PN-AAQ-289
ISN: 35212

not an evaluation

LANDSAT ANALYSIS OF AGRICULTURAL CHANGES IN PAKISTAN

Final Report

Prepared for the U.S. Agency for International Development
USAID/Pakistan, Islamabad, Pakistan

In Fulfillment of Order No.
ASB-0470-0-00-3021-00

April 1983

Prepared By:

IRIS International, Inc.
4301 Garden City Drive
Suite 202
Landover, Maryland, 20785, USA

CONTENTS

Page

Executive Summary iv

Findings of the Study 1

Background 2

Data for the Study 3

Findings 4

Agriculture by Elevation 7

Agriculture and Slope 7

Pressure on Marginal Land 7

Discussion of the Results 10

The Question of Poppies 11

Reliability of the Results 11

Future Improvements 12

The Technology and Its Potential 15

The Technology: A Definition 15

Potential Application of the Technology 16

Project Definition 18

Project Inputs 19

Project Monitoring 19

Project Performance 20

Project Evaluation 21

Requirements of the Technology 21

Training 22

CONTENTS--Continued

Methodology 23

Project Definition 23

Data Types and Sources 25

Base Map/Spatial Coordinate System 28

Data Files 30

Base Map Used in Final Map Production 37

Analysis 38

Results 50

LIST OF FIGURES

Page

Figure 1 - LANDSAT 4 Image 14

Figure 2 - Project Development Plan Flow Chart 17

Figure 3 - Geographic Information System Flow Chart..24

Figure 4 - Data Collection Procedure 29

Figure 5 - Topographic Elevation Map 31

Figure 6 - Slope by Percent Map 33

Figure 7 - Raw 1975 LANDSAT Data Map 35

Figure 8 - Raw 1982 LANDSAT Data Map 36

Figure 9 - Typical LANDSAT Results 40

Figure 10 - Histogram Listing for 1982 LANDSAT data.43

Figure 11 - Histogram Listing for Classified 1975 LANDSAT data 45

Figure 12 - Histogram Listing for Classified 1982 LANDSAT data 46

Figure 13 - Unsupervised LANDSAT Classification Map - 1975 47

Figure 14 - Unsupervised LANDSAT Classification Map - 1982 48

Figure 15 - Histogram Listing for Suitability for Roads 51

Figure 16 - Suitability for Roads and Road Network Map 52

Figure 17 - LANDSAT Clustering Chart 53

Figure 18 - 1975 Classified LANDSAT Landcover Map ...54

Figure 19 - 1982 Classified LANDSAT Landcover Map ...55

Figure 20 - Landcover Change Between 1975 and 1982 Map 57

EXECUTIVE SUMMARY

The purpose of this study is to assist in the development of a project in the Northwest Frontier area of Pakistan. Four maps have also been prepared as part of the report.

The data developed for use in the study is primarily from LANDSAT. The report uses the LANDSAT data and two large scale maps to analyze the changes in land use over a seven-year period. The data shows that agricultural activity has increased by nearly half over the 7-year period. What is more critical is that marginal land at higher elevations and with steep slopes is being cultivated more intensely. As a result, we suspect that the watershed may be threatened.

The report documents the steps necessary to turn a simple map and LANDSAT data into a Geographic Information System (G.I.S.) that can then be used in a number of analyses. In one analysis the study locates and maps the agricultural areas and maps the areas that have changed between 1975 and 1982. It also maps the changes in agriculture by elevation and slope. The G.I.S. is also used to develop corridors that would be most suitable for a road network.

The project has specific applications to Pakistan, but is useful as a general demonstration of the applications of LANDSAT data and Geographic Information System for development, especially where conventional sources of data are not available or current.

FINDINGS OF THE STUDY

This is a report of agricultural activity in an area of approximately 180 square miles centered on 72° 40' North latitude and 34° 20' East longitude, situated in the Northwest Frontier Provinces (NWFP) of Pakistan. The report is intended to assist in preparation of an area development project in the Gadoon area. In preparing the study, we used LANDSAT data and the few available maps.

The report has three sections: the findings of the study, a plan for incorporating the technology into the project, and methods and procedures used. The second section explains in detail how the use of the technology can help identify high priority areas for development and assist in project management, project monitoring, and project evaluation.

In addition to the report, we have produced a series of color maps that show:

- Major changes in land under cultivation from 1975 to 1982 with population settlements, transportation systems, water courses, vegetation, and contours depicting terrain, slope and elevation.
- 1982 landcover with vegetation, agriculture, water areas, etc., and showing population areas, transportation systems and contour lines.
- Suitability for roads with networks linking major population centers and leading into the project area.

Background

LANDSAT data has not been widely used as a tool for development. There is a compelling rationale for utilizing LANDSAT data in many development projects since current data on land use is seldom available. It is even more difficult to establish with some degree of reliability, how land use has been changing. Through use of the LANDSAT data archived since 1972 for much of the world it is possible to determine how land use has been changing.

The rationale for using LANDSAT data for project development in the NWFP of Pakistan is especially strong. Data for this area is scarce, travel is difficult and there are few available maps of the area.

LANDSAT data dating back to 1972 is available for the area. While we have deferred an extensive discussion of the technology in a later section, some basic information will be useful for those with no experience with LANDSAT digital data.

LANDSAT data consists of four bands or channels of data, representing the reflectance of the sun's energy from the earth's surface. Each channel has four numbers that apply to a small area of the earths surface, designated a pixel. The pixel size before any manipulation, is a square 79 meters on a side. The numbers relating to the pixel are the measure of reflected energy in the green, red and near infrared spectral bands.

Since surface features all have different reflective values, the data can be used to group like values in classes. Each group of like reflectance is referred to as a signature. The computer uses the signature to separate each pixel of the area into classes of landcover. This entire process is discussed more fully in the section on methodology.

Data for the Study

The data used in this study were LANDSAT images from April 1975 and April 1982. April is just prior to poppy harvest and when the plants should be in their most vigorous growth period. The only other data used were two topographic maps and a small scale schematic map of the country.

The LANDSAT images of the area displayed on a monitor show a rugged terrain with sparse vegetation and areas of exposed rock and bare soil. Some of the areas, the western facing slopes, are in deep shadow. The areas of cultivated agriculture are clearly identifiable from the bright red coloration.

The 1975 and 1982 images are exceptionally clear with no cloud cover or haze. A comparison of the two images show little change with one exception. The Indus river lying on the east of the study area was a narrow channel of water in 1975; by 1982 the level of water is substantially higher.

Both 1975 and 1982 images were classified and both registered to a topographic map. The images were then classified into landcover types and merged with information on slope and elevation.

In this section we report the results of the classification, examine the agricultural activities in the region and how they have changed, and as the scope of work required determined whether the changes indicate any pressure on marginal lands or cutting of the watershed.

Findings

The study area has over 50,000 hectares of land. Most of the area in both 1975 and 1982 was covered with native vegetation, exposed rock or bare soil. Agriculture has increased substantially in the period from 1975 to 1982. It makes up over 6000 hectares or 12 percent of the area in 1975 and over 8000 hectares or 16 percent of the area. in 1982. (Table 1)

The net increase of a third in land under cultivation represented a combination of land that had remained in agriculture, land that is newly cultivated and land that went out of cultivation. Over 40 percent of the land farmed in 1975 was still in production in 1982. The balance of agricultural activity in 1982 was on lands brought into cultivation since 1975. (Table 2)

Table 1--Landcover of Study Areas, 1975 and 1982

Landcover	1975 Area (Hectares)	1975 Percent of Total Area	1982 Area (Hectares)	1982 Percent of Total Area
Agriculture	6130	12	8153	16
Native Vegetation 1/	39549	78	36927	71
Sandy Areas/ Wetlands	4822	8	4890	10
Water	1113	2	1645	3
TOTAL AREA	51615 Hectares			

1/ Includes native vegetation, bare soil and rock areas of shadow.

NOTE: Throughtout this report the totals in the tables are not always consistent due to rounding in the computer.

Table 2--Land in Agriculture, 1975 - 1982

	1975	1982
Areas cultivated both years	2664	2664
Additions, 1975-1982		5488
Lost 1975-1982	3466	
TOTAL	6130	8152

Agriculture by Elevation

The study area varies in elevation from less than 1500 feet to over 6000 feet. There is very little land that is less than 1500 feet and most of that is river and wetlands. Agriculture occurs at all elevations and surprisingly in the 1975 period is relatively more intense at elevations above 4000 feet. The land at these higher deviations make up 30 percent of the land area and in 1975 accounted for 37 percent of the land in crops.

In 1982, the agriculture at higher elevations increased proportionately. Land over 4000 feet made up 43 percent of the land in cultivation. The land at over 5000 feet had a very high percentage of increase by over half from 1975 although the total amount of land was small, only 1500 hectares (Table 3).

