US Army Corps of Engineers
Hydrologic Engineering Center

Guide Manual for the Creation of Grid Cell Data Banks

September 1978

TD-2

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
September 1978	Training Document	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Guide Manual for the creation of Grid Cell Data Banks	
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Darryl W. Davis, Bill Hodson, R. Pat Webb	
	5e. TASK NUMBER
	5F. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
US Army Corps of Engineers Institute for Water Resources Hydrologic Engineering Center (HEC) 609 Second Street Davis, CA 95616-4687	TD-2

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 manual documents basic procedures that are necessary for the successful creation of grid cell data banks. The manual describes Grid Cell data banks and provides guidance on: 1) selection of data variables, 2) source of information, 3) selection of grid cell configuration, 4) basemap creation, 5) data encoding and editing, 6) processing and file structure of data bank, and, 7) updating files and expanding existing data banks. Software needed and available for creating grid cell data banks is described.

15. SUBJECT TERMS
spatial data management, geographic information systems, grid cell data, polygon data, computer mapping, data management

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	84	19b. TELEPHONE NUMBER

Guide Manual for the Creation of Grid Cell Data Banks

September 1978

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-2

FOREWORD

Spatial Data Management techniques are rapidly becoming practical tools for use by Corps of Engineers field offices in a variety of their responsibilities. Various aspects of these techniques have been applied in traditional Survey and Phase I General Design Memorandum Studies and on a large scale in the Expanded Flood Plain Information (XFPI) studies of the Corps' Flood Plain Management Services program.

The use of grid data, e.g., spatial data stored in computer files in a specific grid cell format, has been determined to be the only spatial data management technique that offers significant analytical opportunities when compared to polygon oriented approaches. The grid structure successfully used in applications to date have included square, rectangular, and triangular cells, the latter being an emerging method with particular potential in the management of terrain (topographic) data

This manual is intended to be initial documentation of the basic procedures that are necessary for the successful creation of a grid cell data bank. The manual was prepared primarily to aid the XFPI pilot studies in which gridded data banks are being created as a major focal point of the studies. However, anyone interested in spatial data management should find that the manual contains valuable information.

Computer software is available to perform most of the tasks described in the manual. The Hydrologic Engineering Center, 609 Second Street, Davis, California, 95616, should be contacted for the latest computer source code and program documentation to implement the guidelines suggested herein

The preparation of this manual was the result of the talents and experience of many people. R Pat Webb (project manager), Planning Analysis Branch, HEC, and Darryl W Davis, Chief, Planning Analysis Branch, HEC, were responsible for the formulation and content of the material included in the manual.

The Environmental Systems Research Institute (ESRI), Redlands, California, prepared this manual under contract to the Hydrologic Engineering Center. Bill Hodson was the principal author with technical assistance from Ray Postma, Glenn Huibregtse and Bill Matteson. Jack Dangermond, Director of ESRI, made major contributions to the content and the perspective of the manual. Special thanks are due to Gene Hodson for typing the manuscript drafts and to Xantharid Virochsiri for the original graphics.

Bill S. Eichert was Director of the Hydrologic Engineering Center during the preparation of this manual. Funding support for preparation of this manual was provided by Flood Plain Management Services, Office of the Chief of Engineers, Washington, D C.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

APPENDICES

LIST OF FIGURES

TABLE OF CONTENTS (Continued)

LIST OF TABLES

CHAPTER I
INTRODUCTION

The use of automated geographic analysis by persons whose work is not primarily with computers has resulted in the need for straightforward methods of preparing geographic data and creating a data bank which can be accessed by the analysis programs. This manual identifies and discusses a series of steps which will result in the creation of a grid cell data bank. The resulting data bank is accessible by a computer, which allows geographers, engineers, planners and others to take advantage of computer analysis.

Objectives

This manual has six principal objectives:

1. Provide a clear definition of grid cell data banks. This is the subject of the Introduction and Overview Chapters (I and III)
2. List possible uses for grid cell data banks. (Chapter II)
3. Describe the process of selecting data variables. (Chapter IV)
4. Provide a listing of types and sources of natural resource and geographic data (Chapter V)
5. Present methods used to prepare data for inclusion in a grid cell data bank. (Chapters VI, VII, and VIII)
6. Describe the steps involved in the creation and management of a grid cell data bank. (Chapter IX)

The following sections of the manual describe a procedure for producing a useful grid cell data bank for studies using automated analysis. This technique can be used for many types of investigations involving spatially located data.

Description of the Manual

This manual is intended to be a self-teaching document directed primarily to the new user of automated geographic systems. It may be used by participants in workshops and also by personnel responsible for data acquisition and management. The manual is not intended to be a treatise on the state of the art of the theory of geographic analysis systems. Rather, it describes the decisions and steps required to successfully create a geographic data bank. In some instances, more than one method is identified which can be used to accomplish a single task. Considerations for choosing between alternatives are identified in these cases.

The technique presented here is one of several possible methods which can be used to prepare geographic and natural resource data for automated analysis. One method is described in order to reduce the confusion which might be generated by a discussion of alternatives and issues while the reader is unfamiliar with the basic methods. Several methods exist which can be used to accomplish the same goal. Selection of one possible method over another is based upon a variety of considerations which need not concern the newcomer who is attempting to learn a basic procedure.

CHAPTER II
GRID CELL DATA BANKS

Description of Grid Cells

Geographic data banks can be created in two ways; using grid cell or x,y coordinate (polygon) methods. The grid cell method involves subdividing the study area by overlaying a uniform grid and using each grid cell subarea as the generalized data unit for all data encountered within it. The x,y coordinate methods use point coordinates to define the locations and boundaries of points, lines and areas (polygons) describing geographic variation.

The use of grid cell data banks in the analysis of geographic data has become widespread in the last few years due to the following advantages which are available from grid cell analysis systems:

1. The grid cell provides an easy way to collect data. In the simplest method, one lays a plastic overlay of grid cells on top of maps or aerial photos and interprets information onto encoding forms for subsequent inclusion in a data bank.
2. Computer storage and the subsequent access and processing of this matrix of information is extremely straightforward using a simple row/column method (matrix array). Coordinate digitizing is not necessary, but may be included when necessary.
3. Editing of stored data is also straightforward using computer graphics output. Grid cells are represented by the standard printer characters.
4. Due to its consistent nature, the grid cell can be used for both discrete and continuous types of data analysis. Simple computer programs allow the user to aggregate many basic data items into the grid cell for different types of display.
5. Data encoded by x,y coordinate methods can be directly translated to a grid system by available computer software programs.
6. Point data of a continuous surface nature (e.g., air pollution records) can be directly generated to a grid cell scheme by interpolation.

The major disadvantage of grid cell representation of map data is the loss of resolution which accompanies restructuring data to fixed grid cell boundaries. Resolution increases as the size of grid cells decreases, however, cost normally also increases. For most users, the efficiency of grid cell analysis more than compensates for loss of resolution. In cases where very high accuracy of map quality is necessary, x,y coordinate systems are often applied.

Description of a Grid Cell Data Bank

A geographic data bank is a stored computer file of spatial resource data (map data), which can be accessed for various types of analysis. The stored data bank is the central feature of automated geographic analysis systems.

Map data occurs both as discrete forms which can be bounded by definite lines (e.g., watershed boundaries) and as continuously changing data (e.g., elevation). Both of these must be provided with an ordinal value (legend) for storage in a data bank. Discrete types of data are classified into groups and legended, while continuous data is assigned a specific value for a given sub area (e.g., a representative elevation for each grid cell).

For each data variable stored in the data bank its values (legends) and the geographic location (grid cell) of each value must be identified. This is accomplished by identifying each grid cell with a specific location (row and column) as the first two values for a grid cell record. All data associated with a particular grid cell is then stored sequentially at the address specified for that grid cell. This record may be stored in a computer memory, on magnetic tape or on a magnetic disc or other machine-readable device The stored data bank exists as a matrix of numbers which identifies data values and location.

Because data is stored for each grid cell, each data variable must be identified according to the same grid cell size, shape and location Choice of the grid cell size and shape is critical in the sense that all subsequent data encoding will be done with the same grid cells. Different data banks can be prepared using different grid cell sizes, but **any given data bank will contain only one grid cell size and shape**

Figure II-1 **SINGLE VARIABLE DATA BANK**, illustrates how a map can be visualized as a numerical matrix with the three-dimensions of row, column, and data value for each grid cell. The row and column numbers describe the spatial x and y location of a grid cell and the third dimension is the value or legend of the variable This type of a stored data bank is called a single variable file.

When more than one data variable is placed in a data bank, a new matrix of numbers is produced The spatial x,y location of each grid cell is stored along with all of the variable values associated with the cell. Figure II-2 **MULTIVARIABLE DATA BANK**, shows a typical grid cell data bank which includes several (m) variables. Note that this process is conceptually similar to overlaying data maps on one another and recording all of the variables (soil type, land use, vegetation, etc.) at each site represented by a grid cell This type of data bank is called a multivariable file.

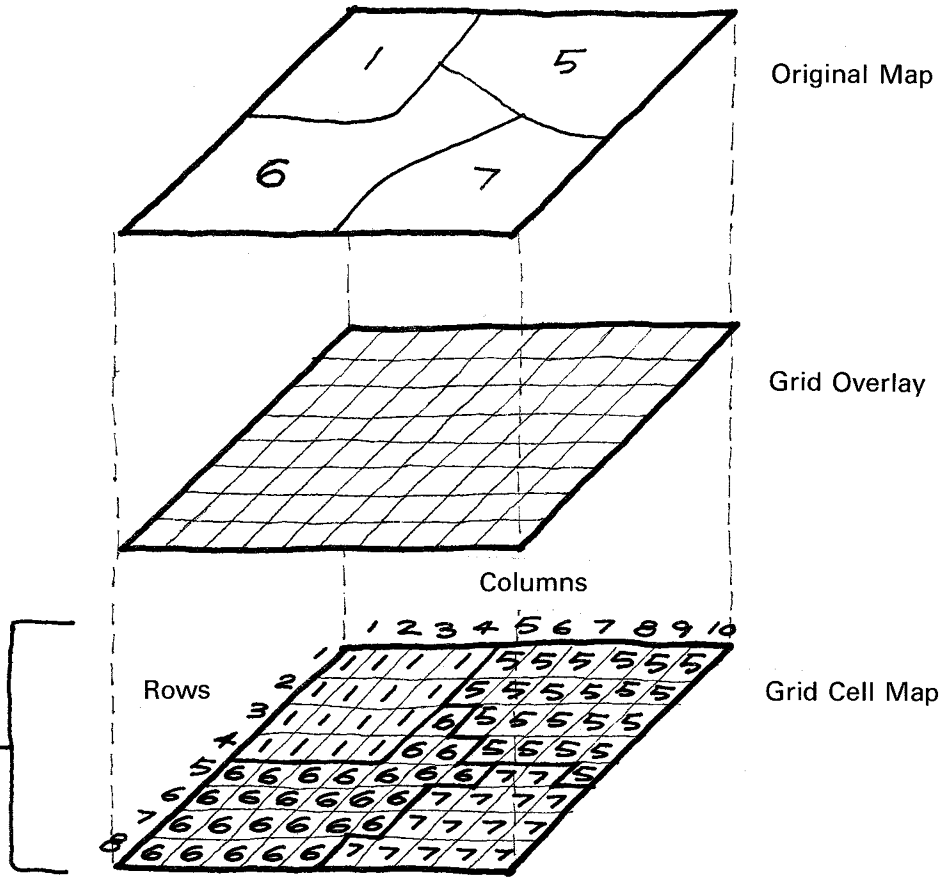
Multivariable files are structured sequentially in the computer so that input analysis and graphic computer output will be straightforward processes. This type of data bank allows the user to access all of the information needed about a particular grid cell.

Grid cell data banks are used in Corps studies for storing geographic and natural resource data which is subsequently analyzed in several ways Examples of these analyses include the following:

1 Hydrologic Analysis
- determination of subbasin area statistics
- determination of runoff coefficients
- computation of subbasin precipitation
- spatial display of erosion, sediment deposition and dredge material sinks

2. Economic Analysis
- determination of land use statistics
- computation of damage to development in floodplain
- formulation of nonstructural plans
- display of assessed values
- determination of proximity to transportation corridor

3 Environmental Analysis
- determination of vegetative cover
- determination of wildlife habitat zones
- computation of land surface erosion parameters
- computation of urban storm runoff quality parameters
- determination of locational attractiveness indexes
- assessment of general geographic impact
- coincidents tabulation of habitat reduction for use in scenarios of potential impacts.

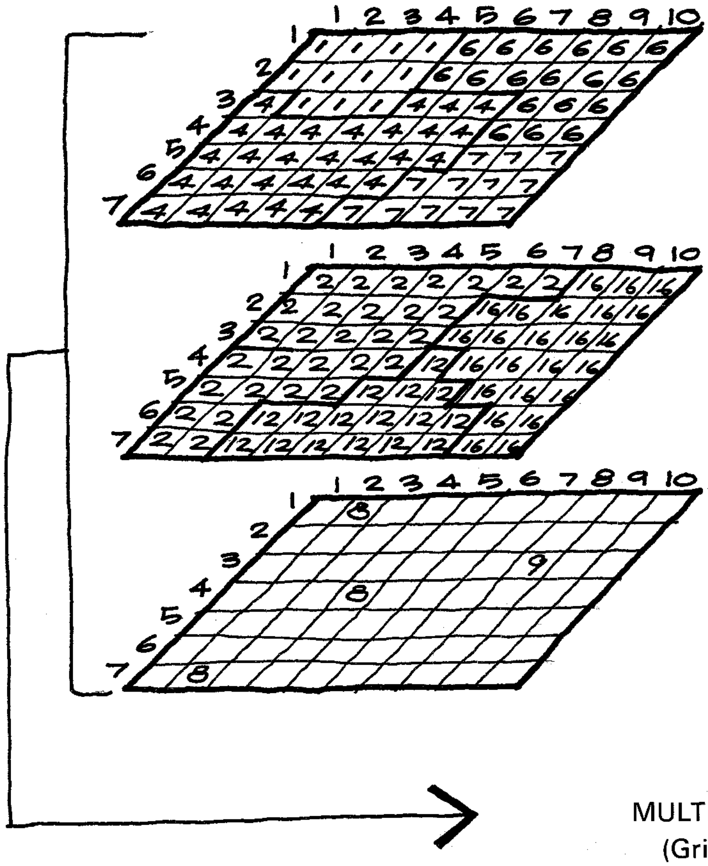


STORED DATA FILE
(single variable file)

Row	Column	Value
1	1	1
1	2	1
1	3	1
1	4	1
1	5	5
1	6	5
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.

SINGLE VARIABLE DATA BANK

Figure II-1



EXISTING LAND USE (DATA VARIABLE 1)

Code 6 is Single family housing
Code 4 is Industrial
Etc.

NATURAL VEGETATION (DATA VARIABLE 2)

Code 16 is Hardwoods
Code 12 is Cleared areas
Etc.

ARCHAEOLOGICAL SITES (DATA VARIABLE m)

Code 9 is Indian mounds
Code 8 is Unclassified
Code 0 or blanks are no sites
Etc.

MULTIVARIABLE FILE STRUCTURE
(Grid Cell Value Assignments)

Grid Cell Location ROW	COLUMN	Data var. 1	Data var. 2	Data var. m
1	1	1	2	0
1	2	1	2	8
1	3	1	2	0
1	4	1	2	0
1	5	6	2	0
1	.	.	.	.
1	10	6	16	0
2	1	1	2	0
2	2	1	2	0
2	3	1	2	0
2	.	.	.	.
2	.	.	.	.
.	.	.	.	.
.	.	.	.	.
k	.	.	.	.
k	.	.	.	.

MULTIVARIABLE DATA BANK

Figure II-2

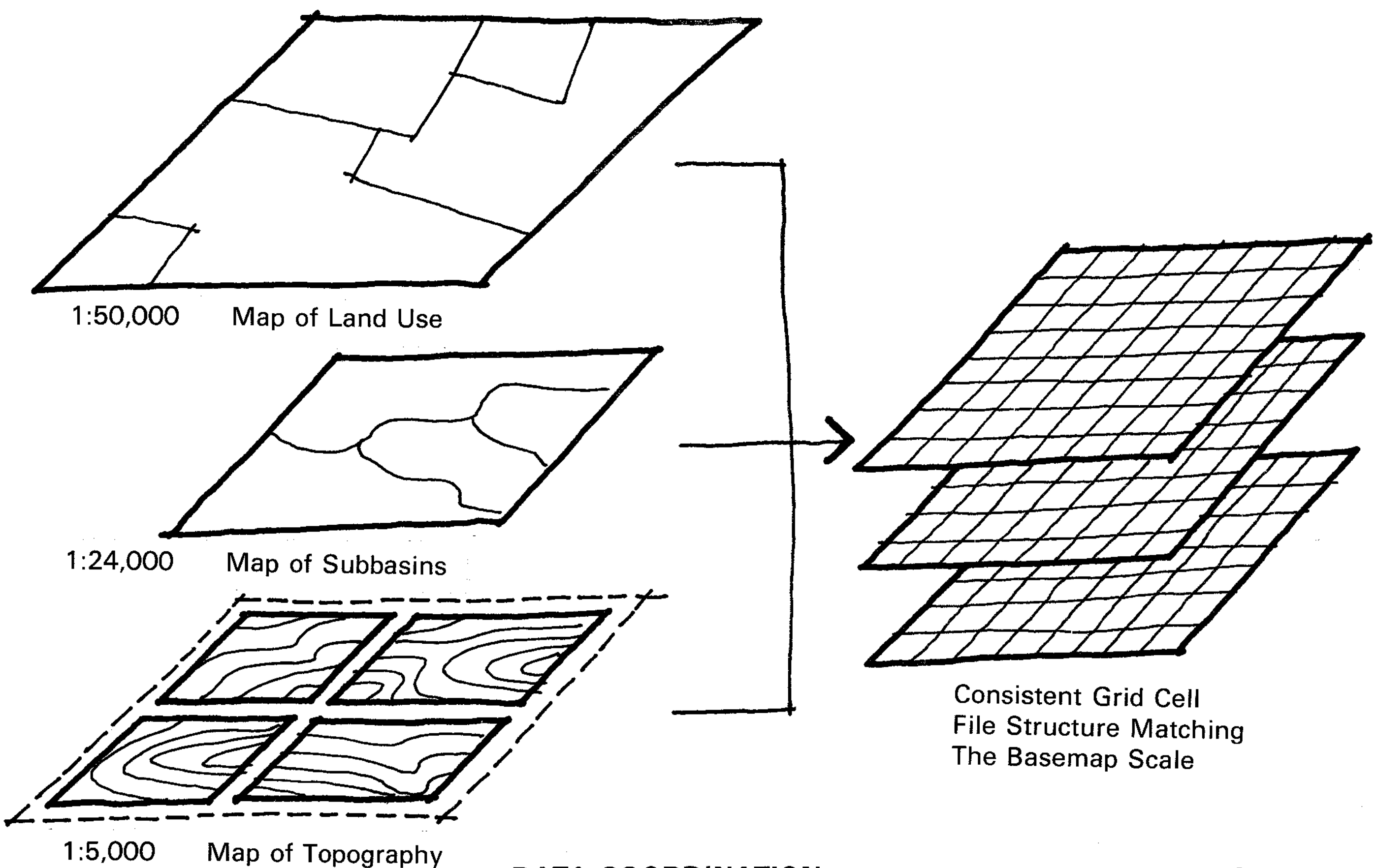
4. Social Analysis
   - display spatial distribution of income level
   - display ethnic zones
   - assess age variations
   - compute census statistics

Use of the Data Bank

Automated geographic analysis is intended to aid studies by enhancing the effectiveness of familiar, time-tested techniques and to provide opportunities for application of new techniques. **It is not necessary or desirable for traditional methods of classification systems to be replaced in order to benefit from the increased speed, versatility and cost-effectiveness of automated systems.** The data bank is of a flexible nature so that it can store any numeric data and any discrete classification of spatial variables required by the user for a particular analysis.

