GEOGRAPHIC INFORMATION SYSTEM
USING ILWIS

NATIONAL INSTITUTE OF HYDROLOGY
JAL VIGYAN BHAWAN
ROORKEE-247667 (U.P.)
1992-93

PREFACE

The demand for the storage, analysis and display of complex and voluminous spatial data has led, in recent years, to the use of computers for data handling and the creation of sophisticated information systems. Effective use of large spatial volume is dependent upon the existence of efficient systems that can transform these data into useable information. Geographic Information System (GIS) is designed to accept large volumes of spatial data, derived from variety of sources, including remote sensors and to efficiently store and retrieve, manipulate, analyse and display these data according to user defined specifications. In this report all of these components are covered in different chapters alongwith a case study.

The present report has been prepared by Sh. S.K.Jain and S.K. Goyal , Scientist 'B' and is a part of training which they obtained at International Institute of Aerospace Survey and Earth Sciences (ITC), Enschede,The Netherlands within the frame work of Indo-Dutch Training Program on water management (WAMATRA PHASE-II).

Satish Chandra
(SATISH CHANDRA)

CONTENTS

List of Figures

List of Tables

1.0 Introduction

A Geographic Information System (GIS) is a computer based system for capturing and processing spatial data of geographic nature. GIS is used for the input storage manipulation and display of geographic data(Fig. 1).The data input subsystem supports digitizing of analog map or importing of spatial and image data in digital form. Data storage and retrieval functions provide the storage and maintenance of both spatial and related nonspatial data files.Data manipulation functions perform analytical functions such as reclassification, overlay and neighbourhood analysis tasks etc. A tabular and graphic output subsystem is used for display of spatial and tabular data.

GIS are rapidly becoming a standard tool for management of resources (Fig.2) Effective utilization of large spatial data volumes is dependent upon the existence of an efficient geographic handling and processing systems that will transform these data into usable informations.

Geographic data have traditionally been presented in the form of a map. However, conventional maps have some springout limitations that render them inadequate for meeting modern needs with respect to spatial data management.One of the constraints associated with conventional maps is the laborious and time consuming nature of tradional cartographic methods of map compilation. Other limitation of these maps is that they are primarily instruments of presentation rather than analysis. While it was relatively easy to retrieve small amount of data or consider the spatial relationship of a few element, these methods become unwidely when large volumes of data involves. The computer

```
IMAGES

MAPS

REPORTS
TABLES
```

