US Army Corps
of Engineers
Hydrologic Engineering Center

A Tutorial on Creating a Grid Cell Land Cover Data File from Remote Sensing Data

June 1985

TD-22

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
June 1985

2. REPORT TYPE
Training Document

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
A Tutorial on Creating a Grid Cell Land Cover Data File from Remote Sensing Data

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Gary E. Ford, Doreen I. Meyer, & V. Ralph Algazi
Signal and Image Processing Laboratory
Department of Electrical and Computer Engineering
University of California, Davis
Davis, CA 95616

5d. PROJECT NUMBER

5e. TASK NUMBER

5F. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Corps of Engineers
Institute for Water Resources
Hydrologic Engineering Center (HEC)
609 Second Street
Davis, CA 95616-4687

8. PERFORMING ORGANIZATION REPORT NUMBER
TD-22

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/ MONITOR'S ACRONYM(S)

11. SPONSOR/ MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
.

14. ABSTRACT

This document describes the procedures involved in linking a remote sensing system and a spatial data management system. Specifically, it identifies and describes a sequence of steps fro the classification of LANDSAT multi-spectral scanner (MSS) data and the creation of a grid cell land cover data file. The resulting grid cell file can be merged into a Hydrologic Engineering Center (HEC) Spatial Analysis Methodology (SAM) spatial database. The procedures are also applicable to remote sensing data acquired from other sensor systems. The presentation is intended for those who are familiar with the SAM system and the characteristics and potential of remote sensor image data.

15. SUBJECT TERMS
remote sensing, Geographic Information Systems, spatial data management

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			
U	U	U	UU	80	**19b. TELEPHONE NUMBER**

A Tutorial on Creating a Grid Cell Land Cover Data File from Remote Sensing Data

June 1985

US Army Corps of Engineers
Institute for Water Resources
Hydrologic Engineering Center
609 Second Street
Davis, CA 95616

(530) 756-1104
(530) 756-8250 FAX
www.hec.usace.army.mil

TD-22

A Tutorial on Creating a Grid Cell Land Cover Data File from Remote Sensing Data

Gary E. Ford, Doreen I. Meyer, and V. Ralph Algazi

Signal and Image Processing Laboratory
Department of Electrical and Computer Engineering
University of California, Davis
Davis, California 95616

ABSTRACT

This document describes the procedures involved in linking a remote sensing system and a spatial data management system. Specifically, it identifies and describes a sequence of steps for the classification of Landsat multispectral scanner (MSS) data and the creation of a grid cell land cover data file. The resulting grid cell file can be merged into a Hydrologic Engineering Center (HEC) Spatial Analysis Methodology (SAM) spatial database. The procedures are also applicable to remote sensing data acquired from other sensor systems. The presentation is intended for those who are familiar with the SAM system and the characteristics and potential of remote sensor image data.

A Tutorial on Creating a Grid Cell Land Cover Data File from Remote Sensing Data

Contents

List of Figures

List of Tables

1. INTRODUCTION

Spatial data management systems, also known as geographic information systems, provide powerful, practical tools for the management and analysis of spatial data. The primary source of data for spatial data analysis studies has been maps and map data. However, remote sensing systems provide another valuable source of spatial data. Many planning studies involving geographic data would benefit from the use of this data, which can be merged with other geographic data in a spatial database.

In this tutorial document, we discuss the steps involved in linking a remote sensing system and a spatial data management system. Specifically, we identify and discuss a sequence of steps for the classification of Landsat digital data and the creation of a grid cell land cover data file that can be merged into a Hydrologic Engineering Center (HEC) Spatial Analysis Methodology (SAM) spatial data base [1]. The procedures are also applicable to remote sensing data acquired from other sensor systems. Our presentation is intended for people who are familiar with SAM methods and the characteristics and potential of remote sensor image data.

1.1. OBJECTIVES

This tutorial has three principal objectives:

(1) Provide an overview of the major steps involved in processing Landsat multispectral scanner (MSS) digital remote sensing data to create a land cover classification map file and in merging this file into a SAM spatial data base.

(2) Discuss the key steps and issues involved in this procedure when it is performed by Corps of Engineers personnel using software and computers available to the Corps, or when it is performed by an outside agency or company under contract to the Corps.

(3) Provide examples of the major steps of the procedure in its application to a Corps of Engineers project.

1.2. DESCRIPTION OF THE DOCUMENT

This document is intended to be a self-teaching tutorial directed primarily to a user of the SAM spatial data analysis methods who is interested in methods for incorporating Landsat remote sensing data in the spatial database. This document will be useful as a training or reference manual, although it does not provide a comprehensive presentation of Landsat remote sensing data analysis. Instead, it describes the major steps and decisions that are important in creating a SAM spatial land cover database derived from the Landsat data. Issues involved in alternative approaches to the analysis and use of this data are discussed.

Illustrative examples are given for the major steps involved in the University of California, Davis (UCD) Classification Procedure [2]. This procedure is only one of several possible procedures that can be used, and it has been chosen because it was developed and used by the Corps of Engineers.

Although the document is intended to be self-contained, the descriptions of Landsat characteristics, land cover classification, and SAM spatial database methods are necessarily concise. Thus, readers who intend to apply the procedures described here will need a stronger background in these areas. Basic information can be found in the following documents. A short overview of the characteristics of the Landsat MSS system is given in an appendix to this document. An excellent tutorial on Landsat data analysis has been written by Short [3]. Users manuals are available for elements of the SAM system, for example [4, 5], and an excellent tutorial guide manual has been written as an introduction to the use of these methods [6]. Additional references on detailed topics are given later in this document.

1.3. ACKNOWLEDGEMENTS

This document was prepared under contract to the Hydrologic Engineering Center. It is part of a broader contractual effort to develop methods for the analysis and use of Landsat remote sensing data for use by the U.S. Army Corps of Engineers. We wish to express our appreciation to Darryl W. Davis and Arlen Feldman of HEC and to Robert Cermak, formerly of HEC, for their support, guidance, and assistance in this effort.

Darryl W. Davis of HEC provided us with the original outline of the content of this document. We wish to thank him for his many helpful suggestions, for providing us with the support and encouragement to prepare the document, and for his thorough review of the drafts of the document.

We wish to thank John F. Gauthier of the U.S. Army Corps of Engineers Detroit District for the information he provided concerning the methods for aggregation of land cover

data into large grid cells.

We wish to thank Richard Rollins for his professional skill in writing, testing, and documenting the software for the UCD Classification Procedure, and in providing his comments on portions of this document.

2. GRID CELL DATA BANKS AND REMOTE SENSING

The use and management of spatial (map type) data with computer technology is an important task for a wide variety of activities in the Corps of Engineers. The increasing availability of remotely sensed (usually imagery) data has further focused attention in this area. The range of spatial data that is used by Corps offices includes topography and other physiographic features, land use, cultural features, environmental habitat, vegetative cover, atmospheric and subsurface features, and many others. The list can be quite long. Well documented and maintained computer-based systems are needed to accomplish the task in an efficient, effective manner. The Spatial Analysis Methodology (SAM), developed by the Hydrologic Engineering Center is a readily useable spatial data management system.

The HEC-SAM system was developed in the late 1970's to provide for integrated use of spatial (map type) data in water resources studies — primarily flood control. The data management features of the system, however, have wider utility than that required for flood control studies. For example, SAM has been used to identify potential dredged-material disposal sites [7]. Computer programs are available for assistance in creating a spatial (grid cell) data bank, editing and managing the files, performing simple combinational analysis, and producing graphic displays. Figure 2.1 shows the functional elements of the HEC-SAM system, with an emphasis on the data management features.

Data from remote sensing scanners, classified and resampled to the desired grid, can be inserted into a spatial data bank and managed by the use of the BANK program [5]. Data maps may be processed into grid form by machine digitizing to grid or polygon data forms and then inserted into the spatial database using the programs BANK (grid data) or GRIPS (polygon data). Plots may be produced of data in the grid or polygon forms. Map data may be manually gridded as an alternative to the use of digitizing equipment. The BANK program has a number of features that assist in processing manually created grid data into a data bank. Manual preparation of grid data is quite practical and economical for smaller data banks on the order of several thousand cells.

Spatial data resident in the data bank may be plotted, combined, analyzed, weighted, with results saved as new data bank variables and plotted (program RIA [4]), or may be processed by applications programs. The applications programs used depend upon the investigation being performed. Programs now exist for several flood control applications, including hydrologic modeling, flood damage, and land surface erosion [8]. Other

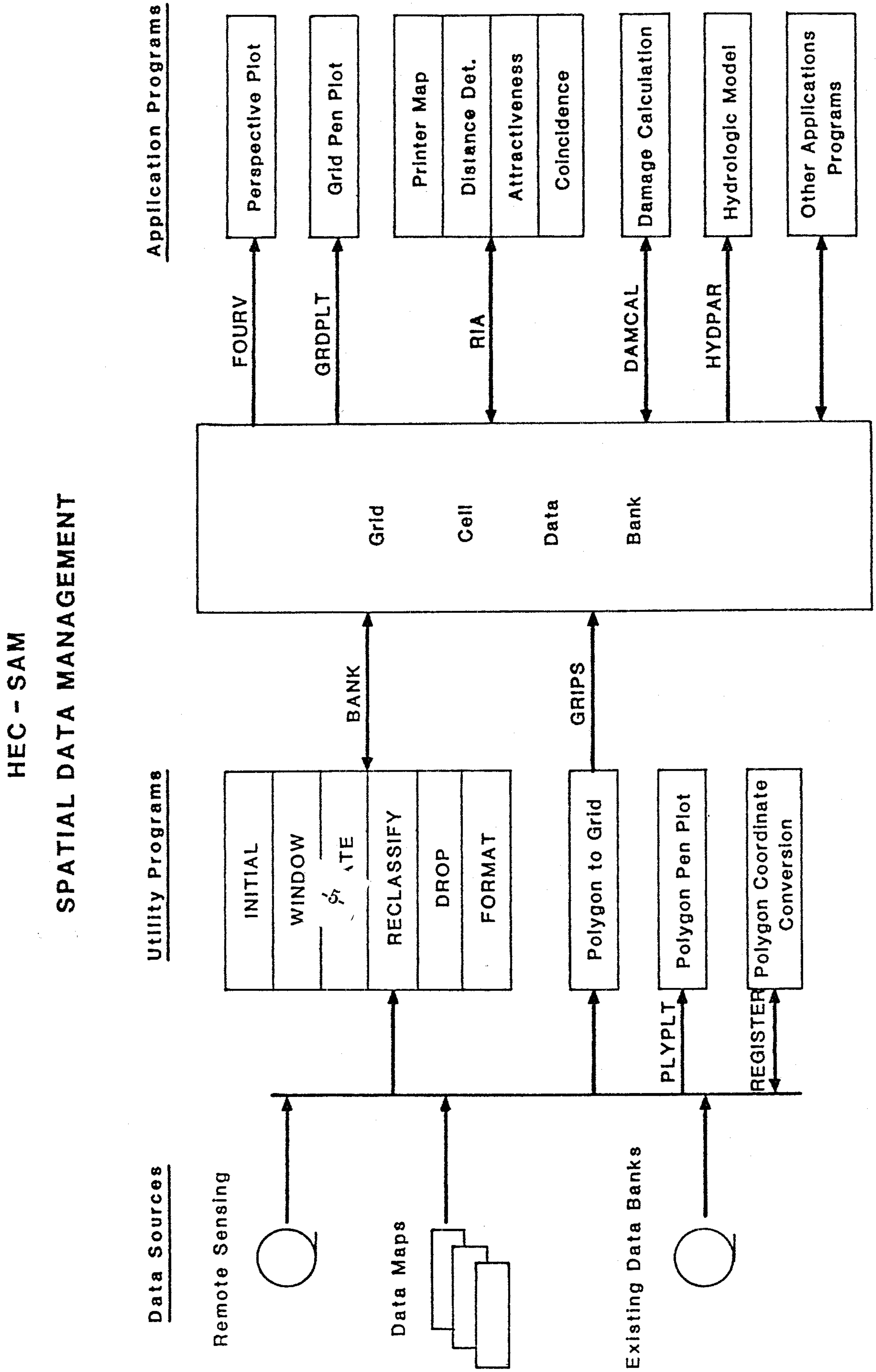


Figure 2.1 HEC-SAM Spatial Data Management System

applications would require development of interfaces to (presumably) existing programs.

An important spatial data variable is land cover or land use. For example, the hydrologic modeling of a watershed, particularly urban or urbanizing basins, requires the determination of land use distribution. The amount and timing of runoff is directly related to infiltration capacity. The extent of pervious and impervious land surfaces is the most important variable in the modeling. Water quality parameters have a similar dependence on land cover. The rate of accumulation of a pollutant is normally expressed as a function of land cover. Water resource planners are interested in not only an assessment of the present state of the water and related resource system, but also its possible future states. By expressing hydrologic parameters as a function of current land cover, a planner is able to predict the impact of land use changes on the quantity and quality of runoff in the future.

Remote sensing data from multispectral scanner systems mounted on satellites or aircraft can be computer processed to provide land cover information. This information can be acquired at a scale similar to that of the resolution of the scanner system with a cost and accuracy competitive with that of conventional methods based on manual interpretation of aerial photography. For example, the Landsat MSS data can provide land cover data at a scale of 1 acre grid cells with accuracy that is acceptable for hydrologic modeling [9, 10, 11, 12, 13]. The digital remote sensing data can be directly analyzed by available classification procedures, and can be resampled for automatic inclusion in a geographic database. Thus, if a geographic data based study requires the determination of land cover at a level of detail, accuracy, and scale suitable for remote sensing data, then an approach using this data should be considered.

3. OVERVIEW OF MAJOR PROCEDURAL STEPS

In this document, we discuss a procedure for the determination of land cover from Landsat remote sensing data and the creation of a grid cell land cover data file. We present the procedure as a sequence of steps, as illustrated in figure 3.1. In this section, we present a brief outline of these procedural steps and identify the sections of the document where the steps are described in detail.

(1) Acquire the data required in the procedure, including a computer-compatible tape (CCT) containing the Landsat multispectral scanner (MSS) data, a photograph of the selected Landsat scene, USGS topographic maps, and high-flight aerial photographs of the study area. (Section 4)

(2) Delineate the study area on the USGS maps and on the Landsat scene photograph, and use this information to read the selected subscene from the Landsat CCT and to create the Landsat image data file. Detect and correct radiometric errors in the Landsat image. (Section 5)

(3) Establish the database coordinate system and the grid cell configuration (size and shape). (Section 6)

(4) Compute the coordinate transformation from the database coordinate system to the Landsat coordinate system. (Section 6)

(5) Choose the set of land cover classes to be used in the classification of the Landsat data. (Section 7)

(6) Apply a computer classification algorithm to the Landsat data to create a land cover map in the Landsat coordinate system. (Section 7)

(7) Resample the land cover map to transform its geometry to that of the database coordinate system. (Section 8)

(8) Assess the accuracy of the geometric rectification and the land cover classification. (Section 9)

(9) Create the land cover data file in a grid cell format that can be read by the HEC-SAM system. (Section 10)

In addition, following the description of the procedure, we discuss the key issues involved in applying this procedure (Section 11).

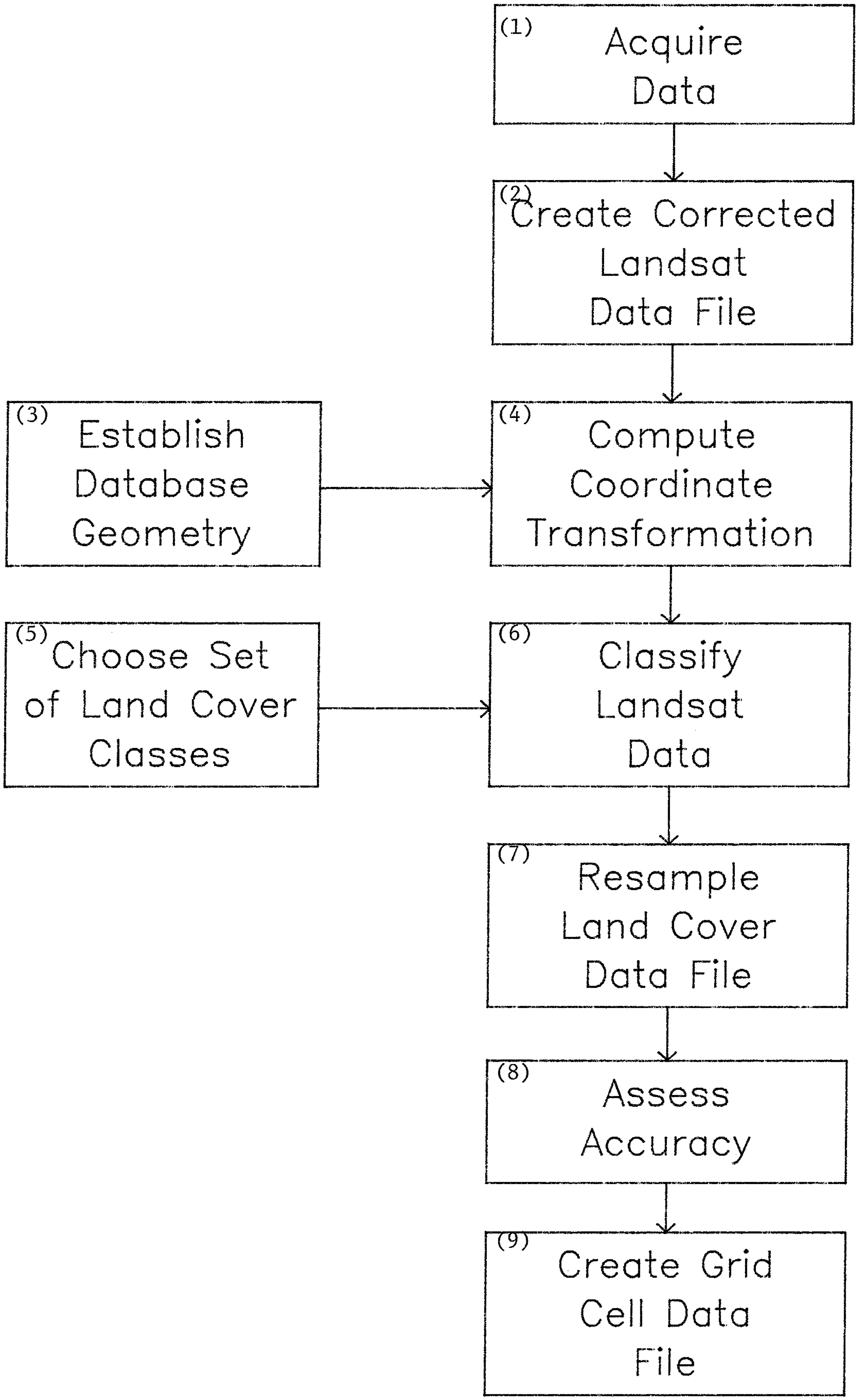


Figure 3.1 Major Procedural Steps

4. DATA ACQUISITION

The following data are required to determine land cover from Landsat remote sensing data:

(1) Computer-compatible digital data for a Landsat scene.

(2) Black and white photographic print of the Landsat scene.

(3) Aerial photographs of the study area.

(4) USGS topographic maps of the study area.

