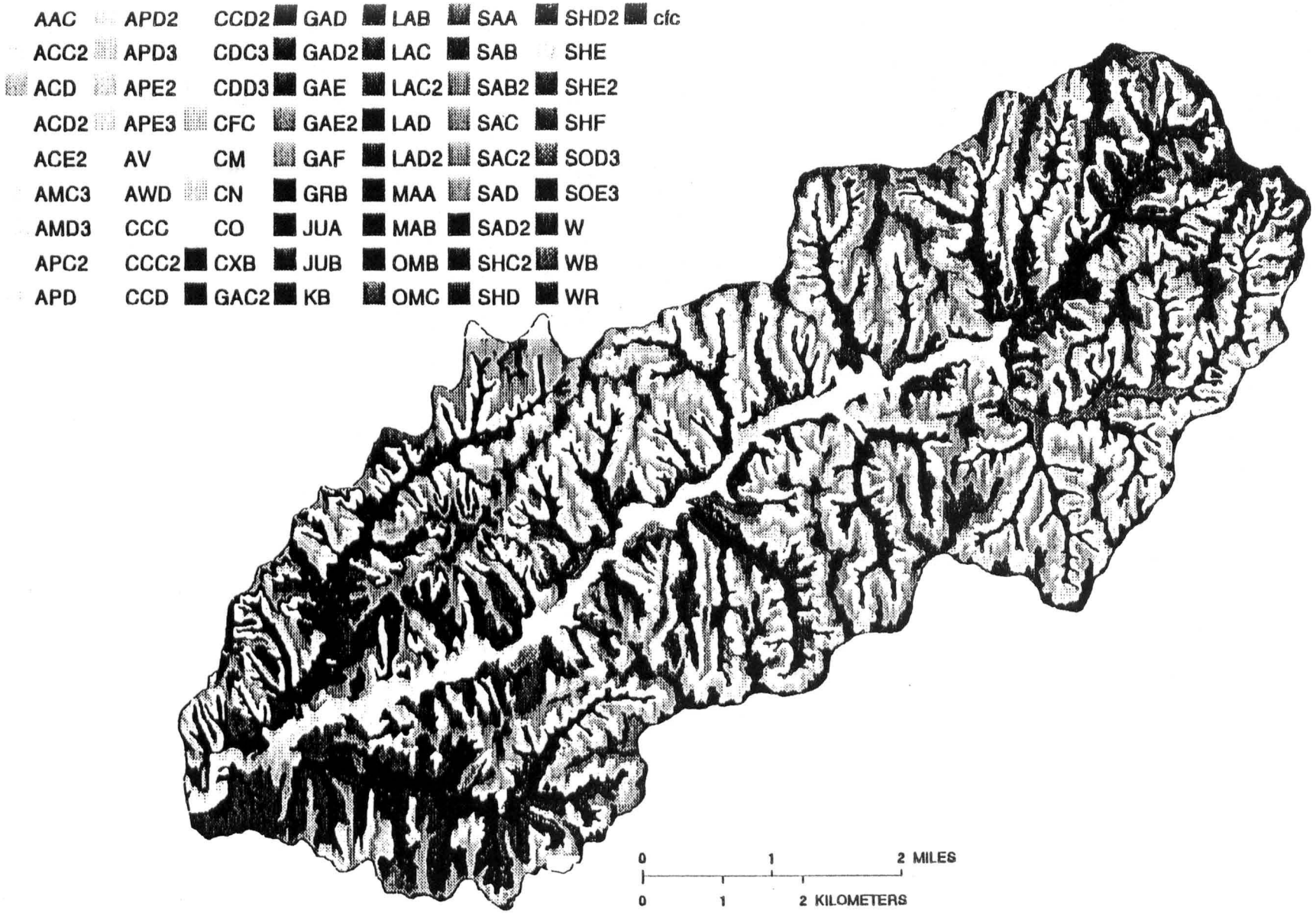


Figure A-5. Lake Icaria Watershed Database, soils

Data set name: TRACTS

Locations: /home/gis5/rocky/thesis/covers

Explanation: This data set is a polygon coverage of the fields in the Icaria Watershed as delineated and numbered by the ASCS in Adams County.