DATA INPUT

SUBSYSTEM

DATABASE

MANAGEMENT

SUBSYSTEM

DATA ANALY-
SIS MODELLING
OUTPUT

SUBSYSTEM

```
IMAGES

MAPS

REPORTS
TABLES
```

RASTER   VECTOR   ATTRIBUTES

GEOGRAPHIC DATABASE

FIG. I   SCHEMATIC REPRESENTATION OF A GIS

DATA RETRIEVAL AND QUERIES
FOR MANAGEMENT

DATA MANAGEMENT   DATA ANALYSIS

DATA INPUT

DATA TRANSFORMATION
AND MODELLING

DATA SOURCES

USERS

DATA GATHERING

DECISION MAKING

SURVEYING   VALIDATION
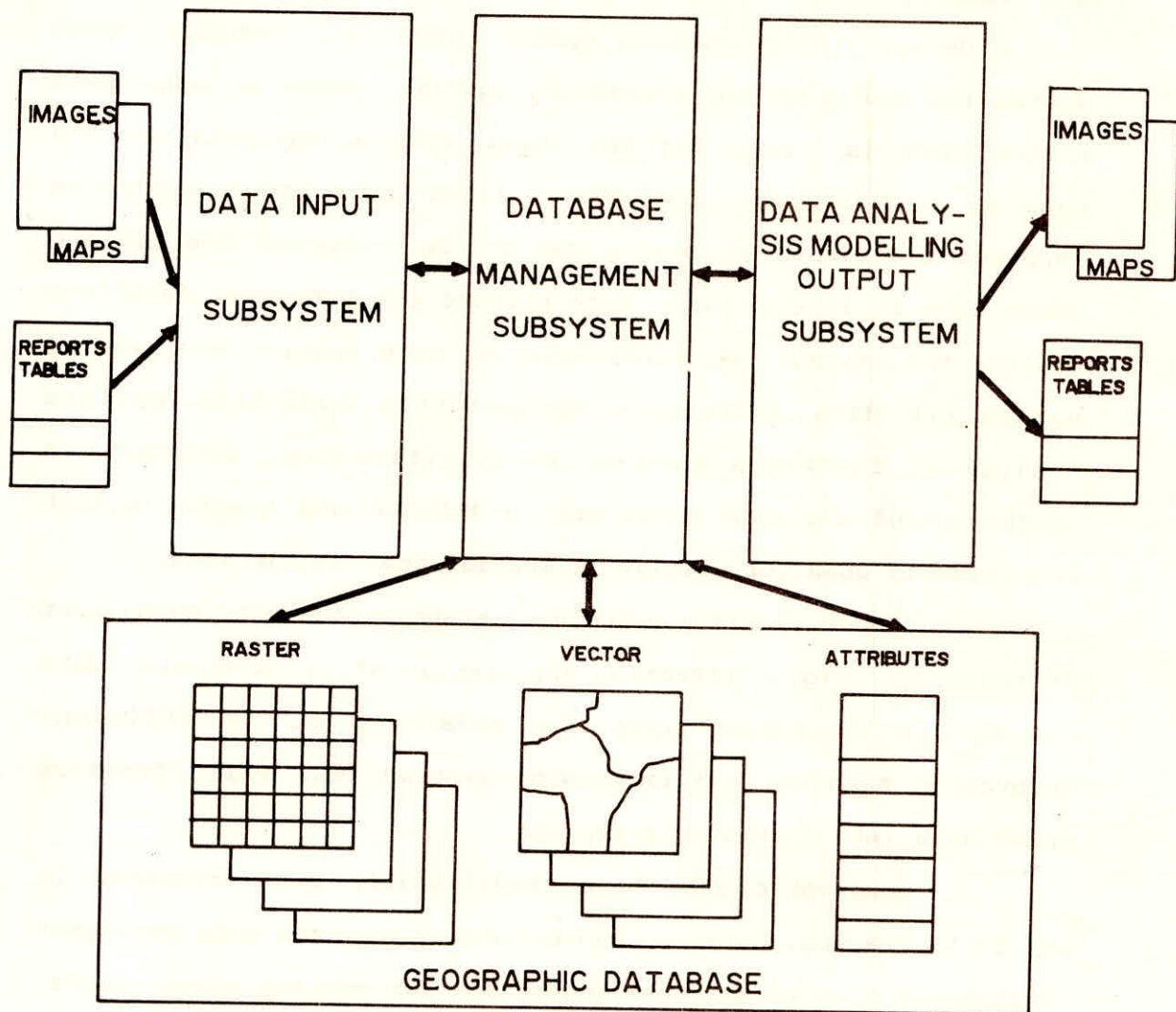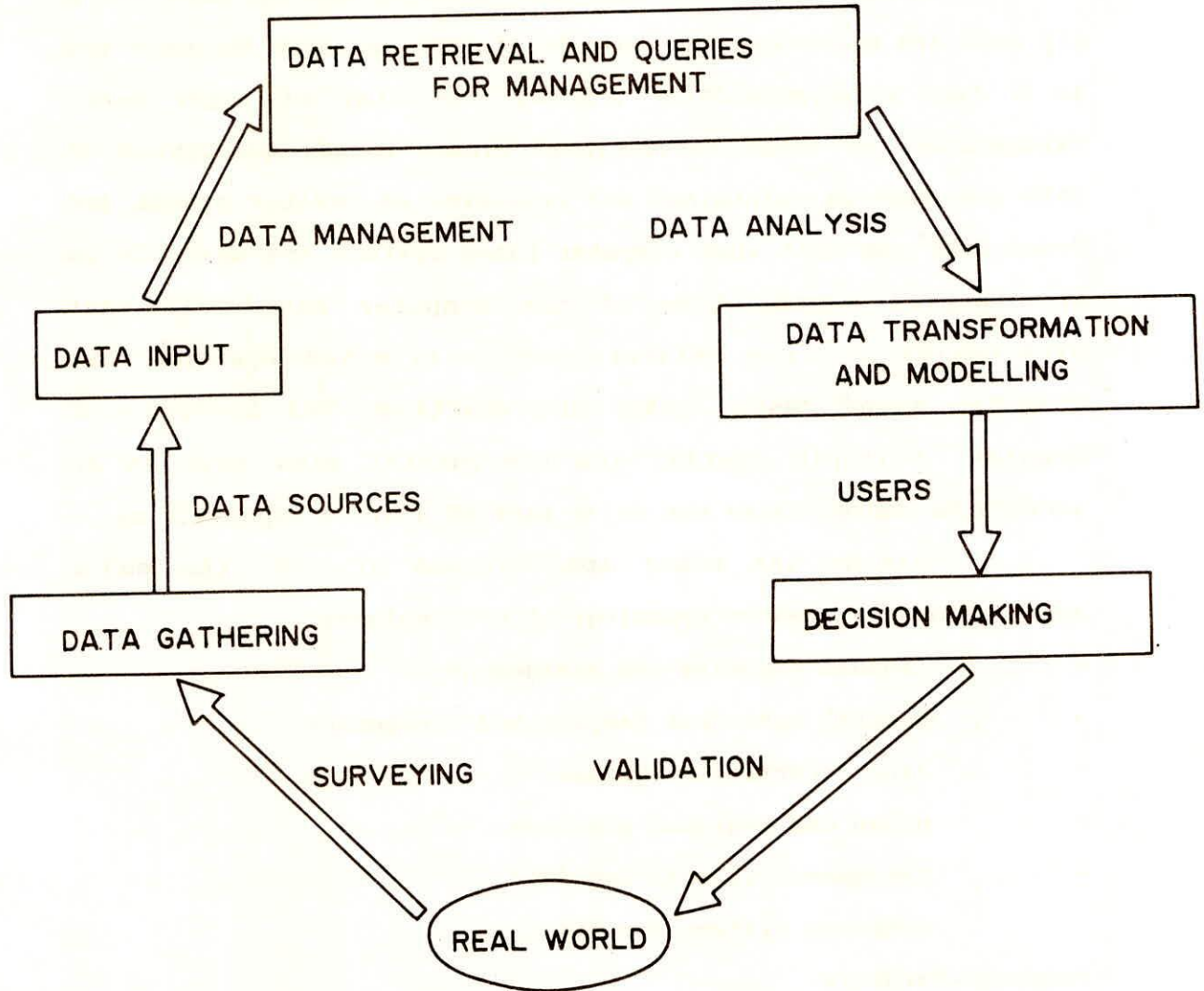
REAL WORLD