This section describes how to request information about data availability, and what criteria to use in selecting data. Sections 4.1, 4.2, and 4.3 describe this process for Landsat products, aerial photography, and topographic maps, respectively.

4.1. LANDSAT PRODUCTS

The primary data source is the digital data from the Landsat MSS system, available as a computer-compatible tape (CCT). A black and white photographic print of the Landsat scene is also required, for use in locating and extracting the data for the study area. These products are available from the Earth Resources Observation System (EROS) Data Center.

Requesting Information On Available Landsat Data

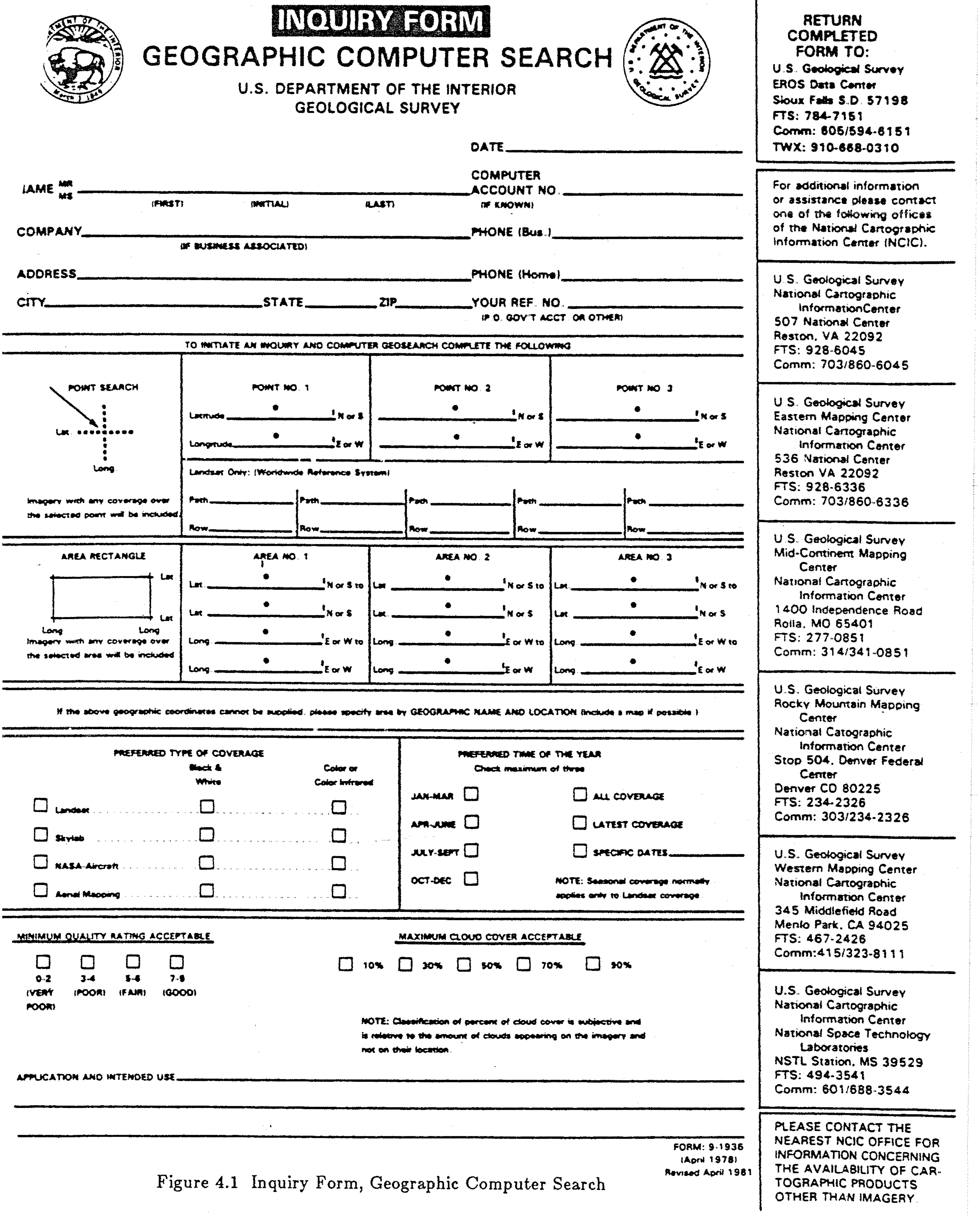
The EROS Data Center maintains a storage-retrieval system for Landsat imagery. To request a computer-generated search for the imagery covering the study area, use the "Inquiry Form, Geographic Computer Search" given in figure 4.1. Use the following criteria in filing out the inquiry form:

Location: Specify the latitude and longitude coordinates of the corner points of a rectangle covering the study area (called the area rectangle by EROS). To minimize the number of scenes to be searched, limit the size of the area rectangle as much as possible.

Preferred type of coverage: Request both black/white and color.

Preferred time of year: Request all coverage available.

Minimum quality rating acceptable: Chose the rating 0-2. This rating refers to the photographic quality of the print of the Landsat band 5 data, and not to the quality of the digital data on the CCT.



INQUIRY FORM
GEOGRAPHIC COMPUTER SEARCH

U.S. DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY

DATE ____________

IAME MR / MS ____________ (FIRST) (INITIAL) (LAST)

COMPUTER ACCOUNT NO. ____________ (IF KNOWN)

COMPANY ____________ (IF BUSINESS ASSOCIATED)

PHONE (Bus.) ____________

ADDRESS ____________

PHONE (Home) ____________

CITY ____________ STATE ____________ ZIP ____________

YOUR REF. NO. ____________ (P.O. GOV'T ACCT. OR OTHER)

TO INITIATE AN INQUIRY AND COMPUTER GEOSEARCH COMPLETE THE FOLLOWING

POINT SEARCH

Imagery with any coverage over the selected point will be included

POINT NO. 1	POINT NO. 2	POINT NO. 3
Latitude ____ • ____ ' N or S	____ • ____ ' N or S	____ • ____ ' N or S
Longitude ____ • ____ ' E or W	____ • ____ ' E or W	____ • ____ ' E or W

Landsat Only: (Worldwide Reference System)

Path ____	Path ____	Path ____	Path ____	Path ____
Row ____	Row ____	Row ____	Row ____	Row ____

AREA RECTANGLE

Imagery with any coverage over the selected area will be included

AREA NO. 1	AREA NO. 2	AREA NO. 3
Lat. ____ • ____ ' N or S to	Lat. ____ • ____ ' N or S to	Lat. ____ • ____ ' N or S to
Lat. ____ • ____ ' N or S	Lat. ____ • ____ ' N or S	Lat. ____ • ____ ' N or S
Long. ____ • ____ ' E or W to	Long. ____ • ____ ' E or W to	Long. ____ • ____ ' E or W to
Long. ____ • ____ ' E or W	Long. ____ • ____ ' E or W	Long. ____ • ____ ' E or W

If the above geographic coordinates cannot be supplied, please specify area by GEOGRAPHIC NAME AND LOCATION (include a map if possible.)

PREFERRED TYPE OF COVERAGE

		Black & White	Color or Color Infrared
☐	Landsat	☐	☐
☐	Skylab	☐	☐
☐	NASA Aircraft	☐	☐
☐	Aerial Mapping	☐	☐

PREFERRED TIME OF THE YEAR
Check maximum of three

JAN-MAR ☐
APR-JUNE ☐
JULY-SEPT ☐
OCT-DEC ☐

☐ ALL COVERAGE
☐ LATEST COVERAGE
☐ SPECIFIC DATES ____________

NOTE: Seasonal coverage normally applies only to Landsat coverage.

MINIMUM QUALITY RATING ACCEPTABLE

☐ 0-2 (VERY POOR)
☐ 3-4 (POOR)
☐ 5-6 (FAIR)
☐ 7-9 (GOOD)

MAXIMUM CLOUD COVER ACCEPTABLE

☐ 10% ☐ 30% ☐ 50% ☐ 70% ☐ 90%

NOTE: Classification of percent of cloud cover is subjective and is relative to the amount of clouds appearing on the imagery and not on their location.

APPLICATION AND INTENDED USE ____________

FORM: 9-1936
(April 1978)
Revised April 1981

RETURN COMPLETED FORM TO:

U.S. Geological Survey
EROS Data Center
Sioux Falls S.D. 57198
FTS: 784-7151
Comm: 605/594-6151
TWX: 910-668-0310

For additional information or assistance please contact one of the following offices of the National Cartographic Information Center (NCIC).

U.S. Geological Survey
National Cartographic InformationCenter
507 National Center
Reston, VA 22092
FTS: 928-6045
Comm: 703/860-6045

U.S. Geological Survey
Eastern Mapping Center
National Cartographic Information Center
536 National Center
Reston VA 22092
FTS: 928-6336
Comm: 703/860-6336

U.S. Geological Survey
Mid-Continent Mapping Center
National Cartographic Information Center
1400 Independence Road
Rolla, MO 65401
FTS: 277-0851
Comm: 314/341-0851

U.S. Geological Survey
Rocky Mountain Mapping Center
National Catographic Information Center
Stop 504, Denver Federal Center
Denver CO 80225
FTS: 234-2326
Comm: 303/234-2326

U.S. Geological Survey
Western Mapping Center
National Cartographic Information Center
345 Middlefield Road
Menlo Park, CA 94025
FTS: 467-2426
Comm:415/323-8111

U.S. Geological Survey
National Cartographic Information Center
National Space Technology Laboratories
NSTL Station, MS 39529
FTS: 494-3541
Comm: 601/688-3544

PLEASE CONTACT THE NEAREST NCIC OFFICE FOR INFORMATION CONCERNING THE AVAILABILITY OF CARTOGRAPHIC PRODUCTS OTHER THAN IMAGERY.

Figure 4.1 Inquiry Form, Geographic Computer Search

Maximum cloud cover acceptable: Choose 30 percent cloud cover, as this is usually the maximum cloud cover that will allow for an adequate land use analysis.

The completed form is sent to:

U.S. Geological Survey
EROS Data Center
Sioux Falls, South Dakota 57198

Selecting a Landsat Scene

Within two weeks of submission of a search form, the EROS Data Center will send a computer-generated listing of the available Landsat scenes, with considerable information about each scene. From this listing, select the Landsat CCT that most closely meets the following specifications:

Path and row: Landsat data is indexed by an orbital path and row. The computer listing will include scenes from several (path, row) coordinates. The listing for each scene will also include the corner point latitude/longitude coordinates of the scene. Compare the corner point coordinates of scenes for a given (path, row) with the area of interest and select the appropriate (path, row) coordinates. Notice that the corner coordinates for a given (path, row) will vary, so check the coordinates carefully if the area of interest is near the boundary of a Landsat scene. Consider all scenes having the chosen (path, row) coordinates. Landsat scenes overlap by about 10% in both directions, so most study areas will be covered by a single scene. For large study areas, it may be necessary to acquire two adjacent scenes. This will increase the cost and complexity of the study, as it is difficult to merge the scenes (referred to as mosaicking).

Imagery-type: Landsat MSS.

Cloud cover: Cloud cover will interfere with the land cover analysis. If there is a possibility of having cloud cover over the area of interest, order a black and white photograph of band 5 to check for the presence of clouds over the study area before ordering the tapes.

Exposure date: In general, the accuracy of land cover classification of Landsat data will be highest for data acquired when vegetative cover is in a growth phase. Grasslands are correctly classified at that time, while they may be misclassified in the late summer when the grass has dried. Also, later in the growth phase, deciduous trees that are fully leafed-out may obscure buildings. The time of this early vegetative growth phase depends on the climate of the study area.

CCT availability: An "N" in the CCT column means that EROS does not have a master tape, and they must order it from NASA Goddard Space Flight Center. This will require an additional six weeks for delivery. A "Y" in the CCT column means that the CCT is available at EROS and can be delivered in three weeks.

In the UCD Classification Procedure [2], we recommend use of the CCT referenced by product code 184-B. This is a 9-track, 1600 bit per inch (bpi) tape containing data from all MSS bands written in the band interleaved by line (BIL) format. Other CCT types include densities of 800 or 6250 bpi, or a band sequential (BSQ) format. These may be of use if other classification procedures are used.

To aid in locating the study area on the Landsat scene, we recommend use of a black and white photographic print of band 5 of the Landsat scene, referenced by product code 26. This is a 29.2 inch square print having a scale of 1:250,000.

4.2. AERIAL PHOTOGRAPHY

High-flight aerial photography is useful in the land cover analysis of Landsat remote sensing data as the principal source of ground-truth data. It is primarily used to locate and select classification training areas. These are regions of the study area having a homogeneous and identifiable land cover.

Requesting Information About Available Aerial Photography

The EROS Data Center also provides NASA aerial photography. The inquiry form used to request a search for Landsat data can also be used to request a search for this photography. Request both black/white and color infrared as the "preferred type of coverage."

The National Cartographic Information Center of the U.S. Geological Survey maintains the same aerial photography database as the EROS Data Center. In addition, they maintain a data base of aerial photography from other vendors. To request information from this data base, write or call an office of the National Cartographic Information Center listed on the right hand side of the form given in figure 4.1. In requesting this information, specify the search area, acceptable cloud cover, and exposure dates of interest.

Selecting Aerial Photography

To determine the land cover from the Landsat data, you will require aerial photographs of all land cover types in the study area. Select photographs that most closely meet the following conditions:

Corner point coordinates: Select photographs that contain representative portions of the study area, and representative samples of all land cover types under consideration.

Imagery type: The easiest photographic product to work with is a 9 inch color film positive (CIR-9.0 inch), or a 9 x 18 inch color film positive.

Scale: We have found that a photographic scale of 1:50,000 is good, 1:100,000 is marginal, and 1:125,000 is inadequate for identifying land cover classes.

Cloud cover: A higher percent cloud cover can be tolerated with the aerial photography, as long as significant features are not obscured.

Exposure date: It is especially critical to coordinate the date of the aerial photography with the date of the Landsat scene, especially in areas undergoing rapid urban development.

For NASA aircraft photography, we recommend the use of 9 inch color film positive photographs (product code 53).

4.3. TOPOGRAPHIC MAPS

The USGS 7.5 minute (1:24,000) series topographic maps are used to delineate the study area, aid in the geometric rectification of the Landsat data with the geographic data base, and aid in the land use classification of the Landsat data. They provide enough detail and description to locate most natural and cultural features of interest, and are photo-revised periodically. Both Universal Transverse Mercator (UTM) and longitude-latitude coordinates are printed on the topographic maps. In addition, a USGS 1:250,000 scale topographic map is used in the extraction of data for the study area from the full Landsat scene.

Requesting Information About Available Maps

Request an index from the appropriate USGS office by specifying the state(s) in which the study area is located. Indexes and maps for states West of the Mississippi River may be ordered from:

Branch of Distribution, USGS
Federal Center
Denver, CO 80225

Indexes and maps for states East of the Mississippi River may be ordered from:

Branch of Distribution, USGS
1200 South Eads Street
Arlington, VA 22202

5. LANDSAT DATA FILE CREATION

The first step in processing the Landsat data is to extract the data for the study area from the CCT. The image quality of the extracted data is then analyzed and if problems are detected, an image restoration algorithm, known as radiometric correction, is applied. In this section, we describe the processes involved in data extraction and radiometric correction.

5.1. STUDY AREA LOCATION AND EXTRACTION

A Landsat scene covers 115 by 115 miles (185 by 185 kilometers), but a typical study area will only cover a small portion of a scene. Thus, the subimage covering the study area must be read from the Landsat CCT to create a computer image data file. This file can then be displayed as an image or printed as a graphics product to verify that the extracted data actually covers the study area.

The first step in extracting the data is to locate the study area on the USGS maps. First delineate the study area boundary on the 7.5 minute series topographic maps, or on other basemaps used in the geographic study. Use this information to draw a rectangle on the 1:250,000 scale topographic map that encloses the study area. Then locate the study area on the 29.2 inch (1:250,000 scale) photographic print of the Landsat scene, and again draw a rectangle that encloses the study area. The sides of the rectangle should be parallel to the sides of the photograph.

The rectangle drawn on the photograph is used to determine the Landsat coordinates of the study area. But before describing this process, we must provide a short description of the Landsat image format. A Landsat image is a four-variable grid cell data file, with a grid cell for each location within the scene. The four data variables in each cell are the radiometric intensities for the four spectral bands. In image processing terminology, the grid cells are known as picture elements, or pixels. If we interpret the Landsat data as a matrix of light intensities, the Landsat coordinates are the line number (vertical, north-south direction), and the column number (horizontal, east-west) of the matrix, with the origin (0,0) at the northwest (upper left hand) corner. To extract the data for the study area, it is necessary to determine the Landsat coordinates of the upper left hand corner of the study area and the horizontal (east-west) and vertical (north-south) dimensions of the study area in pixels.

A full Landsat scene is 3240 pixels (horizontal) by 2400 pixels (vertical). It is not square because the pixel spacing is roughly 79 meters horizontally by 57 meters vertically. This information and measurements made on the photographic print of the scene can be used to estimate the study area Landsat corner coordinates and dimensions.

The procedure is complicated by different subdivisions of scene data into files in the various CCT formats. Recent Landsat data is written in band interleaved by line (BIL) or band sequential (BSQ) format. In these formats, the scene has been divided in half vertically, and one has to determine which half of the scene contains the study area. Pre-1979 Landsat data is written in a band interleaved by pixel-pair (BIP2) format. This format divides the Landsat scene into four vertical strips, so that it is necessary to determine the strip number(s) containing the study area, and then to determine the scene coordinates within the strip. The UCD Classification Procedure [2], provides a computer program to handle the details of determining the Landsat coordinates of the study area. Measurements made on the photographic print are input to the program, and the required Landsat coordinates are computed, and passed to the tape read program that extracts the study area data file.

The size of the study area is an important factor in the procedure. Many land cover classification procedures limit the maximum size of the data files to be processed to 512 by 512 or 1024 by 1024 pixels. Larger study areas require that subscenes be processed independently, and it is then more difficult to maintain geometric and classification accuracy. The data file should be at least 256 by 256 pixels, to locate major features and to define training sets for use in classification.

When the location of study area in Landsat coordinates is known, the four study area data files, one per spectral band, are read from the Landsat CCT. The formats of the Landsat CCTs are sufficiently complicated that it is generally necessary to read the tape using a program written specifically for that purpose, such as that provided by the UCD Classification Procedure [2]. Some computer systems provide general tape read utility programs that can handle the CCT formats, but this approach is more difficult than using a specialized program. Details on the characteristics of the Landsat CCT formats are given in [14, 15].

After creating the four study area data files, it is necessary to verify that the data read actually covers the study area. The image data can be displayed on an image display system, or if this is not possible, an enhanced version of the image can be printed for verification. In the UCD Classification Procedure [2], curvilinear features, such as roads and outlines of water bodies are enhanced for this purpose. These features are used to verify that the extracted data covers the study area.

5.2. RADIOMETRIC CORRECTION

Radiometry refers to the numerical intensity value of the sensed data [16]. In the case of visible and infrared sensors, the digital radiometric value represents the scene radiance or reflectance. Errors in radiometry can be caused by variations in the sensor responses. In Landsat MSS data, the most serious errors are caused by line-to-line variations in the response of the sensors. The MSS system scans six image lines at once, using six separate sensors for each of the four spectral bands. If the sensitivities of the six sensors are not identical, there will be a six-line periodic variation in radiometric intensity. These errors are visible as a periodic horizontal striping, streaking, or banding [17]. Statistically they result in periodic variations of the line-by-line average radiometric intensity [18]. If left uncorrected, these radiometric errors will degrade the accuracy of the classification of the data.