Agriculture and Slope

The study area has little flat land available for cultivation. Only 28 percent is flat and of this nearly half is not arable due to the river and wet areas. Agriculture is independent of slope, with slopes above 40 percent comprising 8 percent of the total land area and 7 percent of the agriculture. (Table 4)

PRESSURE ON MARGINAL LAND

Agriculture at the highest elevation and on the steepest slope increased between 1975 and 1982. Total area over 6000 feet represents only 4 percent of the land area but has 7 percent of the agricultural activity. It is significant that this is the area with the highest percentage of increase in agriculture between 1975 and 1982.

Table 3--Agriculture by Elevations - 1975 - 1982

Elevation (feet)	Percent of total area	1975 Agriculture Hectares	1975 Percent	1982 Agriculture Hectares	1982 Percent
less than 1500	8	500	8	57	1
1500-2500	24	1036	17	1384	17
2500-3000	17	765	13	1252	15
3000-4000	22	1541	25	1965	24
4000-5000	18	1414	23	2031	25
5000-6000	8	576	9	881	11
6000	4	298	5	583	7
TOTAL		6130		8153	

Table 4--Agriculture by Slope

Slope (Percent)	Percent of Total Area	1975 Agriculture Hectares	1975 Percent	1982 Agriculture Hectares	1982 Percent
0 - 5	29	1453	24	1590	20
6 - 10	16	1194	19	1462	18
11 - 15	15	889	14	1371	17
16 - 25	17	1211	20	1696	17
26 - 40	15	1011	17	1428	17
Greater than 40	8	366	6	606	7
Total Agriculture		6130		8153	

Highly sloped land, (over 26 percent slope) at the highest elevation, (over 6000 feet) was also the area that had the least stable patterns of farming. Only 7 percent of the land in crops in 1975 was also cropped in 1982. As slope decreased and land became less sloped, agriculture tended to be more stable.

The shift in agriculture may be caused by the excessive erosion resulting from the combination of fragile land at high elevations that is steeply sloped. Whatever the reason it is clear that there is a pattern of shifting agriculture in the land with the most potential for erosion. To the extent possible, this should be protected from agricultural activity to avoid erosion and disruption to the watershed.

Discussion of the Results

It is clear from this examination of 1975 and 1982 LANDSAT data that there has been a significant change in agricultural activity between 1975 and 1982. Clearly the concerns expressed in the scope of work outlining the study, the cutting of the watershed and the pressure on marginal lands are justified. The increase in agricultural activity from 1975 to 1982 has been at all elevations and is of particular concern on the steep slopes. The overall change is not as significant as where the change is occurring. LANDSAT data not only provided area calculations, but also provided information on where the marginal lands are experiencing pressure.

The Question of Poppies

We have identified a number of distinct classes of cultivated agriculture. Each class varies somewhat in reflective characteristics but retains the predominant feature of high reflective values in the near infrared band. The variation in signature is a function of crop type, stage of development, orientation, vigor and cultural practices. There is no question that at least one of the signatures represents cultivated poppies. The problem is that without specific knowledge of areas of poppy cultivation to correlate with the LANDSAT signatures we have no basis for separating out poppies. The problem is similar to asking an untrained person to identify flower types. She/he could separate the flowers into groups, but could not provide the associated names.

Reliability of the Results

The accuracy and reliability of the results of the analysis require some qualifications. We are confident of the agricultural areas in the valleys and on the east facing slopes, but have some doubts about the accuracy of the data on the western slopes and are concerned with the reliability of the topography and slope maps. (See Methodology Section)

The valleys and the east facing slopes provide unhampered signatures that allow for a separation of agriculture, natural vegetation and other landcover classes. We are confident that the accuracy of the classification in these areas will approach 85 percent.

The classifications on the west facing slope must be considered significantly less reliable. Since the satellite records the data in the morning, the west facing slopes are in deep shadow. As a result, the difference between rapidly growing cultivated areas and native vegetation is less distinct. Classification of the west slopes should therefore be considered less accurate, perhaps only 50-75 percent.

We have little data to estimate the accuracy of the elevation and slope data. Using ground control points we have adjusted the LANDSAT data to the single available base map, but we find some lack of registration between the ridge lines, which can be readily identified on the LANDSAT data and the contours.

Future Improvements

There are developments that will improve the usefulness of satellite data in this project. LANDSAT 4 is scheduled to improve the resolution by nearly seven times. In addition, it will provide seven bands of data, rather than the present four useful bands.

The French are also launching a satellite, SPOT, scheduled to be operational in 1984. SPOT will provide resolution nine times better than LANDSAT 4 although only a single band of data will be provided at this fine resolution.

The improved resolution will allow detection of features such as dirt roads, paths and settlements, that we could not identify with the present data and no ground truth. The increased number of reflective bands will allow increased discrimination between crop types. We have included a sample printout of the LANDSAT 4 data showing the effect of the improved resolution on the ability to distinguish between features (Figure 1).

Figure 1

THE TECHNOLOGY AND ITS POTENTIAL

This section describes the potential of using this technology for identifying high-priority areas for development and for evaluating projects based upon their success in reducing poppy cultivation.

THE TECHNOLOGY: A DEFINITION

The technology is a combination of planning and analysis methodologies, computer systems, data and data relationships. The concept is simply to organize the data in such a manner that when appropriate questions are asked, appropriate answers are produced. The center of the technology is a Geographic Information System (G.I.S). This is the central management system that forms the frame work for both the data organization and analysis.

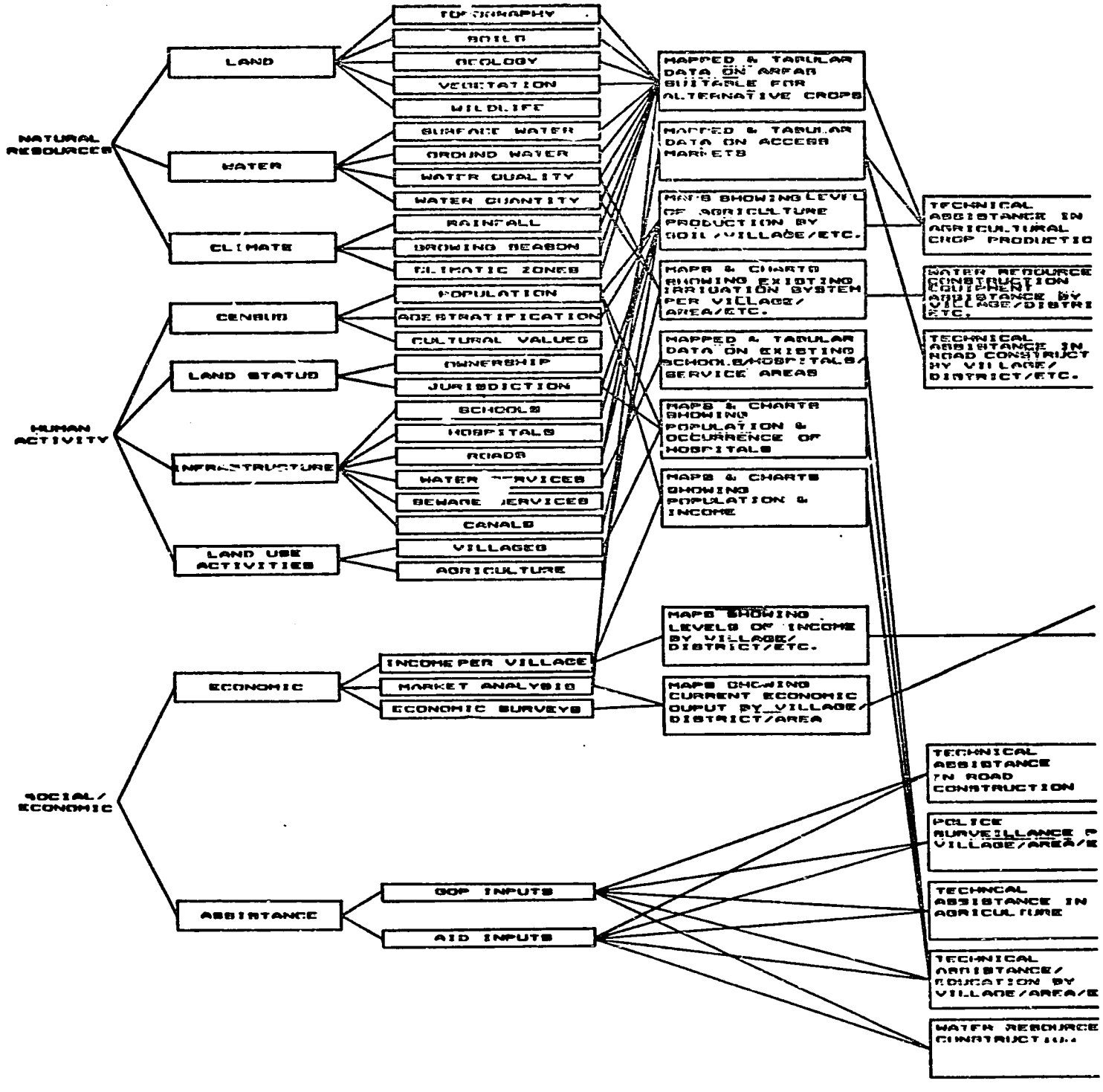
A major component evisioned for a G.I.S. for Pakistan is LANDSAT data. LANDSAT is a source of current and archived data that can provide a history of the spatial changes of a project area over time. A significant part of the proposed area development project will be the identification of poppies. Although this study has not identified those areas in poppy cultivation, such identification will become routine as improved satellite imagery becomes available and as subsequent analyses using known areas of poppy cultivation are performed.