FIG. 2 GEOGRAPHICAL INFORMATION SYSTEM AS A MANAGEMENT TOOL.

based geographic information systems are developed to provide the power to analyze large volume of geographic data.

A GIS is powerful tool for handling spatial data. In a GIS data are maintained in a digital format. As such the data are in a form more physically compact than that of paper maps, tabulations, or other conventional types. Larger quantities of data can also be maintained and retrieved at greater speeds and lower cost per unit when computer based systems are used. It is the spatial capabilities of the computer based GIS that distinguish it from related graphic oriented systems like computer aided design (CAD) and drafting. The analysis of complex, multiple spatial and non-spatial data sets in an integrated manner forms the major part of a GIS's capabilities.

Beside the other applications of GIS, the major applications related to Hydrology of GIS includes

-       landuse planning and management
-       natural resources mapping and management
-       land information Systems
-       urban and regional planning
-       management of well log data
        Computer Systems for GIS

Computer Hardware

In addition to the basic computer hardware such as the C.P.U, storage devices, keyboard and monitor, in GIS, some more hardware are also required and they are as follows:

-       a digitizer and/or a scanner for inputting spatial data from existing maps.

4

-        a high resolution color monitor for displaying maps and
imageries etc.

-        a printer for printing raster maps, remote sensing
images and textural information.

-        a plotter for high - quality production of map in
vector form.

        The computer of C.P.U. is linked to a disk drive
storage unit, which provides space for storing data and
programs. A digitizer or other devices is used to convert data
from maps and documents into digital form and send them to
computer. A plotter or other kind of display device is used to
present the results of the data processing.Presently GIS is
available on wide range of computer hardware starting from micro
computers to maintenance computers.In the following tables a list
of some commercially available GIS is given.