A model for these errors and an algorithm to detect and correct the errors has been developed [17, 19]. The error detection procedure involves computing the scan line mean and variance of the radiometric intensities from each sensor in each of the four spectral bands. If the difference in line means for a given band is statistically significant, then the data for that band should be corrected. A line mean variation greater than fifteen percent of the largest standard deviation for sensors in that band indicates that the radiometric error is severe and requires correction. The correction method involves selecting one sensor per band as a reference and then estimating equalization functions from the image data, using histogram equalization methods. The estimated equalization functions are applied to correct the radiometry of the data file before applying the classification algorithm.

6. DATABASE GEOMETRY

Before the grid cell data file can be created, the desired geometry of the spatial database must be established. This requires the selection of a coordinate system and grid cell configuration, and relating the grid cell coordinates to the database coordinate system. The Landsat coordinate system will generally differ from the database coordinate system, so that it is necessary to determine a mathematical transformation from one coordinate system to the other. This transformation is used to rectify the Landsat-derived land cover data with the database geometry.

6.1. SELECTION OF THE DATABASE GEOMETRY

All data placed in a spatial database has to be registered to a common coordinate system. Three of the most commonly used coordinate systems in the continental United States are the spherical graticule (latitude, longitude), the UTM projection, and the state plane coordinates. These coordinate systems are recorded on the USGS topographic maps. A discussion of the relationship among these coordinate systems is given in [6]. The UTM coordinate system has been employed in previous applications of spatial database methods to Landsat data analysis [2].

Selection of the grid cell configuration involves choosing the grid cell shape and size. This choice determines the unit size of all data variables and feature assignments in the spatial database. The size of the grid cell depends on the spatial resolution required for the most detailed analysis of the database. Thus, the intended applications of the spatial database methods are of primary importance in this selection. However, if Landsat-derived data is to be merged into the spatial database, the resolution of the Landsat data must also be considered.

Guidance for the selection of the grid cell configuration is given in [6], and further discussion of the issues involved is given in [20]. In most applications, grid cell dimensions are based on the area covered by a computer printer character on a USGS 7.5 minute series topographic map. A computer character that measures 1/10th of an inch wide by 1/6th inch high (printer spacing 6 lines per inch) corresponds to 200 feet by 333.33 feet on the ground, and an area of 1.53 acres. A character measuring 1/10th inch wide by 1/8th inch high (8 lines per inch) corresponds to 200 feet by 250 feet on the ground, and an area of 1.14 acres. This choice of rectangular cell size permits the use of line printers for fast and inexpensive paper output at the proper scale for checking data and results. These choices are well suited to the resolution of Landsat data. The instantaneous field of view (IFOV) of the Landsat MSS sensor is about 79 meters (259 feet) square, equal to 1.53 acres. A choice of grid cell size that is significantly different than the Landsat IFOV will lead to problems when the Landsat-derived land cover data is transformed to the grid cell coordinate system. These problems are discussed in section 8, where guidance is given for choices of non-standard grid cell sizes.

Finally, the correspondence between the grid cell coordinates and the database coordinate system must be established. For grid cell data, the coordinate system is generally the row and column indices of each cell. Rows and columns are referenced to a common origin (row 0, column 0), usually located at the upper left corner of the study area. This origin is then related to the geographic coordinates in the database coordinate system. The horizontal and vertical spacing of the grid cell centroids must also be determined in units of the database coordinate system, if a grid cell configuration other than the standard configurations described above has been chosen.

6.2. GEOMETRIC RECTIFICATION

For a variety of reasons to be discussed below, the geometry of the Landsat data file will differ from that selected for the spatial database. Before the Landsat-derived land cover data can be merged into the the spatial database, it must be rectified to the geometry of the database. This is done by determining the mathematical transformation from Landsat coordinates to database coordinates, and then transforming the Landsat data to the grid cell coordinate system, using a process known as image resampling.

Sources of Geometric Error

It is useful to consider first the types of spatial discrepancies between the Landsat and database coordinate systems:

(1) Translation: The origins of the two coordinate systems will differ, as the coarse graphical method used to extract the Landsat data for the study area is not accurate to within the grid cell dimensions. In other words, the upper left hand reference corners of the Landsat data file and the grid cell database will not correspond to the same point on the ground and a translation is required to bring them into registration.

(2) Scale: As discussed in section 6.1, the spacing of the grid cells will differ from the spacing of the Landsat pixels, creating a scale change between the two coordinate systems.

(3) Aspect Ratio: Again as discussed in section 6.1, the aspect ratio, or ratio of horizontal to vertical distance of the grid cells (200 feet to 250 or 333.33 feet) differs from that of the Landsat pixels (259 feet to 187 feet) and different horizontal and vertical scalings must be applied.

(4) Rotation: The Landsat satellite collects data along its orbital path, which is oriented about 3 degrees from the longitude lines. Thus, the data acquired will be rotated with respect to the coordinates of the spatial database.

(5) Skew: As the Landsat satellite collects data as it moves south along its orbital path in the northern hemisphere, the Earth rotates on its axis, causing each successive image line acquired to be shifted westward, causing a skew distortion in the image.

(6) Distortion: The image is also subject to more localized geometric errors, because of the combined effects of the sensor operation, orbit and attitude anomalies, and terrain effects.

Further discussion of these sources of geometric errors is available in [16, 21, 22, 23].

Coordinate Transformation

The mathematical transformation of coordinates commonly used to correct these geometric discrepancies is a bivariate polynomial function having unknown coefficients. The coefficients are determined using the method of least squares.

An automated procedure is available to determine the coordinate transformation. To apply this procedure, several ground control points (GCPs, referred to as matchpoints in [6]) must be located on both the Landsat image and the topographic maps. The GCPs must be stable, easily locatable terrain or cultural features such as road intersections, airports, land-water interfaces, or geological and field patterns. The common method of locating the GCPs in the image is to display the image on a high resolution display system and use an interactive cursor overlay.

In cases where an interactive display is not available, as in the UCD Classification Procedure [2], a noninteractive procedure can be used in which curvilinear Landsat features such as roads and shore lines are digitally enhanced and printed as standard line printer output. The intersections of the curvilinear features are used as GCPs and they are located by fitting smooth curves to sections of these features that are apparent on the printout. Road intersections are generally the most reliable features, although road-water intersections are also useful.

In areas where GCPs are not easily detected, virtual control points can be used, which result from the intersections of projections of the curvilinear features. Examples of virtual control points are shown in figure 6.1. The most effective use of virtual control points is demonstrated by control point C1, where straight sections of a curved road are extended to an intersection. Other well-defined control points are C2, C3, and C4. Extending lines to locate exact points within large freeway intersections is also useful. Control points C5 and C6 are not used since they will not be accurate because of the shallow angles of intersection. Lines cannot be defined accurately by two pixels alone, as there is too much error in the slope of a line defined in this way. Projections far away from observed features are subject to substantial errors and should not be used.

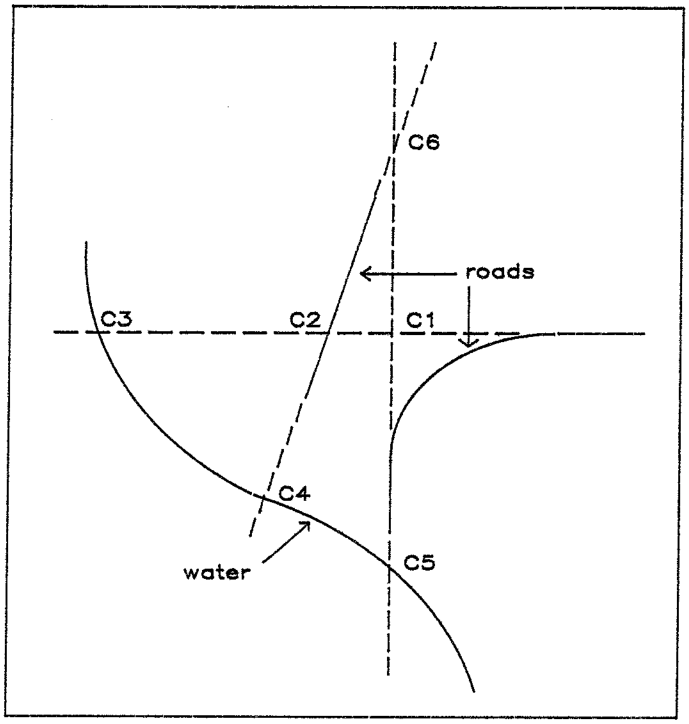


Figure 6.1. Virtual control points

The corresponding locations of the GCPs on the topographic maps are determined using a digital tablet, or by making manual measurements on the maps. These locations should be recorded using the coordinate system of the spatial database.

To rectify the Landsat image to the geometry of the grid cell database to within one pixel, or grid cell, accuracy over the study area it is necessary to accurately determine the locations of 25 to 35 control points that are spatially distributed as uniformly as possible over the study area. These GCPs are used to estimate the transformation coefficients by minimizing the sum of squared errors in the Landsat coordinate system between the image GCPs and the transformed map GCPs. The automated procedure also provides a statistical assessment of the accuracy of the transformation, based on the observed set of GCP measurements, including an estimate of the root-mean-square (RMS) transformation error, and the apparent transformation error for each of the GCPs. The desired accuracy may not be achieved with the first acquisition of GCPs, because of a variety of possible errors in locating GCPs or recording their coordinates. The best way to proceed is to carefully check the five GCPs with the greatest error. Note that if the location of one point is miscalculated, points nearby will be influenced. Points that have been mislocated must be changed and the set of GCPs resubmitted to the program that estimates the transformation coefficients.

This empirically determined mathematical coordinate transformation will be used later in the procedure to rectify the Landsat-derived land cover data, as described in section 8 of this document.

7. LAND COVER CLASSIFICATION

The determination of land cover from Landsat data by a computer classification algorithm is the most critical step in the procedure. In this section, we present a brief discussion of pattern recognition and classification. We emphasize the basic concepts of automatic classification and the key issues of interest to the user of a classification algorithm.

A good overview of pattern recognition and classification concepts is available in [24], which also provides a detailed annotated bibliography. A good introduction to classification of multispectral remote sensing data is provided by Sabins [25]. We will follow his approach in our discussion of the conceptual framework of classification.

7.1. CLASSIFICATION CONCEPTS

Our objective in classification is to determine the land cover of each pixel in the Landsat data file of the study area. We use the spectral responses of the MSS sensors to the land cover to make this determination. At each pixel in the Landsat image, the radiometric value, which represents the reflectance of the sun's illumination, is recorded in each of four bands of the electromagnetic spectrum (see Appendix A for a description of the Landsat MSS system). A pixel is characterized by its spectral signature, determined by the reflectance values in the four bands. Multispectral classification analyzes the spectral signatures and assigns pixels to land cover categories based on the similarity of these spectral signatures.

To illustrate the concept of classification by spectral signatures, consider an application in which the categories of interest are: agricultural, forest, residential, and water. From regions of known land cover, spectral reflectance measurements can be averaged and plotted as spectral reflectance curves, as shown in figure 7.1. Note that there is a "typical" curve for each class, hence the term "spectral signature." This characteristic signature is the basis for classifying the image data. Because these signatures will show some variations for a given land cover type, it is useful to plot a scatter diagram (repeated measurements) of spectral reflectance values. The reflectance values are mathematical vectors with four numerical components in a "feature space," having one axis for each reflectance value measured by the multispectral sensor. A simplified scatter diagram for Landsat bands 4, 5, and 6 is shown in figure 7.2. Note that the samples belonging to each class are grouped and form a cluster

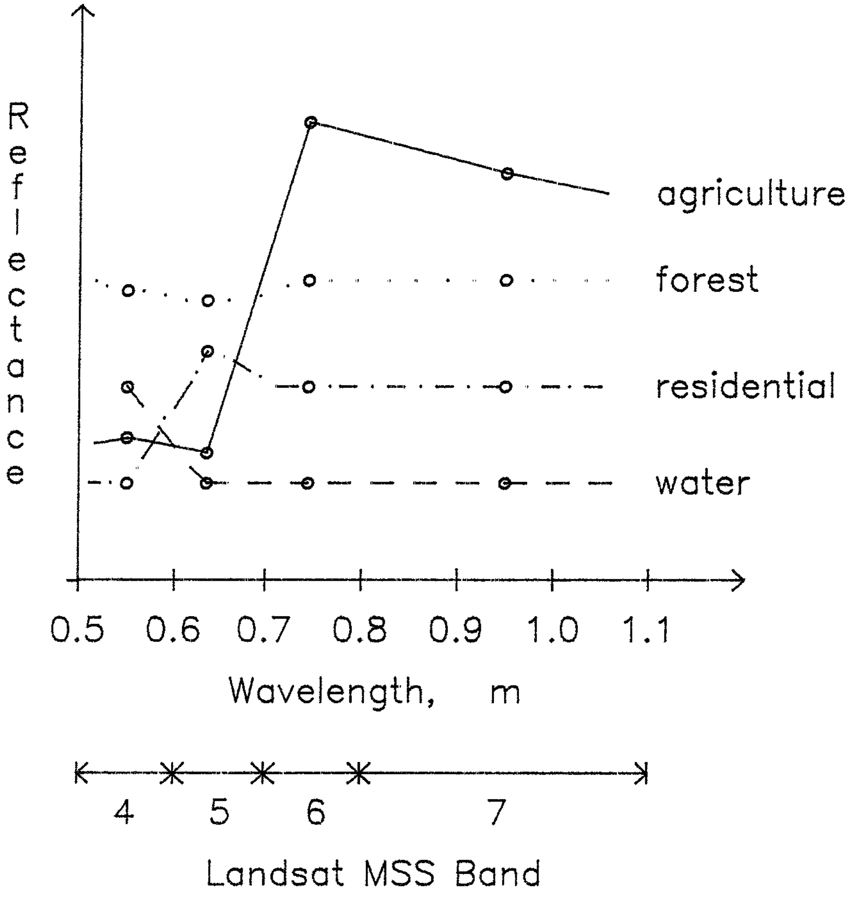


Figure 7.1: Spectral reflectance curves

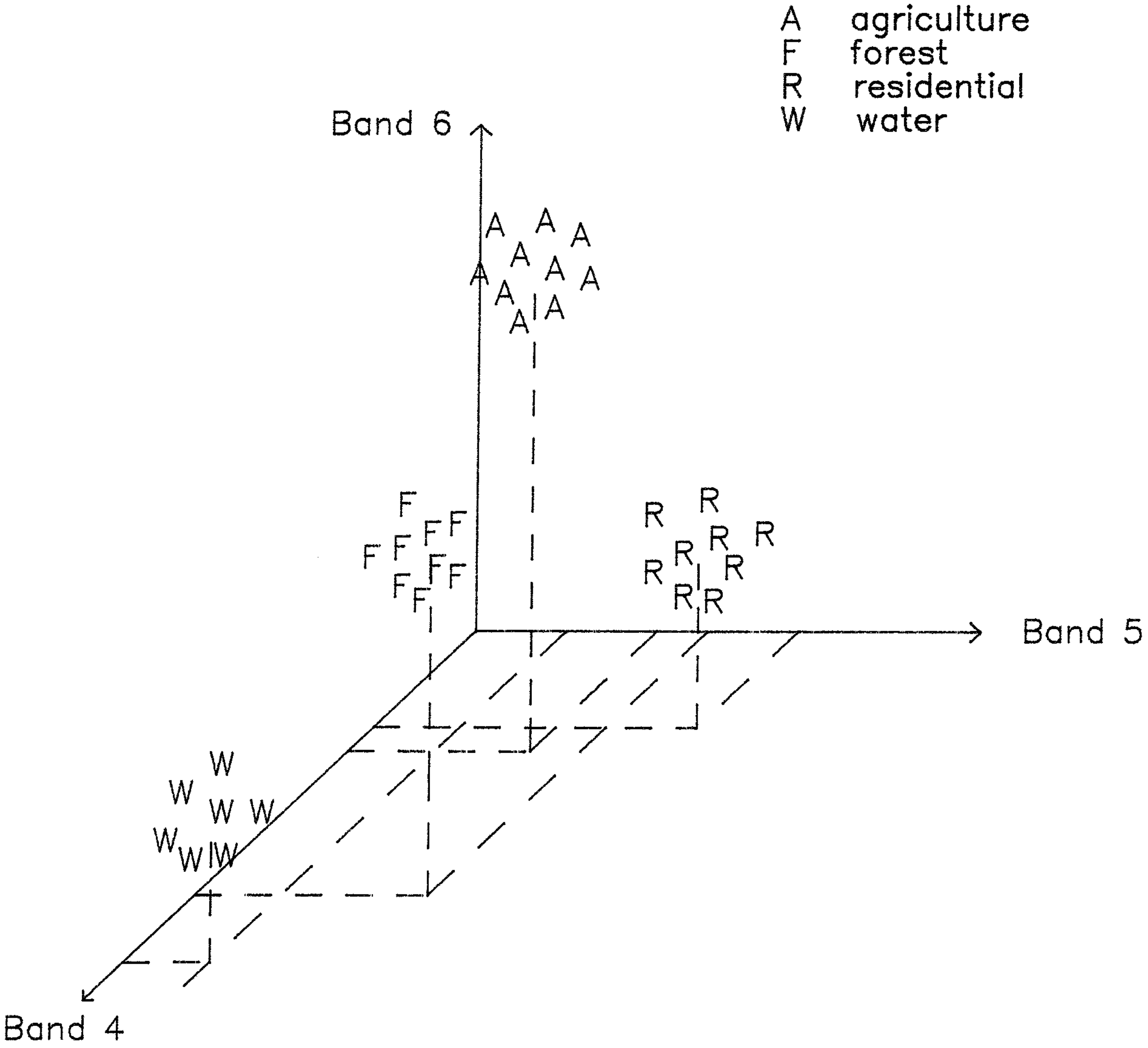


Figure 7.2: Three dimensional feature space scatter diagram

within this feature space. Classification is implemented by partitioning the feature space into "decision regions" assigned to land cover classes. A newly observed pixel will fall within a decision region and will be assigned to the corresponding class. There are conceptual and computational problems in the choice of decision regions and their use in classification. One commonly-used method is to evaluate the distance in feature space from a newly observed reflectance vector to the mean of the known samples collected previously for each class. The unknown sample is then assigned to the class with the nearest centroid in a procedure known as "nearest neighbor" classification. The computer classification of Landsat data actually operates in four-dimensional space using all four spectral bands, but this cannot be shown on the three-dimensional plot of figure 7.2.

There are two general approaches to classification, known as supervised and unsupervised classification. While each approach uses feature space partitioning, the methods used to define these partitions differs considerably.

In supervised classification, the locations on maps and aerial photographs of areas within the study area that typify each of the desired classification categories must be specified. These areas are known as training sets. The classification algorithm first computes sample statistics on the spectral reflectance values in each of the training sets, to characterize the spectral signature of each land cover class. A mathematical decision rule based on the relative locations in feature space of the spectral signatures is used to classify unknown samples. The most commonly used approach is based on Bayesian decision theory, under the assumption that the samples from each training set have a jointly Gaussian, or normal, probability distribution [24, 26]. Supervised classification is the most widely used of the two classification approaches, if dedicated image processing hardware is available.