POTENTIAL APPLICATION OF THE TECHNOLOGY

The technology used in the study can be used for project management, project monitoring and project evaluation. In this section we discuss specific applications of the technology. A chart, Figure 2, is a graphic representation of how a geographic information system can fit into the proposed are development project.

The heart of a G.I.S. is data management. In part, this project will be successful in direct relationship to the amount of data that can be collected, analyzed, and used to direct the project. Some data was developed for this study, some will be collected in project planning, and some will be developed as the project is implemented through monitoring and evaluation. Figure 2 is a suggested list of data that would be useful throughout the project. In the following sections we show how this dtata organized in a G.I.S. would assist in assuring project success.

The log-frame concept requires the formulation of assumptions that relate to the transformation of inputs to outputs; outputs to achievement of project purpose and project purpose to the development goal. The technology can be built into the set of verifiable indicators and can also be used in monitoring and evaluation of the project.


NATURAL RESOURCES
LAND
TOPOGRAPHY
SOILS
GEOLOGY
VEGETATION
WILDLIFE
WATER
SURFACE WATER
GROUND WATER
WATER QUALITY
WATER QUANTITY
CLIMATE
RAINFALL
GROWING SEASON
CLIMATIC ZONES
HUMAN ACTIVITY
CENSUS
POPULATION
AGE STRATIFICATION
CULTURAL VALUES
LAND STATUS
OWNERSHIP
JURISDICTION
INFRASTRUCTURE
SCHOOLS
HOSPITALS
ROADS
WATER SERVICES
SEWAGE SERVICES
CANALS
LAND USE ACTIVITIES
VILLAGES
AGRICULTURE
SOCIAL/ ECONOMIC
ECONOMIC
INCOME PER VILLAGE
MARKET ANALYSIS
ECONOMIC SURVEYS
ASSISTANCE
GOP INPUTS
AID INPUTS
MAPPED & TABULAR DATA ON AREAS SUITABLE FOR ALTERNATIVE CROPS
MAPPED & TABULAR DATA ON ACCESS MARKETS
MAPS SHOWING LEVEL OF AGRICULTURE PRODUCTION BY SOIL/VILLAGE/ETC.
MAPS & CHARTS SHOWING EXISTING IRRIGATION SYSTEM PER VILLAGE/ AREA/ETC.
MAPPED & TABULAR DATA ON EXISTING SCHOOLS/HOSPITALS/ SERVICE AREAS
MAPS & CHARTS SHOWING POPULATION & OCCURRENCE OF HOSPITALS
MAPS & CHARTS SHOWING POPULATION & INCOME
MAPS SHOWING LEVELS OF INCOME BY VILLAGE/ DISTRICT/ETC.
MAPS SHOWING CURRENT ECONOMIC OUPUT BY VILLAGE/ DISTRICT/AREA
TECHNICAL ASSISTANCE IN AGRICULTURAL CROP PRODUCTIO
WATER RESOURCE CONSTRUCTION EQUIPMENT ASSISTANCE BY VILLAGE/DISTRI ETC.
TECHNICAL ASSISTANCE IN ROAD CONSTRUCT BY VILLAGE/ DISTRICT/ETC.
TECHNICAL ASSISTANCE IN ROAD CONSTRUCTION
POLICE SURVEILLANCE P VILLAGE/AREA/E
TECHNICAL ASSISTANCE IN AGRICULTURE
TECHNICAL ASSISTANCE/ EDUCATION BY VILLAGE/AREA/E
WATER RESOURCE CONSTRUCTION
POTENTIAL DATA FILES FOR A GEOGRAPHIC INFORMATION SYSTEM
PROJECT DEFINITION (PURPOSE)
PROJECT INPUT

We see five areas where the existence of a G.I.S. would be useful:

- Project definition (purpose)
- Project inputs
- Project monitoring
- Project performance (on-going evaluation)
- Project evaluation (final)

Project Definition

The technology can provide an important input into project definition and evaluate the reasonability of certain goals. A base line of information is required in order to identify high-priority areas. Many examples are illustrated in Figure 2. For example, before alternative crops can be recommended as a substitute to poppy cultivation, questions such as the following should be addressed:

1. What is the inherent capacity of the land?
2. What is the land capacity in relationship to access to markets?
3. What areas are suitable for certain crops? Are they near existing poppy areas?
4. What are the limiting factors for alternative crops:
   - water
   - soil
   - access to markets - roads
   - slope
   - etc.
5. Do existing economic/demographic conditions support proposed agricultural changes, i.e., is the population base large enough to support specific production

6. Do irrigation systems exist for new crop production? The technology can assist the project manager in developing a series of analyses that can identify high-priority areas as well as assist in determining limiting factors that inhibit the success of the project. But perhaps even more important, the technology can quickly help project managers to visualize the relationship between various data at the initial phases of the project.

Project Inputs

Based upon the project definition, the project inputs could also be organized in a G.I.S. The AID and GOP inputs could be entered into the system as they are implemented. Project inputs could then be analyzed in relationship to the actual monitoring. The analysis in the project definition will assist the AID and the GOP in directing inputs. In addition, the information in the project monitoring will illustrate how well the implementation is being carried out. Figure 2 illustrates some potential inputs and their relationship within the G.I.S.

Project Monitoring

An important asset of the technology is the ability to accept and store information about the project area once the project has begun. The technology will provide an organized place to store this information in a similar manner as the information on inputs is stored. The stored information will then be readily available for rapid retrieval, update, etc. The project manager need only to ask questions

etc. The project manager need only to ask questions relating to the level of inputs and the data produced by field survey, census or LANDSAT data, etc. For example, AID inputs relating to new irrigation systems could be monitored by using LANDSAT data to determine shifts in cultivated areas or shifts in cropping patterns. Another example might be to cross tabulate and map the amount of technical assistance to education by village. Figure 2 illustrates some examples of potential monitoring applications.

Project Performance

Project performance is an analysis of the purpose, inputs and monitoring data in order to evaluate the effectiveness of any one or a combination of inputs. The technology would allow the tracking of inputs over a series of monitoring phases in relationship to the inputs by phase. For example, the LANDSAT data could be used on a seasonal basis to monitor the extent and type of crops. In addition, this information could be modified based upon the integration of rainfall and temperature data during this period in order to provide an accurate evaluation adjusted to yearly fluctuations in climate.

The project manager would be able to correlate the amount of agricultural technical assistance per village or area with the increase/decrease in cropping practices. This would provide important uniform information on the effectiveness of the inputs on whether course corrections are needed.

Project Evaluation

The technology will provide a long term data base in which the project could be evaluated. A time frame analysis could be employed looking at the inputs and a summary of the monitoring. This would provide at a minimum, two outputs. The first is the evaluation of the original project goals in relationship to the final results and a second, a sophisticated project tracking mechanism which would allow better pre-project and on-going evaluation of potential projects with similar characteristics. This would provide important data that could be tabulated and displayed spatially with statistical summary supportive data showing AID projects and results.

Requirements for the Technology

Successful technology transfer requires equipment, training of operators for the equipment and an orientation of all personnel on the opportunities and limitations of the technology. There are a number of small micro-based computer systems that would be suitable for the Pakistan project. The system should be able to classify LANDSAT data and also be the center of a G.I.S. The system should be able to read LANDSAT computer tapes and read and write floppy disks. It should also be interfaced with a digitizing table for entering base map data.

Costs of systems vary widely. Price range for such a system is from $100,000 to $200,000.

Training

Training is always an important factor in technology transfer. The technology is simple to operate and in general, two months of training would allow someone to utilize most of the systems' capability. However, there are two components to the training related to technology transfer. The first is to train the operator on the concept of LANDSAT data and Geographic Information Systems. This is necessary before the system can be used efficiently. This requires at least 15 months orientation, including having a trained person working along side, pointing out opportunities and pitfalls. An additional 9 months would be required to make an operator fully qualified to realize the full potential of the system.

METHODOLOGY

The methodology used in this study is a guide to a LANDSAT study incorporating other spatial data on a computer based Geographic Information System (G.I.S.). There are six important topics to consider in formulating a G.I.S.

- Project definition
- Data types and sources
- Base map/spatial coordination system
- Data files
- Analysis
- Results

An overall framework is illustrated graphically in Figure 3.

Project Definition

The scope of work for the study required the following:

Identify Cultivated Crops

Identify through the use of digital satellite imagery cultivated crops (agriculture) in the study area. This identification is to include present growth patterns (1982) and changes that have occurred over the last decade (the years 1975-1982 were used based upon the available data). Also, the study was to assess the effect of changes in watershed and on marginal lands.

Poppy Identification

Where possible, identify areas of poppy cultivation.