Other uses of a grid cell data bank include the following:

1. Computer mapping of basic data according to the original scale or rescaled values
2. Figure II-3 **DATA COORDINATION**, illustrates that different data for the same area may exist on maps of different sizes and scales and that by using appropriate data registration techniques, a data file of a single scale can be created



DATA COORDINATION Figure II-3

3. Selecting small geographic areas (windowing) from a larger geographic area file (e.g., watershed planning units). Figure II-4 **WINDOWING**, illustrates the selection of a subregion of an area for analysis. If the entire area has been stored as a grid cell data bank, windowing can be done by extracting the desired grid cells from the larger data bank.

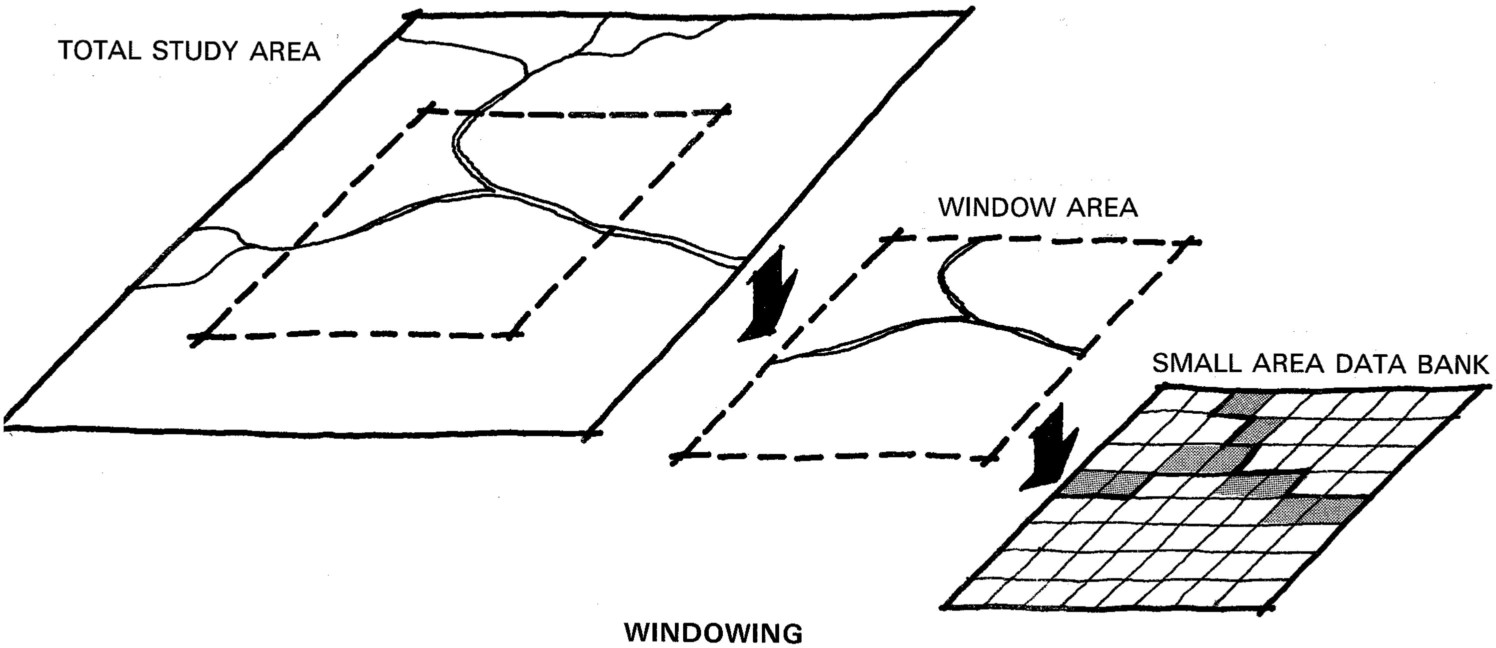
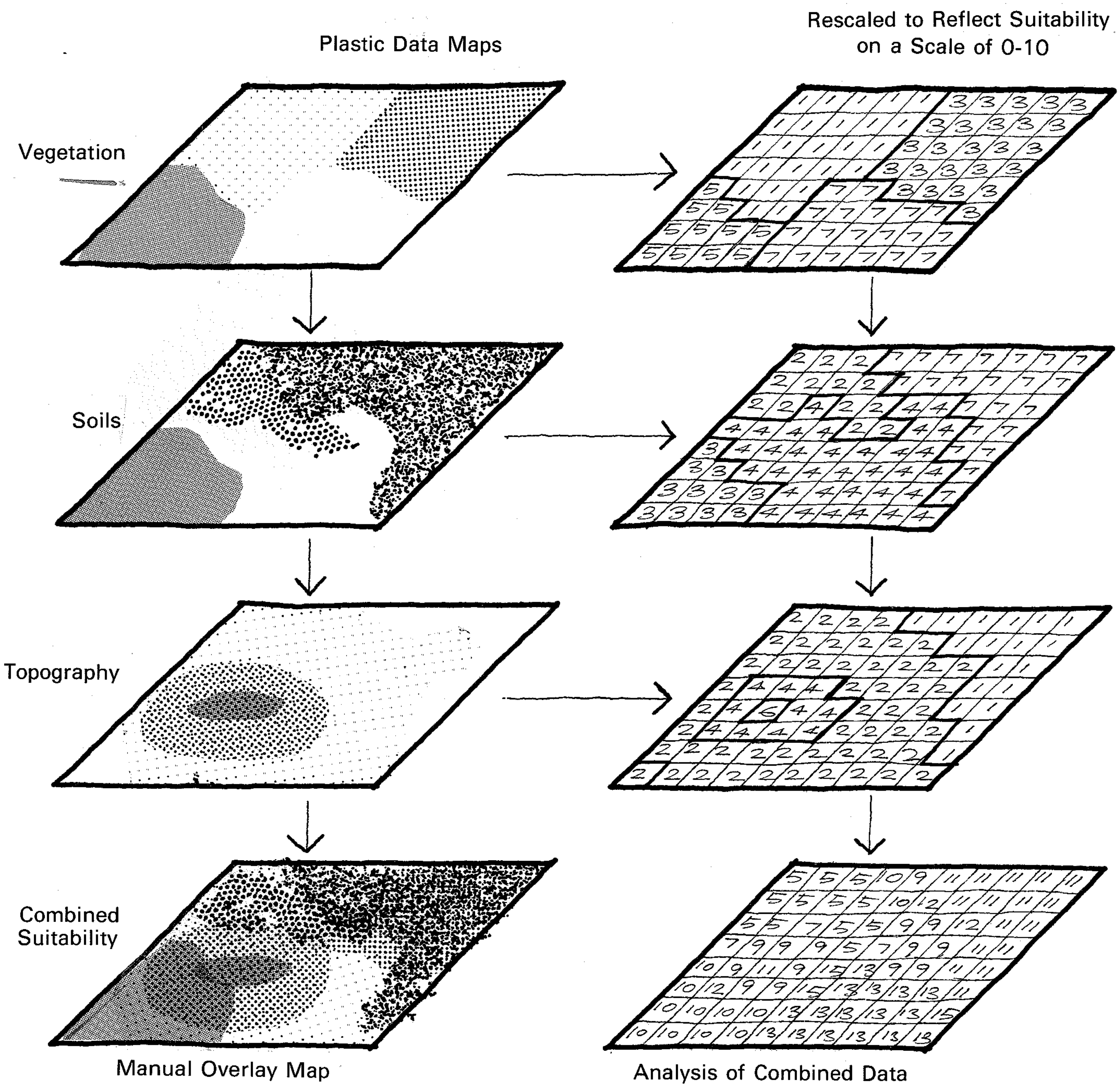


Figure II-4

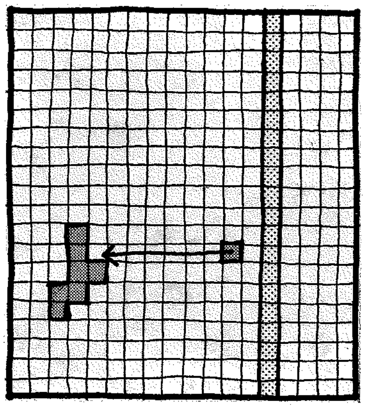
4. Combining two or more data variables to produce new information (this can involve a variety of weighting techniques). A significant advantage of automated combined variable analysis is the flexible use and re-use of the data using various weightings for various application maps. Figure II-5 **DATA VARIABLE COMBINATION**, illustrates that map data may be rescaled for a particular analysis and several rescaled maps may be combined to produce new analytical information.



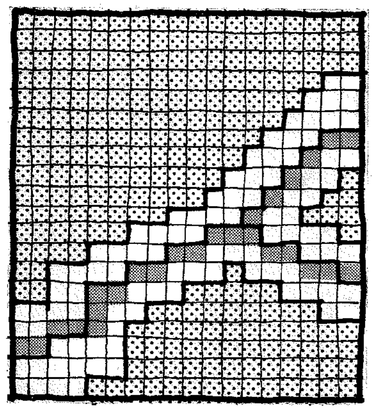
DATA VARIABLE COMBINATION

Figure II-5

5 Searching and calculating distances to or from a specific grid cell location of phenomenon. Figure II-6 **DISTANCE DETERMINATION**, illustrates the types of search which can be performed by current software. Distances from linear features such as streams can be calculated and illustrated as well as calculation of the distance from a given point to the nearest specified feature (e.g., distance from a logging site to the nearest stream).



Distance from Roads

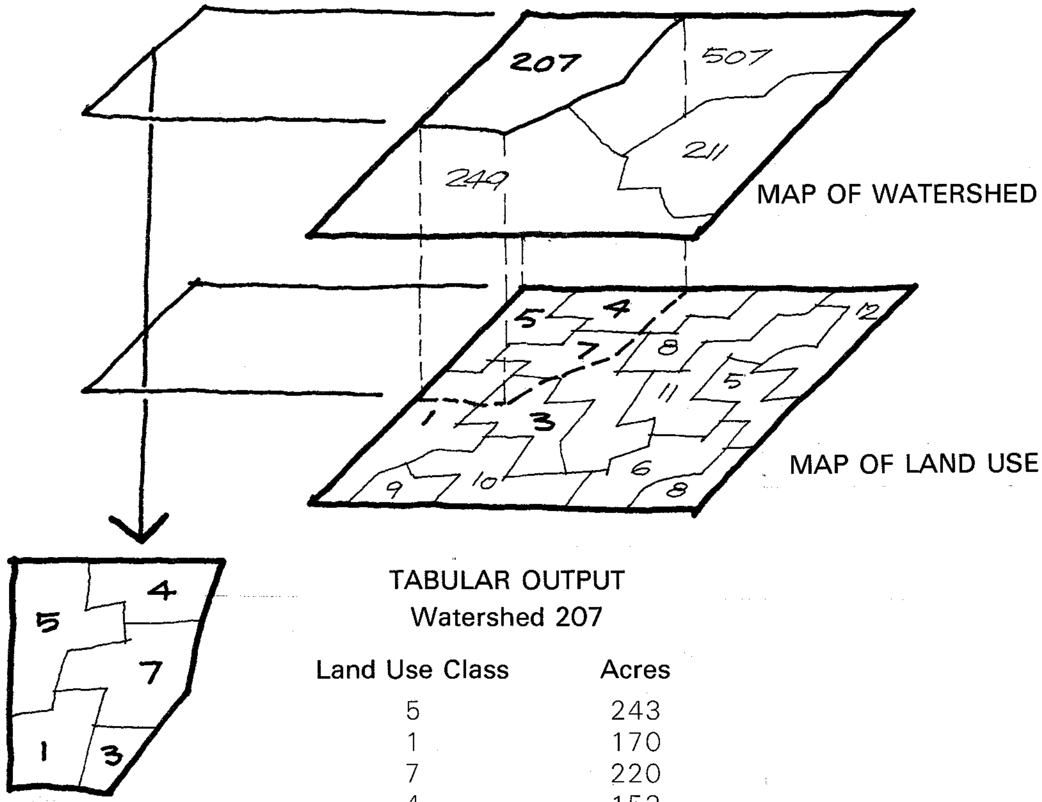


Distance to nearest specified feature, such as surface water.

DISTANCE DETERMINATION

Figure II-6

6. Selecting sub areas and calculating information contained within these (e.g., subbasin, census tract, etc.). Figure II-7 **AREA CALCULATION**, illustrates the calculation of areas occupied by various land uses within individual watershed boundaries.



TABULAR OUTPUT
Watershed 207

Land Use Class	Acres
5	243
1	170
7	220
4	153
3	52

AREA CALCULATION

Figure II-7

Properly applied, these uses of spatial data can aid in a variety of planning studies such as agricultural analysis, suitability studies for given types of construction and/or land use, projection of growth rates and resource requirements, and various evaluations of geographic and environmental relationships. **While use of an automated data bank system for a one-time study is cost-effective, it is anticipated that local agencies and/or the federal government will be able to utilize the data bank and analysis system for additional studies in the future.** This would increase the value and effectiveness of the original effort.

Appendix I is a selected bibliography of available literature which illustrates representative applications of grid cell data banks. This appendix will assist in identifying information useful for planning data banks of general utility

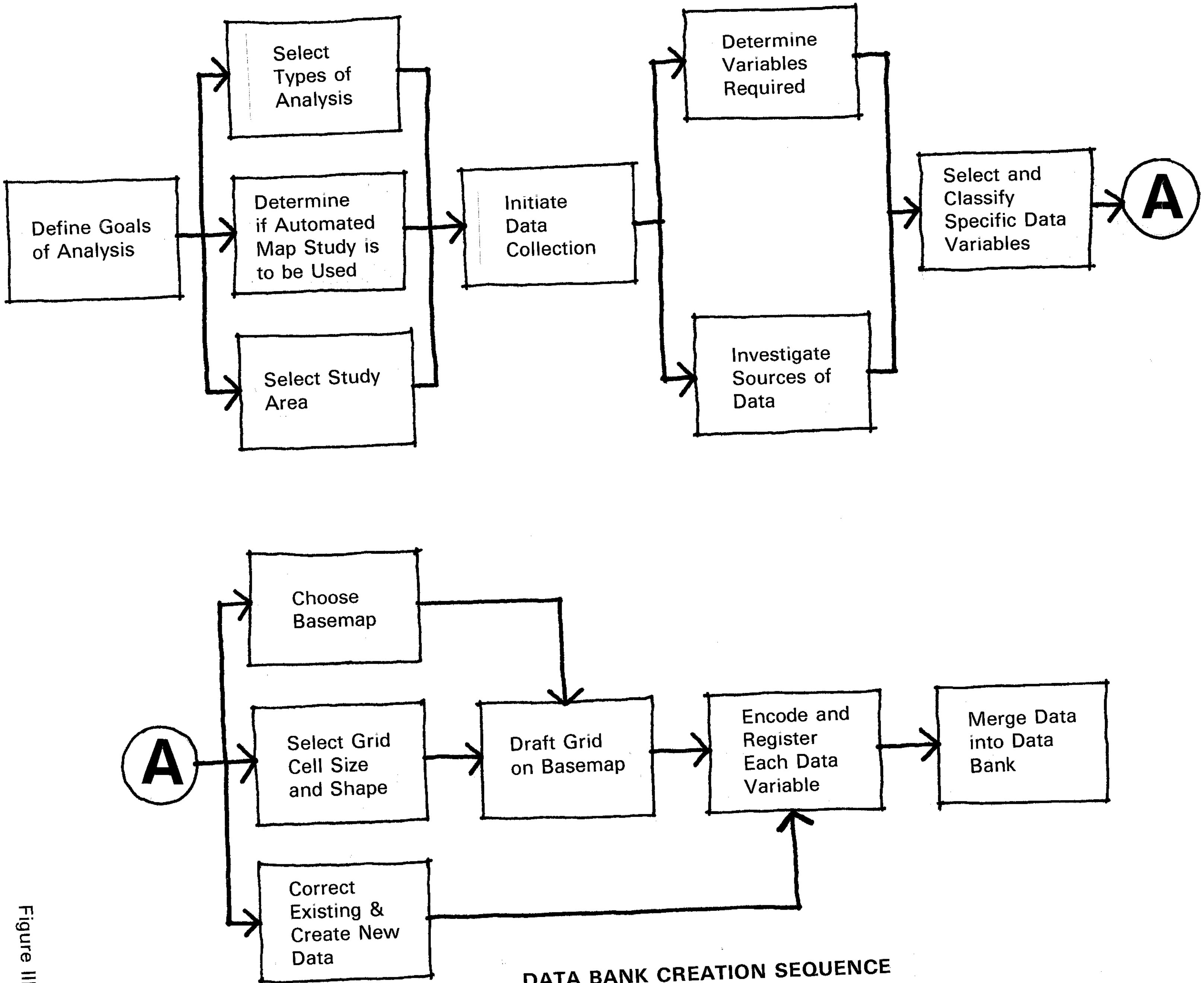
CHAPTER III
OVERVIEW OF DATA BANK CREATION

The method to create grid cell data banks which is described here requires that a series of steps be taken. Each step requires several basic decisions to be made in order to insure the construction of an accurate data bank. Figure III-1 **DATA BANK CREATION SEQUENCE**, illustrates the sequence of these steps. This chapter presents a brief outline of the creation process and identifies the chapters of the manual which discuss these in detail.

1. Select data variables which are appropriate to the types of analysis desired. In some cases, data will not be accurate or available and some remapping may be required. Variables must be classified and given values consistent with the types of analysis planned. (Chapters IV, V)
2. Select the size and shape of the grid cell. This step will determine the grid cell configuration for the entire study. The grid cell must be small enough to capture the required detail, but large enough so that computer storage and analysis can be performed efficiently. (Chapter VI)
3. Create a Basemap which will serve as the standard for the study. All additional data will be spatially identified with respect to the Basemap. This involves insuring the accuracy of all data. (Chapter VII)
4. Encode the data variable in such a way as to produce an accurate representation or machine record of the data maps. (Chapter VIII)
5. Create the data bank by merging each single data variable file into a multivariable file (Chapter IX)
6. Edit data files and expand the data bank as needed in order to enlarge or update the study (Chapter X)

In addition, short sections are provided which outline procedures for using commercial assistance (Chapter XI) and a brief summary of the grid cell method (Chapter XII).