In unsupervised classification, a computer algorithm uses only the statistical properties of the image data to determine a set of natural clusters or spatial groupings in feature space. The computer alone determines the partitioning of the data into clusters and no information on the land cover classes or eventual use of the clusters is used in this step. The user of this algorithm must then use training data, also known as ground truth, to determine the corresponding land cover class for each cluster. A commonly used algorithm in unsupervised classification is the clustering algorithm ISOCLS [27], which is used in the UCD Classification Procedure [2].

7.2. CHOICE OF LAND COVER CLASSES

An important issue in classification is the selection of a set of land cover classes to meet the needs of the geographic analysis project, given the constraints imposed by the data and the computer-based procedure. The set of land cover classes of interest to the Corps of Engineers has been discussed in the report of a pilot study by the Hydrologic Engineering Center on land use classification using large-scale aerial photographs and a grid based data structure [28]. Guidelines for the selection of categories included:

(1) The categories should be reasonably compatible with local and other agency land cover classification schemes.

(2) It must be reasonably possible to classify the land cover within the study area by conventional or automated means.

(3) The land cover categories should allow rational, consistent determination of flood hazard, economic, environmental, and other important effects of land use changes important to the planning study.

(4) The land cover categories should be compatible with those needed by certain available computer models.

The choice of categories is constrained by the multispectral scanner data, the classification algorithm, and the expertise of the operator in land cover identification. Recognizing these constraints, the U.S. Geological Survey [29], has recommended a standardized set of land use and land cover categories for use with remote sensor data. These categories were designed to meet the following criteria:

(1) The minimum level of interpretation accuracy in the identification of land use and land cover categories from remote sensor data should be at least 85 percent.

(2) The accuracy of interpretation for the several categories should be about equal.

(3) Repeatable or repetitive results should be obtainable from one interpreter to another and from one time of sensing to another.

(4) The classification system should be applicable over extensive areas.

(5) The categorization should permit vegetation and other types of land cover to be used as surrogates for activity.

(6) The classification system should be suitable for use with remote sensor data obtained at different times of the year.

(7) Effective use of subcategories that can be obtained from ground surveys or from the use of larger scale or enhanced remote sensor data should be possible.

(8) Aggregation of categories must be possible.

(9) Comparison with future land use data should be possible.

(10) Multiple uses of land should be recognized when possible.

A four-level classification system was specified, where Level I categories may be obtained using Landsat data, and Level II categories may require high-altitude photography having a scale of less than 1:80,000. Level I and II categories are shown in table 7.1. Note that the system has been designed such that Level II classes can be aggregrated to Level 1.

The choice of a subset of classes from this classification system depends on the objectives of a specific study. In applications of the UCD Classification Procedure [2], the objective has been the determination of land cover for use in water resource analysis, in particular for hydrologic modeling [12, 13]. The classes were chosen on the basis of relative imperviousness, which strongly affects runoff. In a study of the Walnut Creek watershed near Austin, Texas, the set of land cover categories included: industrial, commercial, low-density residential, medium-density residential, high-density residential, natural vegetation, developed open space, agricultural, pasture, and water. Note that these classes are primarily Level I classes, with the exception of the first five classes, which are the Level II subdivisions of the Level I urban or built-up land cover class. These subdivisions were important to the study due to their large variability in imperviousness.

For environmental and wildlife habitat studies, the objectives and thus the choice of classes are different. The U.S. Fish and Wildlife Service [30] has developed a habitat-based wildlife evaluation method entitled "Habitat Evaluation Procedure" (HEP). In the HEP method, a cover type classification system has been developed to assist in the development of habitat suitability models [31]. The HEP land cover classes included in the four-level classification system meet many of the criteria developed by the U.S. Geological Survey. The major difference is that the HEP system has much less detail in urban and built-up lands, which are generally not important wildlife habitats. The HEP system also has less

Table 7.1

Land Use and Land Cover Classification System

Level I		Level II	
1	Urban or Built-up Land	11	Residential
		12	Commercial and Services
		13	Industrial
		14	Transportation, Communications, and Utilities
		15	Industrial and Commercial Complexes
		16	Mixed Urban or Built-up Land
		17	Other Urban or Built-up Land
2	Agricultural Land	21	Cropland and Pasture
		22	Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas
		23	Confined Feeding Operations
		24	Other Agricultural Land
3	Rangeland	31	Herbaceous Rangeland
		32	Shrub and Brush Rangeland
		33	Mixed Rangeland
4	Forest Land	41	Deciduous Forest Land
		42	Evergreen Forest Land
		43	Mixed Forest Land
5	Water	51	Streams and Canals
		52	Lakes
		53	Reservoirs
		54	Bays and Estuaries
6	Wetland	61	Forested Wetland
		62	Nonforested Wetland
7	Barren Land	71	Dry Salt Flats
		72	Beaches
		73	Sandy Areas other than Beaches
		74	Bare Exposed Rock
		75	Strip Mines, Quarries, and Gravel Pits
		76	Transitional Areas
		77	Mixed Barren Land
8	Tundra	81	Shrub and Brush Tundra
		82	Herbaceous Tundra
		83	Bare Ground Tundra
		84	Wet Tundra
		85	Mixed Tundra
9	Perennial Snow or Ice	91	Perennial Snowfields
		92	Glaciers

detail in agricultural and barren land categories and it does not include tundra and snow or ice categories. However, the HEP system has more specific classes in the natural vegetation categories for both uplands and wetlands.

Further discussion of classification systems is given in [32].

7.3. USE OF A CLASSIFICATION PROCEDURE

Guidelines for the use of a classification procedure are clearly dependent on the particular procedure being used. However, some general guidelines can be given. These guidelines differ depending on whether a supervised or unsupervised classification procedure is used. In either case, good results require a user who is experienced in the interpretation of aerial photography and maps and who is knowledgeable about the land cover distribution in the study area. It is this person who defines the training sets in supervised classification and who relates the clusters generated by an unsupervised classifier to the chosen classes, using the ground truth data. These decisions will directly affect the decisions implemented by the classifier, and will thus determine the validity of the results.

Supervised Classification

Supervised classification procedures are usually iterative procedures implemented on interactive image processing systems. Sitting in front of an interactive image display system, the interpreter will use a trackball or joystick controlled cursor to define the locations of training sets for each of the selected classes. At least 100 samples are required for each class, and the set must contain representatives from all land cover types included in each class. The computer will compute some statistics for each of the training sets, define a decision function, and use this decision function to assign each pixel in the Landsat image to one of the selected classes. The results of this classification will be displayed as a graphics overlay on the Landsat image, with each class assigned a different color. The adequacy of this classification must be assessed visually, making comparisons with the ground truth data. If the accuracy is not adequate, the training sets must be changed. It is necessary to define sub-classes when the spectral signatures of some representatives of the selected classes are found to differ. For example, a forest class may contain coniferous and deciduous subclasses having substantially different signatures. Training sets can be defined for each of the sub-classes and after classification the two sub-classes can be combined in a single forest class. Another problem often encountered is that the training sets can be found to contain a mix of land cover types instead of the desired homogeneous cover. When this occurs, some training data must be discarded and other training sites selected. This iterative procedure of defining training sets, applying the classification procedure, assessing performance, and modifying training data is performed in a matter of minutes or hours, and is continued until the classification results are satisfactory. The end point in this iteration is when the user believes that the results are consistent with ground truth and are no longer improving. The best performance that can usually be obtained will give typically 80 to 90 percent agreement with classification results based on exhaustive study of high resolution aerial photography.

Unsupervised Classification

An unsupervised classification procedure usually requires considerably less interaction, and can be performed without an image display system, although such a display is preferred. A clustering algorithm is applied to the Landsat image, separating the data into clusters of feature vectors close to each other. An image of the clusters assigned to pixels is then printed on a "cluster map," with several clusters printed per map. On each cluster map, a unique label (printer symbol) is printed for each pixel assigned to a particular cluster. Blanks are used for pixels assigned to other clusters to reduce the clutter and ease the interpretive process. If a display system is available, the clustering results could be presented as a color graphics overlay. Land use classes are then assigned to clusters by comparing contiguous groups of pixels from each cluster (spatial group) with information from maps and aerial photographs. By referring to the aerial photographs and the topographic maps, the appropriate land cover class is determined for each spatial group. If all spatial groups of a given cluster have the same class assignment, the cluster is assigned to a definite land cover category. A cluster having multiple class assignments is considered to be in conflict. The new partitioning into clusters is done for those pixels whose assignment was in conflict, to further subdivide these clusters and allow unequivocal class assignments. Again using cluster maps, the class assignments of these new clusters are determined. Further details of the steps involved in an unsupervised procedure using ISOCLS are given in [2].

Land Cover Classification File

The final product of any classification procedure is a land cover data file. This is a file in the Landsat coordinate system in which each pixel contains a label representing the appropriate class assignment. This label is the numerical value assigned to each land cover class. The numerical assignment of classes is arbitrary, however, it is common practice to use consecutive integers from 1 through N, the total number of classes.

Before this classified file can be merged into a spatial database, it must be transformed to the geometry of the database, and a final accuracy assessment must be made, as described in the next two sections.

8. GEOMETRIC RESAMPLING

The Landsat-derived land cover file must be transformed from its native Landsat coordinate system to the spatial database coordinate system, using a process known as resampling. This process uses the mathematical coordinate transformation from the database coordinates to the Landsat coordinates that was defined in section 6.

While the coordinate transformation was determined from a set of ground control points, it can be applied to any point in the spatial database coordinate system. By successively applying it to each grid cell in this coordinate system, the land use map can be transformed to database coordinates. The grid cell matrix must first be defined in the database coordinate system, by defining the database coordinates of the origin of the grid and the dimensions of the grid cell in units of the database coordinates, as described in section 6. The coordinate transformation is applied to each grid cell location to determine the corresponding location in the Landsat coordinate system. In general, this location will fall between pixels in the Landsat-derived land use file. Some form of interpolation is required to determine the appropriate grid cell value. Since a land use file contains a set of land use labels, the only interpolation method that is appropriate is the nearest neighbor method. As suggested by the name, the grid cell is assigned the land use label of the pixel nearest the transformed location. This process of resampling the land use file is repeated for all grid cell locations, creating the land use file in the spatial database coordinate system.

The process of geometric resampling is illustrated in figure 8.1. The coordinate transformation is applied to the location of the centroid of a grid cell to determine the corresponding transformed location in the Landsat coordinate system. The nearest Landsat pixel is found, and the land cover label assigned to this pixel is then assigned to the grid cell under consideration.

This description of the process of image resampling provides some insight into the effect of the choice of different sized grids in the grid cell database. The finest effective resolution in the grid based coordinate system is that which is roughly the same as the Landsat resolution. If the grid cell size is chosen to be smaller than the Landsat IFOV, then adjacent grid cells would transform to locations in the Landsat coordinate system having the same nearest neighbor. Thus, several adjacent grid cells will be assigned to the classification label from the same Landsat pixel.

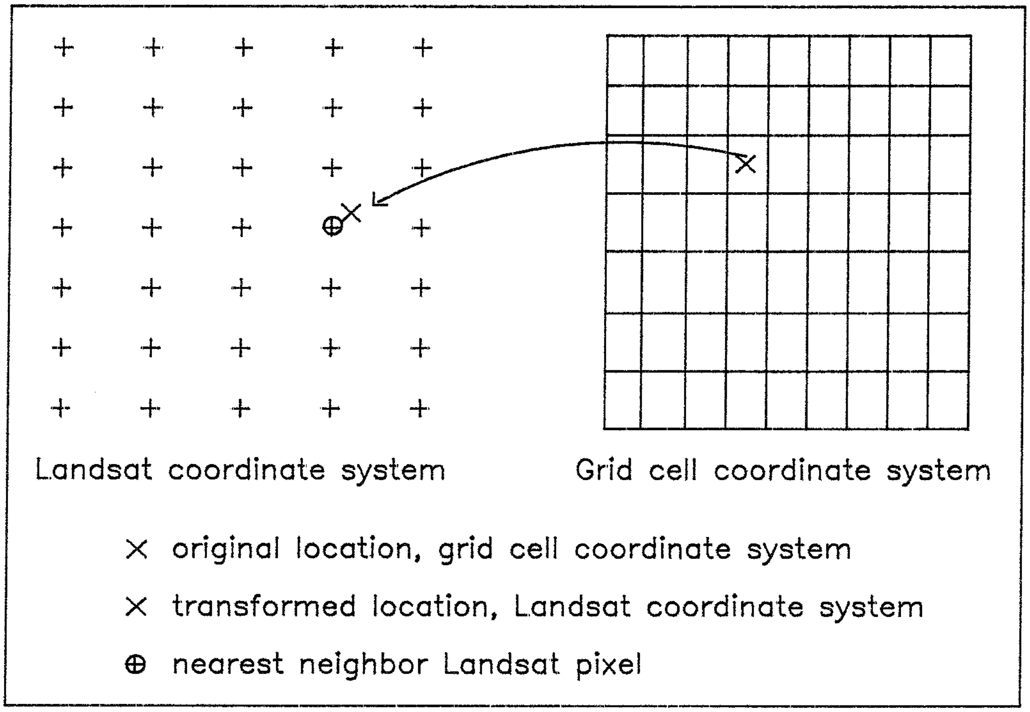


Figure 8.1 Geometric resampling

Conversely, if the grid cell size is chosen to be much larger than the Landsat IFOV, then the land cover data will be strongly undersampled in the resampling step. The effects of undersampling in the Landsat coordinate system are shown in figure 8.2. The open circles show the regular grid of Landsat pixels and the "X's" show the locations of the transformed grid cell centroids. The closed circles are the Landsat pixels that are the nearest neighbors of the transformed grid cells. Only the land cover labels of these nearest neighbors will be transferred to the grid cell data file. Note that in this example, few of the land cover assignments in the Landsat coordinate system will be included in the grid cell file. In planning studies where isolated land cover types are critical, this undersampling could cause serious problems.

In studies of large regions, there is a need to choose grid cell sizes larger than 1 to 1.5 acres to reduce the size of the grid cell files. Thus, there is a need to develop methods to convert grid cell files from one resolution or cell size to another without undersampling the

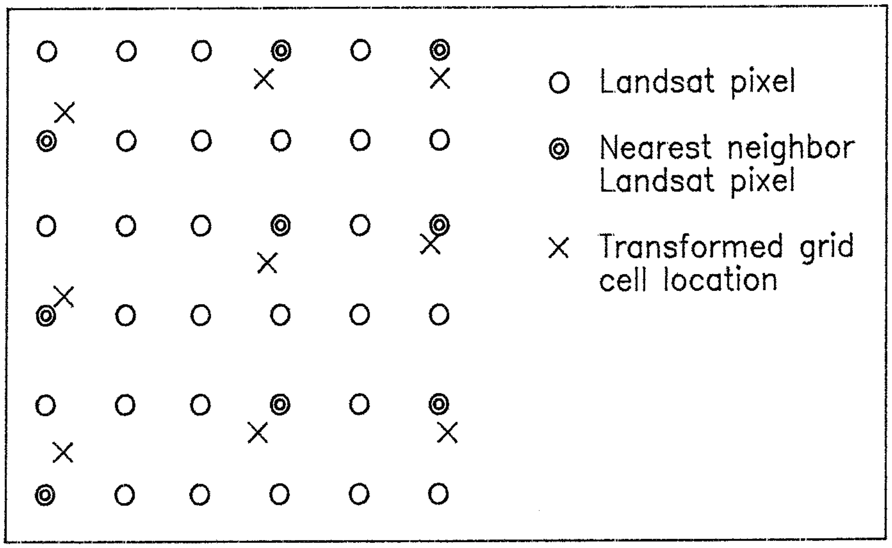


Figure 8.2. Effects of undersampling

data. We will describe one approach to this problem, although we believe that further development of these methods is needed.

When the desired grid size is larger than the Landsat IFOV, the geometric resampling should be done in two steps. The geometry is corrected in the first step and the land cover results are aggregated to the large grid cells in the second step. In the first step, the grid cell size should be chosen to be similar to that of the Landsat pixel spacing: 57 meters (187 feet) between scan lines and 79 meters (259 feet) between pixels on a scan line. In this resampling step, the spacing of the transformed grid cell centroids will be about the same as that of the Landsat pixels. In the second step, the grid cell dimensions should be integer multiples of the previous grid cell dimensions. For example, a grid cell that is 4 times the pixel spacing along scan (316 meters or 1036 feet) and 6 times the pixel spacing in the scan direction (342 meters or 1122 feet) has an area of 26.5 acres. An encoding rule must then be defined to determine the land cover assignment for the large cell from the assignments in the 24 (4 by 6) smaller cells. The choice of a particular rule involves a trade-off between the importance of a particular cover type and the area it occupies. Three possible encoding rules are:

(1) Predominant Type. The large cell is assigned to the class label that occupies a majority of the 24 small cells. Thus, a large cell containing 14 small cell in

agriculture and 10 cells in residential would be assigned to agriculture.

(2) Weighted Predominant Type. The cover types can be assigned numerical weights according to their importance in the planning study. Then the large cell is assigned to the class having the largest summed weight. Thus, if agriculture is assigned a weight of 0.5 and residential is assigned a weight of 1.0, a large cell having 14 small cells in agriculture (summed weight 7.0) and 10 cells in residential (summed weight 10.0) is assigned to residential.

(3) Most Important Type. The large cell is assigned to the cover type that is considered most important to the study. For example, 2 small cells could be natural riparian and the remaining 22 could be agriculture, but in a habitat study it could be appropriate to assign the large cell to natural riparian.

Each of these encoding rules also requires a scheme to resolve ambiguities. If more than one class has the same frequency, summed weight, or importance, then a secondary rule must be applied. In a study of the Saginaw River Basin conducted by the U.S. Army Corps of Engineers [33], a hierarchical assignment scheme was used. In this scheme, the assignments of neighboring grid cells were used to resolve ambiguities. In the study, the 1.1-acre Landsat land cover classification database was converted to 40-acre grid cells (six-by-six blocks of Landsat pixels), using the predominant type encoding rule and the hierarchical scheme for resolving ambiguities.

9. ACCURACY ASSESSMENT

Before creating or using the grid cell data file containing the Landsat-derived land cover information, assessments of the accuracy of the geometric transformation and the classification must be made. Both assessments can be made by generating a printer or graphics representation of the final land cover classification map in the grid cell coordinate system and overlaying this output on the topographic map or maps of the study area. A visual analysis can then be made of the geometric registration, and, with the use of samples of ground truth data, an estimate can be made of the classification accuracy.

9.1. GEOMETRIC ACCURACY

The accuracy of the geometric transformation is assessed by visually checking the alignment of classified features with the map. This is difficult to do, as some readily locatable features, such as roads, are often too small with respect to the resolution of Landsat to be reflected in the classification results. However, there are two types of features that can be used to check the alignment. The first type includes the boundaries of large homogeneous regions, such as agricultural fields. The second type includes water bodies, both lakes and rivers. These are usually classified accurately, so they can be readily located in the classified results. These features should align to within one graphics character if the geometric correction is as accurate as is desired. This should be true over the entire study area. If the geometric accuracy is not satisfactory, it will be necessary to modify the set of ground control points, recalculate the coordinate transformation, and then use this new transformation to resample the land cover file from the Landsat coordinate system to the spatial database coordinate system.