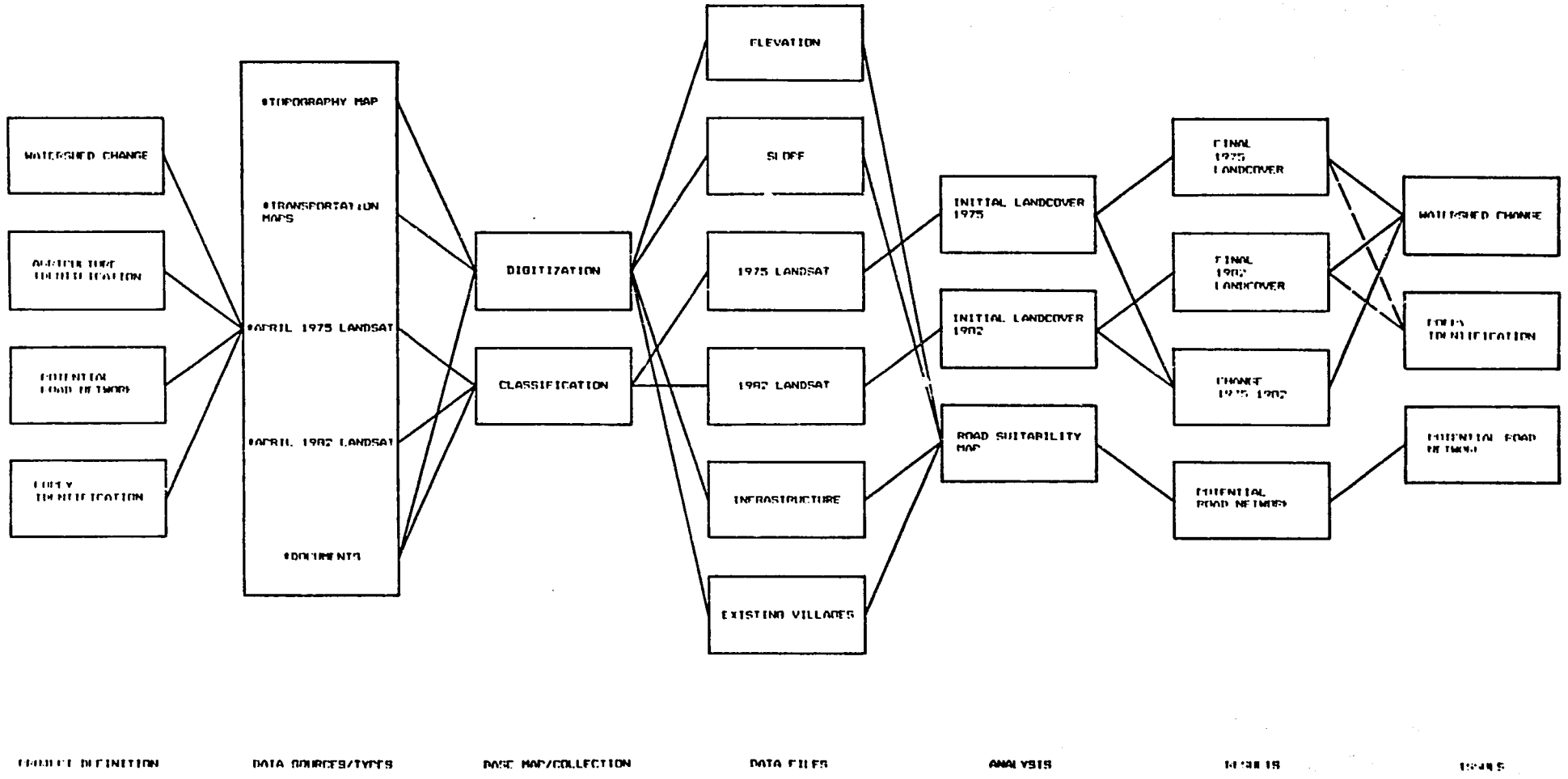

WATERSHED CHANGE
POTENTIAL ROAD NETWORK
#TOPOGRAPHY MAP
#TRANSPORTATION MAPS
#APRIL 1975 LANDSAT
#APRIL 1982 LANDSAT
#DOCUMENTS
DIGITIZATION
CLASSIFICATION
ELEVATION
SLOPE
1975 LANDSAT
1982 LANDSAT
INFRASTRUCTURE
EXISTING VILLAGES
INITIAL LANDCOVER 1975
INITIAL LANDCOVER 1982
ROAD SUITABILITY MAP
FINAL 1975 LANDCOVER
FINAL 1982 LANDCOVER
CHANGE 1975-1982
POTENTIAL ROAD NETWORK
WATERSHED CHANGE
DATA DEFINITION
DATA SOURCES/TYPES
BASE MAP/COLLECTION
DATA FILES
ANALYSIS

Figure 3

Terrain

Prepare maps showing slope, elevation and topographic contour lines. As part of the study we included an analysis of the association between agricultural production in relation to elevation and slope.

Population Settlements/Transportation Systems

The study was to identify and map the population settlements in the study area and the existing networks linking those settlements. Based upon this information and the terrain/landcover data, we were to propose road networks linking major population centers.

In addition, while not part of the methodology, the study was to demonstrate the feasibility of the technology for the proposed area development project.

Data Types and Sources

Data for the region of the study are scarce. We could not locate any air photographs. Our only sources of data were the LANDSAT images and the following Defense Mapping Agency (DMA) topographic maps.

- Tactical Pilotge Chart Map 1:500000 scale, Defense Mapping Agency, Aerospace Center, St. Louis Air Force Station, Missouri, USA.
- Joint Operations Graphic (Ground) Map, 1:250000 scale, Defense Intelligence Agency, Aeronautical Chart and Information Center, U.S. Air Force, St. Louis, Missouri, USA. (used for verification)

The LANDSAT digital data tapes were scenes from April 4, 1975 and April 15, 1982. Both data tapes were of the multi-spectral scanner data which has a ground resolution of 79 x 79 meters. The 1975 tape was resampled to 57 x 79 meters; the 1982 tape to 57 x 57 meters. We used a 57 x 57 meter grid as the resolution for all data storage and analysis. The 1975 LANDSAT data were resampled to this grid size and all other data was entered at this scale.

Computer-Compatible Data Inventory

One of the objectives of the study was to demonstrate the feasibility of the technology in project preparation and later as part of the project itself.

The technology requires a computer system that can store data referenced by a x, y coordinate system, can analyze LANDSAT data and register it to the selected x, y coordinate base map, can do sophisticated spatial analyses and finally, can map the results to scale. We used an ERDAS microcomputer system developed by Earth Resources Data Analysis Systems, Atlanta, Georgia. The system is efficient for handling both small and large study areas. For example, within the Gadoon study area there were approximately 158,865 57 x 57 meter cells, but the system is capable of handling much larger areas. With peripheral data storage and a larger computer, there is no limit to the area that can be stored and analyzed.

Geographic Information Systems

The Gadoon study required the establishment of a small spatially coordinated data base, usually referred to as a Geographic Information System (G.I.S.). The system that we established used a 57 meter square resolution, the cell size of the 1982 LANDSAT data.

Spatial Resolution

The choice of cell size used to collect and record the initial data is critical because aggregation is always possible, where disaggregation is impossible. In deciding on the scale, three factors should be considered (1) the scale of the original data, (2) the use of the data and the subsequent analyses, and (3) the time and cost involved in collecting, coding and manipulating the data.

In general, where there is an uncertainty about cell size, data should be collected and entered at the finest resolution. Once the data are entered they can be readily aggregated while disaggregating them meaningfully is impossible. Analyses requiring finer resolution than has been provided, will require data to be recoded and re-entered. If the scale at which the data were collected is too coarse, the data must be re-collected.

The second factor, scale in relation to use of the data, varies depending on the types of analysis or planning being employed. The data set with too much detail will be costly to enter, process and update. Data with too little detail will not support the desired analyses.