The actual steps taken in a specific study will be related to the specific study area under consideration and the goals of the study. Proper planning in the early phases of the study can result in valuable analysis of geographic data within an appropriate time frame



DATA BANK CREATION SEQUENCE

Figure III-1

CHAPTER IV
SELECTION OF DATA VARIABLES

This chapter describes the process of selecting variables for inclusion in a data bank. Variable selection follows the decision to use automated techniques in the study, as illustrated on the planning flow chart (Figure III-1). The flow chart indicates that **the key to selecting appropriate data variables is a clear prior understanding of the types of analysis which are going to be performed.** This will enable a straightforward selection of the variables necessary to accomplish the study and will avoid the problem of collecting data which will not be used, or data which may be derived from other basic data

By defining the types of analysis which will be done, general variables can be identified which will allow the analysis to take place. As illustrated in Figure III-1, potential sources of data for these variables are then investigated and, if appropriate specific data is available for the study area, it is acquired. If the data required for the analysis is not available, the study may be revised to use another variable or a decision may be made to generate the necessary variables by means of field surveys, interpretation of aerial photography, or remote sensing. Acquisition or generation of the appropriate data variables is critical to the analysis phase of any study

Types of Variables

Any spatially located feature which exists as several classifications or values within the study area is a potential data variable. Variables are used to indicate natural and human-related conditions at a particular place. Some of the features included in this broad definition are:

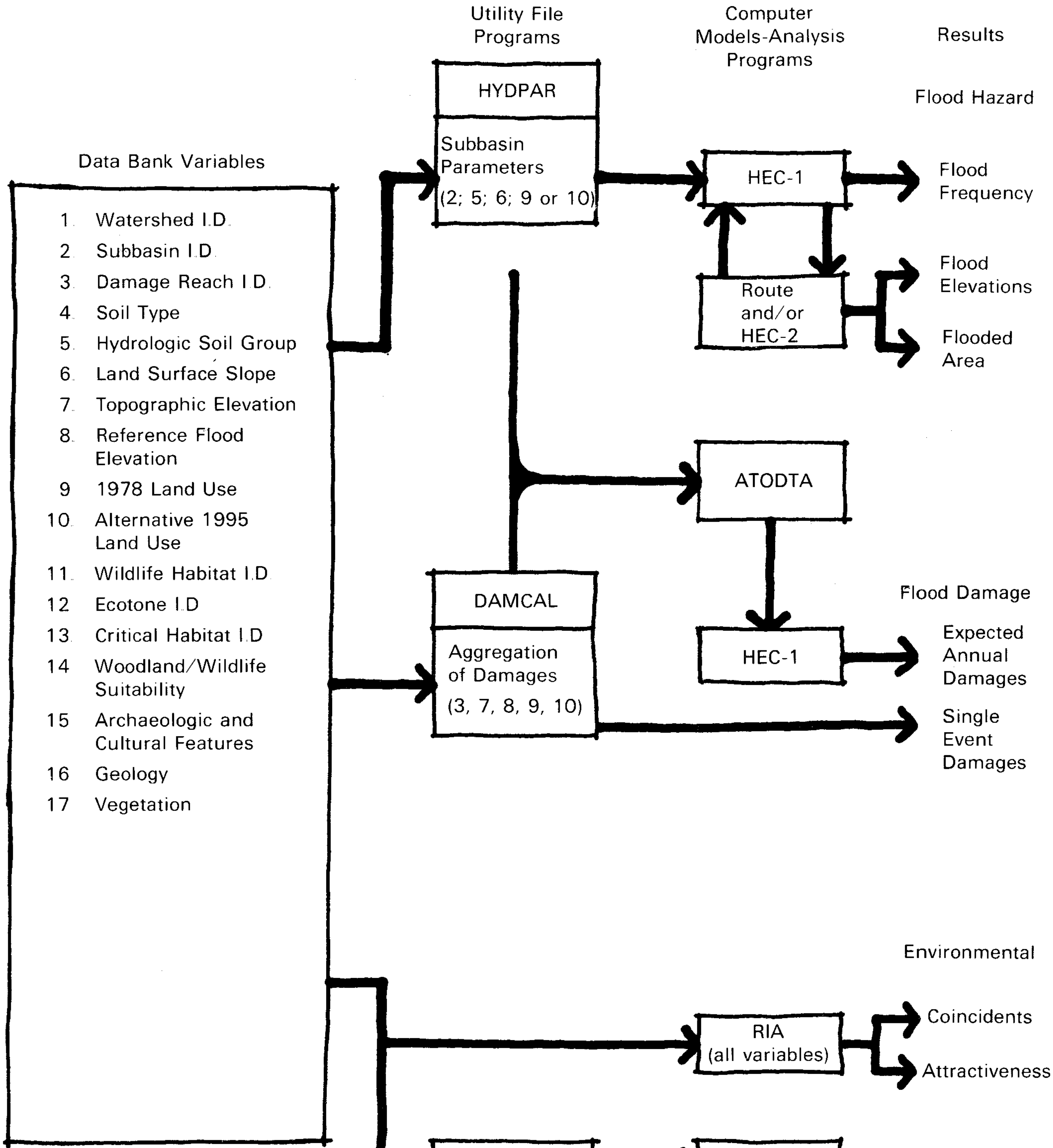
1. Political subdivisions, census tracts, dedicated reserves
2. Terrain features, elevation, slope, water courses
3. Hydrologic subdivisions, watersheds, subbasins, damage reaches
4. Soil types, geology, land form
5. Biological resources, vegetation, location of critical habitats, breeding grounds, protected species
6. Land use, urbanized areas, and industries

Because variables can be located geographically, maps are major sources of variable data. Information can be represented on maps in the form of points (e.g., historic sites), lines (e.g., highways, utilities, streams) or polygons (e.g., land use, vegetation).

Identifying Which Data Variables are Required

Different analytic methods generally require unique sets of data variables for input. For example, an analysis of environmental changes resulting from upstream alteration of runoff characteristics requires a different input than does an economic calculation of the damages expected to be caused by the 100 year flood. The choice of variables to be included in the data bank should reflect the types of analysis for which the data bank will be used. For efficiency of analysis and cost, as well as common sense, only those data variables needed for analysis should be included in a given data bank. Additional variables can be easily added to the data bank in the future if they are needed.

Several analytical interfaces between the data bank and computer modeling programs have been applied in the Corps Expanded Flood Plain Information (XFPI), Survey and Phase I General Design Memorandum (GDM) studies. Figure IV-1 **ANALYTICAL MODEL INTERFACES**, shows the data flow with an indication of the data variables used from the data bank to create the modeling parameters. Land use, for example, is used in more than one analysis. The selection of variables which can be used in more than one analysis contributes to the efficiency of the data bank and to the consistency of the analysis results for the different study interests.



Adapted from Figure VII-1 Data Management and Modeling Process, Phase I Oconee Basin Pilot Study — Trail Creek Test, 1975.

ANALYTICAL MODEL INTERFACE

Figure IV-1

A similar consideration must be applied in the selection of the classification detail used for data variables which have multiple uses. For example, slope gradient categories of 0-10% may be adequate for some types of analysis in a study while other types of analysis in the same study may require more detailed data, such as 0-3%, 4-8%, 9-12%, 13-15%, 16-25%, 26-35%, 36-60% and greater than 60%. In such a situation, the more detailed classification should always be chosen because it can be used to support more types of analysis than the broader classification.

In some cases interpretive variables may be selected for specific types of analysis. If soil type is included as a data variable, it can be used to "calculate" an interpretive soil capability rating. This capability rating can then be used in an analysis without the need for acquiring it as a separate variable. When defining variables to include in a data bank, the emphasis should be given to those which are useful in as many parts of the study as possible. Usually, fundamental variables such as elevation can be applied to more types of analysis than can interpretive variables such as slope. Simple computer programs can be written to interpret fundamental data bank variables in nearly any desired fashion. **Choosing fundamental variables rather than interpretive variables can also increase the usefulness of the data bank for additional studies by local agencies.**

In summary, the following steps will result in the selection of useful data variables to be placed in the data bank:

1. Determine the types of analysis required for the study.
2. Identify the data variables which are required for these types of analysis.
3. When more than one variable can be used for the same analysis, choose the more fundamental variable in the detail necessary for it to be used in all of the analyses.

Classification of Variables

Once the variables which will be used in the analysis have been selected, they need to be organized and legended in a useful manner. For example, the land use variable may be classified for some purposes in a general way to include vegetation, bare ground, suburban, urban and industrial categories. For other purposes, such as calculation of flood damages, a more detailed classification of each of these categories would be necessary.

Because various classification schemes are used in different areas of the country and for different types of studies, the data bank and interface computer programs have been designed to adapt to virtually any classification system. Examples of two possible classification schemes for land use are given in Tables IV-1 **LAND USE CATEGORIES — ROWLETT CREEK** and IV-2 **LAND USE CATEGORIES — SACRAMENTO RIVER**. Table IV-1 is the land use classification used in the Rowlett Creek Study and is a relatively simple breakdown of land use characteristics. A different land use classification is illustrated by Table IV-2 from a study done by Pacific Gas and Electric Company in San Francisco. The more detailed classification of agricultural data reflects the different variety of analyses for which data was gathered, in this case for forecasting power needs in non-urban areas.

When selecting a classification scheme, the following points should be considered:

1. What detail of classification is necessary to capture the data needed for the analysis?
2. What classification systems have given satisfactory results in the past?
3. Does the proposed system include all important classifications which exist within the study area?
4. Can the classification be expanded easily to include possible future changes in the study area?
5. Is the classification consistent with local government agency policies?

TABLE IV-1

LAND USE CATEGORIES — ROWLETT CREEK
(Lovell, Smith, 1977)

VALUES	DESCRIPTION
	URBAN/BUILT-UP LAND
	Residential
1	**Low density.** Single family 1/2 — 2 units per acre; average 1 unit per acre.
2	**Medium density.** Single family: 2 — 3-1/2 units per acre; average 3 units per acre.
3	**High density.** Single family: greater than 3-1/2 units per acre; average 4 units per acre.
4	**Multifamily.** Row houses, apartments, townhouses, etc.
5	**Mobile home parks.**
	Commercial/Services
6	**Central Business District.** Intensive, high density commercial.
7	**Strip commercial.**
8	**Shopping centers.**
9	**Institutional.** Schools, churches, hospitals, etc
10	Industrial — Industrial centers and parks; light and heavy industry.
11	Transportation — Major highways, railroads.
12	**Communication.** Microwave towers, etc.
13	**Public Utilities.** Transformer stations, transmission line right-of-way, sewage treatment facilities, water towers, and water treatment facilities.
	Mixed
14	**Strip Settlement.** Densities less than 2 units per acre; average 1 unit per 3-5 acres.
15	Parks and Developed Open Space — Parks, cemeteries, etc.
16	**Vacant urban land.**
	AGRICULTURAL LAND
17	Cropland
18	Grassland — Pasture, short grasses.
19	Confined Feeding — Feed lot for cattle.
20	FOREST LAND
21	WATER BODIES — Lakes, large ponds.
22	BARREN LAND — Bare exposed rock, strip mines, gravel pits

TABLE IV-2

LAND USE CATEGORIES — SACRAMENTO RIVER (PG&E, 1978)

	POSSIBLE CODES			CLASSIFICATION
		Alternate Hierarchy Codes		
Basic Codes	Primary	Secondary	Tertiary	
	1	100	10100	**Agriculture**
1	1	102	10101	Irrigated Grazing Land
2	1	102	10102	Non-Irrigated Grazing Land
3	1	103	10103	Perennial Crops
4	1	104	10104	Irrigated Perennial Crops
5	1	105	10105	Non-Irrigated Perennial Crops
6	1	106	10106	Annual Crops
7	1	107	10107	Irrigated Annual Crops
8	1	108	10108	Non-Irrigated Annual Crops
9	1	109	10109	Rice Fields
10	1	110	10110	Intensive Stock Farming
11	1	111	10111	Specialty Crop Areas
12	1	112	10112	Timber Lands
	2	200	10200	**Industrial**
13	2	201	10201	Light Industry
14	2	202	10202	Heavy Industry
	3	300	10300	**Residential**
15	3	301	10301	Low-Density Residential
16	3	302	10302	High-Density Residential
17	3	303	10303	Vacant Subdivided
18	3	304	10304	Rural Residential
19	4	400	10400	**General Urban**
	5	500	10500	**Dedicated Reserves**
20	5	501	10501	Military
21	5	502	10502	Airports
22	5	503	10503	Cemeteries
	6	600	10600	**Intensive Recreation**
23	6	601	10601	Major City Parks
24	6	602	10602	Golf Courses and Marinas
25	6	603	10603	County Regional Parks
	7	700	10700	**Conservation Recreation**
26	7	701	10701	National and State Parks
27	7	702	10702	National and State Forests
28	7	703	10703	Wildlife Areas
29	7	704	10704	Publicly Owned Game Refuges and Wildlife Areas
30	7	705	10705	Privately Owned Game Refuges and Wildlife Areas

TABLE IV-2 (Continued)

POSSIBLE CODES				CLASSIFICATION
	Alternate Hierarchy Codes			
Basic Codes	Primary	Secondary	Tertiary	
31	7	706	10706	Wilderness Areas
32	7	707	10707	Primitive Areas
33	7	708	10708	Publicly Dedicated Marsh and Wetlands
	8	800	10800	**Mineral Exploitation/Extraction**
34	8	801	10801	Gravel Pits and Stone Quarries
35	8	802	10802	Dredger Tailings
36	8	803	10803	Gas Extraction and Associated Facilities
37	9	900	10900	**Water**
38	10	1000	11000	**Unused Lands**
39	11	1100	11100	**Other Grazing Lands**

CHAPTER V
SOURCES OF INFORMATION

This section identifies sources of information which can be encoded into a data bank. Actual choices among sources must be made on the basis of accuracy, completeness, age of data, and other factors which influence the suitability of these data for inclusion in the data bank.

After defining the data variables which will be placed into the data bank, potential sources of these data should be investigated. Sources of useful information for geographic studies include maps and map data, aerial photography, remote sensing, and available digital files (pre-existing data banks). Each of these is discussed in this section.

Maps and Map Data

Maps are available from a variety of sources and represent a variety of geographic information and interpretations at several scales. Table V-1 **COMMON MAP SCALES AND THEIR EQUIVALENTS**, lists several common map scales and their linear equivalents. Much of the most common and useful map data is found on USGS topographic quadrangle sheets. These are available for nearly all of the United States at a scale of 1:62,500 (1 inch = 5,280 feet). Several other scales are available for many areas. These include maps produced at scales of 1:24,000 (1 inch = 2,000 feet) and 1:250,000 (1 inch = approximately 4 miles). A useful scale for most planning studies are the 1:24,000 series because of the greater amount of detail shown. These maps are printed on both paper sheets and transparent stable-base mylar plates. Details concerning these and other maps are given in Table V-2 **SOURCES OF MAP DATA**, along with ordering information. USGS maps for some areas may be based on old information, and more recent data, especially regarding land use and vegetation, will probably be required.

TABLE V-1

COMMON MAP SCALES AND THEIR EQUIVALENTS

Map Scale	One Inch Represents (ft)	One Centimeter Represents	One Mile is Represented by	One Kilometer is Represented by
1:2,000	167	20 m	31.68 in	50 cm
1:5,000	417	50 m	12.67 in.	20 cm
1:10,000	834	0.1 km	6.34 in.	10 cm
1:20,000	1668	0.2 km	3.17 in.	5 cm
1:24,000	2000	0.24 km	2.64 in	4.17cm
1:25,000	2086	0.25 km	2.53 in.	4.0 cm
1:31,680	2640	0.317km	2.00 in.	3.16cm
1:50,000	4166	0.5 km	1.27 in.	2.0 cm
1:62,500	5206	0.625km	1.014in.	1.6 cm
1:63,360	5280	0.634km	1.00 in.	1.58cm
1:75,000	6230	0.75 km	0.845in	1.33cm
1:80,000	6653	0.80 km	0.792in.	1.25cm
1:100,000	8342	1.0 km	0.634in.	1.0 cm
1:125,000	10402	1.25 km	0.507in.	8.0 mm
1:250,000	20856	2.5 km	0.253in.	4.0 mm
1:500,000	41659	5.0 km	0.127in	2.0 mm
1:1,000,000	83318	10.0 km	0.063in.	1.0 mm

Several factors must be considered when selecting maps for variable data. These factors include:

1. The age of the data on the map
   - Some variables change over brief periods of time. Transient physiographic variables such as soil moisture are very poor variables. Other transient variables such as location of roads and areas of urbanization are necessary variables for many types of analysis.
2. Cartographic accuracy
   - The material on which the map is drafted may stretch, shrink or otherwise distort in time.
   - Data should be registered to known landmarks or an identifiable coordinate projection.
   - The care with which data was originally drafted can be checked by comparison with other maps and field checks.
3. Resolution of detail
   - The classification system used on the map should be adaptable to the study.
   - The scale of the map should allow the desired resolution.
   - The source of basic data should be compatible with the desired resolution of the data.
4. Compatibility of maps
   - Variable data recorded on different maps may be drafted at different scales and registered to different coordinate projections, which will not allow the maps to be directly overlayed.

These considerations should be accounted for whenever maps are being evaluated as sources of data. Even those maps generally considered to be of high cartographic quality may contain fundamental discrepancies or errors which can result in inaccurate location of variable data.

TABLE V-2

SOURCES OF MAP DATA

SOURCE	TYPES OF DATA
Local Agencies —town halls, county offices, local commercial firms	Archaeologic sites, historic sites, vegetation, park boundaries, well records, soil borings, soil surveys, census tracts, land use zones, future plans, highways and utility rights-of-way, declared nature preserves.
U.S. Geological Survey (West of Mississippi River) **U.S. Geological Survey** Branch of Distribution P.O. Box 25046 Denver Federal Center Denver, Colorado 80225 303-234-3832 (East of Mississippi River) **U.S. Geological Survey** Branch of Distribution 1200 South Eads Street Arlington, Virginia 22202 702-557-2781	**7½ minute topographic quadrangles** (1 inch = 2,000 feet) Contour intervals (typically 10 or 20 feet, occasionally 5 feet), road types, bridges, tunnels, railroads, power transmission lines and towers, dams, large buildings, urbanized areas, storage tanks, quarries, mines, borrow pits, campsites, shoreline rock or reef features, horizontal-vertical control points, spot elevations, municipal boundaries, fence or field lines, section corners, boundary monuments, levees, washes, sand areas, gravel beaches, glaciers, water wells, springs, rapids, channels, depth soundings, dry lake beds, intermittent streams, aqueduct tunnels, waterfalls, intermittent lakes, marshes, inundated areas, submerged marshes, woodlands, orchards, vineyards, mangrove swamps, scrub vegetation, and wooded marshes.
	Geology Maps (scale varies) Bedrock geology, sections, faults
	Surficial Geology (scale varies) Surface geology, fault traces, sections
	Metropolitan Area Maps (1 inch = 2,000 feet) Geographic and land use features of major metropolitan areas
	National Park Series (scale varies) Geographic and land use features of national parks, monuments and historic sites
	State Base Maps (1:500,000 and 1:1,000,000) Large scale maps of general geographic features and political boundaries, urban areas, and major transportation routes
	United States Maps (scale varies)

TABLE V-2 (Continued)

SOURCE	TYPES OF DATA
	Large scale maps showing state and county boundaries, major metropolitan areas, waterbodies, boundaries of national forests and parks, Indian reservations, and national wildlife refuges
	Status Index Maps* (1 inch = abt. 80 miles) Maps showing status of mapping phases and areas covered by aerial photography
National Cartographic Information Center Additional information can be obtained from the publication **Topographic Maps.** This and other information is available from the National Cartographic Information Center Regional Offices listed here: **NCIC — Eastern** U.S. Geological Survey 507 National Center Reston, Virginia 22092 703-860-6336 **NCIC — Mid-Continent** U.S. Geological Survey 1400 Independence Road Rolla, Missouri 65401 314-364-3680, ext. 107 **NCIC — Rock Mountain** U.S. Geological Survey 510, Box 25046 Denver Federal Center Denver, Colorado 80225 303-234-4553 **NCIC — Western** U.S. Geological Survey 345 Middlefield Road Menlo Park, California 94025 415-323-8111, ext. 2427	NCIC is a center for cartographic information. Maps, aerial photographs, geographic data banks and other information can be located by contacting the main office in Reston, Virginia, or the nearest regional office
U.S. National Ocean Survey U.S. Department of Commerce	Maps showing shore features which are similar to USGS topographic maps and include offshore features along the U.S. coast line

TABLE V-2 (Continued)

SOURCE	TYPES OF DATA
Soil Conservation Service Information Division U.S. Department of Agriculture Washington, D.C.	Soil survey maps plus documentation, capability interpretations of soils 1/3 of the U.S. has been mapped in detail
or	
Soil Conservation Service Local Office U.S. Department of Agriculture (see local telephone directory)	Local offices can provide unpublished information about soils in their area
State Geological Survey (scales vary) (see local State Directory)	Types of maps vary. Geologic and topographic maps may be available

*Available free on request from U.S. Geological Survey, Reston, Virginia 22092

Aerial Photography

Aerial photographs are a source of a variety of raw data for interpretation and mapping. Basic photos are subject to all of the distortion problems which accompany projection of a three dimensional surface onto a flat plane. The USGS has developed an orthophotoscope which can optically rectify aerial photography to compensate for this distortion. The resulting images are corrected photographs on which the scale is uniform in all directions.