           GIS   in Mini   and   main   frame   computers
                    (from GIS World,1989)

System Name	Hardware	Geometric Storage	Attribute Storage
Arc/Info	VAX,Prime	Vector	relational
	IBM	Vector	relational
Delta map	HP,Sun	Vector	relational
Infomap	VAX	Vector	relational
Integraph	VAX	Vector	relational
Maps	VAX,Prime	Vector	relational
Sysscan	VAX	Vector	relational
Geobased	VAX	Vector	relational
System 600	VAX	Vector	relational

//

GIS in PC/micro computers (From GIS World, 1989)

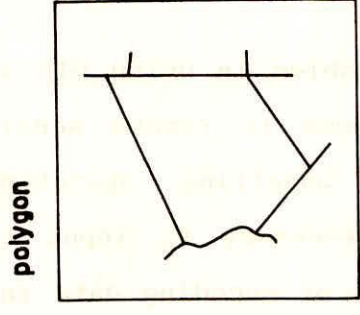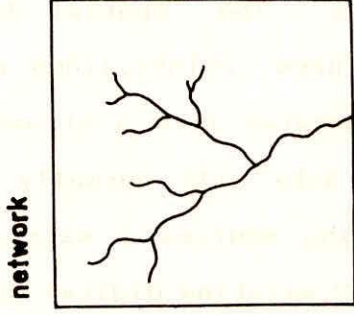| System Name	Hardware	Geometric storage	Attribute Storage
Arc/Info	IBM PC/AT SYSTEM 2	vector	relational
SPANS	IBM PC/AT SYSTEM 2	quadtree vector	
INFORMAP II	IBM/AT	vector	relational
ERDAS	IBM/AT	raster	
ILWIS	IBM/AT	vector raster	relational
PAMAP	IBM/AT	raster	relational
IDRISI	IBM/AT	raster	

2.0 DATA INPUT AND OUTPUT

2.1 DATA INPUT

Data collection and input are major problem in using GIS and these are partially solved by the advances in remote sensing technology.Before any spatial analysis or modelling operations can be carried out in a GIS, it is necessary to input the requisite data.Data input is the procedure of encoding data into a computer readable form and writing the data to the GIS data base.The data to be entered in a GIS are of two types - spatial and associated non spatial attribute data . The spatial data represent the geographic location of features .Points,lines and areas are used to represent geographic features like a street,a lake and a forest stand (fig. 3).These data will normally be obtained from one or more of the following sources : existing maps, aerial photographs,satellite imagery,existing digital data and other GIS data bases(fig. 4).The nonspatial attribute data provide description information like the name of a street,the salinity of a lake or the composition of a forest stand.

The method of spatial data input depends primarily on the source the data. The two principle approaches for inputting data from maps are :Manual digitizing and Scanning.

2.1.1 MANUAL DIGITIZING

The most useful method of converting existing map into digital form is by digitizing. In manual digitizing, digitizing table and pointing device (digitizing tablet) is used to trace the map features. As the map elements are traced, the coordinate data generated from the digitizing table are either processed