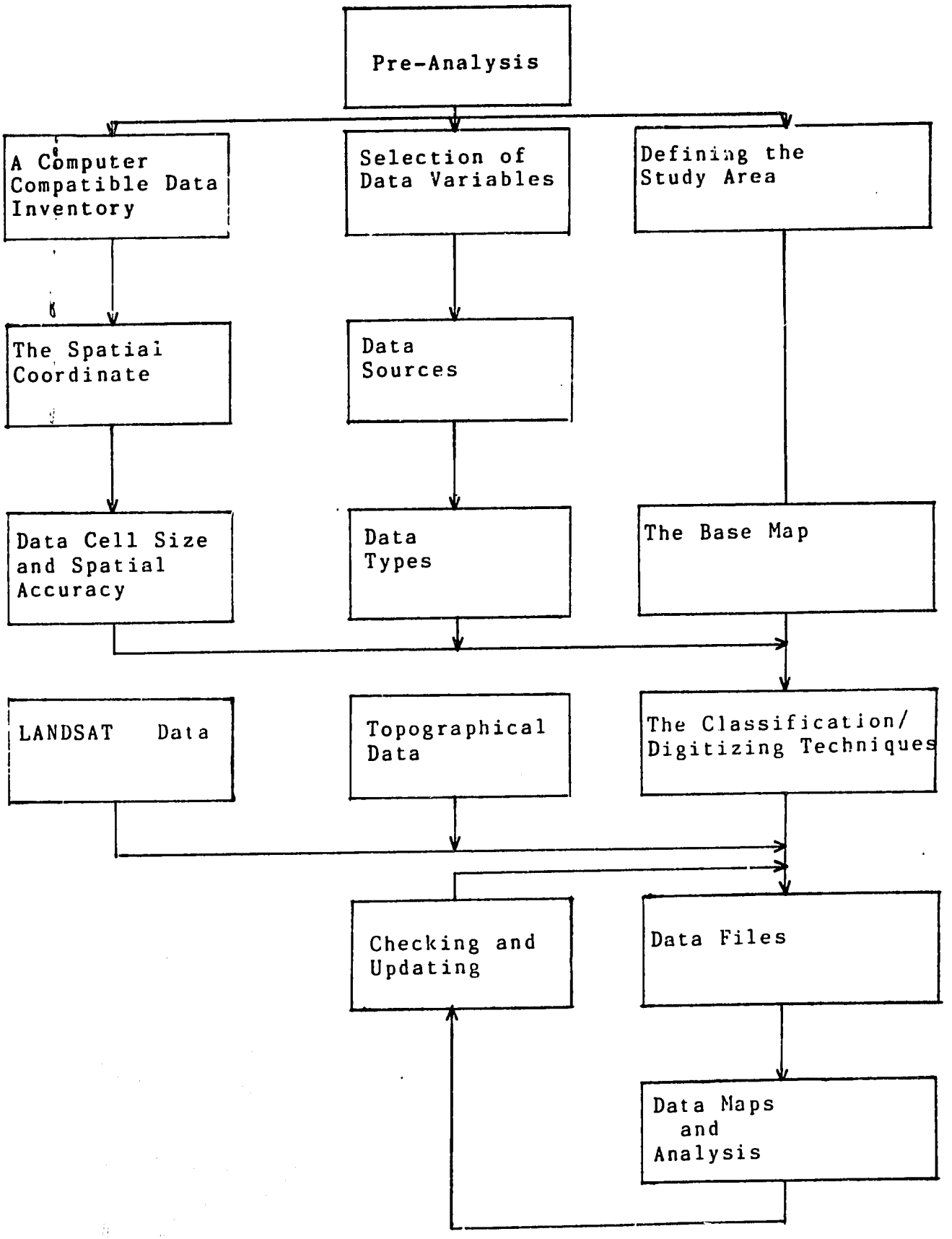
The time and costs of collecting and manipulating data is dependent on both cell size and the types of analyses. The finer the resolution, the more coding is required to enter the data, more computer storage is required and longer computer runs to do the analyses become necessary.

Base Map/Spatial Coordinate System

The preparation of a base map with the physical data on elevation and slope and the location of the details of landcover required three steps: preparation of the base map, digitizing the slope and elevation and entering the LANDSAT data.

A base map was prepared at 1:63560 (1"=mile). The map was derived from a 1:500000 scale, DMA Tactical Pilotage Chart topographic map. The DMA map was used to register the two LANDSAT scenes. Our base map was scaled to fit the 57x57 meter LANDSAT data. (Figure 4 illustrates the data collection procedure).

Figure 4


Pre-Analysis
A Computer Compatible Data Inventory
Selection of Data Variables
Defining the Study Area
The Spatial Coordinate
Data Sources
Data Cell Size and Spatial Accuracy
Data Types
The Base Map
LANDSAT Data
Topographical Data
The Classification/ Digitizing Techniques
Checking and Updating
Data Files
Data Maps and Analysis

A table digitizer was used for entering elevation and slope data. Both data sets were digitized using a latitude-longitude coordinate system.

The two LANDSAT scenes required an extensive amount of classification. The procedures used include a supervised approach using a maximum likelihood classification and unsupervised clustering techniques. The concepts behind the classification will be discussed in greater detail in the section on data analysis.

Data Files

The section describes each data variable used in the study.

Topographic Elevation (Figure 5)

- Data variable name. Topographic elevation in feet.
- Source - Tactical Pilotage Chart Map, Defense Mapping Agency, Aerospace Center, St. Louis Air Force Station, Missouri, USA.
- Scale of original map - 1:500000
- Date of original map - February, 1981
- Definition of data variable - The elevation map depicts the mean standard elevation above sea level in feet, of the ground area covered in the study.
- Major assumptions - The topographic elevation in feet data variable is assumed to be as accurate as the source material from which it was derived. The topographic contour lines were interpreted and grouped into the categories with area counts as shown:

Figure 5

TOPOGRAPHIC ELEVATION IN FEET

- LESS THAN 1500 FEET
- 1500-2500 FEET
- 2500-3000 FEET
- 3000-4000 FEET
- 4000-5000 FEET
- 5000-6000 FEET
- OVER 6000 FEET

Description	Number of Points	Number of Hectares	Percentage of Study Area
Less than 1500 feet	12462.	4048.935	8.11
1500 - 2500 feet	37584.	12211.039	23.71
2500 - 3000 feet	27051.	8788.868	17.07
3000 - 4000 feet	34367.	11165.836	21.08
4000 - 5000 feet	27852.	9049.113	17.57
5000 - 6000 feet	12458.	4047.604	7.86
Over 6000 feet	6713.	2181.053	4.24

Total non-zero points = 158,487
Total non-zero hectares = 51,492.41
Percentages are based on non-zero points

Slope (Figure 6)

- Data variable name - Slope by percent

- Source - Tactical Pilotage Chart Map, Defense Mapping Agency, Aerospace Center, St. Louis Station, Missouri, USA.

- Scale of original map - 1:500000

- Date of original map - February, 1981

- Definition of Data Variable - The slope by percent map depicts the relative flatness or steepness of the study area terrain. That flatness or steepness is described on terms of percentages where 0 percent is equal to flat conditions and 180 percent is equal to a vertical condition, perpendicular to the horizontal ground surface.

- Major assumptions - The slope by percent data variable is assumed to be as accurate as the source material (Tactical Pilotage Chart Map) from which it was derived. The topographic contour lines were interpreted and grouped into the following categories and area counts as follows:

Figure 6

SLOPE BY PERCENT

0-5%
6-10%
11-15%
16-25%
26-40%
40% AND OVER

Description	Number of points	Number of Hectares	Percentage of Study Area
0 - 5 percent	45811.	14883.992	28.91
6 - 10 percent	25081.	8150.440	15.83
11 - 15 percent	23372.	7593.562	14.75
16 - 25 pecent	27679.	8992.906	17.46
26 - 40 percent	23848.	7746.590	15.04
41 percent or greater	12696.	4124.930	8.01

Total non-zero points = 158487
Total non-zero hectares = 51492.41

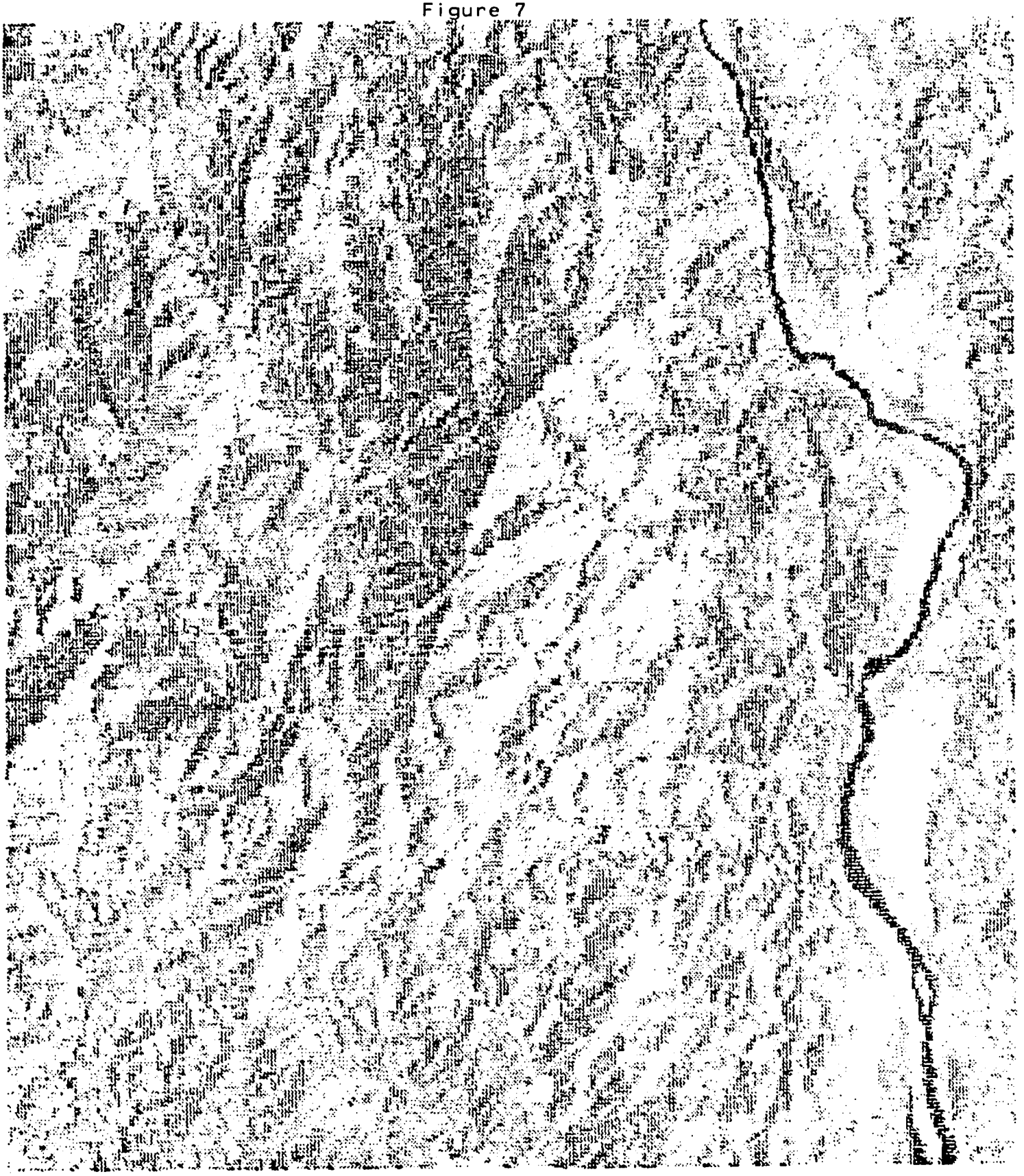
Raw 1975 LANDSAT (Figure 7)