If the UCD Classification Procedure has been used, then the image file of enhanced curvilinear features that was used in computing the coordinate transformation can be transformed and used to check the geometric registration. Since this image is dominated by features such as roads and water bodies, it is ideal for use in evaluating the registration.

Another method of estimating the magnitude of the residual geometric misregistration is to compute a histogram of minimum distances between grid cells of a chosen class in the Landsat land cover file and a manually classified land cover file. For example, a histogram of distances from Landsat residential to actual residential can be computed [34]. If the geometric registration is adequate, a large fraction of the Landsat residential grid cells should be within one or two grid cells of an actual residential area. The Landsat residential cells located beyond this distance can be assumed to be errors in classification. This histogram can be computed using the Distance Determination Package of the Resource Information and Analysis (RIA) program [4]. This program processes SAM grid cell data files, so it is necessary to create the Landsat land cover grid cell file, as described in section 10 of this document, before using the program. Obviously, this method also requires that a manually classified land cover file be prepared.

9.2. CLASSIFICATION ACCURACY

If the geometric registration is acceptable, then the printer or graphics representation of the classification results in the grid cell coordinate system can also be used to assess the classification accuracy. This can first be done qualitatively by visually checking to determine if major homogeneous regions such as water bodies or large agricultural fields are classified properly. If this cursory check shows that the classification results are generally correct, then a quantitative assessment can be made.

The classification accuracy is assessed quantitatively by closely checking a random sample of grid cells for which the land cover classification is known with confidence. By locating the selected grid cells on the aerial photographs, the proper classification of the grid cell can usually be determined. The training samples that were used in the design of the classifier can be used in this accuracy assessment. An estimate of the probability of error of the classifier is obtained by dividing the number of training samples erroneously classified by the total number of points.

The probability of error estimated from the training samples is accurate only if the training samples are truly representative of the overall data set to be classified. In practice, the training sample classification results are usually optimistic; the error rate estimated from the training samples is lower than the true error rate for the entire data set.

A more accurate method of estimating the classification accuracy is to use a set of test samples that are independent of the training samples. As in the case of the training samples, it is important that the test data be representative of the overall data set. Large visually homogeneous areas are often used in the test set, as they are most often likely to represent single classes, as opposed to mixtures of classes. While the larger the test set, the better the assessment of accuracy, there are practical constraints on how much test data can be made available. The most important constraint is the cost and logistics of collecting and manipulating detailed and thus voluminous test sets.

In addition to an estimate of the probability of classification error, a concident or confusion matrix can be used to assess the classification performance. A concident matrix is a cross tabulation table in which each element contains the number of grid cells in the test that have the concurrent Landsat and ground truth land cover assignment. Elements down the diagonal of this matrix represent the grid cells classified correctly and off-diagonal elements are incorrectly classified. For most studies, 80 percent or more of the grid cells should be classified correctly, and each class containing a significant percentage of the cells in the study area should also have a similar accuracy. Large off-diagonal elements indicate a significant degree of confusion between the corresponding pairs of classes. If the probability of this confusion is large, it may be necessary to return to the classification step and work further with the classification of these classes.

A coincident matrix can be computed using the Coincident Tabulation Package of the RIA program [4]. As described in the previous section, this program processes SAM grid cell data files, so the Landsat land cover grid cell file must be created before using the program.

More elaborate methods for quantitative assessment of classification accuracy are described in [32].

10. CREATING THE GRID CELL DATA FILE

When the accuracy of the geometric transformation and the land cover classification have been judged to be adequate, the grid cell land cover data file is ready to be merged into the SAM database. To do this, the land cover file that was geometrically resampled to the coordinate system of the spatial database (as described in section 8) must be reformatted to be compatible with the SAM database.

A reformatting program is used to store the grid cell land cover data file on a computer tape or disk file in a systematic fashion so that it can be read into the SAM database. This file is a single-variable file in which each record contains the land cover labels for a row, or a portion of a row, of the land cover data file. The single-variable file has the following specifications:

(1) The file is sequential, containing ASCII character representations of the numeric land cover labels, with no formatting characters. The file does not contain the row or column numbers.

(2) Each record (line) of the file is restricted to 80 characters or less.

(3) The first record (line) contains the number of data elements per line and the number of lines in the file, written in Fortran 2I5 format.

(4) Two ASCII characters are used to represent each land cover label. Therefore, only 40 grid cell land cover labels can be written on each record of the file. If a data file contains more than 40 columns, then more than one record must be written for each row of the input file. For example, if the land cover file contains 100 columns, then three records are written for each row. The first record contains the land cover labels for columns 1-40; the second record contains the labels for columns 41-80; and the third record contains the labels for columns 81-100. This data is written in Fortran 40I2 format.

(5) The land cover labels are assumed to be integers in the range 1 to 99. If a label in the file is outside this range, it is written as -1.

A program is available in the UCD Classification Procedure to create this single-variable file. The file can be merged into an existing SAM database using the single-variable package of the SAM program BANK [5].

11. DISCUSSION OF KEY ISSUES

In this section, we discuss the key issues involved in successfully applying the procedure described in sections 4 through 10 of this document. We first discuss four key technical issues in applying the procedure, and then we discuss issues involved when the procedure is executed by Corps of Engineers personnel using software and computers available to the Corps, or when it is performed by an outside agency or company under contract to the Corps.

11.1. TECHNICAL ISSUES

Regardless of the specific procedure used to create a grid cell data file from analysis of Landsat data, or who performs the procedure, there are four key technical issues that must be handled properly for the resulting file to be of value in a geographic analysis project. While these issues have all been discussed in previous sections of this tutorial, it is of use to briefly discuss them again, to highlight their importance.

1. Choice of Land Cover Classes. As discussed in section 7.2, an important issue in land cover classification is the selection of a set of land cover categories to be used in the computer-based procedure. This choice is a often a compromise between the needs of the geographic analysis project and the limitations imposed by the Landsat MSS data, the classification algorithm, and the expertise of the remote sensing analyst. If too few classes are defined, or if these classes are too broad, then it will not be possible to perform a detailed geographic analysis. For example, if all urban land uses were to be lumped into a single category, it would not be possible to use this land cover data to perform an analysis of the economic study of potential flood damage. On the other hand, if too many classes are defined, or if these classes are too narrow, it will not be possible to classify the data into these classes with adequate accuracy. Refer to the guidance given in section 7.2 in selecting the set of land cover classes.

2. Definition of Ground Truth and Training Areas. As discussed in section 7.3, the accuracy of the land cover classification is dependent on the adequacy of the ground truth information available for the study area. This is true both for supervised classification, where the ground truth is used to define classification training areas, and for unsupervised classification, where the ground truth is used to assign clusters to land cover classes. This ground truth consists of aerial photographs and topographic maps. It is important that the photographs and maps be accurate representations of the study area for the same date as the Landsat exposure date or classification errors will be made in rapidly changing areas. In addition to this data, it is important that a person who has experience in remote sensing interpretation and has knowledge of the land cover patterns in the study area be actively involved in the classification procedure. The decisions of this remote sensing analyst in interpreting the ground truth and relating it to the intermediate results of classification are critical determining factors in the final classification accuracy.

3. Geometric Rectification. For the Landsat-derived land cover file to be of value as a source of data in a spatial data management system for use in geographic analysis, it must be possible to accurately rectify it to the geometry chosen for the spatial database. This means that each pixel in the portion of the land cover file representing the study area must register with the corresponding grid cell of the spatial data base to within about one half of the dimensions of the grid cell. This accuracy should be checked when the geometric coordinate transformation is computed, as described in section 6.2, and again in a final accuracy assessment, as described in section 9.

4. Classification Accuracy. The objective of the procedure described in this tutorial is the creation of a land cover data file, and thus, the most critical issue is the accuracy of this file. There are limitations to this accuracy, imposed by limitations of the data sources and the procedure. Landsat digital data has geometric and radiometric inaccuracies as well as limitations in spatial and spectral resolution. Due to the limitation in spectral resolution, not all representatives of a given class will be spectrally separable. The choice of land cover classes, the quality of the ground truth data, and the expertise of the remote sensing analyst, as described above, will affect the classification accuracy. Furthermore, the remote sensing data inherently represents land cover, which differs in some cases from land use. For example, a residential water reservoir would be classified from Landsat as water, while a conventional classification may assign it to public utilities. However, if a land cover classification procedure is properly applied to Landsat data, the resulting land cover should have 80 to 90 percent agreement with the results of a conventional classification of land use. For some studies, such as wildlife habitat, the land cover acquired from Landsat data should actually be more appropriate to further analysis.

11.2. PERFORMING THE PROCEDURE IN-HOUSE

Aside from the technical issues, there are a number of resource availability and allocation issues that must be addressed before any land cover classification project is undertaken. These issues involve personnel expertise and training, the cost and time required to apply the procedure, and the availability of computer resources, both hardware and software. These issues must be weighed in deciding whether to perform a procedure in-house or by contract. We consider these issues for a procedure performed in-house and then consider the case of a procedure performed by contract.

The personnel involved in the project must have the proper expertise and training. The key person is one who has experience in analysis and interpretation of aerial photography and maps. When the procedure is performed in-house, this person should have the primary responsibility for the performance of the procedure. Thus, this person must be given adequate time and must have the proper aptitude and interest to learn and apply the computer-based procedure proficiently.

The second resource issue involves the cost of the procedure. These costs include data, computer time, supplies, and labor for an in-house procedure. The costs involved in the application of the UCD Classification Procedure to two separate study areas are described in [12]. Five weeks of engineer/technician labor were required for each study. For small study areas, conventional methods of land cover classification using only maps and aerial photographs can be shown to be more cost effective. Procedures based on the use of Landsat data become more cost effective when the size of the study area exceeds about 10 square miles.

The third issue involves the availability of computer resources. The computer hardware and software on which the procedure is based must be available to the person responsible for the project. Specialized peripherals such as graphics terminals, graphics tablets, or image display systems must be available as required by the procedure. Adequate support for the use of the computer-based procedure must also be available. This includes computer support personnel, documentation and users manuals for the computer system and the procedure, and consultation with the organization that developed the procedure.

Use of the UCD Procedure

The resource issues can be defined more explicitly if the UCD Classification Procedure is used [2]. This procedure was designed such that it does not require specialized expertise in data analysis, computer programming, or image processing. The procedure has been implemented on a Harris 500 minicomputer, and the required peripherals include only a standard computer terminal and a line printer. The costs cited in [12] are for an earlier version of the procedure that was implemented on a Control Data Corporation computer, but the costs associated with the current version of the procedure should be closely related, except for the computer usage charges if the computer is owned by the user.

If procedures other than the UCD Classification Procedure are used, the resource issues must be analyzed in further detail. It will be necessary to determine the cost of acquiring the required computer hardware and implementing and verifying the software-based procedure. Availability and cost of training, documentation, and consultation on the procedure must be investigated. Finally, the labor costs associated with learning and applying the procedure must be estimated.

11.3. PERFORMING THE PROCEDURE BY CONTRACT

Commercial assistance in classification of Landsat data may be appropriate in the following cases:

(1) Personnel with the needed expertise in aerial photography and map analysis and interpretation may not be available to work on the project.

(2) The services of a company with experience in computer-analysis of remote sensing data can be valuable when using new technology.

(3) Hiring a firm to analyze the Landsat data can reduce the computer hardware requirements of the project.

It is possible for Corps personnel to manage the project while having no direct involvement other than writing the contract and receiving the final data product, but this is not recommended. Instead, it is recommended that Corps personnel be involved in every step of the procedure, as it is likely that this approach will lead to a better final product. Specifically, Corps personnel should acquire the data as described in section 4 of this tutorial, assess the geometric accuracy as described in sections 6.2 and 9, select the set of land cover categories as described in section 7.2, be actively involved in the classification process by interpreting ground truth and defining training areas as described in section 7.3, and assess the accuracy of the classification results as described in section 9.

When preparing a contract or a request for proposal, the following items should be specified:

(1) The size and location of the area to be studied.

(2) The geometry and coordinate system of the spatial database.

(3) The requirement for the accuracy of geometric rectification.

(4) The set of land cover classes, with an expanded description of the elements of each class.

(5) The requirement for the accuracy of classification.

(6) The specifications of the final data file to be delivered (generally magnetic tape readable by the analysis system used by the study team).

(7) The specifications of required graphical output products, such as a land cover map printed with a line printer.

(8) The contract time allowed for completion of the project.

If possible, the study manager should visit potential contractors and talk with them. It is also instructive to interview previous clients of the contractor. Of particular importance is the contractors experience in performing a Landsat land cover classification study with similar requirements for geometric and classification accuracy.

12. EXAMPLES OF MAJOR STEPS

In this section, we highlight some details involved in the actual process of using Landsat data for land cover classification. Examples are taken from each step of the classification procedure, as explained in sections 4 through 10.

12.1. DATA ACQUISITION

As a first step in data classification the user should obtain the necessary data sources, including topographic maps, high flight aerial photography, and the Landsat CCT, as described in section 4. The following example describes the computer listings sent by EROS on available Landsat data.

After requesting information about the availability of Landsat scenes from a particular point or area, the user will receive a computer listing. These listings from the EROS Data Center are grouped by orbital path and row, and include both MSS, RBV, and TM data. The listings include many details about each Landsat scene entry, including items A through I, found in Table 12.1.

Table 12.1 Landsat CCT entry categories

A imagery-type: Landsat satellite number and type of sensor used
B scene identification: the number used to refer to a Landsat scene
C film source and film quality
D %cloud cover
E exposure date
F scene center point and scale
G microform identification number
H the status of the false color composite image, the MSS sensor gain when the image was recorded, the data acquisition mode, and the availability of the CCT
I corner point coordinates: the corner points of the Landsat scene in longitude,latitude units

The same A through I identifiers are used in Table 12.2.

Table 12.2 Landsat CCT entry

path=28, row=39
A Landsat-2 (MSS)
B 82253616170X0
C B&W-06.7" 8888***
D 10%
E 1/01/82
F N30D32M00S W096D12M00S 1:100,000
G B7900930294
H P L C E
I #1:N30D45M20S W095D30M34S #2:N31D03M06S W097D21M44S
#3:N30D11M17S W096D59M42S #4:N29D18M54S W095D57M03S

In this example, the Landsat scene was taken by the Landsat-2 satellite using the multispectral scanner sensors. The orbital path and row, (28,39), describes the approximate location of the scene, based on standard operating altitude of Landsat-2. When ordering this scene, the identification number, 82253616170X0, would be used. The film type and size for the photographic products associated with this Landsat scene is 06.7 inch black and white film with the quality of each band being very high (8, on a scale of 0 - 9). The asterisks following 8888 mean no data was available for bands other than 4, 5, 6, and 7. This quality rating is subjective and does not necessarily indicate image usefulness for our purposes. The percent of the image obscured by clouds and their shadows is 10%, a generally acceptable level. The image was taken January 1, 1982, and the geographic center point of the image is North 30 degrees 32 minutes 00 seconds and West 96 degrees 12 minutes 00 seconds. The scale of the master reproducible film on file is 1:100,000. The microform identification number may refer to either a microfilm or microfiche number. In this entry, the first digit, B, indicates that the microfilm copy is available, and the cassette and frame number of this copy is 7900930294. If the first digit is D, the next ten digits describe the zone, path, year, month, and day of the microfiche copy. The microform data is maintained by NCIC facilities throughout the country. The letter, P, indicates that a false color composite image of this scene has not been processed, though it could still be requested. The letter, L, indicates that all MSS bands are in low gain mode, and all sensor response voltages up to a preset saturation limit are digitized. The letter, C, indicates that a compression mode was used to acquire MSS bands 4, 5, and 6. The letter, E, indicates that the computer-compatible tape is in EDIPS high density format and can be requested from the

Eros Data Center. The corner point coordinates are in latitude-longitude units, and are subject to a 10 percent error.

12.2. LANDSAT DATA FILE CREATION

After receiving the Landsat CCT and band 5 photograph, a Landsat data file of the study area must be identified and copied. Recall from section 5 that the study area generally represents a small portion of the scene. First, the study area is identified on a 1:250,000 topographic map. Then, this area is located on the band 5 photograph (scale 1:250,000). The location of the region of interest on the photographic print may be related directly to the location of the region of interest on the Landsat CCT. The details of this process as well as a figure (12.1) are included below. First, a sheet of tracing paper is placed over the photograph.

identify the boundaries of the photograph: As shown in Figure 12.1, draw two vertical lines, parallel to the edges of the photograph and touching the upper right and lower left corners of the photograph. The horizontal edges of the photograph connect to the two vertical lines, outlining the photograph measurement boundaries.

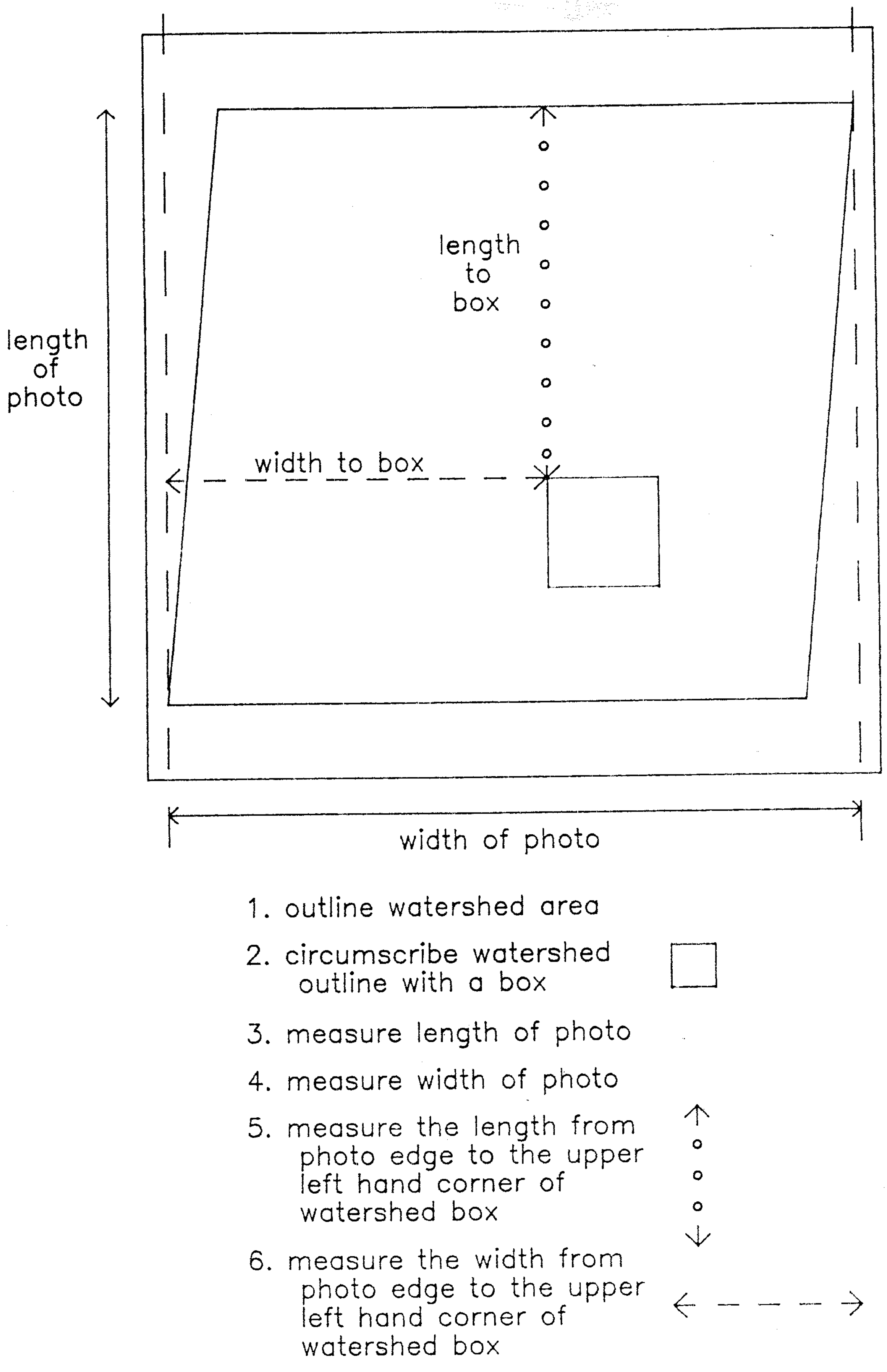
identify the boundaries of the watershed: Draw a rectangle around the region of interest, where the sides of the rectangle are parallel to the traced boundary of the photograph.

measure the lengths in cm: Using the traced photograph and region of interest boundaries, make the following measurements.

length of the photograph
width of the photograph
distance from the photograph to to the ROI top
distance from the left vertical line of the photograph to the left vertical line of the ROI
width of the ROI
length of the ROI

The first record on the Landsat CCT will identify the data type of the CCT. The program, PRETAP, therefore, can verify the format type. PRETAP also uses the photograph measurements to determine the location of the region of interest in Landsat units. The tape can now be read by either the RDBIL or the RDBIP2 program, depending on the data format.