Rectified aerial photographs can be used as Basemaps on which data derived from field investigations can be marked (e.g., Soil Conservation Service soil maps). Similarly, they can be used to correct cartographic errors found on manuscript data maps by an overlay process. Both of these processes use rectified aerial photography as a base to accurately locate mapped features with respect to the photographic images.

A great deal of experience and technology is now available which enables one to use aerial photography as a reliable source of basic geographic data in addition to locating features spatially. The use of improved film emulsions and film-filter combinations which allow isolation of spectral elements has given aerial photography a versatile role in the development of many types of data maps. When combined with other sources of data and field checking, the interpretation of aerial photographs by an experienced person can provide a rapid, efficient and relatively accurate means of defining and locating map data. Maps are usually produced from aerial photographs by overlaying the photo with transparent material on which the data is drafted.

Types of aerial photography which are available include black-and-white panchromatic, black-and-white infrared, visible color, and false-color infrared. Each type produces a unique "signature" of ground features which depends on the portion of the spectrum recorded, and the properties of the film emulsion. For detailed studies, more than one film type may be used, and the features cross-checked with one another and other data sources.

1. Black-and-White Panchromatic: This has proved to be the most versatile film for mapping and interpretive purposes, and is the most widely used. Panchromatic film records the reflected visible wavelengths of sunlight and produces an image in continuous black, gray, and white tones. Emulsions with excellent contrast and resolution capability are available which capture a variety of spectral and physical detail. Interpretation techniques are based upon identification of textural and tonal

differences between objects or patterns of differing sizes, shapes, and positions. One advantage of black-and-white panchromatic film is the ability to filter out blue light scattered by haze. This is done by photographing through a yellow filter, and enables successful photographic missions to be flown even during relatively dense haze conditions. Stereo pairs are often photographed using black-and-white panchromatic film and elevation can be interpreted from 3-dimensional reconstructions of these photos. **The Manual of Photographic Interpretation**, published by the American Society of Photogrammetry, includes a comprehensive background for panchromatic interpretation techniques and applications. Many illustrations of different landforms interpreted by these techniques are included in **Terrain Analysis** by Douglas S. Way.

2. Infrared Black-and-White: This film records reflected light from the near-infrared portion of the spectrum. These wavelengths are not normally visible, and have different reflective properties than does visible light. For example, on infrared aerial photographs, broad-leaved vegetation appears much lighter in tone than conifers because broad-leaved plants reflect much more infrared energy. This quality is useful for forest inventories, and this film is used extensively by the U.S. Forest Service and many timber companies. Drainage patterns, swamps and waterbodies are also emphasized due to the high reflectivity of water. Because contrast in shaded areas is poor on black-and-white infrared film, these photos are not usually suitable for elevation contour analysis.

3. Visible Color: This film type adds familiar color signature to panchromatic photographs. Additional interpretation of vegetation and soil features is possible using the color correlations on these photographs in some areas. Many qualified interpreters have suggested that the additional detail can be confusing because of color changes which do not correlate with actual data variables. In addition, the advantages of color over black-and-white photographs is often offset by the stringent visibility-haze requirements for obtaining the imagery. In northern latitudes, an optimal day for a color mission may not occur during the spring (though black-and-white film could be used), and vegetation often covers terrain features during later flights.

4. Color Infrared: This film is also sensitive to wavelengths in the near-infrared spectrum. Rather than producing a black-and-white image, however, the different wavelengths and intensities are represented by color tones and hues. Healthy broad-leaved vegetation, for example, is represented by a brilliant red or pink color. Differences in vegetation density, general soil types and wetlands can be detected with relative ease from CIR imagery. This film is being used with increasing frequency for many types of interpretations

Table V-3 **SOURCES OF AERIAL PHOTOGRAPHIC COVERAGE**, lists several sources of aerial photography available in the United States.

TABLE V-3

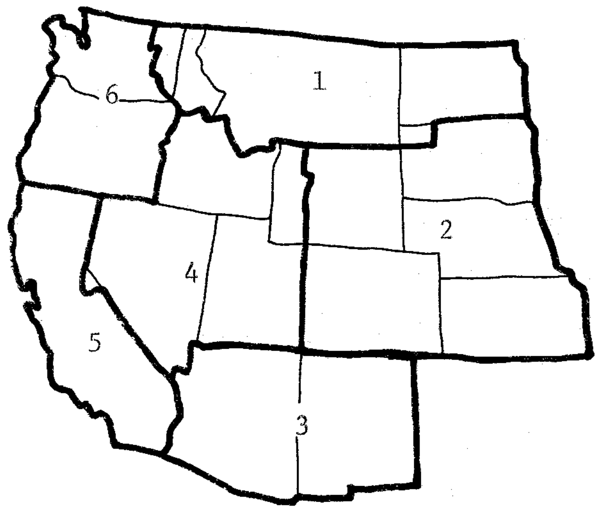
SOURCES OF AERIAL PHOTOGRAPHIC COVERAGE

SOURCE	DESCRIPTION
Publication: **Summary of Aerial Photography** National Cartographic Information Center U.S. Geological Survey 507 National Center Reston, Virginia 22092	A booklet which indicates the best of the Federal Government's aerial photographic coverage available for the U.S. Lists addresses of the agencies holding the original negatives.
Agricultural Stabilization Conservation Service —Eastern Laboratory: Aerial Photography Division ASCS — USDA 45 French Broad Avenue Asheville, NC 28802 —Western Laboratory: Aerial Photography Division ASCS — USDA 2505 Parley's Way Salt Lake City, Utah 84109	ASCS holds the greatest aerial photo coverage of the U.S. by a single agency. Most photographs are available at a scale of 1:20,000. Recent 1:40,000 photos are also available for some areas, but only general interpreting can be done from these. Contact prints of the photos available (ASCE — USDA, 1978) are in the following scales: (see table below)
ASCS Publications: **Code Symbols for the Identification of Aerial Photography** (1975)	Lists all symbols listed by the ASCA, the Soil Conservation Service, and the U.S. Forest Service.
Comprehensive Listing of Aerial Photographic Coverage (1975)	Lists all U.S. counties with ASCS symbol, year(s) of photography and number of photographic indexes.
Status Maps of Aerial Photography of the Agricultural Stabilization and Conservation Service (1975)	Coverage shown on State-County maps.
Eastern Region **U.S. Forest Service** Division of Engineering U.S. Department of Agriculture Washington, D.C. 20250	Most photos are at a contact print scale of 1 inch = 1320 feet. Much of U.S. Forest Service coverage is available in black-and-white infrared.

Paper size	Approximate scale from 1:20,000 photography	Approximate scale from 1:40,000 photography
9½" x 9½"	1" = 1667'	1" = 3334'
13" x 13"	1" = 1320'	1" = 2640'
17" x 17"	1" = 1000'	1" = 2000'
24" x 24"	1" = 660'	1" = 1320'
24" x 24"	1" = 330'	1" = 660'
38" x 38"	1" = 400'	1" = 800'

*This size and scale relationship exists if one-quarter of the original photograph is ordered.

TABLE V-3 (Continued)

SOURCE	DESCRIPTION
Inquiries and orders concerning photographic coverage of the western United States must be addressed to each region shown in the adjacent map (1) Federal Building, Missoula, Montana 59801. (2) Federal Center Building, No. 85, Denver, Colorado 80225. (3) Federal Building, 517 Gold Avenue, S.W., Albuquerque, New Mexico 87101. (4) Forest Service Building, Ogden, Utah 84403. (5) 630 Sansome Street, San Francisco, California 94111. (6) Post Office Box 3623, Portland, Oregon 97208.	
Soil Conservation Service Cartographic Division Federal Center Building Hyattsville, Maryland 20782	SCS aerial photos cover much of the U.S. at a scale of 1 inch = 1667 ft. Other (variable) scales are available for scattered locations. Ask for publication: **Summary Record of Aerial Photography**
National Cartographic Information Center U.S. Geological Survey 507 National Center Reston, Virginia 22092	The U.S. Geological Survey has extensive coverage of the U.S. at a variety of scales. Requests for information should specify the particular area of concern.
National Ocean Survey Department of Commerce Washington Science Center Rockville, Maryland 20852	Photographic coverage of most of the U.S. coastal regions in black-and-white, color, black-and-white infrared and color infrared.
Bureau of Land Management Department of the Interior Washington, D.C. 20242	Wide coverage at a variety of scales for the areas under jurisdiction of the Bureau of Land Management.

Remote Sensing by Scanners

Two types of non-photographic sensors are currently used to produce photograph-like images using a cathode ray tube These sensors produce images using far-infrared wavelengths (thermal energy) and reflected radar. The images produced by the two methods are sources of raw data for interpretation, and are not directly useable as maps.

Thermal imagery has been used by the U.S. Forest Service to detect forest fires or hot spots at night when these are obscured by clouds of smoke. Ecological studies have used thermal imagery to locate and trace warm water plumes from power plant discharges, subsurface volcanic activity, and glacier crevasses and snow bridges.

Side-Looking Airborne Radar (SLAR) has been used to map geologic terrain features at night and through clouds, thin ice and vegetation. Because vegetation is relatively transparent to some radar wavelengths, this imagery has been particularly useful for mapping jungle and heavily-forested terrain.

Interpretation of non-photographic remote imagery usually requires training and experience with the techniques involved. Field checking is necessary to insure accurate interpretation of the various image "signatures" Further information on non-photographic imagery and considerations for its use can be found in "Photogrammetry," published in **Photogrammetric Engineering**, 29: pp 761-799, 1963.

The LANDSAT remote sensing scanner produces an electronic record of changes in spectral intensity rather than a photographic image. A multispectral scanner (MSS), consisting of several channels sensitive to specific wavelengths, is rotated through an arc so that it senses a different patch of ground (pixel) at each successive instant. The output voltage from the photoelectric cell channels is recorded on magnetic tape as a continuously varying signature for each channel.

LANDSAT tapes can be analyzed by a computer to determine similarity, dissimilarity and type of each pixel signature The actual interpretation of the signatures and their relation to ground conditions must be done by an experienced interpreter, either before or after they are sorted by a computer program. While this method can cover large areas in a short time and provide data in a form legible to a computer, three concerns must be identified when considering MSS tapes as a source of data.

1. Boundaries between land areas of different spectral signatures are recorded as a continuous change of output voltage as the scanner passes over the boundary. The sharp boundary is therefore "blurred" and the intermediate signature may be misinterpreted by the analysis program.
2. Different ground conditions may not be separable by their signatures. Heavy industry and commercial areas both are viewed as roofs and parking lots. Dry grass may have the same signature as freeway concrete. Rural areas may have a variety of signatures depending upon the amount of vegetation and roof area sensed by the scanner.
3. Each pixel signature is an average of all conditions sensed within it. Small but important features may not be identified by this method.

More information concerning LANDSAT imagery, and automated analysis of such imagery, is discussed by Link and Aaron (1977).

Other Digital Files

Several government agencies and private companies are investigating and assembling their own automated geographic data banks for assessment and planning purposes. These data files for specific geographic areas include variables which have originated as census data, remote sensing imagery, maps or comprehensive study programs, and are often available for use or purchase from the agency or company. While these sources of data are not generally used by the Corps at the present time, their increasing availability and coverage in the future may make the acquisition of existing data files an attractive source of data for many study areas.

Agencies which have geographic data banks available or in preparation include (but are not limited to) the U.S. Bureau of the Census, U.S. Geological Survey, USDA Soil Conservation Service, National Cartographic Information Service, Planning Commissions in the States of Massachusetts, Maryland, Utah, Washington, Alabama, South Carolina, Georgia, and the Territory of Puerto Rico, and several county planning commissions throughout the U.S. Private companies using automated systems include power companies (Pacific Gas & Electric, Southern California Edison, Pennsylvania Power & Light), land development companies, and lumber industries. **The National Cartographic Information Service is establishing a clearinghouse for geographic data bank information.**

When considering the purchase of existing files, a primary concern will be the ability to interface these files with the user system. Referencing of the data file to geographic features or a known projection is necessary, as is the use of grid cells compatible with other data files being generated for the study. The accuracy of existing files depends upon all of the constraints identified in this manual for creating new data files. Before purchasing a data bank, or even a single data file, compatibility of the file with the system being used must be assured, as well as the accuracy and classification of the original data sources which were used to create the data bank.

CHAPTER VI
SELECTION OF A GRID CELL CONFIGURATION

The choice of grid cell size and shape for the study determines the unit size of all data variables and feature assignments made thereafter. The cell shape is related to the output format. The size of the cell must be small enough to insure accurate data, but large enough that computer storage and analysis are efficient. This section discusses the major considerations involved in the choice of grid cell shape and size

Shape of the Grid Cell

A mesh of grid cells is used to represent discrete analytical subunits of the study region. Assigned to each cell are the values for all of the variables included within its boundaries. The shape of the cell is important with respect to the type of graphic output used in the study. In many studies, working maps are produced on a line printer which has rectangular characters (1/10 of an inch wide by either 1/8 or 1/6 of an inch high) The use of a rectangular print character to represent a square grid cell would result in a distorted output map. For non-distorted output from a line printer, a rectangular grid cell of similar side-ratio to the printer should be used. Plotters, electrostatic printers, cathode ray tubes and other devices allow scaling in the x and y directions, producing an undistorted map for either square or rectangular cells.

Size of the Grid Cell

The size of the grid cell chosen for the study depends upon the data resolution required for the most detailed analysis Larger grid cells may include more than one data value which must be aggregated or prioritized and given a single value, therefore data resolution will decrease. Smaller grid cells allow greater resolution and accuracy, but require more time for coding and analysis, as well as increased computer storage space.

The optimum grid cell size to capture the appropriate detail varies from study to study This is due to differences in total area covered in the study and the heterogenity of the different data variable values. The more homogeneous an area is for critical variables such as topography and land use, the larger the cell size that can be used with accuracy.

The actual size chosen for the grid cell is generally the result of a compromise between data resolution and processing efficiency. For a given area, a change to grid cells of half the size requires four times the storage space **In general, grid cells should be small enough to insure that data variations are recorded with accuracy satisfactory to the particular study. When grid cell encoding techniques are employed, the grid cell size is fixed, and recoding is necessary if smaller cell size is subsequently required. If x,y coordinate encoding is used, grid cell data banks may be generated at several grid cell sizes.** Table VI-1 **AREAS REPRESENTED BY VARIOUS GRID CELL SIZES**, can be used to determine appropriate relationships between grid cell sizes and actual land area.

TABLE VI-1

AREAS REPRESENTED BY VARIOUS GRID CELL SIZES

MAP SCALE	RECTANGULAR GRID CELL SIZE 1/6 x 1/10 inch	RECTANGULAR GRID CELL SIZE 1/8 x 1/10 inch	SQUARE LAND AREA DIMENSION .2 cm x .2 cm	SQUARE LAND AREA DIMENSION .4 cm x .4 cm
1:12,000	.383 acre .155 ha	.287 acre .116 ha	.099 acre .040 ha	.395 acre .160 ha
1:24,000	1.530 acres .619 ha	1.148 acres .465 ha	.395 acre .160 ha	1.581 acres .640 ha
1:62,500	10.377 acres 4.200 ha	7.787 acres 3.151 ha	2.680 acres 1.085 ha	10.720 acres 4.340 ha
1:250,000	166.03 acres 67.19 ha	124.59 acres 50.42 ha	42.882 acres 17.361 ha	171.528 acres 69.444 ha
1:1,000,000	2656.5 acres 1075.1 ha	1993.5 acres 806.76 ha	686.111 acres 277.778 ha	2744.444 acres 1111.111 ha

CHAPTER VII
BASEMAP

Creation of the Basemap

The Basemap is the official map of the study area. The Basemap selected for a particular study depends on the size and scope of the study and will serve as the standard for scale and location of all points and lines. The locations of other data variables must be corrected, if necessary, to match the Basemap.

A data map which includes a variable of basic importance to a particular analysis would be one for primary consideration in choosing a Basemap. A map which includes a large number of important variables would be the logical choice as this would minimize errors of location of variables relative to each other. For many planning purposes it is suggested that the USGS 7½ minute mylar plates be used as the Basemap as they include several variables located with a relatively high degree of accuracy, especially topographic elevation.

In the larger study areas, one quadrangle may not be sufficient, and the Basemap must be constructed by combining two or more 7½ minute plates. When this is done, linear features or contour lines may not match exactly along the map edges and may need some adjustment. This mismatch may result from the following causes:

1. The original mylar maps may be of a slightly different size.
2. Distortion of the maps may result from projecting the spherical surface of the earth on the flat surface of a map. This causes the slight offset which is observed when matching USGS 7-1/2 minute plates at the northern and southern edges. This problem is usually minor when only two or three sheets are being used.
3. Cartographic inaccuracies may have been included during drafting.

The following methods can be used to create the Basemap from two or more plates which do not exactly match:

1. Ignore the discrepancies. This method works when very slight shifts are apparent (as with vertical USGS 7-1/2 minute plates) or when data is encoded by a process of digitizing x,y (polygon) coordinates. It is possible to align topography and linear features with a computer program during generation of the single variable data file, so that the final grid cell data bank is automatically edge-matched.
2. Align topography as closely as possible. Redraft offset linear features where necessary by referring to aerial photography or other map data.
3. Rephotograph the plate at the desired size and scale. This method can be used to produce overlays of maps which were originally drafted at different sizes or scales.