- Data variable name - Raw 1975 LANDSAT
- Source - LANDSAT digital multi-spectral scanner (MSS) data tapes, 9-track tapes, 1600 Bpi.
- Scale of original source - 1:1000000 space oblique, mercator projection.
- Data of original source - April 4, 1975
- Definition of data variable - The LANDSAT data is in "raw" form, meaning it has not been classified into defineable land cover classes. The data contains digital reflectance values of the ground surface as measured through the LANDSAT multi-spectral scanner.
- Major assumptions - The LANDSAT data is assumed to be as accurate as the original digital imagery. The raw data resolution is 57 meters x 79 meters.
- Data categories - The raw digital LANDSAT data was received in its original condition in 4 spectral bands, covering the red, green, and two near-infrared bands.

Raw 1982 LANDSAT (Figure 8)

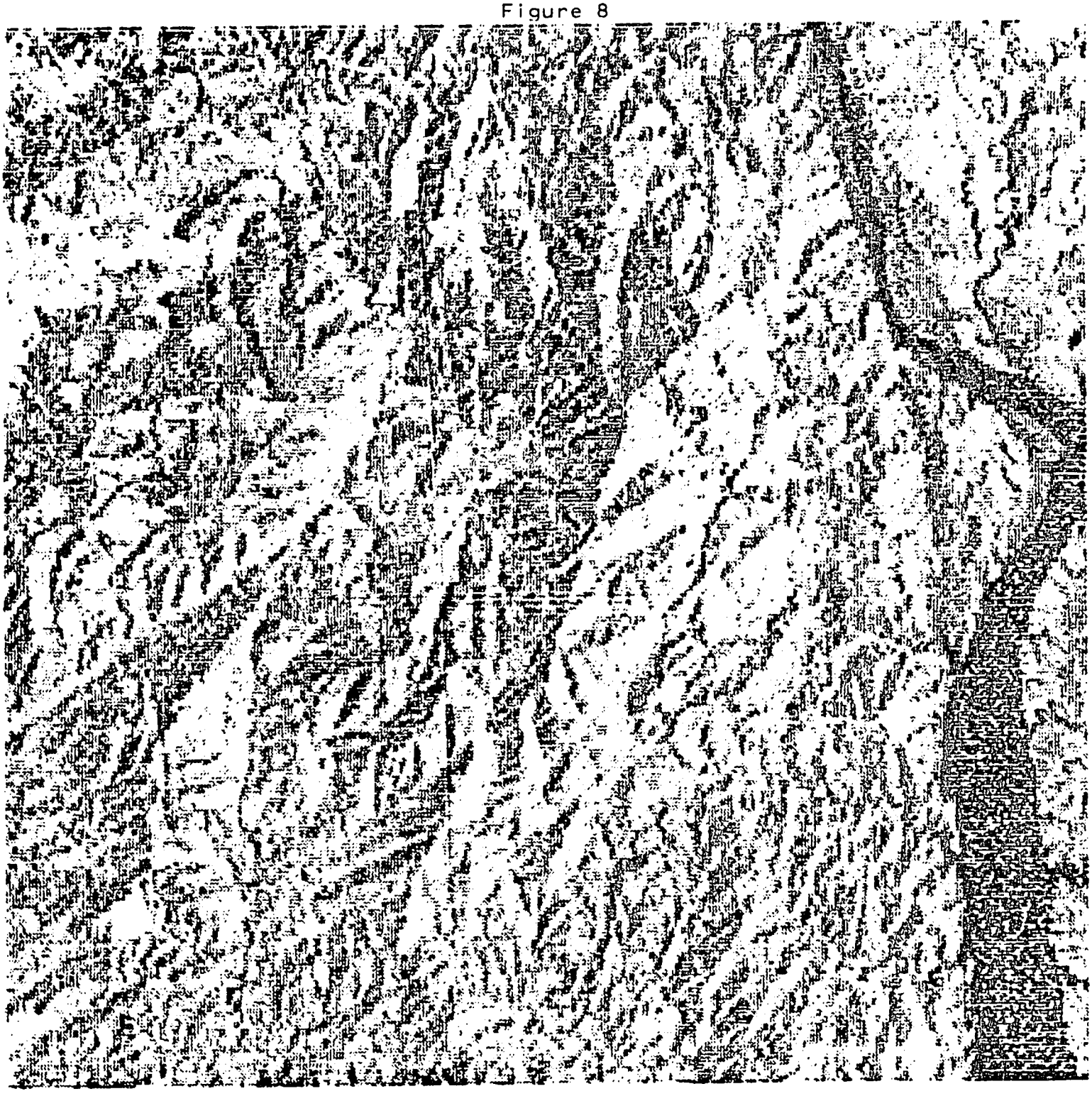
- Data variable names - Raw 1982 LANDSAT
- Source - LANDSAT digital multi-spectral scanner (MSS) data, 9-track tape, 1600 Bpi.
- Scale of original source - 1:1000000 space oblique, mercator projection

Figure 7



RAW 1975 LANDSAT DATA

Figure 8



RAW 1982 LANDSAT DATA

- Date of original source - April 15, 1982

- Definition of data variable - The LANDSAT data is in "raw" form, meaning it has not been classified into defineable land cover classes. The data contains digital reflectance values of the ground surface as measured through the LANDSAT multi-spectral scanner.

- Major assumptions - The LANDSAT data is assumed to be as accurate as the original digital imagery. The raw data resolution is 57 meters x 57 meters.

- Data categories - The raw digital LANDSAT data was received in its original condition in 4 spectral bands covering red, green and two near infrared bands.

Base Map Used in Final Map Production

A base map was developed for the project to be used in the final color map production process. This base map was derived from the interpretation of three major maps:

- Tactical Pilotage Chart Map (see description of topographic elevation map above for specifications)

- Joint Operations Graphic (Ground) Map 1/.

- Pakistan Burmah Shell Northwest Frontier Province Map 2/.

1/ Joint Operations Graphic (Ground) Map, Defense Intelligence Agency, Aeronautical Chart and Information Center, U.S. Air Force, St. Louis, Missouri, USA. Original map scale - 1:250000. Original date - December, 1966.

2/ Pakistan Burmah Shell Northwest Province Map, completed by Pakistan Burmah Shell, with permission of the Surveyor General of Pakistan; printed by Elite Publishers Limited Karachi. Original map scale - one inch equals approximately 12 miles. Original date - 1979.

The data derived from the interpretation was subsequently used to develop the base map and contains the following information:

- Topographic elevation contour lines (in feet)
- Major water courses
- Population areas
- Roadways
- Trails and cart tracks
- Latitude/longitude coordinates

Assumption

The base map is assumed to be as accurate as the original maps from which it was derived.

Analysis

The primary task of the analysis was to develop significant landcover classes from LANDSAT data and attempt to determine areas of agriculture and if possible areas of poppy cultivation. A second task was to prepare a map showing suitability for roads using data on the elevation, slope, land cover and existing transportation.

Classifying the LANDSAT Data

Concept:

The LANDSAT satellites are equipped with a multiple spectral scanner (MSS). This device is designed to collect data from the frequency bands representing red, green, and two near infrared colors. The data is actually a record of the amount of energy reflected from these four segments or channels of the electromagnetic spectrum.

LANDSAT'S level of spatial detection is approximately 1.1 acre/per unit. These units are called pixels (picture elements), and for a 5.5-acre parcel of land, LANDSAT provides five pixels, each containing four pieces of information: a value for each relative reflectance in each of the four channels. See Figure 9.