FIG. 3. TYPES OF SPATIAL DATA

polygon

property boundary

network

drainage pattern

line

proposed skyline

point

raingauge location

raster – discrete

forest stand map

raster – continuous

digital elevation model

raster – image

scanned photograph

area

forest stand

MAPS FIELD OBSERVATIONS DIGITAL PRODUCTS

TERMINALS DIGITIZERES FILES FILES SCANNERS TAPES
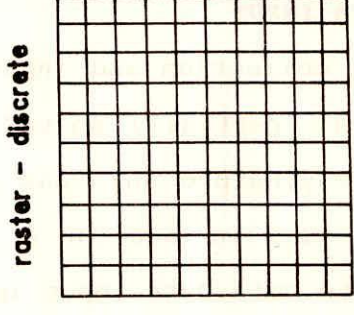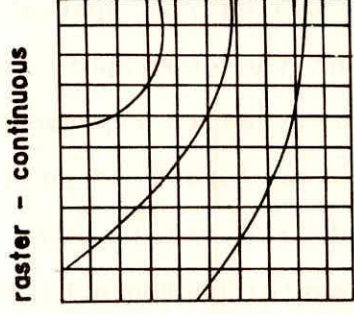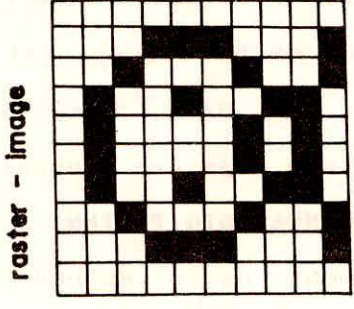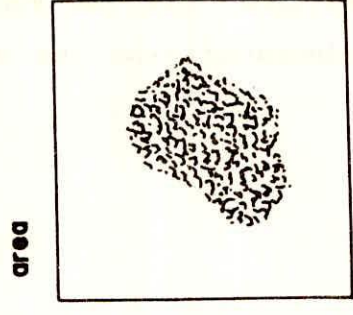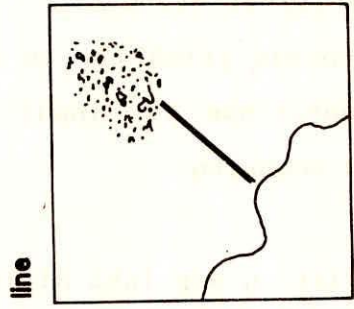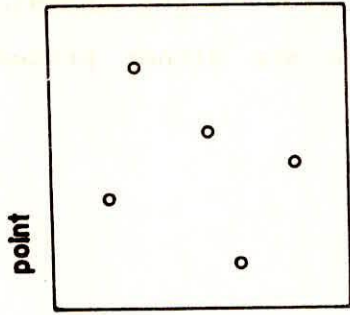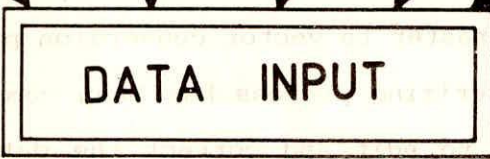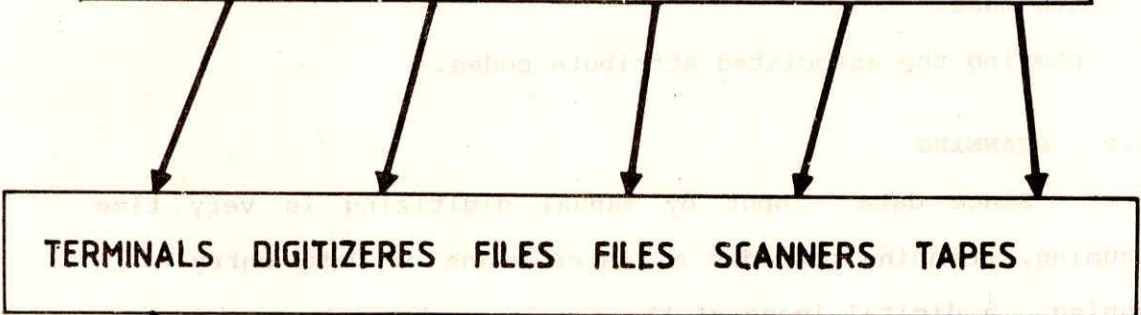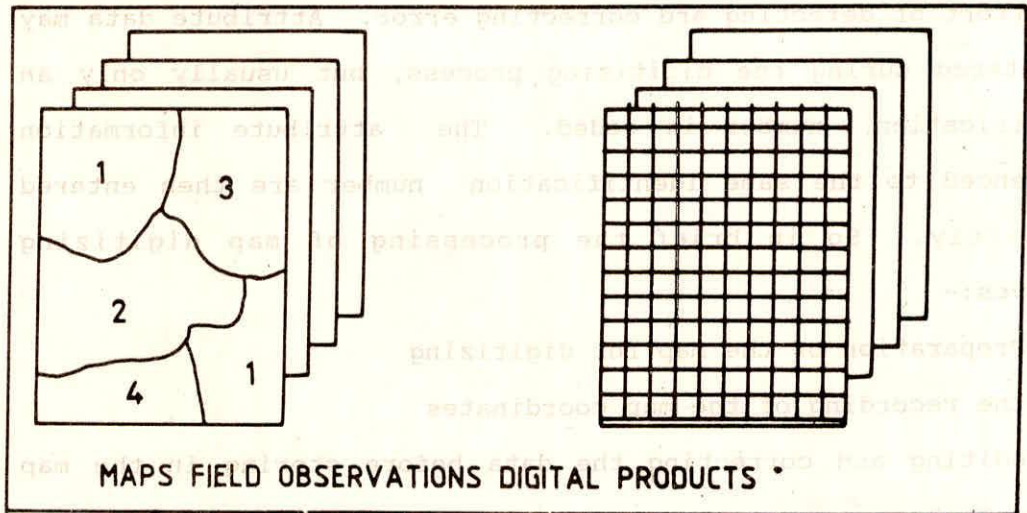
DATA INPUT

FIG. 4. DATA INPUT

immediately by the GIS or are stored for later processing.