Relating Data Variables to the Basemap

Each data variable must be represented in the data bank as a true and correct scaled overlay to the Basemap. Inaccuracies of identification or spatial location of one variable can result in significant errors during data analysis. **Accuracy of analysis depends upon consistency and accuracy of the data variables.**

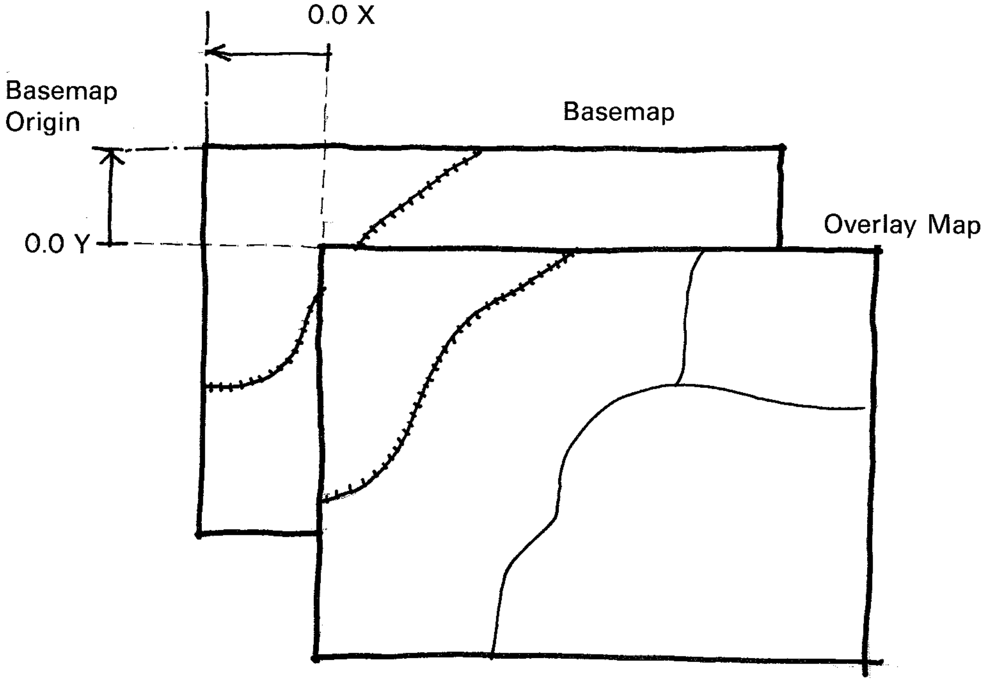
When maps are overlayed, there are several sources of potential errors in spatial location. These include:

1. Errors in the drafting of the data.
2. Different cartographic projections used for data maps.
3. Different photographic projections used to draft original data.

4. Physical changes in the materials used for the maps (shrinkage or swelling).

Spatial discrepancies resulting from these sources fall into the following four general categories of registration problems:

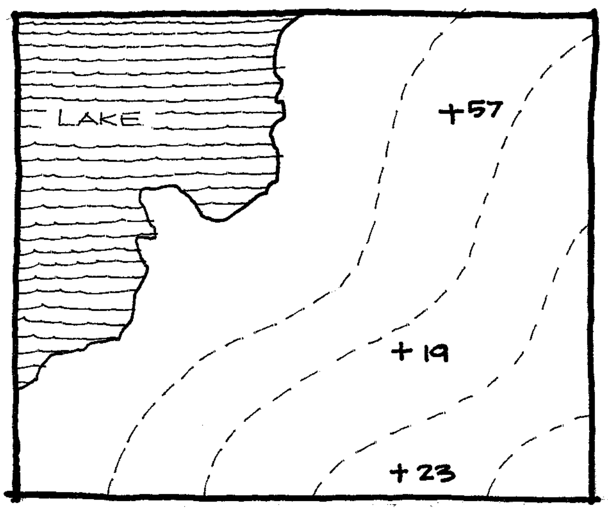
1. Translation: As illustrated in Figure VII-1 **TRANSLATION**, the same site is identified with different x,y locations on different maps. The x,y designations must be shifted to yield consistent locational information



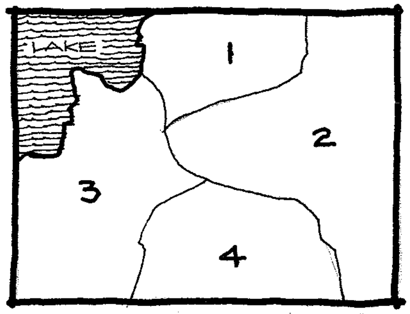
TRANSLATION

Figure VII-1

2. Scale: Figure VII-2 **SCALE**, demonstrates that maps may exist at different scales. These must be reformatted to the common scale chosen for the study.



1:5,000 Mean Annual Rainfall

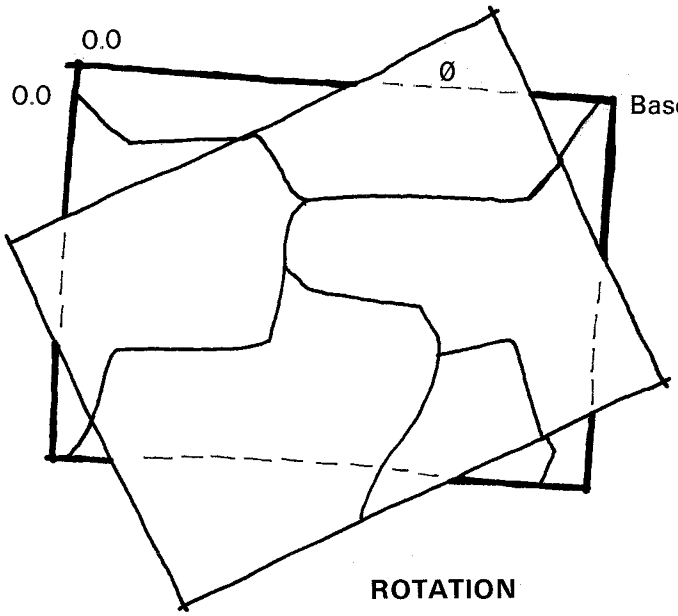


1:10,000 Watersheds

SCALE

Figure VII-2

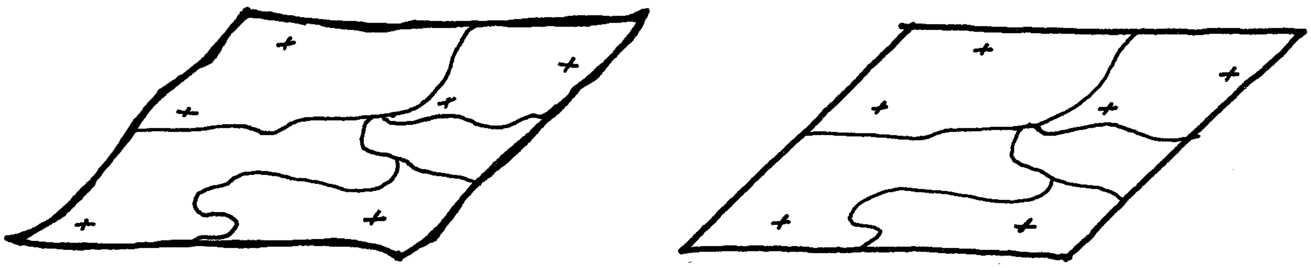
3. Rotation: Figure VII-3 **ROTATION**, illustrates that the identification of direction may not be the same on different maps. These maps must be oriented to a common coordinate system to yield consistent locational information.



ROTATION

Figure VII-3

4. Distortion: As illustrated by Figure VII-4 **DISTORTION**, portions of a map may be offset in several directions resulting in internal distortion and disagreement with the Basemap



DISTORTION

Figure VII-4

It is not uncommon to find data maps which display more than one type of registration problem. For example, soils maps drafted on aerial photographs are often distorted at the edges and do not align exactly with adjacent maps. In addition, the map scale of soils maps is frequently not the same as the scale chosen for the Basemap.

Registration problems can be dealt with in a number of ways. The most accurate method is to encode all of the data variables from the same Basemap. If each variable is accurately located on the Basemap, it will be accurately represented in the data bank. Data which are not included on the Basemap may be added before encoding by the following methods:

1. Remapping the missing data variables onto the Basemap or an accurate Basemap overlay.
2. Using a zoom transfer scope to redraft or produce a photographic overlay of the data variable.
3. Using a terrain unit mapping approach to combine natural geographic features on a single map. This approach involves overlaying data maps to produce areal units which are homogeneous for the data encoded. Appendix III is a detailed description of this approach to mapping geographic information.

The data can then be either grid cell x,y encoded directly from the Basemap.

An automated registration process can be used to register a data map to the Basemap when data is encoded by a polygon (x,y) method. In order to use this method, **a number of matchpoints must be located on both the data map and the Basemap. The matchpoints must be stable, easily locatable terrain or cultural features such as road intersections, monuments or other permanent points common to both maps.** A computer program requires a minimum number of seven matchpoints that should be scattered over the map surface including the region outside the borders of the study areas. A greater density of matchpoints can be chosen in critical or distorted areas. **All matchpoints should be identified and their locations on the Basemap recorded in a matchpoint directory for use in the registration process.**

The registration program operates by determining the coefficients of a higher order polynomial function (using least squares methods) that define mathematically the relationship between the coordinates for each Basemap matchpoint and the (different) coordinates for the same matchpoints on the variable map to be registered. The x,y coordinates of the variable data being registered are then adjusted by application of the polynomial function. Automated registration can correct map registration inaccuracies including: map data which is displaced from the chosen coordinate system (translation and rotation), drawn to a different scale, or distorted due to incompletely rectified photos, stretching or shrinking of unstable base material.

Coordinate Systems

The study area must be given a coordinate system in order to identify data locations. For grid data the coordinate system is generally the row and column location of each grid cell. The rows and columns may be related to distance from a common origin point (row 0, column 0). The origin is located at the upper left of the study map in order to relate the map to the printer output device. For polygon data, points are located by their distance from the origin. (The distance may be recorded in inches, feet or any convenient unit.)

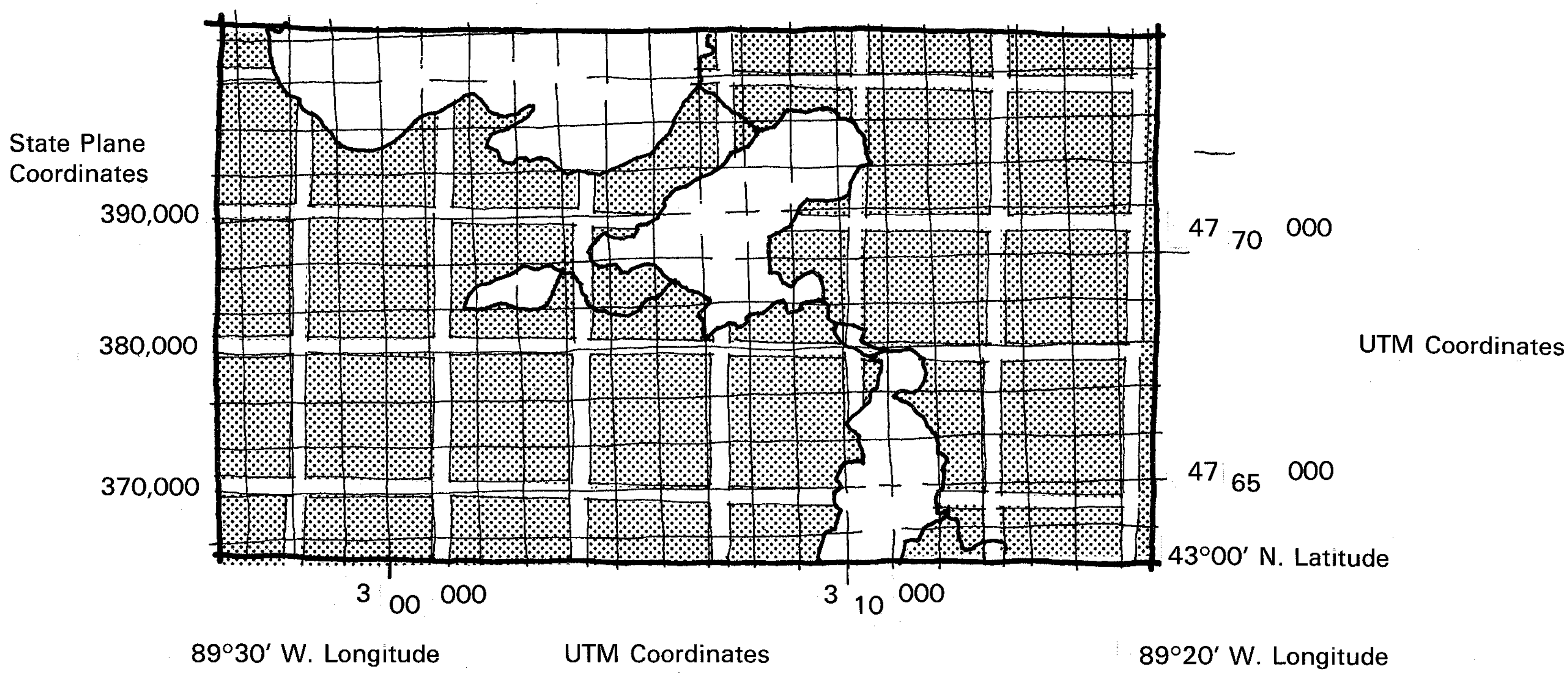
In order to facilitate accurate location of information, data should be registered to a common coordinate system which can be referenced during data transfer. Coordinates may be referenced to one of several commonly used coordinate systems. Figure VII-5 **RELATIONSHIP BETWEEN COORDINATE SYSTEMS**, illustrates the relationships between three coordinate systems most commonly used in the continental United States as they are recorded on the USGS 15 minute Madison quadrangle.

Spherical Graticule: A portion of the spherical coordinate system adopted for the earth is indicated on the outer margins of Figure VII-5. It is formed by degrees of latitude measured from the equator (0°) to each Pole (90°) and degrees of longitude measured from the observatory at Greenwich, England (0°) both East and West to the International Date Line (180°). The graticule thus subdivides the earth into grid segments based on angular measurement from the earth's center. This graticule is used internationally.

UTM Projection: The Universal Transverse Mercator (UTM) projection is employed between 84° N and 80° S latitudes. This projection is the black grid in Figure VII-5. The UTM system divides this area of the earth into 60 North-South columns of 6° longitude called zones. Distances within the zones are measured northward in meters from the equator (arbitrarily given a value of 0 for the Northern Hemisphere and 10,000,000 meters for the Southern Hemisphere). East-West distances in each zone are measured from the central meridian of each zone (which is given a value of 500,000 meters East). The UTM zones are further subdivided into grid zones of 6° longitude x 8° latitude, which are identified by numbers and letters. These are further subdivided into 100,000 meter squares, also given identifying letter combinations. This subdivision of the earth into 2-dimensional projection zones allows mapping of each zone with relatively little distortion resulting from the transfer of points on a curved surface to a flat plane.

State Plane Coordinates: In order to provide the convenience of plane coordinates for each state, the USGS has established a separate projection system for each state. An example is the white grid in Figure VII-5. These are based on the UTM projection of Lambert's Conic projection, depending on the shape of the state. To keep scale distortion to a minimum, each state usually has two or more overlapping zones, each of which has its own projection system and grid. The units used are feet. USGS topographic maps indicate the location of the 10,000 ft. grid corners by tic marks.

Notice that each of the grids superimposed on Figure VII-5 are oriented differently. This is a result of the directions of the meridians chosen for orientation of the rectangular UTM and State Plane Systems. The graticule indicates true North-South and East-West. In the UTM



RELATIONSHIP BETWEEN COORDINATE SYSTEMS

Figure VII-5

zone 16, the meridian of 87° W longitude is the central North-South axis about which the rectangular grid is arranged. In the Wisconsin Coordinate System the North-South axis of the rectangular grid is 90° W longitude.

Orienting the Basemap for a particular study area to one or another of the common grid systems is not absolutely essential, but is advised because this enables the accurate location and overlay of map features recorded with reference to these projections. During encoding, features will be located according to the grid cell in which they occur. The grid cell matrix will be recorded in the computer by its row and column position.

Applying the Grid to the Basemap

There are two methods which can be used to overlay the study grid to the Basemap.

1. A grid of the appropriate shape and size (see Chapter V) can be drafted on clear plastic (mylar) stable material. This grid is then overlayed on the Basemap, being certain that the entire study area is contained within the grid boundaries and that the grid is aligned with the coordinate system or geographic features which have been chosen as the origin and reference points for the study. The same grid can be used on other data maps of the same size and scale if care is taken to insure that the locations of the same features occupy the same grid cells as on the Basemap

2. The second method reduces the chance of error in locating grid cells and is recommended when several variables are being mapped separately for the study area.

   A grid of the appropriate shape and size (see Chapter V) can be permanently marked or "burned" on the Basemap. A grid overlay can also be aligned by using precisely located pins or holes punched near the margins which allow the grid to be registered to the desired coordinate system. Any overlays are then produced with the grid in place and this insures that grid cells are in the same location for each variable. It is generally useful to have a separate copy of the Basemap or overlay for each variable.

   If an automated digitizer is being used, each overlay map should be provided with accurate geographic locaters or tic marks. The tic marks can then be aligned by a computer program to produce an accurate overlay of each data map as it is entered into machine storage.

   In the preparation of maps for encoding the following steps should be taken:

   a. All maps should be on stable base material to minimize distortion.

   b. All maps should be complete with ploygon, line or point data as well as annotation assigning code variations to each feature

   c. Each map should be examined for missing or extraneous polygons, lines, points or codes, or other inaccuracies.

   d. A unique number (or color) should be assigned to each variable to aid identification and avoid confusing one variable with another The identification number (or color) should be clearly specified for encoding.

   e. When the study area covers more than one map module, these are edge-matched to insure that the boundaries match and that the various data types match. Discrepancies should be checked in the field or on aerial photos and the map edge redrafted accordingly.

   If x,y encoding techniques (digitizing) are being used, the steps above should be followed by sequence numbering.

   f. The map elements should be numerically sequenced for digitizing. Sequence numbers are a permanent address to each map element in the data bank.

—Polygons should be numbered using a "nearest neighbor" concept in which the closest polygons are numbered consecutively. This allows efficient digitizing and aids subsequent computer processing. A convention is required for sequence-numbering polygons totally included within larger polygons. When using **AUTOMAP II** and **GRIPS** (ESRI, 1977), these polygons should be numbered after their surrounding (donut) polygon.

—Linear data should be assigned sequence numbers hierarchically by order (e.g., all first order streams are numbered first, followed by second order streams, etc.).

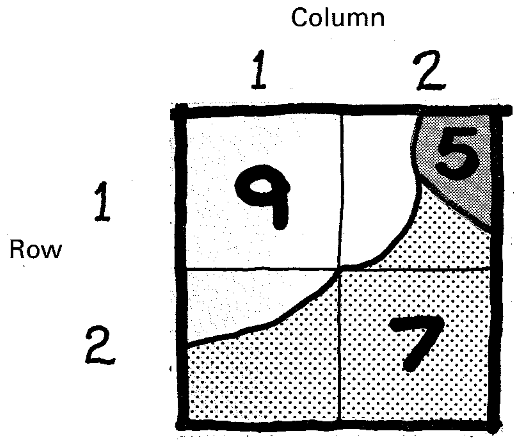
CHAPTER VIII
DATA ENCODING

Encoding is the process of converting data variables from map form into a machine-readable numeric form. Several methods are commonly used for encoding variables and the user should keep in mind the purpose of each variable and hardware requirements when choosing a particular method. The end product of encoding variables is not simply the creation of the data bank. Rather, encoding is done to provide accurate data which will be combined or overlaid with a number of other variables to produce interpretive data and analyses. These may be end products or they may also be added to the data bank for additional use.