The overall response range of the LANDSAT MSS electromagnetic spectrum is between .5 to 1.1 micrometers (MM):

	Frequency (MM)	
Channel 1	.5-.6	Green
Channel 2	.6-.7	Red
Channel 3	.7-.8	Near Infra-Red
Channel 4	.8-.11	Near Infra-Red

The multi-spectral scanner has the ability to distinguish unique cover types in one, two, three, or all four of the channels. For example, research has shown that a single near infrared channel alone discriminates between many types of vegetation. Using only the single band analysts can usually identify agriculture and forest. The ability to discriminate increases when the four present channels are used and will be further improved with the seven bands available in LANDSAT 4.

Figure 9--Typical Results From LANDSAT for a 5.5-Acre Field of Agriculture (five pixels)

	Channel 1	Channel 2	Channel 3	Channel 4
	micrometers .5-.6	micrometers .6-.7	micrometers .7-.8	micrometers .8-1.1
	green relative value	red relative value	near infrared relative value	near infrared relative value
1.	20	20	39	16
2.	25	14	30	23
3.	31	27	26	17
4.	24	19	33	22
5.	29	21	29	18

The goal of classification is to determine a range of values in one, two, three or four of the channels that best describes a specific landcover type. The problem requires sophisticated techniques because values are not absolute. Reflectance values differ because of atmospheric and temporal conditions. For example, reflectance values of a field just after a rain storm might be different than the reflectance value from the same field after a long, dry period. Also, seasonal and yearly changes, clouds, fog, and haze cause enough variation in the scene that the analyst must analyze each scene individually, rather than rely on standard values.

Two types of procedures are used to identify the relationship of what is known on the ground and what the values are for each pixel from LANDSAT. The goal is to determine which range of data values in each of the four data channels represent a certain landcover type. For example, agriculture on the average may have a range of values of 20-31 in channel 1, 14-27 in channel 2, 26-39 in channel 3, and 15-23 in channel 4.

There are two methods used in classification, supervised and unsupervised. In the supervised classification, an area of the pixels known to be landcover are identified and the values are recorded. These values are used to classify all other like pixels. The second, unsupervised classification, is a statistical procedure which groups the data by

natural breaks in the values for each channel. This method produces a map of geographical areas containing similar reflectance values. The analyst then must identify these areas using known or assumed spatial patterns.

This study employed both techniques. Because we had no knowledge of the actual conditions, the unsupervised clustering technique proved to be the most useful. The supervised method was used for special cases such as large shadow areas.

Figure 10 is a histogram of the values from channel 4 (infrared) in the 1982 LANDSAT scene. The high frequent counts represent the very healthy vegetation, while the peak at the low end of the histogram represents the river or water (since water absorbs most of the infrared energy).

One major problem in the study was caused by the terrain and the resultant shadows. Since LANDSAT passes over the area about 9:30 a.m., the shadows are extensive. Because of the high relief, the Gadoon study area was particularly troublesome. In some cases it was impossible to determine a landcover in the shadow using the unsupervised classification. For these areas we use the supervised maximum likelihood method. Areas in the LANDSAT data that were in the shadow area were magnified and training samples were taken. The resultant classified data was then compared against the unsupervised class. The problem is illustrated in Figure 10 where there seems to be a

Figure 10

Histogram listing for file RAW 1992

This image has 480 ROWS , 500 COLUMNS

This is a LANDSAT data file
There are 4 bands in this data set
This is for BAND no. 4

Minimum data value = 1
Maximum data value = 69

Mean value = 39.541
Median = 41.000
Mode = 44.000
Standard Deviation = 10.61943

DATA VALUE	NUMBER OF POINTS	%	
1	7	.01	
2	29	.05	
3	6	.01	
4	6931	1.16	
5	3556	.59	
6	3368	.56	
7	1480	.25	
8	511	.08	
9	238	.04	
10	311	.05	
11	437	.07	
12	660	.11	
13	641	.11	
14	936	.16	
15	947	.16	
16	1339	.22	
17	1383	.23	
18	1681	.28	
19	2100	.35	
20	1627	.27	
21	1655	.27	
22	1644	.27	
23	3041	.51	
24	3065	.51	
25	3864	.64	
26	4667	.77	
27	6930	1.15	
28	7330	1.22	
29	9800	1.63	
30	10255	1.71	
31	11314	1.89	
32	11800	1.97	
33	12668	2.11	
34	13760	2.29	
35	14944	2.49	
36	15600	2.60	
37	16400	2.73	
38	16800	2.80	
39	17400	2.90	
40	17900	2.98	
41	18400	3.07	
42	18800	3.13	
43	19400	3.23	
44	19800	3.30	
45	20000	3.33	
46	20400	3.40	
47	20800	3.47	
48	21200	3.53	
49	21600	3.60	
50	21800	3.63	
51	22000	3.67	
52	22400	3.73	
53	22800	3.80	
54	23200	3.87	
55	23600	3.93	
56	24400	4.07	
57	24800	4.13	
58	25200	4.20	
59	25600	4.27	
60	26000	4.33	
61	26400	4.40	
62	26800	4.47	
63	27200	4.53	
64	27600	4.60	

significant range of frequencies that have a lower value than the agriculture on the east facing slopes. The only way to really determine the agriculture present in the shadows would be to emplr y the use of ground truth or airphotos. However, statistically significant classes were represented in the shadow and based upon patterns of agriculture in non-shadow areas, we found some agricultural classes in the shadow areas.

Figures 11 and 12 are the histograms of the 27 classes produced by the unsupervised clustering algorithm. These classes were then grouped into the 5 major classes described in the results section. These classes were displayed over the raw LANDSAT data in order to group and then finally name them, based upon their spatial patterns. Figures 13 and 14 are the computer generated black and white maps showing the relationship of the 27 classes.

Suitability for Potential Road Networks

In deriving a suitability index the analyst must synthesize a number of variables. The relationship between the variables must be clearly understood. Where there is a theoretical basis for the relationships between the variables, the analyst can be assured of the results. If no theory exists then the relationship must be hypothesized as we did in deriving the suitability index for roads. In this case, the relationships were simple, it is more difficult to build roads over rivers, than up steep slopes. Also roads

Figure 11

Histogram listing for CLASSIFIED 1975

This image has 435 ROWS , 352 COLUMNS

This is a LANDSAT data file
Minimum data value = 0
Maximum data value = 25

Mean value = 10.117
Median = 9.000
Mode = 4.000
Standard Deviation = 6.87066

DATA VALUE	NUMBER OF POINTS	%	
0	9.	.05	%I
1	1316.	7.44	%IXXXXXXXXXXXXXXXXXXXXXXXXXXXX
2	653.	3.71	%IXXXXXXXXXXXX
3	1445.	8.17	%IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
4	1930.	10.91	%IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
5	1641.	9.28	%IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
6	1062.	6.00	%IXXXXXXXXXXXXXXXXXXXXXX
7	96.	.54	%I
8	111.	.63	%IXX
9	1140.	6.44	%IXXXXXXXXXXXXXXXXXXXXXXXX
10	103.	.58	%IXXX
11	211.	1.19	%IXXXX
12	1674.	9.46	%IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
13	1621.	9.16	%IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
14	206.	1.16	%IXXXX
15	381.	2.14	%IXXXXXXXX
16	780.	4.39	%IXXXXXXXXXXXXXXXXX
17	1224.	6.92	%IXXXXXXXXXXXXXXXXXXXXXXXXXX
18	97.	.55	%IXX
19	1[illegible]	2.1[illegible]	%IXXXXXXXX
20	1[illegible]	3.16	%IXXXXXXXXXXXX
21	78.	.44	%IX
22	85.	.48	%IX
23	1[illegible]	.74	%IXX
24	4[illegible]	.24	%I
25	645.	3.64	%IXXXXXXXXXXXXXX
26	339.	1.91	%IXXXXXXX

Figure 12

```
   Histogram listing for file CLASSIFIED 1982

This image has     489 ROWS ,     362 COLUMNS

This is a LANDSAT data file
Minimum data value =      0
Maximum data value =     27

Mean value          =      10.218
Median              =       9.000
Mode                =       8.000
Standard Deviation =     7.00418


DATA VALUE  NUMBER OF POINTS    %
==========  ================  ====== ========================================
     0                  55.     .33 %IX
     1                1049.    5.90 %IXXXXXXXXXXXXXXXXXXXXXXX
     2                 152.     .85 %IXXX
     3                1551.    8.72 %IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     4                1474.    8.29 %IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     5                1377.    7.74 %IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     6                1137.    6.39 %IXXXXXXXXXXXXXXXXXXXXXXXXXX
     7                 411.    2.31 %IXXXXXXXXX
     8                1435.    8.07 %IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     9                1751.    9.84 %IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    10                 942.    5.30 %IXXXXXXXXXXXXXXXXXXXXX
    11                1195.    6.72 %IXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    12                 187.    1.05 %IXXXX
    13                 123.     .69 %IXX
    14                 412.    2.32 %IXXXXXXXXX
    15                 628.    3.54 %IXXXXXXXXXXXXXX
    16                 105.     .59 %IXX
    17                 157.     .88 %IXXX
    18                 184.    1.03 %IXXXX
    19                 383.    2.15 %IXXXXXXXX
    20                 807.    4.54 %IXXXXXXXXXXXXXXXXXX
    21                 420.    2.36 %IXXXXXXXXX
    22                  83.     .47 %IX
    23                 872.    4.90 %IXXXXXXXXXXXXXXXXXXX
    24                 273.    1.53 %IXXXXXX
    25                 343.    1.93 %IXXXXXXX
    26                  30.     .17 %I
    27                 246.    1.38 %IXXXXX
```

Figure 13

UNSUPERVISED CLASSIFICATION-1975

Figure 14

UNSUPERVISED CLASSIFICATION-1982

at the highest elevations would serve fewer people, while those at the lowest levels or in wet areas could be subject to flooding. All these conditions were ranked low on the analysis. Other conditions were ranked higher.