After creating the study area files, the radiometric quality of the data can be determined. Radiometric correction was discussed in section 5.2; an example of radiometric detection and correction follows.



1. outline watershed area
2. circumscribe watershed outline with a box
3. measure length of photo
4. measure width of photo
5. measure the length from photo edge to the upper left hand corner of watershed box
6. measure the width from photo edge to the upper left hand corner of watershed box

Figure 12.1 Band 5 Photograph Measurements

In the UCD Classification Procedure, the output from the computer program DETECT lists the means and variances of the six sensors from bands 4, 5, 6 and 7. DETECT also evaluates whether any sensor may need to be adjusted. Within a given band, a sensor statistic that differs widely from the other sensor statistics may be in error. Specifically, a difference between means of the sensors from any given band greater than 15% of the largest standard deviation will generate errors in data classification. The program, CORECT, will equalize the sensors, as needed.

In the Table 12.3, band 4 has a bad sensor, sensor 2. Its variance is much larger than the other variances in band 4. Bands 5, 6, and 7 are within an acceptable range. Table 12.4 shows the calculations performed on band 4, evaluating sensor 2. Reference sensor 3, band 4, was selected as the reference sensor for use during the correction of sensor 2, since its mean and variance are close to the average mean and variance for the band. Using the program, CORECT, therefore, sensor 2 will be corrected, relying on reference sensor, sensor 3.

Table 12.3. Radiometric Detection Program Results

BAND	SENSOR	MEAN	VARIANCE
4	1	17.73	5.62
	2	18.24	15.48
	3	18.24	7.39
	4	18.96	5.62
	5	18.17	6.39
	6	17.55	8.26
5	1	28.85	129.99
	2	28.56	124.98
	3	28.94	125.97
	4	28.77	125.30
	5	29.23	129.73
	6	28.91	128.70
6	1	55.14	109.23
	2	55.50	101.92
	3	55.44	96.03
	4	55.93	107.48
	5	54.71	102.23
	6	54.70	113.08
7	1	26.40	31.84
	2	27.00	30.84
	3	26.46	31.48
	4	26.62	32.70
	5	27.08	35.40
	6	26.92	34.19

Table 12.4 Radiometric Detection Evaluation

For band 4,
maximum mean: 18.96
minimum mean: 17.55
difference: 1.41
standard deviation of sensor 2: 3.93
1.41/3.93*100 = 35.8%
35.8% is greater than 15%
reference sensor: 3

12.3. DATABASE GEOMETRY

After generating a Landsat data file and verifying its radiometric quality, the geometry of the data file needs to be addressed. Database geometry was introduced in section 6. In the following two examples, the focus will be on the generation of a set of coefficients for later data resampling. In the UCD Classification Procedure, the initial goal is to be able to resample the Landsat data so that it will overlay a 1:24000 USGS topographic map, printing the data at the line printer of interest. Since an image display system is not used during this procedure, an image of the potential identifiable curvilinear features, such as roads, as well as the water bodies is generated. The program, ROADS, can generate this image using data from bands 5 and 7. The program, DMAP, prints the resulting file. Figure 12.2 is an example of output from the programs, ROADS and DMAP. In Figure 12.3, the roads were identified, and drawn in. Ground Control Points (GCP's) were marked (shown as X's) and labeled on this output. The corresponding points were identified and labeled on the topographic map.

After a set of 25 to 35 GCP's have been selected, the computer program, GEOCOR, is used to generate a set of coefficients for subsequent use in resampling the study area. Points are selected from the topographic maps in UTM coordinates and from the Landsat data file in pixel coordinates.

GEOCOR was used to generate the transformation coefficients; its output is shown in Table 12.5. This detailed output is used to evaluate the accuracy of the transformation of the GCP's. The column, MAP GCP's, lists the UTM coordinates; the column, LANDSAT MAP GCPs, lists the corresponding Landsat coordinates. The MAP GCPs are transformed to pixel coordinates in the third column; the fourth column lists the error associated with the transformation. The following set of points was used to correct the geometry of the Walnut Creek watershed, Texas.

```
      120.00    130.00    140.00    150.00    160.00    170.00
         |         |         |         |         |         |
50.00    @@@@       @   @@    @        @@          @ @@  @@       @
51.00       @@@         @@        @                  @ @ @@@@ @@@@@ @
52.00            @@@@   @@@   @@                              @@  @@
53.00    @      @@       @   @         @                     @   @
54.00    @  @             @       @      @                       @
55.00    @                 @      @@      @@                      @
56.00        @   @@          @@@@         @@           @@@
57.00     @  @    @@@@           @@                                 @
58.00    @@@           @@            @@                   @
59.00            @@                 @  @@@
60.00            @@@@              @    @@@
61.00      @@@@         @           @    @@@                  @@@
62.00        @   @       @         @ @    @@           @     @@ @@@@
63.00            @       @        @  @                @            @@
64.00    @@@@     @      @     @     @@@@   @                       @
65.00    @@@@@@@@@       @            @@ @@  @         @@   @   @
66.00       @ @   @     @@   @   @   @       @          @@@@@   @@
67.00       @          @@    @@                         @     @ @@@@@
68.00                 @@        @      @     @@                @@
69.00                 @@          @@    @         @ @        @
70.00       @   @@@    @            @     @        @ @      @  @@@@@
71.00                 @              @   @   @@@@          @    @@@@@
72.00      @         @@          @@     @       @       @@@  @     @@
73.00       @        @              @   @@  @@  @@    @@@    @    @@@
74.00            @   @                 @@  @@   @@  @@      @@ @@ @
75.00       @       @@           @           @ @@@@@@       @
76.00    @          @@        @          @@   @@    @@@          @ @@
77.00      @@@      @      @           @@@@@@    @   @@   @@   @@@@
78.00                     @        @@@@@              @@   @@@
79.00           @@    @@          @           @@       @@         @@@
80.00    @  @@@   @   @      @@@@@@       @  @     @@@  @
81.00    @   @@          @@@@          @@   @@@@@@@@     @        @
82.00             @   @@ @              @   @@@  @@         @     @@
83.00      @      @  @@                  @@          @  @   @@@    @@
84.00      @@@   @         @      @   @@       @@    @        @    @@
85.00      @                      @@@      @@@  @@@            @    @
86.00    @      @ @            @@  @     @   @@@@@@            @@@
87.00    @          @      @@@@                                  @@
88.00               @@  @@          @@                   @@@@    @@@@
89.00             @    @       @@    @     @ @      @@@@@   @@  @@@
90.00      @@  @             @@   @   @          @@
91.00                   @             @   @   @@@@@@@         @   @
92.00         @     @@  @@  @          @  @@@@      @      @  @     @
93.00     @@@      @@   @    @       @@@@    @            @@
94.00              @  @@           @@ @     @@         @@   @   @@@@
95.00         @     @      @   @@@@@         @    @@  @    @    @@
96.00    @    @             @@@@    @            @@       @    @
97.00        @@         @@@@                   @         @@   @@@   @
98.00       @@@      @@@        @@@       @@@      @ @@  @ @
99.00    @        @@        @@@   @   @@@@                @      @
```

Figure 12.2 Band 7 Roads, Raw Image

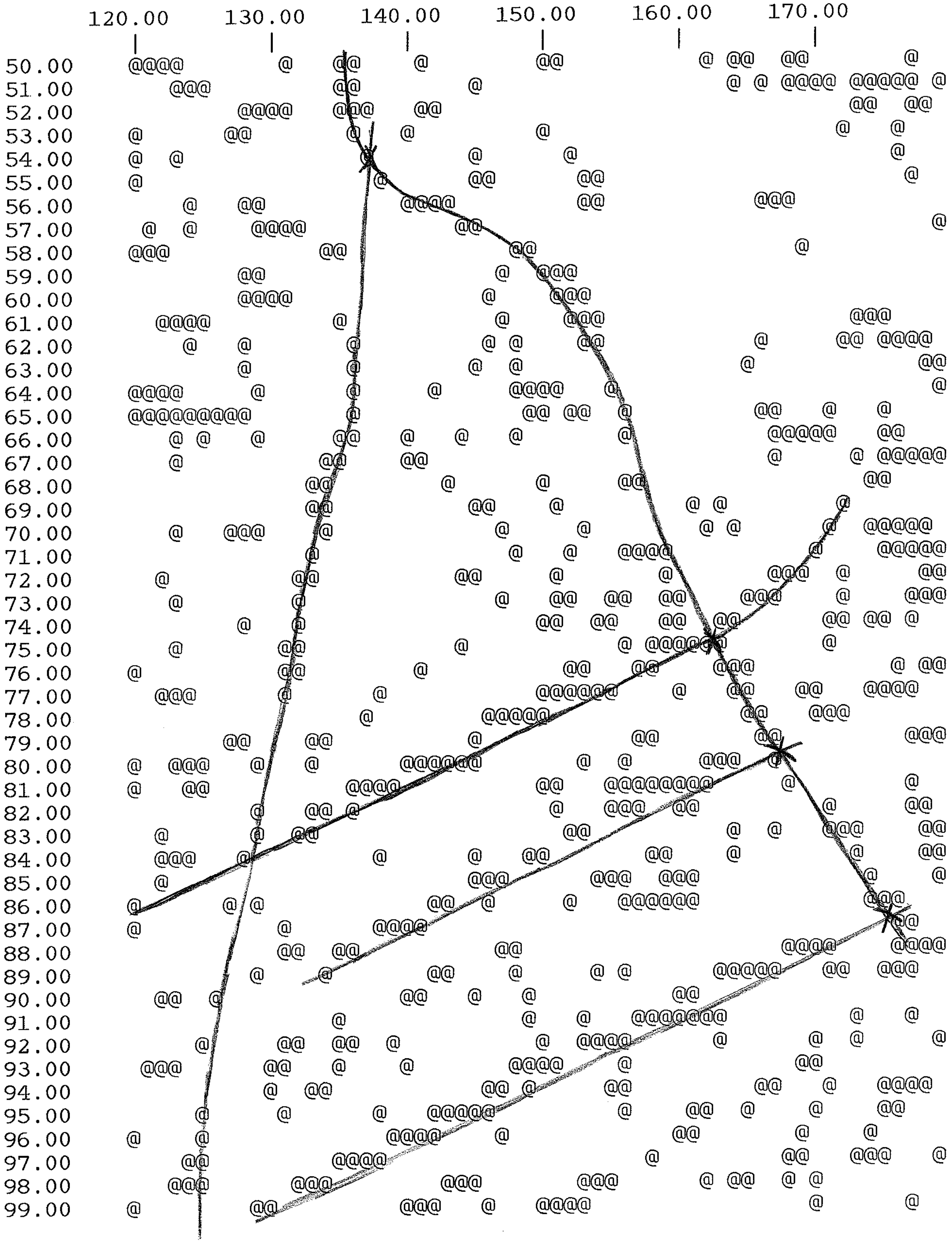


Figure 12.3 Band 7 Roads with Marked Roads and GCPs

Table 12.5 Geometric Correction Coefficient Evaluation

MAP GCPs UTM km		LANDSAT GCPs pixels		TRANSFORMED MAP GCPs pixels		TRANSFORMATION ERROR pixels	
616.767	3358.165	597.000	1103.000	596.173	1102.760	0.827	0.240
616.645	3359.747	570.000	1086.000	570.487	1085.736	-0.487	0.264
615.512	3371.078	526.000	946.000	525.090	945.052	0.910	0.948
616.138	3368.945	543.000	970.000	543.392	970.576	-0.392	-0.576
623.703	3368.668	674.000	958.000	675.068	958.511	-1.068	-0.511
622.646	3368.969	655.000	957.000	655.772	956.961	-0.772	0.039
620.538	3363.651	641.000	1026.000	639.345	1027.598	-0.345	-1.598
624.981	3356.866	744.000	1102.000	744.151	1101.811	-0.151	0.189
628.553	3354.321	816.000	1125.000	817.062	1125.112	-1.062	-0.112
627.489	3363.581	761.000	1015.000	759.488	1013.735	1.512	1.265
634.655	3356.809	910.000	1082.000	910.788	1081.122	-0.788	0.878
634.076	3355.632	907.000	1097.000	906.093	1096.800	0.907	0.200
636.027	3354.808	944.000	1101.000	943.234	1102.382	0.766	-1.382
637.374	3357.821	952.000	1063.000	952.865	1062.525	-0.865	0.475
634.313	3370.092	852.000	917.000	851.784	917.668	0.216	-0.668
636.841	3366.583	907.000	955.000	907.908	955.716	-0.908	-0.716
634.676	3364.483	879.000	987.000	879.116	986.681	-0.116	0.319
632.849	3370.250	827.000	920.000	826.228	918.954	0.772	1.046
627.085	3360.826	764.000	1048.000	763.660	1048.643	0.340	-0.643
627.607	3362.398	766.000	1029.000	766.238	1028.114	-0.238	0.886
627.638	3366.376	751.000	978.000	751.233	978.717	-0.233	-0.717
632.377	3361.733	852.000	1026.000	850.827	1025.826	1.173	0.174

Mean squared errors (pixels): 0.594 0.576
Total rms error (pixels): 0.82

12.4. LAND COVER CLASSIFICATION

Establishing a set of land cover classes that meets the needs of a project as well as the constraints of a remotely sensed data source is a major task of both the agency requiring the study as well as the data interpreter. The land use and land cover classification system, developed by Anderson and Hardy and shown in Table 7.1, is one example of a classification system, and land cover classification, in general, was also discussed in section 7. In the Table 12.6, three different classification schemes for the same region are included: USACE, HEP, and UCD Landsat. This comparison is from a habitat study of the Walnut Creek watershed, Texas, conducted by UC Davis.

In the USACE study, the land use was originally determined by manual interpretation of 1:12000 scale intrared aerial photographs, using 79 land use types [35]. These results were further analyzed to correlate the habitat types to the 79 land use classes and to combine these classes to a reduced number of habitat types [36]. On the basis of faunal utilization in association with the natural plant communities, the original 79 classes were aggregated into 21 habitat types having common biological characteristics. Then, during the final habitat evaluation phase of the study, several categories were combined to create 14 habitat categories [37].

The HEP classes are from a Wildlife and Fisheries classification system emphasizing habitat type and the use of remotely sensed data. The Corps of Engineers participated with the U.S. Fish and Wildlife Service in the development of a habitat-based wildlife evaluation methodology entitled the Habitat Evaluation Procedure (HEP). HEP is used to assist in planning projects that may have an impact on wildlife and their habitat [30]. A habitat suitability model has been designed for key indicator wildlife species, and a major input to this model is the distribution of land cover in the study area. A comprehensive list of the HEP classes may be found in [31].

The UCD Landsat classes were developed during a project that examined the feasibility of determining land cover classes suitable for habitat classification, using Landsat as a data source. An interactive version of the UCD Land Use Classification procedure was used.

Looking at Table 12.6, it can be seen that it was possible to identify 10 of the USACE classes and 8 of the HEP classes. The principal difficulties encountered were in attempting to subdivide the tree classes into the 8 species or mixed species classes of the USACE study, and in differentiating between cropland and herbaceous rangeland.