Software: ARC/INFO

Date digitized: December - January 91-92

Source agency: Adams County ASCS

Source: Aerial photographs (photocopies of)

Date source created: 1983

Source contact: Bill Bartenhagen

Source scale: approximately 7,000

Projection information:

Projection: utm

Units: meters

Zone: 15

Geographic extent: North central Adams county, Iowa.

Lineage: Generated by manual digitizing into Arc/Info by Mike Welch.

Notes: A tract is a collection of fields owned by one person. The fields composing the tracts are usually continuous.

Attribute files:

PAT file

Items:

AREA: The area of the polygon

PERIMETER: The perimeter of the polygon

LANDCOVER#: The internal polygon number

LANDCOVER-ID: The polygon user id

FIPS: The 5 digit fips code for the county

TRACT: The tract ID

FIELD: The 2 digit field ID number

HOMESTEAD: Indicates if polygon is a homestead

LANDCOVER: Indicates vegetative coverage

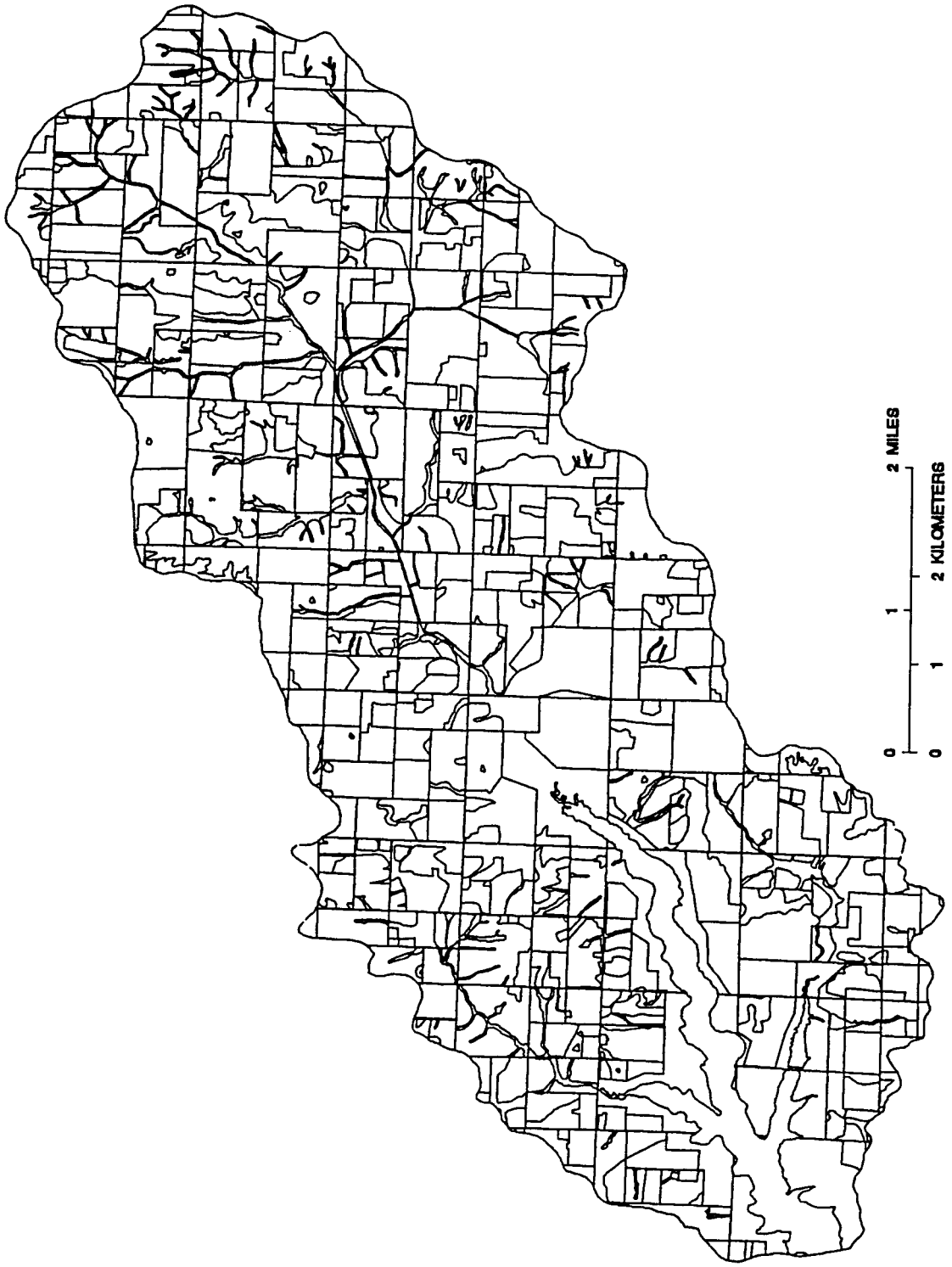


Figure A-6. Lake Icaria Watershed Database, tracts

APPENDIX B: INTRODUCTION TO ARCVIEW

Arcview is a software environment to view and query a spatial database. It is not a typical GIS software but it fills a gap between the GIS developers and the end-users of spatial information. Arcview works on the existing spatial database and focuses on providing a user-friendly mechanism for spatial query process. A GIS specialist can work on a complete GIS package such as ESRI's ARC/INFO to develope and manage GIS database, while the end-user only needs to deal with the simple tasks to view the database using Arcview.

The foundation of Arcview is the concept of a "view". A view in Arcview points to the database. Therefore, if the database changes, the view reflects the change, but the view does not change. The user can easily create or modify the view by adding, removing, or altering themes of spatial data on the current view. To the aspect of display, Arcview can display multiple graphic windows with different map extent. This function makes it easy to zoom in and out on the displayed data. Also the user-friendly spatial and attribute query tools allow the user to retrieve and display geographic information in an easy and efficient way. Its cartographic capabilities make it possible to produce professional map output without extensive effort in map layout design and Arc Macro Language programming.

Arcview was applied in this project to meet user defined requirement. Tracts, soils, transportation, hydrography, topography, and SPOT composite imagery were incorporated to the Arcview. The created view is called THESIS.AV and located in the directory: /home/gis5/rocky/thesis/covers/. The Arcview main menu is shown in Figure B-1, and Arcview graphic display window is shown in Figure B-2.

Arcview is currently installed on Project Vincent at Iowa State University campus. More details about environment set up will be discussed in Appendix D.