The road suitability analysis was accomplished by using 3 variables. These included slope, elevation and 1982 land-cover. Most favorable conditions were assigned higher values.

The model is as follows:

Slope	Assigned Ranking
0 - 5%	6
6 - 10%	5
11 - 15%	3
16 - 25%	2
26 - 40%	1
40% or greater	0

Elevation	
less than 1500 feet	3
1501 - 2500	6
2501 - 3000	6
3001 - 4000	6
4001 - 5000	2
5001 - 6000	1
6000 feet or greater	1

1982 Landcover	Assigned Ranking
water/river	1
sandy/wet areas	1
agriculture	6
native vegetation/bare	6

Figure 15 illustrates the frequency of the model after it has been reduced to 5 levels. The lowest levels are those that are less suitable for roads while the high numbers illustrate the most suitable areas. Figure 16 is a computer map showing the relative shades of gray for suitability (lighter is the best).

Results

The analyses described above lead to the resultant maps. These included 1982 landcover, 1975 landcover, change between 1975-1982 and potential road network.

1982 and 1975 Landcover Maps

The final 1982 landcover map was generated from the cluster analyses employed during classification. Figure 17 illustrates how the original clusters were grouped to produce the final maps with the 5 classes. Figure 18 and 19 are computer printouts showing these 1975 and 1982 scenes with the 5 classes.

Figure 15

Histogram listing for file SUITABILITY FOR ROADS

This image has 439 ROWS , 362 COLUMNS

This is a LANDSAT data file
Minimum data value = 0
Maximum data value = 5

Mean value = 3.114
Median = 3.000
Mode = 3.000
Standard Deviation = .94315

DATA VALUE	NUMBER OF POINTS	%	
0	102.	.06	I
1	12130.	7.55	IXXXXXXX
2	22650.	14.09	IXXXXXXXXXXXXXX
3	61765.	38.42	IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
4	62279.	37.74	IXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
5	3038.	1.91	IX

Figure 16

SUITABILITY FOR ROADWAYS

SUITABLE

MODERATE

UNSUITABLE

WATER/RIVER

POTENTIAL ROAD NETWORK

Figure 17

1975 landcover		1982 landcover	
Original cluster	Final class	Original cluster	final class
1	agriculture	1	shadow
2	agriculture	2	shadow
3	native	3	native
4	native	4	native
5	shadow	5	native
6	shadow	6	sandy areas
7	shadow	7	water
8	native	8	sandy areas
9	native	9	native
10	native	10	agriculture
11	native	11	water
12	sandy areas	12	native
13	water	13	native
14	agriculture	14	sandy areas
15	sandy areas	15	agriculture
16	sandy areas	16	agriculture
17	sandy areas	17	native
18	water	18	sandy areas
19	agriculture	19	agriculture
20	agriculture	20	agriculture
21	native	21	sandy areas
22	native	22	water
23	native	23	native
24	native	24	native
25	water	25	water
26	water	26	water
27	water	27	water

Figure 18

1975 CLASSIFIED LANDCOVER

SHADOW

NATIVE VEGETATION

AGRICULTURE

SANDY

WATER

Figure 19

1992 CLASSIFIED LANDCOVER

SHADOW

NATIVE VEGETATION

AGRICULTURE

SANDY

WATER

Change 1975 - 1982

The method used to assess change in the cultivated areas was through the use of a pair-wise analysis illustrated below:

		1982	
		non-cultivated	cultivated
1975	non-cultivated	non-cultivated	cultivated in 1982 only
	cultivated areas	cultivated in 1975 only	cultivated in both years

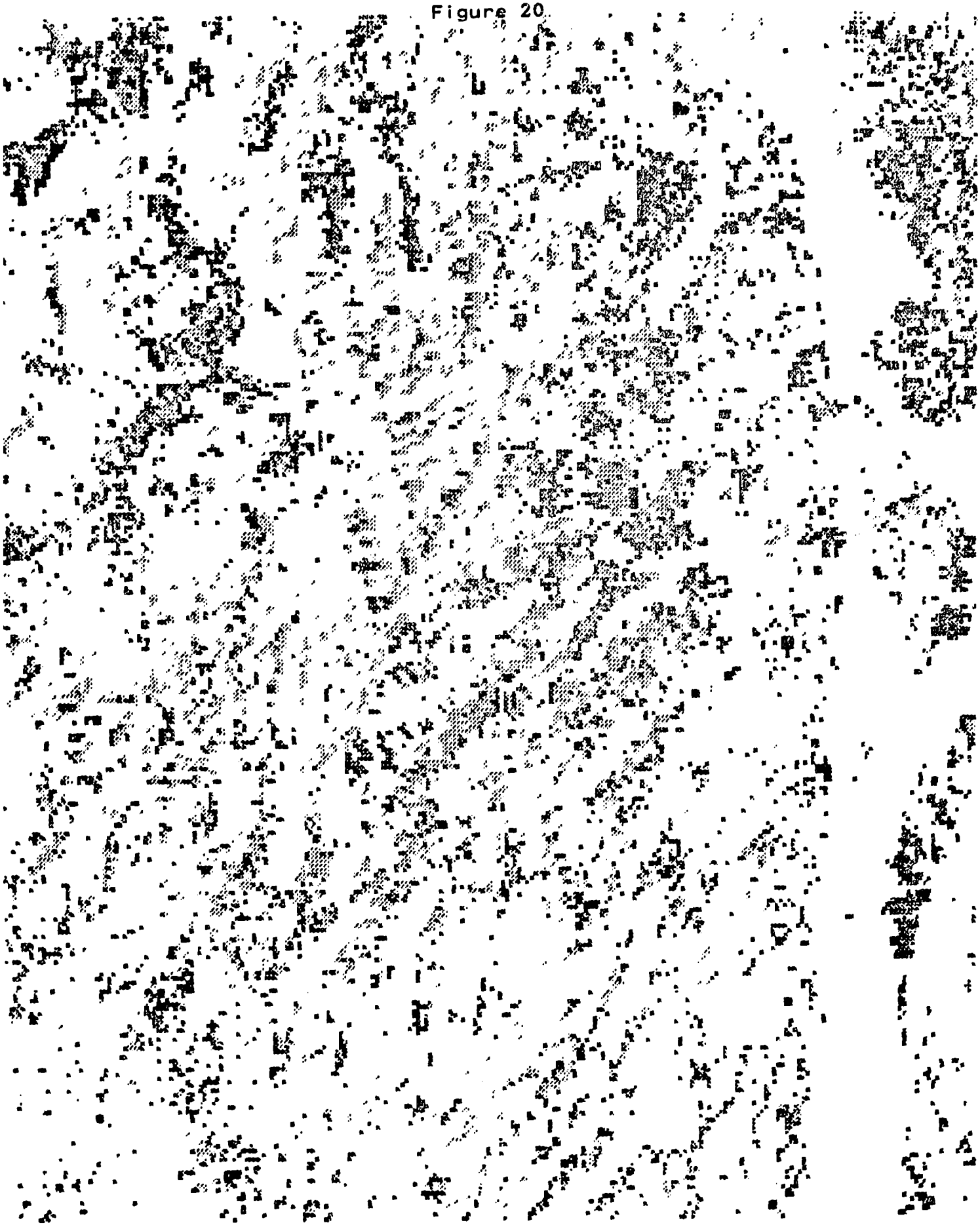
Figure 20 is a computer map showing change in cultivation.

Potential Road Network

The potential road network was created from the road suitabilty map. The map was superimposed upon existing transportation maps and a potential network was designed avoiding areas of low suitabilty whenever possible. Figure 16 is a computer printout showing the road suitability map grouped into three categories with the potential road network superimposed.

Methods of establishing a G.I.S. for an area development project may seem complex, but much of the complexity relates to a theoretical understanding of the process. In practice the process is routine. The computer programs for analyses of LANDSAT and for entering data on slope and topography are readily mastered and taught.

Figure 20



LANDCOVER CHANGE BETWEEN 1975 AND 1982

AGRICULTURE 1975-1982
GAIN 1975-1982
LOSS 1975-1982