The types of encoding vary according to the use of a grid or polygon format, the accuracy desired for variable boundaries, and the type of computer storage available. Both grid cell and polygon (x,y) encoding methods can be used to generate a grid cell data bank. These encoding methods are described below:

Cell Encoding Strategies

Grid cell encoding involves assigning to each cell a numeric value representing the variable class contained within that cell. The numeric value may be an actual value (e.g., 321.4 for elevation) or an assigned legend value (e.g., 14 designating a specific habitat type). If the value or classification of the variable is homogeneous throughout the area of the cell, no difficulty occurs in selecting the value to be encoded. Many times, however, more than one value for a variable will exist within the boundaries of the cells (e.g., 2 or more soil types within a cell soils data), and a decision must be made about which value to assign to the cell. The cell in Row 1, Column 2 of Figure VIII-1 **ENCODING STRATEGIES**, for example, contains data types 5, 7 and 9. Four different strategies are described here which allow encoding of such cells. Once an encoding strategy is adopted for a particular variable, the variable should be encoded by the same method throughout the study area. Different strategies can be used for different variables.



Row \ Col	1	2
1	9	7
2	7	7

CENTROID METHOD

Row \ Col	1	2
1	9	9
2	7	7

PREDOMINANT TYPE

Row \ Col	1	2
1	9	5
2	7	7

MOST IMPORTANT TYPE (Assume highest priority of 5, then 7, 9)

Data Value

Row	Column	1	2	3	4	5	6	7	8	9
1	1	0	0	0	0	0	0	0	0	100
1	2	0	0	0	0	25	0	30	0	45
2	1	0	0	0	0	0	0	70	0	30
2	2	0	0	0	0	0	0	100	0	0

PERCENTAGE BREAKDOWN

ENCODING STRATEGIES

Figure VIII-1

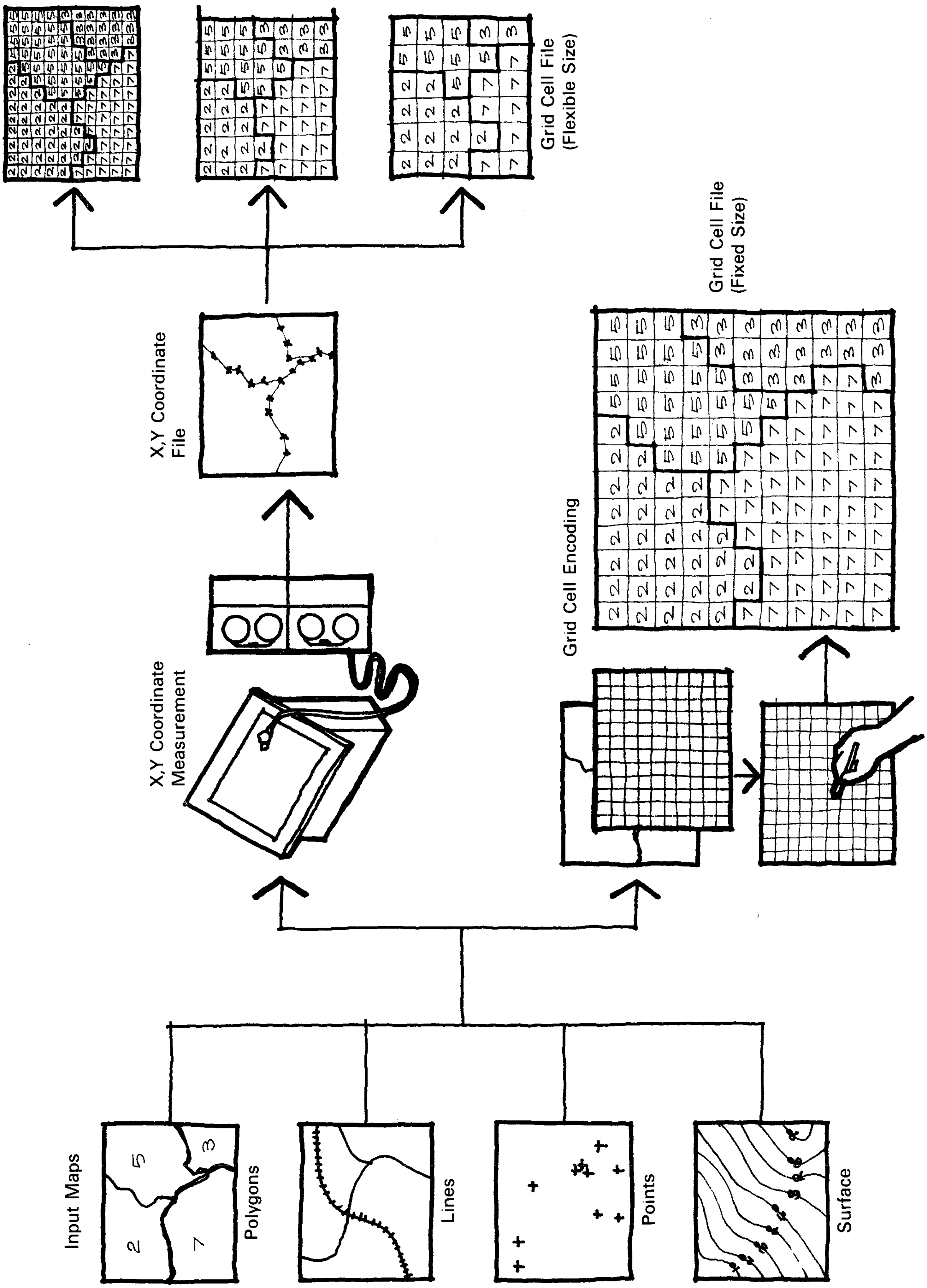
Encoding sheets should be designed to facilitate both the strategy selected and the entry of the data into computer storage. Keypunching directly from the encoding sheets reduces the chance of error when data is transferred to another form. Note that the example encoding sheets in Figure VIII-1 can be used directly for keypunching.

1. Centroid Method — each cell is assigned the value which exists at the centroid (center) of the cell. This method can be used for any variable type, and is particularly useful for continuous variables such as elevation. Using this technique, the example grid cell Row 1, Column 2 would be given the value of 7
2. Predominant Type — each cell is assigned the value for the variable which occupies a majority of the cell. Thus, a cell containing 20% silty sand and 80% organic clay soil types would be coded organic clay. This method is most often used for discontinuous variables. With this technique, the example grid cell Row 1, Column 2 would be given a value of 9.
3. Most Important Type — each cell is assigned the value for the variable which is considered most important to the study. For an environmental quality study, a cell in which 20% of the area is a wildlife preserve and 80% is in agriculture may be coded wildlife preserve. This method is valuable when clear distinctions of importance can be made. Using this technique, the example grid cell Row 1, Column 2 could be given a value of 9, 7 or 5, depending on their relative importance.
4. Percentage Breakdown — each cell is assigned several values according to the percentage of the cell occupied by each value. While this method is useful for storing statistical data, this strategy is not recommended because it does not permit straightforward geographic analysis and it requires additional storage and analysis.

For a majority of planning studies, cells containing two or more classes of data should be encoded using the centroid, predominant or most important type strategies. The choice of a particular strategy involves a trade-off decision between the importance of a particular data class and the area which it occupies.

Cell Encoding Techniques

Each variable should be encoded separately. Several techniques can be used in the manual phase of encoding variables. One grid encoding technique and one using x,y coordinates are illustrated in Figure VIII-2 **MAP ENCODING TECHNIQUES**. This figure shows that the basic data maps can be encoded by an x,y technique (using an electro-mechanical digitizer in the figure) or by a grid cell technique. Both methods result in a grid cell file. The x,y encoding method produces an intermediate x,y data file which can yield grid cell files of any dimensions.



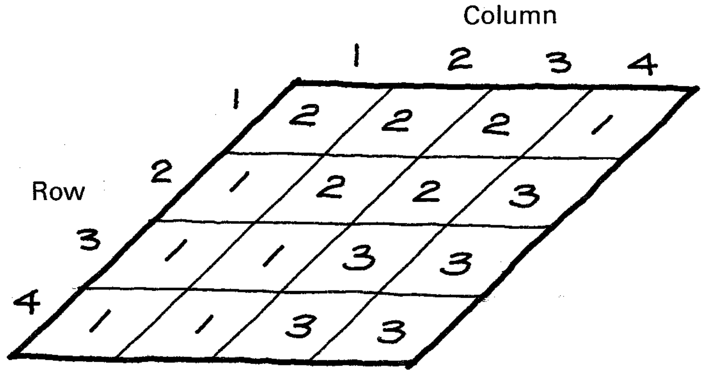
MAP ENCODING TECHNIQUES

Figure VIII-2

1. 1:1 Encoding — using this method, each grid cell is coded with the value assigned to it by the strategy selected above. Each cell is then processed into storage, resulting in a numeric matrix in the computer. Cells outside the study area are encoded "0" or left blank.

2. Run Encoding — using this method, **only those cells in a row where a variable changes value are encoded**, the others are left blank. Processing software "fills in the blanks" for the stored record so that analysis occurs just as though 1:1 encoding has been used. This method is useful for variables which do not change value over a space of several cells (i.e., large polygons).

   Keypunching this format requires that each row be entered and the column and new value for the grid cell where the value change occurs be entered into storage, as illustrated by the encoding sheet in Figure VIII-3 **RUN-ENCODING SCHEME**. Polygon files from digitized data can be transformed into run-encoded records as well. Because two entries (column and new value) are required for each entry, storage space is saved only when the variable is homogeneous over an average distance greater than two grid cells.



ENCODING SHEET

Row	Column	Value
1	3	2
1	4	1
2	1	1
2	3	2
2	4	3
3	2	1
3	4	3
4	2	1
4	4	3

RUN-ENCODING SCHEME

Figure VIII-3

The 1:1 encoding method is directly related to the actual map features, and requires the smallest amount of computer storage space for continuous variables such as elevation. Run-encoding allows a significant saving of coding time and storage space for discontinuous variables, where homogeneous areas are of a large size.

Polygon (x,y) Encoding Techniques

Polygon encoding can be used as an intermediate step to developing grid cell data banks. This method involves recording the x,y coordinates of points or strings of points to identify the locations of map data. The boundary between data types can be recorded as accurately as it is drafted on the original maps. Point data is recorded as individual x,y coordinates; lines are recorded as strings or series of points; and polygons are recorded as closed strings of point coordinates. Curved lines are generalized as a series of short straight lines, accuracy depending on the length of these straight segments. Polygon coordinates may be recorded by three strategies:

a. Common Boundary Digitizing — each polygon is digitized completely. This results in common borders being digitized twice. The arrows in Figure VIII-4 **ENCODING X,Y COORDINATES**, show that line a-b is recorded once with polygon D and again with polygon E. Common boundary digitizing is the preferred technique for reasons of editing, storage space, and final map accuracy, but involves more digitizing time.

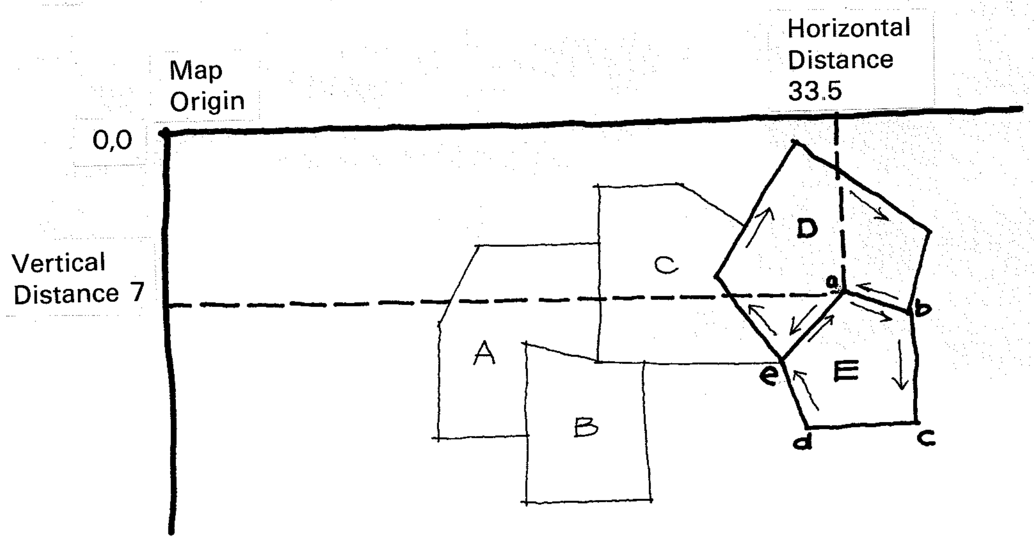
b. Vertice Point Digitizing — this strategy results in a file for the location of each vertice or point (e.g., point 1 at x = 33.5, y = 7). Polygons are then encoded by the points which comprise them rather than by their x,y locations (e.g., polygon E, value 185, points 1,4,10,12,22,410, 1). This method can save substantial analysis time in the computer by separating polygon values from polygon locations. Another version requires that each point be recorded along with all data values for which it applies. A point at the corner of three polygons, such as point e in Figure VIII-4, would include a record of its location and three data values. The polygons are reconstructed by special software which requires the user to specify all points defining a polygon along with the code or data value of the polygon.

c. Line Segment Digitizing — each line segment is recorded as a series of points (to define the line) and two values (which the line separates). In Figure VIII-4 line a-b would be recorded as the two points a and b plus the two data values E and D. Polygons are reconstructed by specifying the polygon identifying code plus all of the lines which identify the polygon. Line segment digitizing can also be done by creating separate location and value files as described for vertice point digitizing. In this case the location of line segments, rather than points, is used to identify the polygon. This strategy is useful when proven computer software is available. It does not allow for the ease of editing provided by common boundary digitizing.

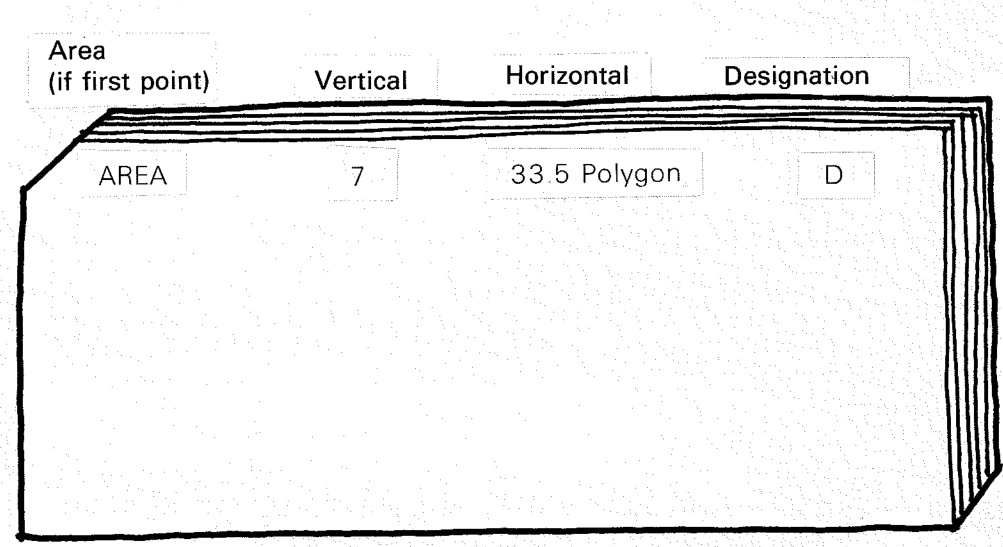
Polygon (x,y) Data Capture Techniques

Two methods can be used to encode x,y coordinate data into machine-readable form:

a. Manual encoding: In Figure VIII-4 each point where the line being coded changes direction is identified and its coordinates with respect to the coordinate system chosen are recorded, along with an identifying code for the point, line or polygon.



The point illustrated would be encoded:



ENCODING X,Y COORDINATES

Figure VIII-4

b. Electro-Mechanical Digitizing: Polygon or x,y encoding can be done with an electromechanical device called a digitizer. The digitizer converts movements of a mechanical cursor or "point locater" into electronically identified locations which can be read into the computer. Digitizers are usually designed to be accurate within a few thousandths of an inch. When using this method of encoding variables, the coordinate system is also recorded by digitizing three or four matchpoints or tics of known location, usually on the outer portions of the map. This enables accurate merging of adjacent maps. Each point, line or polygon is encoded as for manual x,y encoding along with an identifying code. The digitizer automatically records the position of each point, and converts it to a machine-readable form.

Table VIII-1 **COMMON METHODS FOR ENCODING SPECIFIC VARIABLES**, indicates coding procedures which have been effective for certain variables in past studies. The actual choice of a particular technique depends upon the amount of variation in a variable within the study area.

TABLE VIII-1

COMMON METHODS FOR ENCODING SPECIFIC VARIABLES

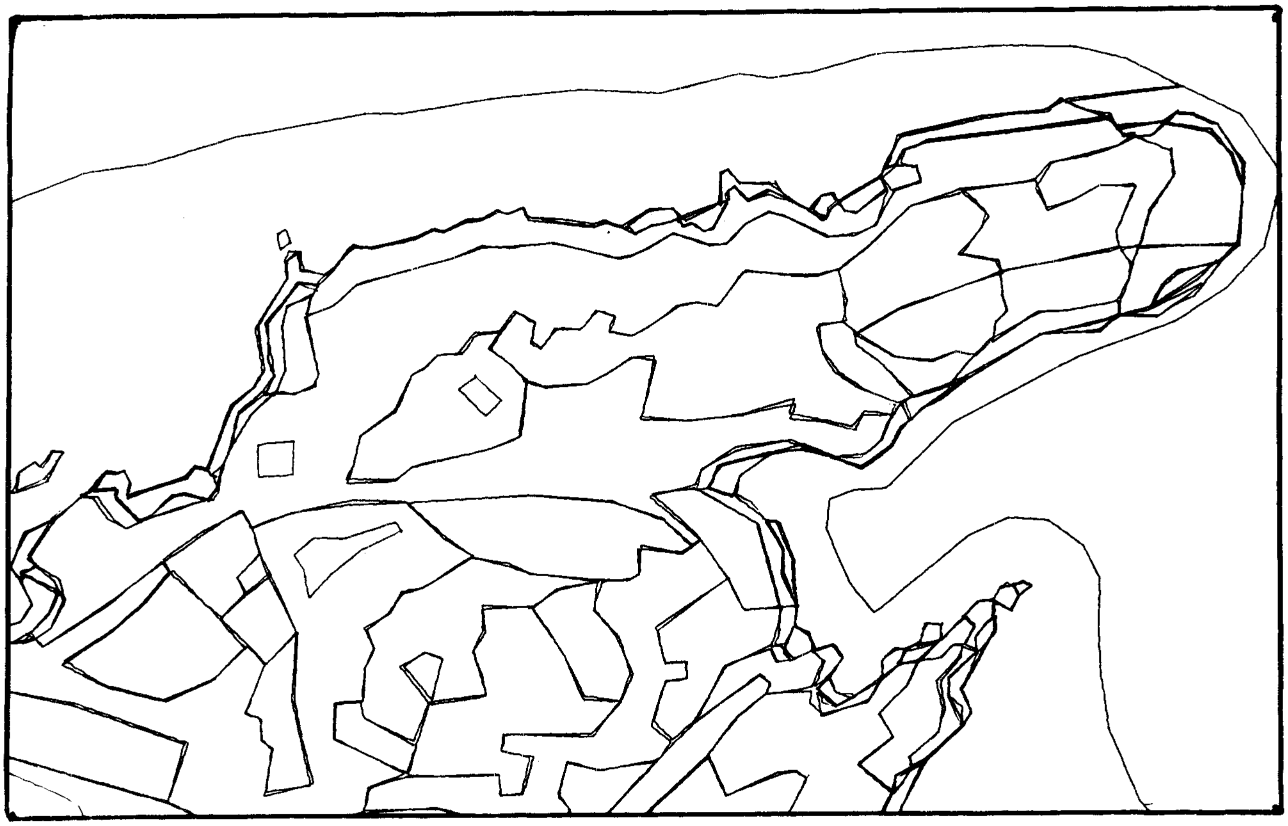
Variable	Encoding Technique
Elevation	1:1 Grid
Base Land Use	Polygon, Run-Encoding
Future L.U.	Run-Encoding for Changes
Soil Data ● Soil Types ● Hydrologic Group ● Slope Group ● Erosion Index	Polygon, Run Encoding Derive by Computer Processing from Soil Type
Reference Flood Elevation	Computer Surface Generation 1:1 Grid
Areal Units ● Subbasins ● Damage Reach ● Political Boundaries, etc.	Polygon
Linear Units ● Transportation Systems ● Stream Networks ● Pipelines and Utilities, etc.	1:1 Grid, Polygon
Habitat I.D.	Polygon, Run-Encoding
Critical Habitat I.D.	Polygon, Run-Encoding
Archaeological/Cultural Features	Polygon, Run-Encoding
Geology	Polygon, Run-Encoding
Vegetation	Polygon, Run-Encoding