Table 12.6 A Comparison of 3 classification studies

UCD Landsat	USACE	HEP	%Cover
water	aquatic	riverine	0.15
residential	residential urban/rural other Urban	urban and built-up land	19.18
commercial/ industrial	commercial industrial transportation	urban and built-up land	8.62
barren	barren land	mining/barren	2.53
trees with grassy understory	mesquite post oak-live oak live oak ashe juniper live oak-ashe juniper mesquite-juniper	forest	6.02
crop/pasture	cropland herbaceous range	cropland pasture and hayland forbland	17.55
irrigated vegetation	cropland other urban	cropland urban and built-up land	2.69
trees with shrubby understory	elm-hackberry pecan-elm	forest	3.58
shrub	shrub-brushland	shrubland	4.18
tree savanna	herbaceous range	tree savanna	35.50

In unsupervised classification, a clustering algorithm is applied to the Landsat image, separating the data into clusters of feature vectors, as discussed in section 7.3. In unsupervised classification, an image of the clusters assigned to pixels is then printed on a cluster map, with several clusters printed per map. Figure 12.4 is an example of a symbol map, where the symbol, @, identifies cluster 1 and the symbol, %, identifies cluster 2. Cluster 3 is a *, and all other clusters are identified as blank spaces. Figure 12.4 is only a small portion of a cluster map, for a much larger area is needed for analysis in land cover identification. For each cluster, groupings of pixels (called spatial groups) are identified. In Figure 12.4, the cluster @ is associated with two spatial groups and cluster % is associated with one spatial group. During this step, 6 spatial groups are selected for each cluster. The spatial groups will vary in size and location, for these spatial groups will be used to represent the land cover possibilities for the entire cluster.

```
       200.00    210.00    220.00    230.00    240.00    250.00
          |         |         |         |         |         |
200.00               *    * *                C    AAAA@BBB  B   B*B *
201.00    *     A                A            CC AA          **   B
202.00     B**  @           B*  A        CCC      A        B C  *B*  *
203.00    B *     **   *     B    *  *          A AA      A B  CC  *
204.00    B ****  * * ***   C      *     CCC     AAA A   A A * *  *
205.00         *   ****              C  C        A    AA AA@ @
206.00        *  *    B****   @@    B      CC  @ A AA@@@@AA@         C
207.00        *  *    CCCC     @@ C B   A A      @@@AAAA  A   C  C
208.00       * **   B*   CC         *BB* @@  @@       @AAA C
209.00    B     *  BB  B          * B             C    @AA         CC
210.00    *             BB**       **  @       @    CCCC @ C     *
211.00     B*        B            @  @AA     @@@@    C C CC
212.00      B *       B    @       @    B C     @    C C C  **
213.00       B  B*          @@  @    C     C         CC  C   **B
214.00    B    B  B*   **  *    @        CC *   *       A
215.00    BB  B    B           @@@      CC  ****B       A  @@ *      B
216.00        B     **     B       @@@   CC        B   @@      *
217.00               *     *B  *  A@      C           @@@
218.00            B                          @A@@          C CC
219.00                   *** B  B*    B    @@@A@@    *  C C
220.00            **   *B  *******  *BB*B    @@A@ *  B CC C        C
221.00                 BBB* *****    *B*   *   @@    ***          CCCC
222.00        B           B ****BB*       *BB *
223.00           BBB    B* ***  **  * *BBB*   B   @         @    C
224.00    A  *B    BBB  *  B B****     BBB  ***     @  CC   @
225.00     @        B*    B BB**     *  B BB  *     @ CC CC        **
226.00     @BBB  BB  B       BB**B*  *BB   BB  B**** **  C  *
227.00    B B      B*B  BB*   B*B* * BBBBBBBB            C
228.00      *B  BB  BB  CC *   B  B *BBB*B   BB    *
229.00          * * BB**  ***B C   BBBBB BBBB*BB **   *          B*
230.00    * ****BB  B      * C     ***BB B**BB** *                 B
231.00     *****BB *B     *   * BB  ** * BBB**B*           C C  *
232.00    ** BBB**BBB B   B   B*  A BB*   *BBBB          *         **
233.00       BBB B* B   BB *B      * *   B** B  B    * B            BB
234.00        *B** *      ** C C *  * **BBBBBBB* B  *B    *          *
235.00       * B   @**        CC    *BB   **BB*   B*     B *
236.00      B**  B    *        *  ***BB    *     *        **        *B
237.00        *B    ***    B *  *   *      B     *B     B  B       B B
238.00    B    B B      **     *       B  B      *BB  BBB  B*     * BB
239.00    BBBBB  B*         *   B   B            B    B*
240.00    BB   BBBBB*B* **  *    B **BB              B
241.00      B  BBB BB*****B  B*B B  *      B        B  *
242.00       BB*BBB B B*BBB*    B    B         *     B   C   @   *
243.00       **BBBBB*   *   *  ***    *  CCC         B*CC   C    CC
244.00    B  * *  BB* C*    *   B    BB  CC C    *      C
245.00      BBBBBB**B***   B ** **  B* B*  C   C        CC C
246.00    B    B BBB*B*BB C*  B* BB   *    *  C C          C       BBB
247.00     BB B B  **  B***BBBBB BBBB *   * C        C     C
248.00     BBB*B   ***** B** B BB **B* B      CCC         C
249.00        *B ** ******B* B**BB***B   * A  CC              *
```

Figure 12.4 A Symbol Map

Each spatial group should be assigned to a land cover class. The following guidelines may be used when deciding on a land cover class.

- Use any available ground truth photographs, USGS maps, and personal knowledge of the area when making a decision.
- Pixels near the boundary of the spatial group may not be representative of the spatial group.
- Consider the relative size and shape of the spatial group, as well as the scale of your ground truth data.
- Be aware of any differences in exposure dates among your ground truth data sources.

After identifying spatial groups and assigning land cover classes to cluster numbers, there may be a few clusters with more than one class assignment. Another aid in assigning classes to clusters is the feature space plot, shown in Figure 12.5 and 12.6.

In Figure 12.5 and 12.6, clusters have been transformed using a linear transformation to feature space. They been transformed in such a way that adjacent clusters have similar land cover characteristics. In this feature space, each axis represents one data component. Initially, each point is associated with a cluster number; we have not numbered these cluster means in these figures. The points represent the location of the clusters after the linear transformation from four dimensions to two dimensions. Adjacent clusters generally have similar class assignments. After clusters 1 and 2 from figure 12.4 have been assigned to a land cover class, their assignment would be noted on a plot such as these. Cluster 1 (from Figure 12.4) was subsequently assigned to the class, residential, and cluster 2 was assigned to the class, commercial.

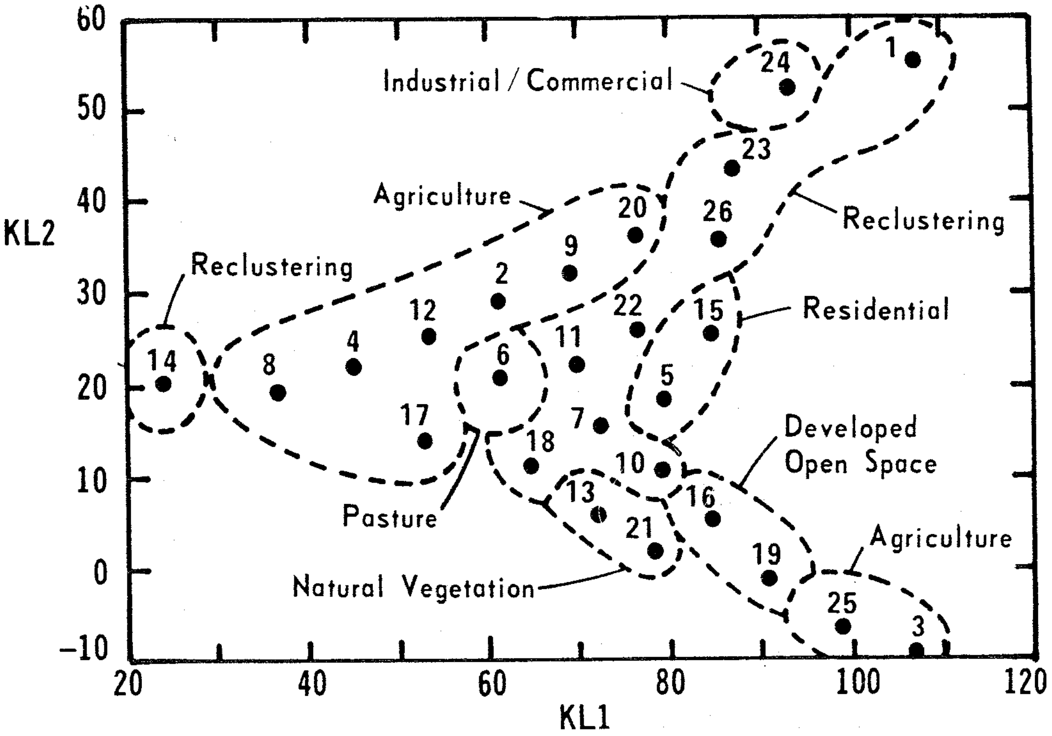


Figure 12.5 Cluster Transformation, Intermediate Plot

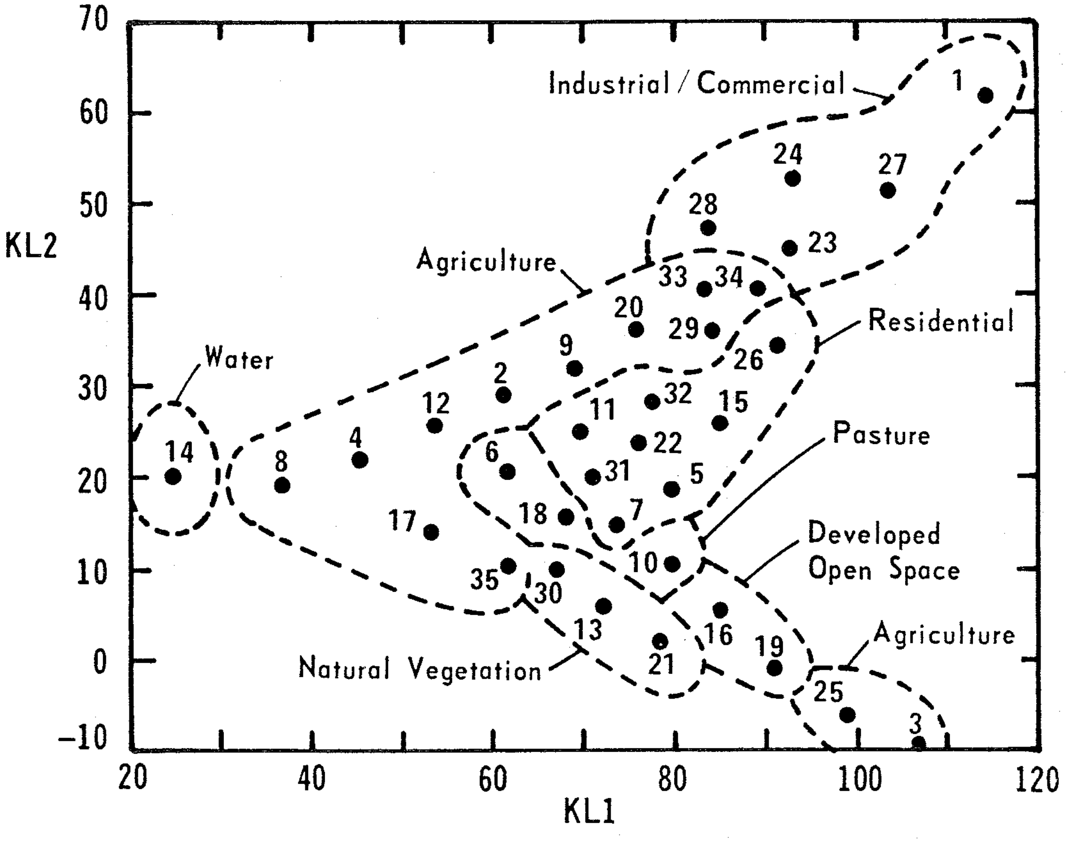


Figure 12.6 Cluster Transformation, Final Plot

12.5. GEOMETRIC RESAMPLING

In the UCD Classification Procedure, after selecting a set of ground truth control points and generating the geometric transformation coefficients (section 6, section 12.3), the Landsat data is resampled from pixel coordinates to UTM coordinates. The Figure 12.7 is a resampled version of Figure 12.2. Figure 12.7 will now overlay the 1:24000 topographic map, shown in Figure 12.8. In order to resample the roads file (12.7), the program, RESAMP, was used. RESAMP uses a nearest neighbor interpolation technique. Inputs to RESAMP include the size of the output cell (in this case, the printer character), the scale of the USGS map, the directions in which the output coordinate system increases, and the upper left corner point of the output file in UTM coordinates.

```
        30.00     40.00     50.00     60.00     70.00     80.00
          |         |         |         |         |         |
  0.00      @   @       @@    @     @ @   @    @@@   @@ @@  @@       @
  1.00      @@@           @@@ @@     @@@@@ @            @    @ @@    @
  2.00          @@ @@@@   @  @@   @@            @@  @           @  @
  3.00    @@     @  @     @  @@    @@@        @ @     @@@   @
  4.00          @@  @       @@@    @@ @        @            @     @
  5.00             @     @   @@    @@    @        @@@@@@@@@@@   @@   @
  6.00         @@      @     @@                         @@   @ @@@ @
  7.00    @@@@@@  @@ @@@@@  @@       @        @   @         @ @@ @ @ @
  8.00         @@@     @@   @  @@@@@ @         @@@@@@@       @@@
  9.00    @@        @@     @@@    @        @     @  @@  @         @@@
 10.00    @@  @   @@   @   @@@        @    @@          @@         @@
 11.00    @                 @   @ @                     @ @@        @
 12.00        @             @             @               @ @ @@
 13.00    @@  @    @         @      @@      @                 @@@@ @@@
 14.00               @@        @@@@          @@                   @@@
 15.00           @@     @@         @@       @@                       @
 16.00    @@@@@  @@@@                 @@@                 @          @
 17.00      @   @       @            @  @@@                 @        @
 18.00    @     @        @               @@@                @
 19.00    @@@    @      @          @ @    @@@
 20.00    @ @ @@        @     @  @@  @    @@
 21.00          @     @@            @@@@@               @      @@@@
 22.00               @@    @    @    @@ @@ @@         @       @@ @@@@
 23.00             @@       @       @       @                       @@
 24.00      @@@@    @         @      @                 @@   @        @
 25.00            @            @@@    @    @@          @ @@@@   @
 26.00           @@              @      @       @            @ @@@@
 27.00      @   @            @@@  @   @   @      @            @@    @@
 28.00         @@               @    @       @            @
 29.00        @@@                  @@   @@@  @       @@  @    @@@@@  @
 30.00        @          @             @@   @@@    @@@    @@     @@@@@
 31.00    @         @               @@@  @ @@@@@@        @@    @@@@
 32.00     @   @ @           @@@@@@@@@@@   @    @@@      @  @@ @
 33.00            @    @@@@ @               @    @@               @@@@
 34.00    @   @@@ @ @@              @    @@      @@    @   @@@@@@@@@ @
 35.00        @                 @@@   @@      @@  @@    @        @   @
 36.00             @             @   @@@@@@@@@    @@         @@@@@
 37.00                    @    @@ @@      @@@      @              @
 38.00     @      @    @@  @@           @@    @   @   @     @@
 39.00         @@   @@           @   @   @@@  @       @@@    @@     @
 40.00       @             @@          @@@@                  @@@     @
 41.00        @    @@       @                           @@   @@@ @@ @@
 42.00    @  @@             @     @ @            @@       @         @@
 43.00       @              @   @       @@  @@@@@   @    @@@@        @
 44.00     @              @@@@  @@@@ @@@@@@@            @@@   @    @
 45.00         @    @@  @@ @       @      @          @       @@  @@@@@
 46.00      @@@ @@@@   @@        @@             @@@  @   @  @@
 47.00    @                       @    @@    @@            @@@@     @
 48.00          @@ @@@@               @@         @    @@@@   @@@
 49.00                @   @@  @@@              @     @@@       @
```

Figure 12.7 Band 7 Roads, Resampled

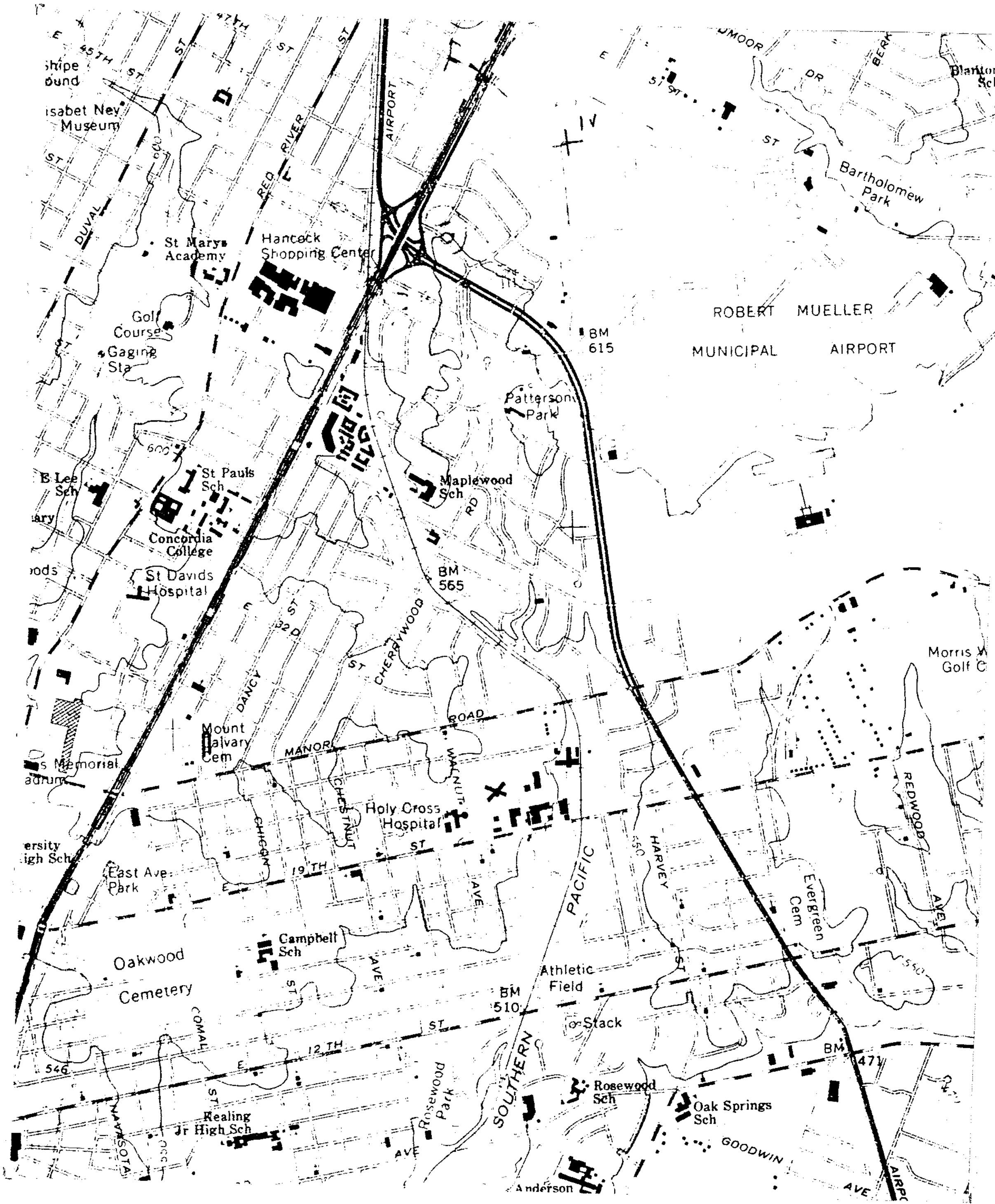


Figure 12.8 Topographic Map Aligned with Figure 12.7

12.6. ACCURACY ASSESSMENT

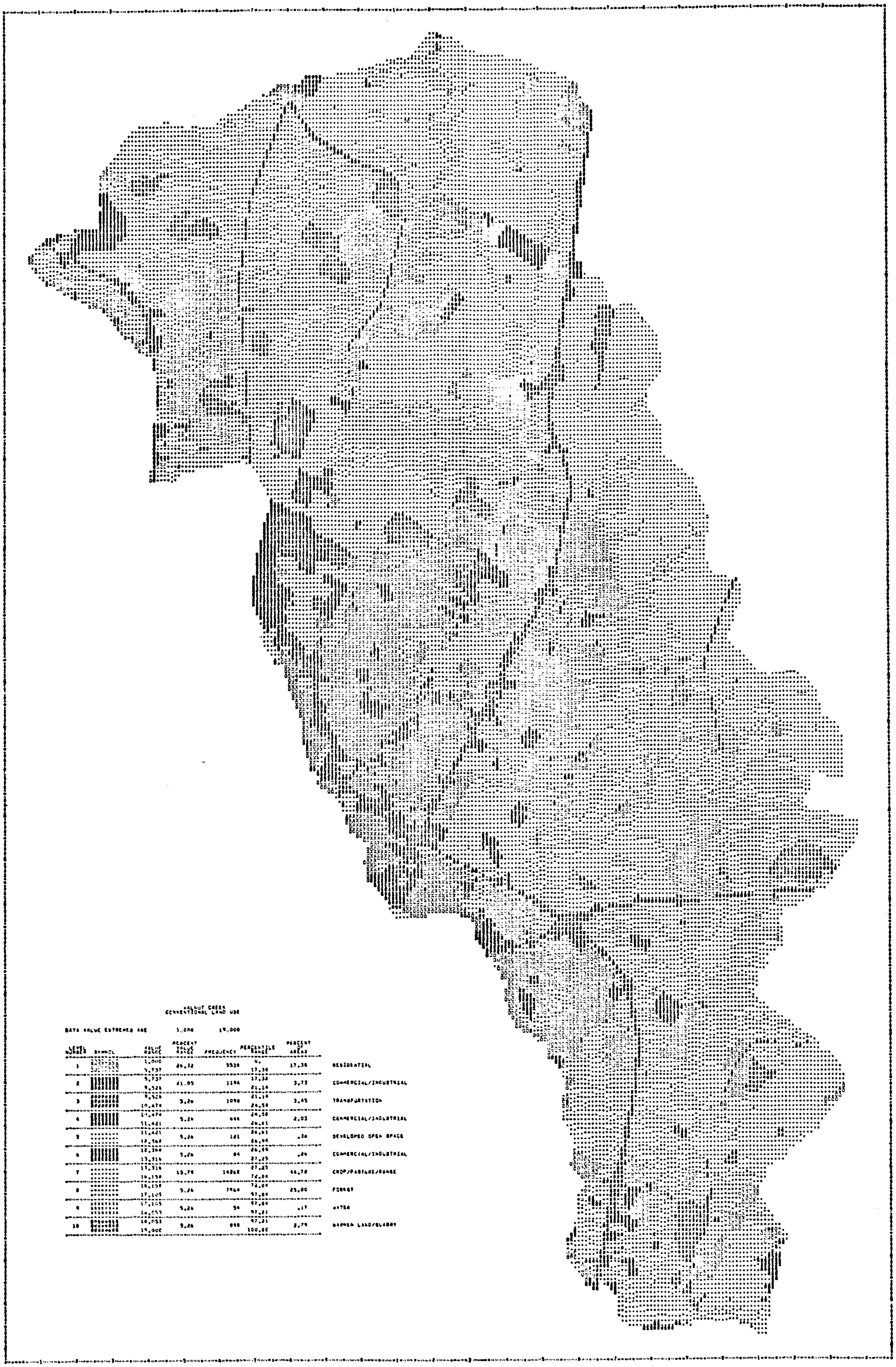
As discussed in section 9.2, a cell-by-cell comparison of one classification method with another may be helpful in assessing classification accuracy. Table 12.7 was generated by HEC using the RIA programs discussed in section 10. Columns represent the conventional results from Walnut Creek, Texas, and the rows represent the Landsat-based results from Walnut Creek, Texas. The top number in each box is the number of acres, and the bottom number in each box is the percent of acres.