The process of tracing lines is time consuming and error prone. The software can provide aids that substantially reduce the effort of defecting and correcting error. Attribute data may be entered during the digitizing process, but usually only an identification number is coded. The attribute information referenced to the same identification number are then entered separately. So in brief the processing of map digitizing involves:-

- Preparation of the map for digitizing

- the recording of the map coordinates

- editing and correcting the data before storing in the map data base

- entering the associated attribute codes.

2.1.2   SCANNING

Since data input by manual digitizing is very time consuming, scanning provides a faster means of data entry. In scanning , a digital image of the map is produced by moving an electronic detector across the map surface. The lines of the original map are recorded as a series of small pixels forming a binary image(image, no image). If the data are required in vector format, additional raster to vector conversion processing is done. Once this vectorizing process has been completed it will normally be necessary to edit and correct the data. Data entry using scanning is claimed to be 5 to 10 or more times faster than digitizing.

2.1.3 INPUT OF NONSPATIAL DATA

Input of non spatial data to the GIS data base is generally performed by typing at the keyboard. Many systems also accept data files imported from commercial data base and spread sheet software packages.

2.2 REMOTELY SENSED DATA

In areas in which map data are scarce and even in areas in which rapid change is occurring but that are otherwise well mapped, the remote sensing images may be a significant source of map data. In extreme case, imagery may be the only source of map data. In fact, most of the spatial data now used in a GIS are produced with remote sensing technology. Since 1972, with the launch of Landsat-1 high resolution digital remotely sensed data have become widely available. These satellite base systems now produce publicly available data for virtually all of the Earth's surface at resolutions as high as 10 meters equivalent to high altitude aerial photography.

The Images produced by the sensor system represent measurement of the reflected and emitted radiation from the earth's surface. Because the imagery is in digital form, the data for single wave length or multiple wavelengths used in combination can be computer processed to directly generate geographic information. Landuse, crop condition, flooding, water quality are few examples of geographic data that can be produced using computer methods. Computer enhancement of digital imagery enables other resource information to be visually interpreted and used to update the information in a GIS.

2.3 DATA OUTPUT

Output is the operation of presenting the results of data manipulation in a form suitable to the user. Data are output in one of the three formats - Hardcopy, Softcopy and electronic.(fig.5)

2.3.1 HARDCOPY OUTPUT

Hard copy output are permanent means of display. The information is printed on paper, photographic film or similar materials. Maps and tables are commonly output in this format. The hardcopy output takes longer to produce and requires more expensive system. However it is a permanent record that is easily transported and displayed. The most common and cheapest hardcopy output device is the dot matrix printer which can print alphanumeric character at very high speed. Each letter or number is formed by a series of tiny dot produced by needles on the printing head. Striking an inked ribbon. Ink jet plotters give better quality colour output, because both the range of colours and the resolution is higher. Ink jet plotter work by squirting tiny drops of colour ink out the plotting paper. The pen plotter is the most common type of vector based equipment. This device comprises a pen which is driver over the plotting surface by x-and y- step motors. There are few more output devices which can be used to produce outputs from GIS system.Such as matrix camera,colour laser printer etc.

2.3.2 SOFT COPY OUTPUT

Soft copy output is the format as viewed on a computer monitor. It may be text or graphics in monochrome or colour. The soft copy display is generally not used for final output