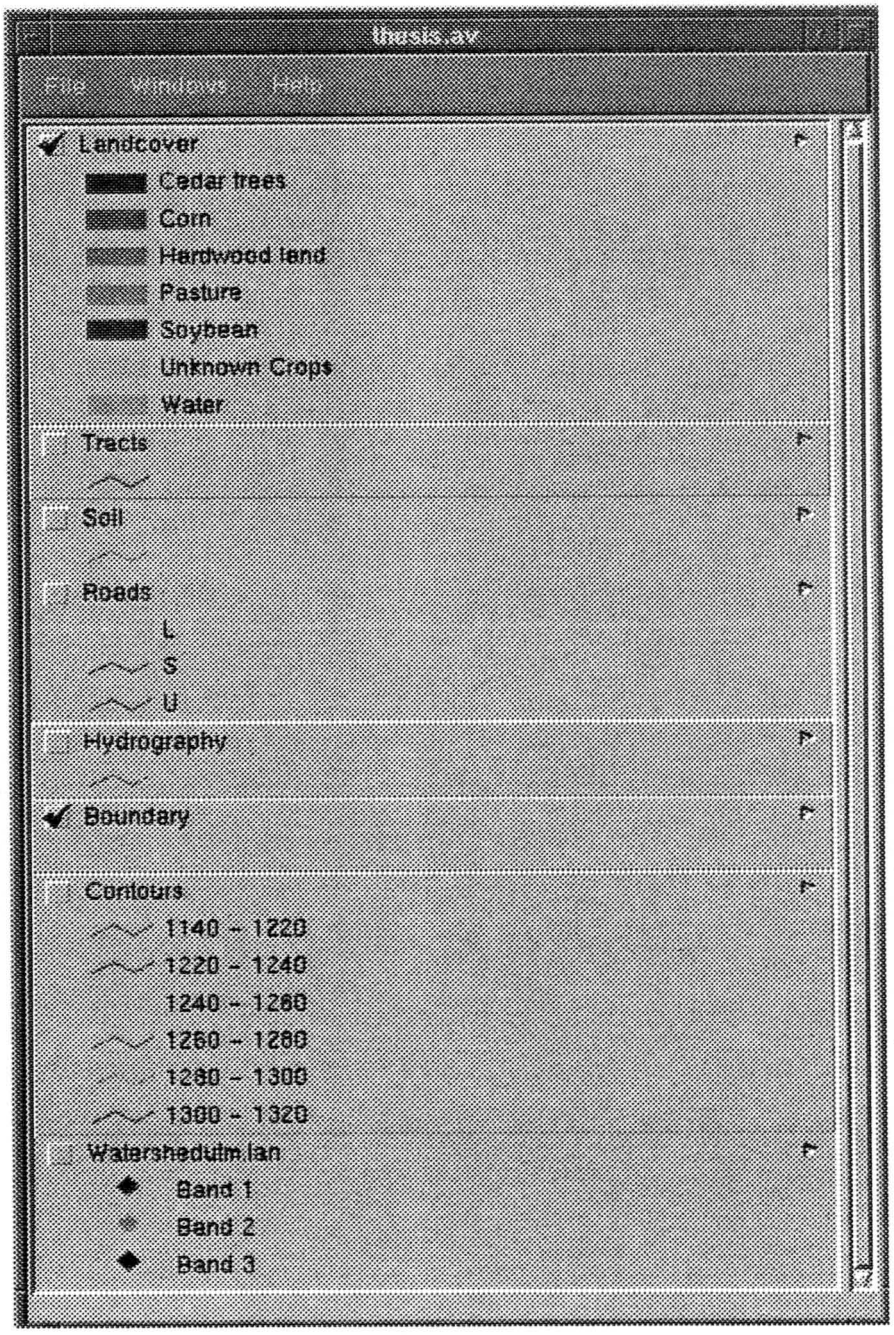


Figure B-1. Arcview Main Menu

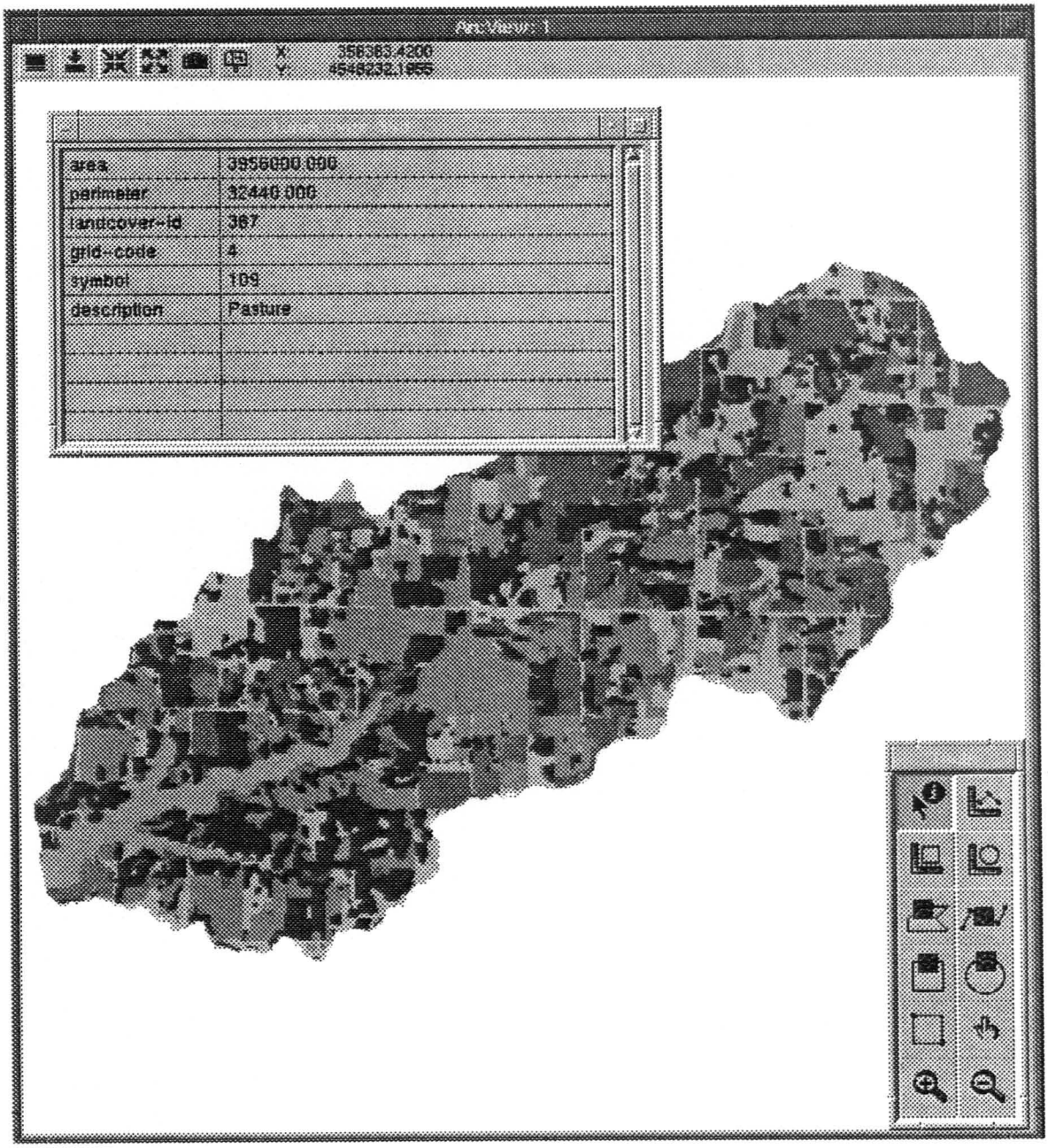


Figure B-2. Arcview Graphic Display Window

APPENDIX C: ARC MACRO LANGUAGE PROGRAMS

Program 1. Plot Output

```
/*this program is used to plot the 4 sections which was used as ground control
data in classification process.*/
    clear
    clearselect
    pagesize 11 8.5
    box 0 0 11 8.5
    startmap section
    maplimits 1 1 10 7.5
    mapposition cen cen
    shadeset colornames.shd
    mapextent /home/gis5/rocky/thesis/covers/boundary
    /*polygons /home/gis/div/pls/adams
    reselelect /home/gis/div/pls/adams poly township = 'T72N' and range = 'R33W'
and section = 5
    aselelect /home/gis/div/pls/adams poly township = 'T72N' and range = 'R33W'
and section = 6
    aselelect /home/gis/div/pls/adams poly township = 'T73N' and range = 'R33W'
and section = 31
    aselect /home/gis/div/pls/adams poly township = 'T73N' and range = 'R33W'
and section = 32
    polygonshades /home/gis/div/pls/adams 4
    polygons /home/gis/div/pls/adams
    arcs /home/gis5/rocky/thesis/covers/boundary
```

```
textquality prop
textfont 'Helvetica'
textsize .1
textoffset 0 0
labeltext /home/gis/div/pls/adams township
textoffset 0. -0.1
labeltext /home/gis/div/pls/adams range
textoffset 0 -.2
labeltext /home/gis/div/pls/adams section
textsize .2
move 5.5 1.2
text 'Figure 6-1 Four Sections used as Ground Truth Data' cc
map end
&if ^ [query 'Make a p.s. file'] &then &return
&type Making the ps file....
&sv delvar [delete section.gra -file]
mkgraphic section section 2
&sys rotate section.ps sectionr.ps
&return
```