Editing Data Files

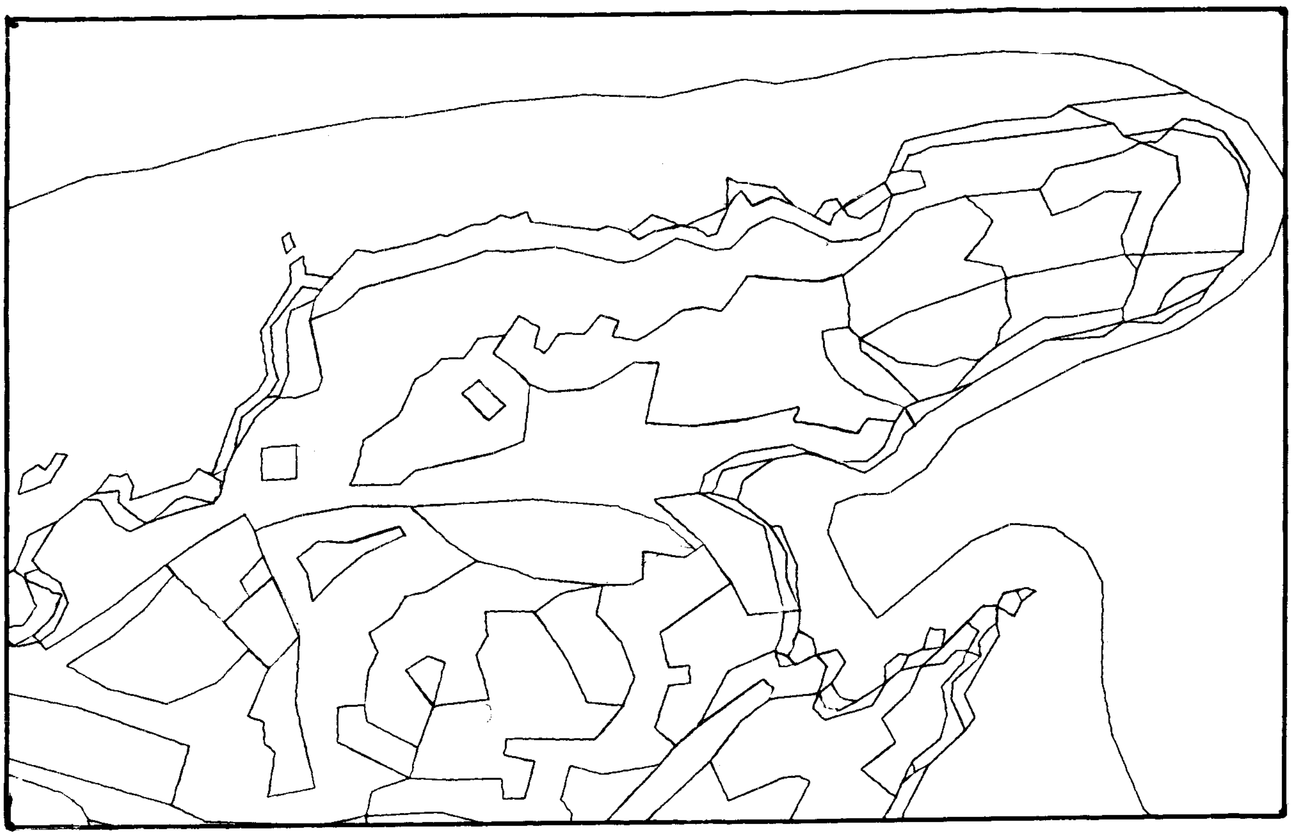
Editing variables encoded by grid cell methods is done on a cell-by-cell basis. The code sheets should be verified by a person other than the original coder. Printer output maps can be visually compared with the original data maps, and values which were recorded for each cell can be listed for verification. Several edit computer programs can be written to correct data during the creation of the single variable file.

Manual and machine editing of encoded data to remove errors is probably one of the most time consuming tasks in creating a data bank. It is, however, one of the most important facets of any study. Editing is required to assure accurate, systematic, consistent assessment of land use and geographic patterns. It is almost impossible to over-check the data bank for accuracy.

Editing polygon data involves graphic review. The common boundary digitizing technique recommended (i.e., digitizing polygons individually means that contiguous borders are digitized twice), will result in identifiable sliver errors wherever points were improperly digitized. Figure VIII-5 **COMPARISON BETWEEN INITIALLY DIGITIZED FILE AND CLEAN FILE**, shows the gaps which occur when boundaries are digitized twice. **Grid cell conversion absorbs many of these slivers, but large ones may remain.** Different tolerances are sometimes set, based on scale and desired accuracy standards. While line segment digitizing is possible, the editing benefits of double digitizing is sacrificed.



a



b

COMPARISON BETWEEN INITIALLY DIGITIZED FILE AND CLEAN FILE

Figure VIII-5

Recognizing that common boundary digitizing spotlights errors, two edit steps can be accomplished prior to generating a plot or printer map for a visual edit. The first is to check that all polygon boundaries return to their starting points, or are "closed". The other edit involves identifying contiguous line segments and matching them point for point to determine points to be added or deleted. These edits eliminate most of the graphic and statistical problems that occur when common lines do not coincide exactly. Following these edits, each file is then edited to resolve polygon hierarchy (i.e., polygons contained entirely within other polygons).

When using a polygon system, maps can be generated on a pen plotter for each variable. These maps are used to visually check the accuracy of the digitized and machine edited x,y coordinates.

Graphically editing polygons is the simplest and least costly method currently available. If each of the plots from the automated files matches the original map within acceptable limits, then the files are assumed to be sufficiently accurate

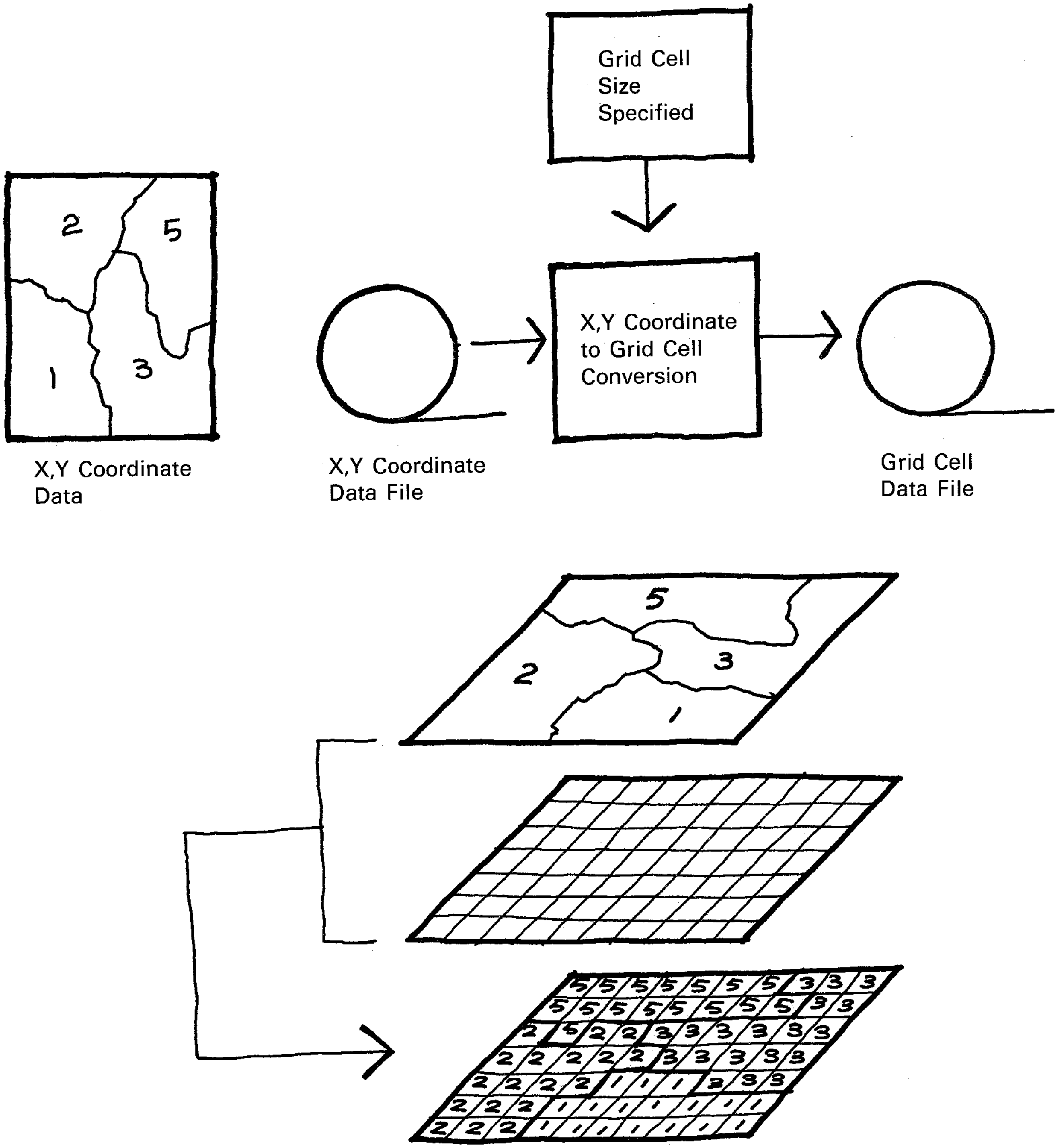
If x,y coordinate software is available, data errors discovered in the edit are relatively easy to correct. Polygon edit software allows code corrections, point changes, etc. In more complex cases, such as missed polygons, the entire polygon must be redigitized. This involves remounting the original mylar map and redigitizing the entire polygon in error. This redigitized information can be readily merged with the previous information because the tic marks are also redigitized, insuring perfect alignment when the data is transformed to the study coordinates. After each variable has been digitized and edited, the file is ready to be transformed into a grid file which can be subsequently merged into a grid cell data bank.

Converting Polygon Files to Grid Cell Files

The advantage of x,y coordinate encoding is the ability to capture accurately the boundaries of polygons and shapes of lines. The use of grid cell analysis is more desirable when rapid cost-efficient processing is required. **For multiple-purpose studies and those of a continuing nature, it is often advantageous to digitize the variables by the polygon method and retain these x,y single-variable files separately off-line.** When polygon coding techniques are used, grid cell matrices of various shapes and sizes can be generated from the x,y files as individual studies or analyses require them. The exact grid cell size and shape is determined by the size of the output map and the number of grid cell rows and columns if a line printer is used as an output device.

Several software processes can be used to convert x,y coordinate files to grid cell files. The Hydrologic Engineering Center currently makes use of the **AUTOMAP II** (ESRI, 1977) software for this purpose. This software generates grid cell data as a by-product of polygon data.

Figure VIII-6 **CONVERTING X,Y COORDINATE DATA TO A GRID CELL DATA FILE**, illustrates the sequence of steps normally required to generate a grid file from polygon data. An x,y data file is read by the conversion program and the data is aggregated to a grid of specified dimensions (specified numbers of rows and columns). The grid specified must account for width of the output map (strip records), horizontal and vertical scale factors to relate the grid cell size and shape to a known map unit, and the grid cell size chosen for the study.



CONVERTING X,Y COORDINATE DATA TO A GRID CELL DATA FILE

Figure VIII-6

CHAPTER IX
CREATING THE GRID CELL DATA FILE (DATA BANK)

Data variables which have been registered accurately, encoded into machine-readable form and stored initially as a single-variable file are ready to be processed into the data bank. The single-variable files have been encoded as a grid matrix directly or an x,y polygon file and converted to grid cell files as described above.

A copy of each of the edited or "clean" single-variable data files is merged into a multivariable file which becomes the grid cell data bank, and will include the values for each variable assigned to each grid cell. Figure IX-1 **THE CONCEPT OF A MULTIVARIABLE GRID CELL FILE**, illustrates that conceptually the various grid cell maps are stacked together, mathematically, a new matrix is generated to contain all values for each cell. The software reads the record of the variable value for a grid cell and stores it sequentially with the records previously entered for other variables entered for the same grid cell. A computer program does this for each cell in the data bank.

The first pass through a data bank computer program establishes the sequence and number of variables and their formats. Space is then available for addition of the remaining variables.

The matrix is stored on a computer tape or disc in a systematic fashion so that efficient access and manipulation is possible. The sequential convention of grid cell data storage most frequently adopted consists of storing all data variables for each cell sequentially cell by cell from the upper left coordinates to the lower right coordinates. This method permits efficient retrieval of multiple data variables for each cell so that, regardless of the number of variables in the data bank to be used in an analysis, or the number of grid cells for each variable, only one search of the data bank is necessary. Computer storage capacity to process only the number of variables involved in the analysis is required by this method.

A method now in the experimental stage allows multivariable files to be structured according to a run-encoded method. This structure includes a position indicator for the grid cell, the variable number, and the new value for that variable which occurs in that cell. As with other run-encoding concepts, this method would increase storage and analysis efficiency for variables which are of relatively homogeneous classification (large polygons).

Each run-encoded record is larger than the corresponding records for 1:1 representation, but in this case records are entered only for those cells where a variable changes value. The usefulness of this coding/storage method in a given study depends on the frequency of changes in value which are recorded for the variable being considered.

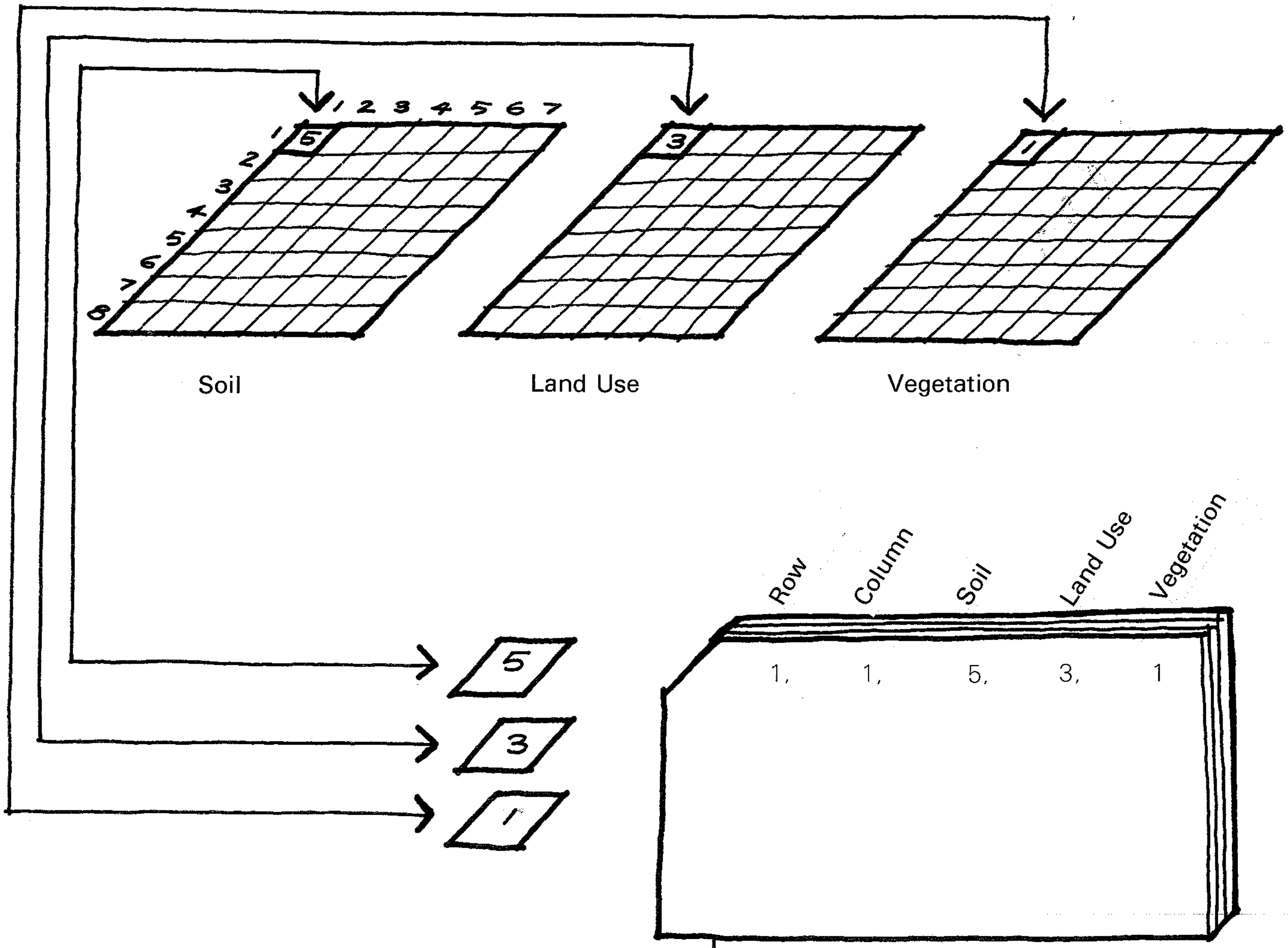
If this method is used, two data banks must usually be created, one using 1:1 records for such continuous variables as elevation, and the other using run-encoded records for discontinuous variables such as land use.

Initial editing of the multivariable file is accomplished through examination of print-outs of data values and locations.

Corrections of the multivariable file may be made by reentering corrected single-variable files to replace existing data, or by replacing erroneous data in a particular cell with new data.

Packing the Data Bank

The grid cell data bank may be "packed" to exclude those grid cells which lie outside the study area. Grid cell locations (row and column) are specified only within the study area boundary, requiring less computer storage space and analysis time.



Typical Multivariable
Grid Cell File Layout

Location		Variables					
Row	Col	Soil	LU	Veg	Elev	Slope	...n
1	1	5	3	1	27	3	8
1	2	5	2	1	28	2	7

THE CONCEPT OF A MULTIVARIABLE GRID CELL FILE

Figure IX-1

CHAPTER X
UPDATING FILES AND EXPANDING THE DATA BANK

Because data variables in an area may change through time and the accuracy of some data values may require refinements, procedures are necessary to update, correct or change information which has been stored in a grid cell data bank. In addition, new variables may be added to those already stored in the data bank. There are three basic strategies which are commonly used to update old data or add new variables.