Table 12.7 Coincidence Matrix

	Residential	Commercial/ Industrial	Quarry	Crop/Pasture/ Range	Forest	Water	Totals
Residential	3801 10.4	1098 3.0	225 0.6	3147 8.6	859 2.3	6 0.0	9136 25.0
Commercial/ Industrial	471 1.3	816 2.2	130 0.4	366 1.0	78 0.2	0 0.0	1861 5.1
Quarry	13 0.0	61 0.2	111 0.3	17 0.0	21 0.1	0 0.0	223 0.6
Crop/Pasture	1581 4.3	1106 3.0	313 0.9	9324 25.5	2947 8.1	28 0.1	15229 41.8
Forest/Range	493 1.3	361 1.0	234 0.6	3655 10.0	5233 14.3	13 0.0	9989 27.3
Water	0 0.0	31 0.1	8 0.0	9 0.0	6 0.0	16 0.0	70 0.2
Totals	6359 17.4	3473 9.5	1021 2.8	16518 45.2	9144 25.0	63 0.2	36478 100.0

12.7. CREATING THE GRID CELL DATA FILE

After the grid cell data file has been generated, the RIA programs can also output coded maps, such as the one produced by HEC in figure 12.9.



WALNUT CREEK
CONVENTIONAL LAND USE

DATA VALUE EXTREMES ARE 1.000 19.000

LEVEL NUMBER	SYMBOL	VALUE RANGE	PERCENT VALUE RANGE	FREQUENCY	PERCENTILE RANGE	PERCENT OF AREA	
1	░	1.000 5.737	26.32	5530	0. 17.38	17.38	RESIDENTIAL
2	▮	5.737 9.526	21.05	1194	17.38 21.14	3.75	COMMERCIAL/INDUSTRIAL
3	▩	9.526 10.474	5.26	1098	21.14 24.58	3.45	TRANSPORTATION
4	▮	10.474 11.421	5.26	646	24.58 26.61	2.03	COMMERCIAL/INDUSTRIAL
5	░	11.421 12.368	5.26	121	26.61 26.99	.38	DEVELOPED OPEN SPACE
6	▮	12.368 13.316	5.26	84	26.99 27.25	.26	COMMERCIAL/INDUSTRIAL
7	░	13.316 16.158	15.79	14248	27.25 72.04	44.78	CROP/PASTURE/RANGE
8	░	16.158 17.105	5.26	7956	72.04 97.04	25.00	FOREST
9	▒	17.105 18.053	5.26	54	97.04 97.21	.17	WATER
10	▩	18.053 19.000	5.26	888	97.21 100.00	2.79	BARREN LAND/QUARRY

Figure 12.9 Land Cover Map of Walnut Creek, Texas

References

1. Darryl W. Davis, "Comprehensive flood plain studies using spatial data management techniques," *Water Resources Bulletin*, vol. 14, pp. 587-604, 1978.
2. Gary E. Ford, V. Ralph Algazi, and Doreen I. Meyer, "A noninteractive procedure for land-use determination," *Remote Sensing of Environment*, pp. 1-16, 1983.
3. Nicholas M. Short, *The Landsat Tutorial Workbook,* National Aeronautics and Space Administration, Washington, DC, 1982.
4. U.S. Army Corps of Engineers, *Resource Information and Analysis Using Grid Cell Data Banks: Users Manual,* The Hydrologic Engineering Center, Davis, CA, 1981.
5. U.S. Army Corps of Engineers, *Grid Cell Data Bank Update Program (BANK): Users Manual,* Hydrologic Engineering Center, Davis, CA, 1980.
6. U.S. Army Corps of Engineers, *Guide Manual for the Creation of Grid Cell Data Banks,* The Hydrologic Engineering Center, Davis, CA, September 1978.
7. T. B. Heverin and G. R. Rohn, "Computer models for dredged material disposal planning," *Proceedings of the ASCE Dredging Conference*, Clearwater Beach, FL, November 14-16, 1984 (to appear).
8. David T. Ford and Darryl W. Davis, "Hydrologic Engineering Center Planning Models," Technical Paper No. 92, U.S. Army Corps of Engineers, Hydrologic Engineering Center, Davis, CA, 1983.
9. R. H. Burgy and V. R. Algazi, "An assessment of remote sensing applications in hydrologic engineering," Research Note No. 4, The Hydrologic Engineering Center, Davis, CA, 1974.
10. T. J. Jackson, R. M. Ragan, and W. N. Fitch, "Test of LANDSAT-based urban hydrologic modeling," *Journal of the Water Resources Planning and Management Division , ASCE*, vol. 101 (WR1), pp. 141-158, 1977.
11. R. M. Ragan and T. J. Jackson, "Use of satellite data in urban hydrologic models," *Journal of the Hydraulics Division, ASCE*, vol. 101 (HY12), pp. 1469-1475, 1975.
12. U.S. Army Corps of Engineers, "Determination of land use from Landsat imagery: applications to hydrologic modeling," Research Note No. 7, The Hydrologic Engineering Center, Davis, CA, 1979.

13. Robert J. Cermak, Arlen D. Feldman, and R. Pat Webb, "Hydrologic Land Use Classification Using Landsat," Technical Paper No. 67, U.S. Army Corps of Engineers, Hydrologic Engineering Center, Davis, CA, 1979.

14. P. F. Holkenbrink, *Manual on Characteristics of Landsat Computer-Compatible Tapes Produced on the EROS Data Center Digital Image Processing System,* U.S. Geological Survey, EROS Data Center, Sioux Falls, SD, 1978.

15. U.S. Geological Survey, *Landsat Data Users Handbook,* EROS Data Center, Sioux Falls, SD, 1979.

16. Ralph Bernstein, Charles Colby, Stephen W. Murphrey, and John P. Snyder, "Image geometry and rectification," in *Manual of Remote Sensing, 2nd Ed.*, ed. Robert N. Colwell , American Society of Photogrammetry, Falls Church, VA, 1983.

17. V. Ralph Algazi, Gary E. Ford, and James A. Kazakoff, "Radiometric correction and equalization of satellite digital data," *Proceedings of the Twelfth Annual Asilomar Conference on Circuits, Systems and Computers*, pp. 336-340, 1978.

18. V. Ralph Algazi and Gary E. Ford, "Radiometric equalization of non-periodic striping in satellite data," *Computer Graphics and Image Processing*, vol. 16, pp. 287-295, 1981.

19. B. K. P. Horn and R. J. Woodham, "Destriping Landsat MSS images by histogram modification," *Computer Graphics and Image Processing*, vol. 10, pp. 69-83, 1979.

20. U.S. Army Corps of Engineers, *Proceedings of a Seminar on Variable Grid Resolution - Issues and Requirements,* The Hydrologic Engineering Center, Davis, CA, 1977.

21. Gary E. Ford, V. Ralph Algazi, and Brian G. Agee, "On the correction of geometric distortion in satellite- acquired images," *Proceedings of the Twelfth Annual Asilomar Conference on Circuits, Systems and Computers*, pp. 341-344, 1978.

22. Ralph Bernstein, "Digital image processing of earth observation sensor data," *IBM Journal of Research and Development*, pp. 40-57, 1976.

23. P. Van Wie and M. Stein, "A Landsat digital image rectification system," *IEEE Transactions on Geoscience Electronics*, vol. GE-15, pp. 130-137, 1977.

24. Robert M. Haralick and King-Sun Fu, "Pattern recognition and classification," in *Manual of Remote Sensing, 2nd Ed.*, ed. Robert N. Colwell, American Society of Photogrammetry, Falls Church, VA, 1983.

25. Floyd F. Sabins, *Remote Sensing Principles and Interpretation,* W.H. Freeman and Company, 1978.

26. Philip H. Swain, "Fundamentals of pattern recognition in remote sensing," in *Remote Sensing: The Quantitative Approach*, ed. Philip H. Swain and Shirley M. Davis, McGraw-Hill, New York, 1978.

27. E. P. Kan, W. A. Holley, and H. D. Parker, Jr., "The JSC clustering program ISOCLS and its applications," *Proc. 1973 Symp. on Machine Processing of Remotely Sensed Data*, pp. 4B-36 to 4B-50, Purdue University, West Lafayette, IN, 1973.

28. U.S. Army Corps of Engineers, *Phase 1 Oconee Basin Pilot Study: Trail Creek Test,* The Hydrologic Engineering Center, Davis, CA, 1976.

29. J. A. Anderson, E. E. Hardy, J. T. Roach, and R. E. Witmer, "A land use and land cover classification system for use with remote sensor data," Professional Paper 964, U.S. Geological Survey, 1976.

30. U.S. Fish and Wildlife Service, "Habitat as a basis for environmental assessment," Publication 101ESM, Division of Ecological Services, U.S. Department of the Interior, Washington, D.C., 1980.

31. U.S. Fish and Wildlife Service, "Standards for the development of habitat suitability index models," Publication 103ESM, Division of Ecological Services, U.S. Department of the Interior, Washington, D.C., 1981.

32. John R. Jensen, "Urban/suburban land use analysis," in *Manual of Remote Sensing, 2nd Ed.*, ed. Robert N. Colwell, American Society of Photogrammetry, Falls Church, VA, 1983.

33. Harlan L. McKim, Stephen G. Ungar, Carolyn J. Merry, and John F. Gauthier, "Integration of Landsat land cover data into the Saginaw River Basin geographic information system for hydrologic modeling," Special Report 84-1, U.S. Army Corps of Engineers, Cold Regions Research & Engineering Laboratory, Hanover, NH, 1984.

34. R. Pat Webb, Robert Cermak, and Arlen D. Feldman, "Determination of Land Use from Satellite Imagery for Input to Hydrologic Models," Technical Paper No. 71, The Hydrologic Engineering Center, Davis, CA, 1980.

35. R. D. Baker, "Land Use Study and Soil Series Map - Walnut and Williamson Creek Watersheds," Report to U.S. Army Corps of Engineers, Ft. Worth District ,

Department of Forestry Science, Texas A&M University, College Station, TX, 1978.

36. M. D. Springer and D. L. Wilkinson, "Habitat Identification and Community Description for Walnut and Williamson Creeks - Austin, Texas," Report to U.S. Army Corps of Engineers, Ft. Worth District, LGL Limited-U.S., Inc., 1978.

37. U.S. Army Corps of Engineers, *Walnut and Williamson Creeks: Expanded Flood Plain Information Study, Environmental Consideration,* 3, pp. E3-E31, USACE Ft. Worth District, Ft. Worth, TX, 1980.

Appendix: The Landsat Multispectral Scanner

Landsat is the name of a series of NASA satellites intended to provide Earth resources information to the general scientific and environmental planning community. The first three satellites in the series, Landsats 1 through 3, had near-polar orbits 570 miles (920 km) above the Earth's surface, circling the earth every 103 minutes, or about 14 times per day. The sun-synchronous orbit allowed the spacecraft to view a given latitude at the same local time on each pass, crossing the equator at about 9:30 am local time. Repeat coverage of a given location from a single satellite occured every 18 days. When two satellites were active at once, their orbits were staggered to provide 9 day repeat coverage. The most recent satellites in the series, Landsats 4 and 5, have similar sun-synchronous polar orbits, but the altitude has been reduced to 435 miles (700 km), with a period of 99 minutes and a repeat cycle of 16 days. Landsat 1 was launched in July of 1972 and was active for 4 years. Subsequent satellites in the series have been launched about every 3 years. Landsats 4 and 5 are currently active (January 1985), although failure of some of the solar panels in Landsat 4 have limited its performance.

The primary sensor system is the multispectral scanner (MSS). The scanner senses reflected electromagnetic energy in four spectral bands. The MSS uses line scanning devices that continually scan the Earth in a 115 mile (185 km) swath perpendicular to the Landsat orbital track, as shown in figure A-1 (taken from [1]). Scanning is accomplished in the cross-track direction by an oscillating mirror; six lines are scanned simultaneously in each of the four spectral bands for each mirror sweep. The forward motion of the satellite provides the along-track progression of the scan lines. The resolution of the sensors, also known as the instantaneous field of view (IFOV), is about 79 meters square. The pixel spacing is 57 meters along scan by 79 meters across scan. The spectral bands of the MSS and their characteristics are:

Band 4, the green band, 0.5 to 0.6 micrometers, emphasizes movement of sediment-laden water and delineates areas of shallow water, such as shoals, reefs, etc.

Band 5, the red band, 0.6 to 0.7 micrometers, emphasizes cultural features, such as urban areas.

Band 6, the near-infrared band, 0.7 to 0.8 micrometers, emphasizes vegetation, the boundary between the land and water, and landforms.

Band 7, the second near-infrared band, 0.8 to 1.1 micrometers, provides the best penetration of atmospheric have and also emphasizes vegetation, the boundary between land and water, and landforms.

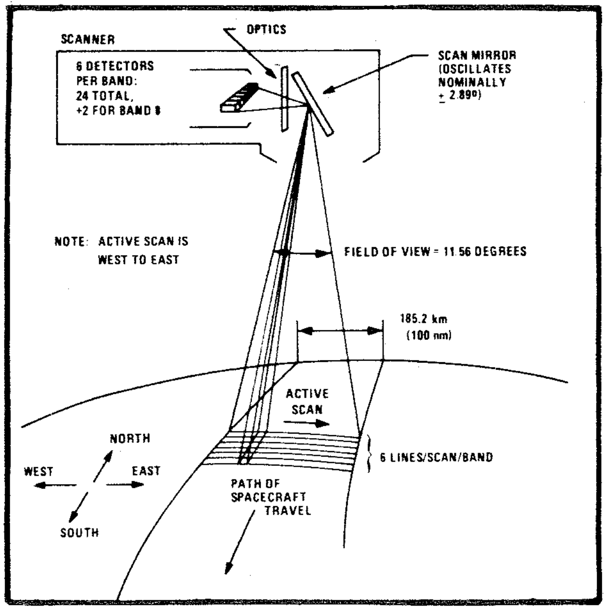


Figure A-1. MSS sensor system

The reflected energy sensed is converted into an analog voltage, which varies in proportion to the energy intensity in the respective spectral band. The analog video signals from each detector are sampled, multiplexed, and digitized by a constant-rate, on-board analog-to-digital system. The resulting digital signal is then transmitted to receiving stations on the ground.

The digital data may be either reformatted into computer compatible tapes (CCTs) or reconverted into sets of black and white or color photographs. A Landsat scene is an imaged ground area normally 115 miles (185 km) square. A Landsat scene stored on a CCT contains 2340 scan lines of 3240 columns (samples) each, an aggregate of over 7.5 million data elements, called pixels. Each pixel has a numerical reflectance value for each of the four spectral bands.

Landsat photographic imagery is derived from the digital data. Relative reflectance values are assigned gray levels for the production of band 4, 5, 6, or 7 scene photographs. Color photographs are made from combinations of individual black and white images by projecting each band through a particular optical filter. In this rendition, called a false color image, band 4 is commonly assigned to blue, band 5 to green, and band 6 or 7 to red, producing a photograph having characteristics similar to that of a color infrared photograph. Growing vegetation will appear in various shades of red, rocks and soils will normally show colors ranging from blue through yellow and brown, water will stand out as black to blue depending on depth and amount of suspended sediment, and cultural features (towns and roads) will normally be recognized by bluish-black tones arranged in characteristic patterns.

The current satellites were developed under contract and were launched by NASA. The National Oceanic and Atmospheric Administration (NOAA) has responsibility for the operational management of the system and distribution of the collected data. Legislation enacted in July 1983 authorized the transfer of U.S. land remote sensing capabilities from the Federal Government to the private sector, and a proposal process for this tranfer was initiated. At this writing (January 1985), negotiations are continuing with the one remaining competitor in the proposal process, the Earth Observation Satellite Company (EOSAT). Technically, EOSAT is a joint venture of two companies, RCA Corporation and Hughes Aircraft Corporation, with four major subcontractors: Santa Barbara Research Center (SBRC), Computer Sciences Corporation (CSC), Earth Satellite Corporation (Earth-Sat), and RCA Astro-Electronics. At this writing, a final draft of the contract that would govern the Landsat commercialization process was being circulated both within the Department of Commerce and at EOSAT headquarters in Arlington, Virginia.

Further information on the Landsat system is available in [2], and technical details are provided in [1].

References

1. U.S. Geological Survey, *Landsat Data Users Handbook,* EROS Data Center, Sioux Falls, SD, 1979.

2. Nicholas M. Short, *The Landsat Tutorial Workbook,* National Aeronautics and Space Administration, Washington, DC, 1982.