Program 2. Display the Histogram of SPOT Data

```
/*this program is used to plot the histogram of three bands SPOT data.*/
clear
clearsel
pagesi 8.5 11
box 0 0 8.5 11
```

```
startmap histogram
textquality prop
textfo 'Helvetica'
textsi .15
shadeset colornames.shd
gridnodatasy 26
maplimits 1 1 7.5 4
mappos cen cen
&r /home/gis4/rocky/3d/histogram adams1
/*graphbar adams1.vat info value count
maplimits 1 4 7.5 7
&r /home/gis4/rocky/3d/histogram adams2
/*graphbar adams2.vat info value count
maplimits 1 7 7.5 10
&r /home/gis4/rocky/3d/histogram adams3
/*graphbar adams3.vat info value count
textcolor 1
move 5.3 2.8
text 'Band 1'
move 5.3 6.
text 'Band 2'
move 5.3 9.1
text 'Band 3'
move 4.25 .9
text 'Figure 4-1 Histograms of each band' cc
map end
```

```
&if ^ [query 'Make a p.s. file'] &then &return
&type Making a graphic file...
&sv delvar [delete histogram.gra -file]
display 1040
histogram
plot histogram.map
display 9999 1
&sv delvar [delete histogram.ps -file]
&sys arc postscript histogram.gra histogram.ps
&return
```

Program 3. Plot the Color Composite Image with SPOT Data

```
/*this program is used to plot the color image of color infrared SPOT image*/
clear
clearsel
pagesi 11 8.5
box 0 0 11 8.5
startmap infrared
maplimits 1 1 10 7.5
mappos cen cen
textquality prop
textfo 'Helvetica'
shadeset colornames.shd
shadesy 17
gridnodatasymbol 26
mape /home/gis5/rocky/thesis/spot/boundary
```

```
gridcomposite rgb adams3 adams2 adams1 linear
resel /home/gis5/rocky/thesis/spot/bd poly lake ne 1
polygonshades bd 26
arcs /home/gis5/rocky/thesis/covers/boundary
textsi .2
textcolor black
move 4.3 7.2
text 'Raw SPOT Imagery' cc
move 4.3 6.95
text 'False Color Composite (infrared)' cc
map end
&if ^ [query 'make a ps file'] &then &return
&type The arc/info is working for you....
mkgraphic infrared infrared 2
&sys rotate infrared.ps inf.ps
&sys rm infrared.ps
&return
```

APPENDIX D. HOW TO SETUP ARC/INFO AND ERDAS

In order to use ARC/INFO and ERDAS on Iowa State University campus, a user must have a Project Vincent account first. This can be done by applying in room 197 in Durham Center. After the user has the access to the Project Vincent, a few system files must be set up before ARC/INFO and ERDAS can be invoked.

In the /home/user/.cshrc.mine file, add following lines to run ARC/INFO:

setenv AVHOME /local/gisproagram/esri61/arcview

setenv SAMPLESHOME /local/gisproagram/esri61/samples61/

setenv ARCHOME /local/gisprogram/esri61/arcexe61

setenv SDLHOME /local/gisproagram/esri61/arcsdl/arc_sdl

setenv ATHOME /local/gisproagram/esri61/arctools

and add the following lines to run ERDAS:

setenv ERDAS '/home/gis2/erdas/'

setenv IMAGAINE '/home/gis2/erdas/imagine/'

Also, the user needs to set a new /home/user/.Xresource file to support ARC/INFO and ERDAS. The new resource file can be created by:

add rocky

cp /home/rocky/.Xresources /home/user/.Xresources

This file will take care the system environment needed to run ARC/INFO and ERDAS.The last thing the user has to remember is that there is a GIS user list on ISU campus. Any new user has to be added to this list before he or she can run ARC/INFO and ERDAS. In another words, he or she must see the GIS administrator first in order to have the ARC/INFO or ERDAS access. The GIS administrator can be reached at 294-2279 or 218 Durham Center.