Window Updating

This strategy requires identification of the variable and the cell or a contiguous group of cells (a "window") to be updated. The individual cell or group of cells is then isolated by a computer program and the new data value is entered to replace the value which was originally stored in the data bank.

General Updating

When it is necessary to update many cells scattered throughout the study area (rather than in a few discreet "windows"), a general update program can be used to read each cell's record of the variable being changed, make the stipulated change in value where necessary, and copy those cells which do not require a change.

Adding New Variables

In some cases, a new data variable may be required for a study. The new variable may be prepared by following the steps in this manual for data acquisition, classification and encoding to create a single variable file. During these steps, particular care should be paid to the registration of the new data with the original Basemap for the study area. The new single variable grid cell file can then be merged into the data bank by the same computer software which was used to create the multivariable data bank (described in Chapter IV).

Using update techniques, a study may include analysis of the results of projected future changes in land use, vegetation, urbanization, pollution sources, or other activities. The analysis of the results of future patterns can then be compared to present and past conditions to provide planners with the probable effects of alternatives. If this is done, special care must be taken to assign logical future changes to individual grid cells, as the data bank will accept even illogical changes of value. For example, it is possible to change a land use code in the data bank from a present heavy industrial use to a future use as pasture. This particular change rarely occurs in real life, and the results of any analysis based on this change would probably not reflect actual future conditions. A printout of a table of coincidents, such as the one in Table X-1 **COINCIDENTS MATRIX**, will indicate what types of changes have been made by the update. This table displays existing land use categories (rows) which will change to accommodate a future land use pattern (columns) for a specific geographic part of a study area that was catalogued in the data bank as a separate variable. In this example, row 7, column 9 indicates that 4.59 acres in the study area have been changed from present industrial use to future pasture, an unlikely change. The ability to find such discrepancies allows the user to correct the update to reflect potential changes which are more realistic.

TABLE X-1

COINCIDENTS MATRIX
(Areas in Acres)

	1	2	3	4	5	6	7	8	9	10	TOTAL
1	105.57	0.00	0.00	6.12	0.00	0.00	41.31	27.54	91.80	0.00	272.34
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	3.06	0.00	0.00	0.00	0.00	0.00	0.00	3.06
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	6.12	0.00	0.00	0.00	0.00	0.00	6.12	0.00	7.65	0.00	19.89
7	4.59	0.00	0.00	0.00	0.00	0.00	13.77	0.00	4.59	0.00	22.95
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	24.48	0.00	0.00	1.53	0.00	0.00	10.71	9.18	24.48	0.00	70.38
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TOTAL	140.76	0.00	0.00	10.71	0.00	0.00	71.91	36.72	128.52	0.00	

Row Categories are Existing Land Use

1 Natural Vegetation
2 Developed Open Space
3 Low Density Housing
4 Medium Density Housing
5 High Density Housing
6 Agriculture
7 Industrial
8 Commercial
9 Pasture
10 Waterbody

Column Categories are 1990 Land Use

1 Natural Vegetation
2 Developed Open Space
3 Low Density Housing
4 Medium Density Housing
5 High Density Housing
6 Agriculture
7 Industrial
8 Commercial
9 Pasture
10 Waterbody

CHAPTER XI
COMMERCIAL ASSISTANCE

Commercial assistance in data bank preparation may be required in the following cases:

1. Special expertise may be needed to acquire certain types of data in an accurate form and efficient manner.
2. The services of a company with experience designing automated data base studies can be valuable when using new technology
3. Hiring a firm to digitize map data and produce x,y data files can reduce the hardware requirements of the study team.
4. Automated data files already exist for some areas. These may be purchased from a commercial source.

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

1. The size of the area to be studied, and its location. These should be determined as accurately as possible, as the cost of the study relates directly to the area and travel required to gather data.
2. The type of analysis for which the data bank is being prepared will indicate the detail and accuracy which will be required.
3. Which variables will be contracted out, how they are to be acquired, what level of accuracy is desired and the classification scheme to be used will influence the amount of effort required of the contractor.
4. The type of encoding desired should be specified, including the accuracy of data resolution and the final size and shape of the grid cell output.
5. The specifications of the final data file delivered (generally magnetic tape readable by the analysis system used by the study team) should be indicated.
6. The number and quality of output maps or tables which will be required to check the contractor's performance and the accuracy of the data bank
7. What provisions should be made for updating or correcting the data bank
8. The time period anticipated for completion of the contract should be indicated

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 for creation of a useful data bank are:

1. Experience creating data banks of the quality desired. Possession of a digitizer or hardware system does not in itself indicate that a contractor can produce a meaningful or accurate data file.
2. Demonstrated ability and willingness to follow-up. The contractor should be able and willing to update data files if necessary in the future, rather than performing one large project with no interest in continued smaller-scale efforts.
3. Thorough understanding of the type of data bank desired and the purposes for which it will be used.

When chosen, the contractor should be provided with the items necessary to complete the contract.

1. Basemap of the study area.

2 Classification and legending scheme of the data (unless the contractor is responsible for this).

3 Data maps with registration marks. These should be accurately marked with the points, lines or polygons to be encoded. Each data mark should be clearly associated with an identifying code

Editing of the contract deliverables can be done by comparing the output maps and tables with the original data provided to the contractor. If the two do not match within the level of accuracy specified in the contract, the errors should be rectified by the contractor. If the contractor has acquired or generated original data, this should be checked by field sampling to insure the level of accuracy specified in the contract.

CHAPTER XII
SUMMARY

The following list summarizes the activities which need to be undertaken and the decisions which need to be made which will lead a user to the successful creation of an accurate and useful grid cell data bank.

1. Define the problem to be studied.
2. Define the area to be studied.
3. Select the types of analysis required to solve the problem in the study area.
4. Decide on whether or not to use a grid cell data bank.
5. Identify and collect the variables to be placed in the data bank.
   —what is available
   —what is required for the analysis
   —what can be obtained or generated
6. Classify the variables into a useful system.
7. Acquire or generate data maps.
8. Select or create a Basemap.
9. For grid encoding, decide shape and size of grid cell and draft grid on Basemap.
10. Select encoding method(s) and strategy.
    —accuracy desired
    —capability
11. Locate matchpoints and record in matchpoint directory.
12. Encode or digitize each variable.
13. Register each digitized variable not encoded from the Basemap to Basemap matchpoints.
14. For x,y data, determine size of grid cell required.
15. Convert each variable to grid cell single variable file.
16. Merge all single variable grid files into the multivariable grid cell data bank.
    —Data bank is now available for use in analysis.

Continued use of the data bank may require the following steps:

17. Update single cells and windows as necessary.
18. Update entire data bank as necessary.
19. Add new data variables as necessary.

REFERENCES

Agricultural Stabilization and Conservation Service, 1975 **Code Symbols for the Identification of Aerial Photography**, United States Department of Agriculture, Salt Lake City, Utah

Agricultural Stabilization and Conservation Service, 1975. **Comprehensive Listing of Aerial Photographic Coverage**, United States Department of Agriculture, Salt Lake City, Utah.

Agricultural Stabilization and Conservation Service, 1978 **Prices for Aerial Photographic Reproductions**, United States Department of Agriculture, Salt Lake City, Utah.

Agricultural Stabilization and Conservation Service, 1975. **Status Maps of Aerial Photography of the Agricultural Stabilization and Conservation Services**, United States Department of Agriculture, Salt Lake City, Utah.

American Society of Photogrammetry, 1975. **The Manual of Remote Sensing**, Falls Church, Virginia.

Environmental Systems Research Institute, 1977. **AUTOMAP II Users Manual**, Redlands, California.

Environmental Systems Research Institute, 1977 **GRIPS Users Manual**, Redlands, California

The Hydrologic Engineering Center, 1977. **Annual Report**, United States Army Corps of Engineers, Davis, California.

The Hydrologic Engineering Center, 1975. **Phase I Oconee Basin Pilot Study — Trail Creek Test**, United States Army Corps of Engineers, Davis, California

Link, L. E and Aaron, G., 1977. **Remote Sensing of Land Use: Pitfalls and Potential**. A paper presented at the International Symposium of Urban Hydrology, Hydraulics and Sediment Control, University of Kentucky, Lexington, Kentucky.

Lovell, Troy and Smith, Milburn, 1977. **Data Management for Floodplain Dynamics Studies**, Fort Worth, Texas.

National Cartographic Information Center, 1977 **Topographic Maps**, United States Geological Survey, Reston, Virginia

Pacific Gas and Electric Company, 1978 **The California Automated Resource Inventory System**, San Francisco, California.

"Photogrammetry", Photogrammetric Engineering, 29, pp 761-799, 1963.

Way, Douglas J., 1973 **Terrain Analysis: A Guide to Site Selection Using Aerial Photographic Interpretation**, Dowden, Hutchison, and Ross, Inc., Stroudsburg, Pennsylvania.

APPENDIX I
SELECTED BIBLIOGRAPHY
OF
AUTOMATED DATA BANK APPLICATION

APPENDIX I
SELECTED BIBLIOGRAPHY OF AUTOMATED DATA BANK APPLICATIONS

1. Davis, Darryl W., May, 1976. **Analytical Tools in Planning Nonstructural Flood Control Measures**, paper presented at Seminar on Nonstructural Flood Plain Management Measures, Hydrologic Engineering Center, Davis, California.
2. Davis, Darryl W., September, 1976 **Comprehensive Flood Plain Studies Using Spatial Data Management Techniques**, paper presented at Twelfth American Resources Conference, Chicago, Illinois.
3. Lovell, Troy L. and Smith, M. D., April 1977. **Data Management for Flood Plain Dynamic Studies**, Corps of Engineers, Fort Worth, Texas.
4. Maryland Department of State Planning, **The Maryland Automated Geographic Information System**, Baltimore, Maryland
5. Nelson, T. L., September, 1970. **Synthetic Unit Hydrograph Relationships — Trinity River Tributaries**, Fort Worth-Dallas Urban Area, paper presented at Seminar on Urban Hydrology, Hydrologic Engineering Center, Davis, California.
6. Pacific Gas and Electric Company, 1978. **The California Automated Resource Inventory System**, San Francisco, California.
7. Corpozulia (CIUR), 1977. **Zulia Regional Planning Study**, Faculdad de Arquitectura, Universidad del Zulia, Apartado 526, Maracaibo, Venezuela.
8. Webb, R. P. and Burnham, M. W., 1976. **Spatial Analysis of Nonstructural Measures**, paper presented at Rivers 76 Symposium, ASCE meeting, Fort Collins, Colorado.
9. Statewide Air Pollution Research Center, 1971. **Air Pollution Computer Mapping of South Coast Air Basin**, Project Clean Air, Statewide Air Pollution Research Center, University of California at Riverside, Riverside, California, (Dr Joe Behar).
10. Pacific Gas & Electric Company, 1971. **Development of Grid Based Biological Impact Models for Power Plant at Davenport, California**, Pacific Gas and Electric Company, 77 Beale Street, San Francisco, California 94106, (Mr. Shiraz Kaderali).
11. Ministry of Transportation, 1974 **Computer Studies of New Town Impacts for Suzuran-ko, Japan**, Ministry of Transportation, 9-8, 7-Chome, Ginza Chuo-ku, Tokyo, Japan.
12. Office of Water Quality, Planning, and Management, 1975. **Automated Environmental Resource Information System — AERIS** — New Castle County, Delaware, Office of Water Quality, Planning, and Management, Peddlers Village, One Peddlers Row, Newark, Delaware, (Ms. Merna Hurd).
13. Comprehensive Planning Organization, 1971. **Soils Mapping System for Planning (Polygon Information Overlay System)**, San Diego, California, Comprehensive Planning Organization, 1200 3rd Avenue, Security Pacific Plaza, San Diego, California 92101, (Mr. Ross Hall).
14. City of Los Angeles, 1972. **Land Use Planning and Management System (LUPAMS)** — Los Angeles, California, City of Los Angeles, Planning Department, Los Angeles, California 90012, (Mr. Glen Johnson).
15. Southern California Edison Company, 1973, 1975, 1976. **Land Use Mapping and Overlay Studies**, Southern California Edison Company, P.O Box 800, Rosemead, California, (Mr. Ralph Crouch).

16. Department of Housing and Urban Development, 1973. **Social Area Analysis for 30 U.S. Cities**, Department of Housing and Urban Development, 451 7th Street, S.W., Washington, D.C. 20410, (Mr. Harold Goldblatt).

17. Puerto Rico Planning Board, Territory of Puerto Rico, 1971. **Computer Mapping System of Social Data, San Juan, Puerto Rico**, Puerto Rico Planning Board, Territory of Puerto Rico, San Juan, Puerto Rico, (Mr David Snyder).

18. Southern California Edison Company, 1977. **Palo Verde/Devers Transmission Routing** (3300 sq mi.), Southern California Desert Areas, Southern California Edison Company, P.O. Box 800, Rosemead, California 91770, (Ms. Mona Myatt).

19. Department of the Army, 1975 **Demonstration of Automated Mapping and Cartographic Techniques** for U.S. Army Corps of Engineers, Washington, D.C., Department of the Army, Office of the Chief of Engineers, Washington, D.C 20314.

APPENDIX II
DATA BANK CREATION SOFTWARE SUMMARY

APPENDIX II
DATA BANK CREATION SOFTWARE SUMMARY

Program	Type of Data	Program Developer[1]	Description
AUTOMAP II	Polygon, point or line	ESRI	Produces line printer overprint graphics of polygon, line, point or continuous surface data. Used primarily in data verification process.
	Polygon		May be used to create a "strip" grid cell file of polygon data. This file is accessed by the **BANK** program for insertion of the data into the data bank.
	Point (Continuous surface)		Used to convert point values into a continuous surface grid cell format. Used primarily to generate reference flood elevations and topographic elevations from point samples.
GRIPS	Polygon, point or line	ESRI	Special adaptation of **AUTOMAP II** for more efficient conversion of polygon data into grid cell data. The program has the added capability of inserting linear and point data into the data bank.
AUTOPLOT	Polygon, point or line	ESRI	Used to verify digitized data and to produce report graphics by a pen plotter, 35 mm, microfilm or CRT plot of polygon, line or point data.
REGISTER	Polygon, point or line	HEC	Used to register x, y coordinate (polygon) data to the Basemap prior to generation of grid representation.
BANK	Grid cell	HEC	Used to place cell by cell, strip grid cell or run length encoded data into data bank.
RIA	Grid cell	HEC	Graphics portion is used in the data verification process to generate line printer overprint graphics of data variables from a grid cell data bank.
GRIDPLOT	Grid cell	ESRI	Used primarily for final report quality graphics; generates pen plotter, 35 mm film, microfilm or CRT plots of data variables from a grid cell data bank.

Program	Type of Data	Program Developer[1]	Description
4VIEW	Grid Cell	WES-HEC	Used in data verification and in the preparation of final report quality graphics to display plots of data variables from the four major compass headings. Used primarily for topographic and demographic displays. Hidden line capability.

[1]All programs that are listed and their supporting documentation are maintained by the Hydrologic Engineering Center and will be furnished upon request. The programs developed by ESRI are the proprietary property of the Environmental Systems Research Institute and will not be distributed outside the U S. Army Corps of Engineers

APPENDIX III
TERRAIN UNIT MAPPING

APPENDIX III
TERRAIN UNIT MAPPING

For increased fidelity of related variables, remapping by terrain units can produce a single map which includes each of the natural geographic variables defined for the project. Integrated Terrain Unit Mapping (ITUM) is a method of producing comprehensive maps of natural geographic variation based upon identifiable patterns visible on aerial photographs plus additional natural resource data. The method is a recent adaptation of Unit Terrain Analysis (UTA) which was developed for use in less developed areas of the world where little or no geographic data is available By applying the ITUM method to developed regions, land areas can be identified which are homogeneous for selected categories of the abundant data generally available. The terrain unit mapping approach involves dividing the landscape into areal map units (polygons) which possess common or homogeneous sets of characteristics. These characteristics are pre-selected to allow specific interpretive capability and are used to define the terrain unit polygons.

This technique is successful because a large number of natural resource variables coincide with one another, particularly under natural conditions. For example, areas of similar slope and soil type in a region tend to be covered by a similar vegetation type which, in turn, provides a habitat for similar animal species. Boundaries are drawn between areas which reflect natural groupings of resources, rather than producing a separate map for each individual variable. (This method is known as the parametric mapping and overlay technique.) These natural areas are homogeneous for the resources used to define them, and are called Terrain Units. The following steps used to produce integrated Terrain Unit maps are similar to the steps outlined for preparation of other data variables for encoding.

1. **Basemap**

   A suitable Basemap for referencing aerial photography and data overlay of the study area is first selected. The Basemap is usually a USGS topographic quadrangle. The Basemap is covered with a clear plastic overlay and reference tics are drafted on the overlay. Map borders are also carefully marked on the overlay to insure accurate edge-matching of contiguous polygons on adjoining map modules. Terrain units are mapped within these module boundaries with the use of collateral maps, high altitude imagery and field investigation.

2. **Mapping of Overlays**

   The second step in production of terrain units is the mapping of information overlays at the Basemap scale. Such information as topographic slope, vegetation type, soil characteristics, landform, geology and hydrology can be used to define terrain units. The system of classification of variables is important, as the terrain units will be defined by the differences between variables. Once defined, further subdivision of a variable classification would require an additional mapping effort. The overlay maps ~~are produced as described in~~ this section, and are used to define the Terrain Units.

3. **Map Overlay Integration/Photo Interpretation**

   The final step in the production of terrain units begins with interpretation of aerial photography (generally color infrared (CIR)). A mylar overlay is placed directly over one CIR image of each stereoscopic pair. Viewing through magnifying stereoscopes, the interpreter divides the image into polygons of apparently homogeneous slope and vegetation. The interpreted polygons are then transferred to the Basemap modules by using an optical pantograph. The polygon pattern is compared with enlarged aerial photos to ensure accuracy of line detail

Each polygon module is then placed over the information overlays on a light table to ensure that all polygons are sufficiently divided to allow characterization by a single code for each variable. New lines are added where necessary, and each module is edited to detect problems such an unclosed polygons and unclear linework. Each polygon is then assigned a discrete sequence number and a code which describes the characteristics for which it is homogeneous. A final edit of each module is made to assure proper edge matching and coding of polygons. Contiguous polygons with the same code number are merged during this edit by removing the common boundary line which separated them.

Terrain unit mapping results in map units which are based upon the interaction of geographic processes. This allows for an integrated interpretation of geography and spatial definition of complex natural systems. Terrain Units are normally visually definable in the field. This "imageability" of landscape makes management of specific areas and planning decisions easier to visualize and communicate than most other methods.

Two important reasons for developing integrated terrain units are accuracy and efficiency:

1. **Accuracy**

The integration of existing resource maps into a terrain unit structure spotlights cartographic inaccuracies or variations among the maps. Resolving these inaccuracies and variations results in more accurate data than normally results from using the existing maps directly.

2. **Efficiency**

There is enormous efficiency in automating of landscape parameters using the terrain unit maps as compared with automation of each variable separately. This is due to the drastic reduction in the number of maps handled and the significant reduction in the total number of polygon map boundaries which require digitizing. The decrease in number of polygon boundaries is a result of the close similarity of polygon boundary lines among separately defined variables. In reality these separately defined lines are often the same line which has been defined differently by specialists in different fields. Resolving these differences by using boundaries defined on aerial photography allows map integration and reduces the total number of lines.

GPO 689-166/024