DISPLAY AND OUTPUT

C.R.T.S. PRINTERS PLOTTERS TAPES DICS

MAPS   IMAGES   TABLES   GRAPHS

FIG. 5. DATA OUTPUT

13
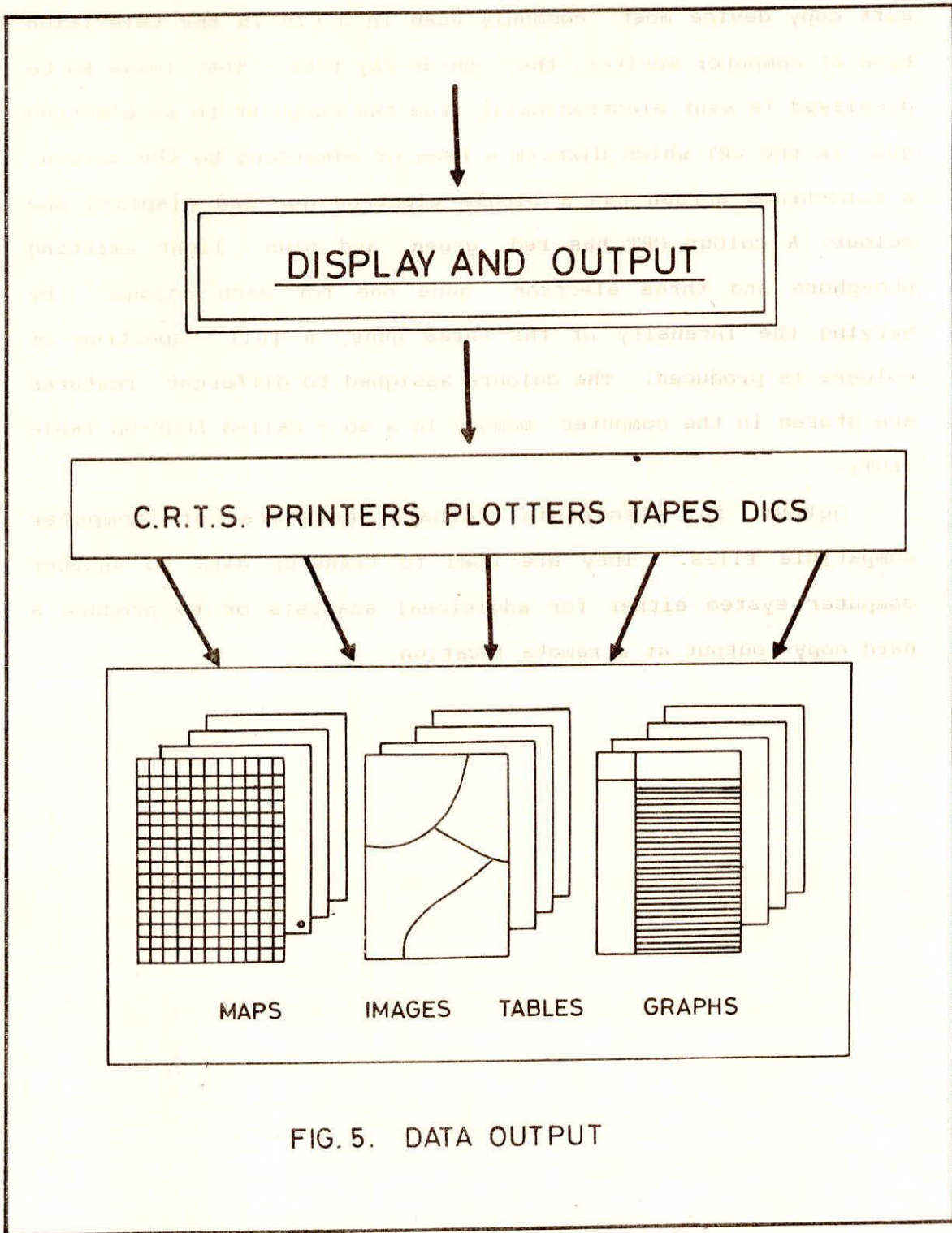
because of its small size and the loss in quality when the screen image is photographed or electronically captured. The soft copy device most commonly used in a GIS is the television type of computer monitor, the cathode ray tube. The image to be displayed is sent electronically from the computer to an electron gun in the CRT which directs a beam of electrons to the screen. A monochrome screen has a single electron gun and displays one colour. A colour CRT has red, green, and blue light emitting phosphors and three electron guns one for each colour. By varying the intensity of the three guns, a full spectrum of colours is produced. The colours assigned to different features are stored in the computer memory in a so - called Look-Up Table (LUT).

Output in electronic formats consists of computer compatible files. They are used to transfer data to another computer system either for additional analysis or to produce a hard copy output at a remote location.

3.0 DATA BASE MANAGEMENT

3.1 INTRODUCTION

A data base is a collection of information about things and their relationships to each other. For example, a database may consist of names and addresses. The names may themselves be categorized by ocher relationships, such as 'client', 'friend' or 'family'.

Geographic Information Systems (GIS) deal with data related to location in real world space called as geographic or spatial data.GIS must store large amounts of data and make them available on demand. Geographic or spatial data base is a collection of information about things and their relationships to each other (Fig.6). The objective in collecting and maintaining information in a data base is to relate facts and situations that were previously separate. Geographic data has three major components: its geographic position, attributes or properties, and time or dynamics.

Position- Position refers to the fact that each feature has a location that must be specified in a unique way.The location can be absolute or relative. Absolute such as coordinate position using a coordinate system like cartesian (X,Y) or global geographic (latitude/longitude), row and column position describing a position on a grid etc. Relative which is location with reference to some other object e.g. topologic position which is described through relationship to other objects-"adjacent to", "intersects with".

Attributes- Attributes are often termed as non spatial data since

FIGURE 6- GMOGRAPHIC DATABASE

16
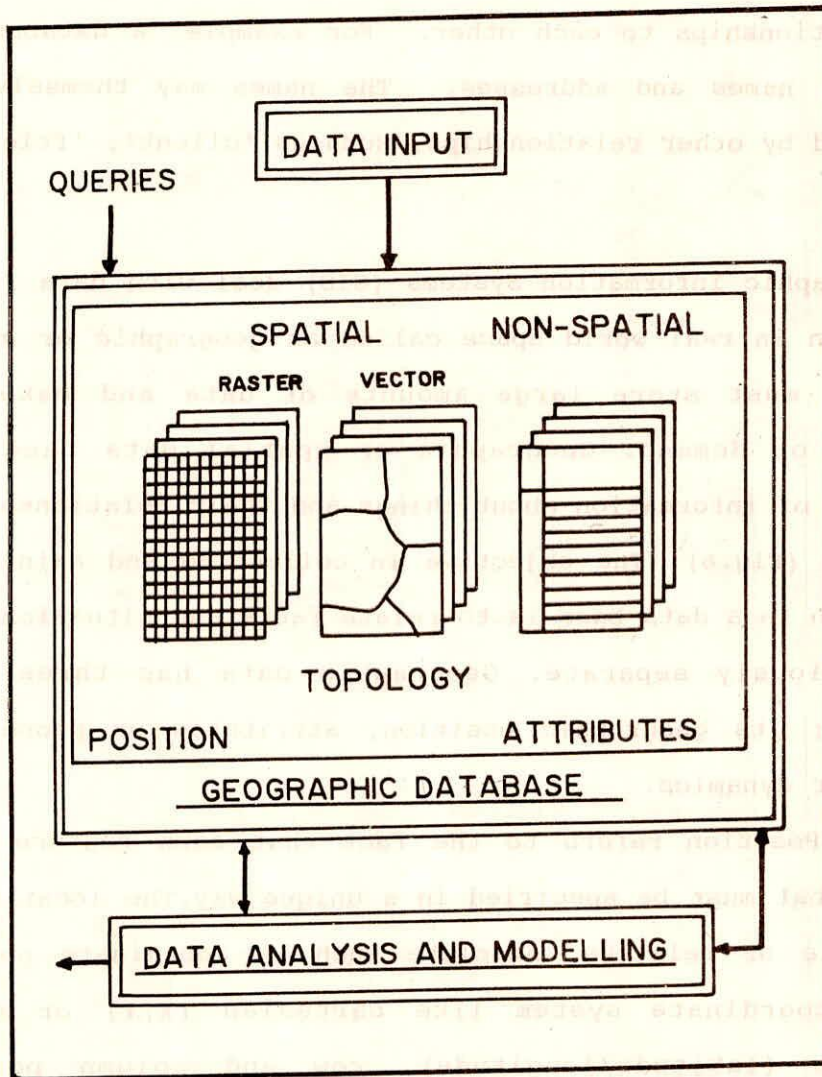
they do not in themselves represent locational information. An attribute is a characteristics of an entity.

Time - Time is important and critical since geographic information is referenced to a point in point in time or a period of time. Knowing the time where geographic data were collected is of utmost importance to using these data appropriately e.g. an area may be covered by forest one year and have been clear out the next.

3.2 Data Base Concepts

GIS are increasingly being built around existing data based management system (DBMS) (Aronoff, 89).

Data base management systems (DBMS) are designed to facilitate storage and retrieval of large data collections. They include facilities to protect and secure data, enforce consistency of the data stored, and make data available to multiple users at the same time. These services are necessary for GIS and GIS should therefore be built using DBMS.

There are two approaches using DBMS

(a) They handle both, the spatial and non-spatial data. Non-spatial data is accessed through the DBMS, while spatial data is handled directly by the GIS. Several applications involve the use of only the nonspatial data before any spatial analysis is considered. For example, a land evaluation can be performed without having any geographic access (Spatial data,) provided a good attribute database for soil characteristics, crop requirements and social economic characteristics is available. Geographic access and spatial analysis at a later stage in the process will add a new dimension to the land evaluation model.

The data base approach increases the benefits obtained from a GIS.Among the advantages are:-

- Reduction in data redundancy: Sharing data by different applications instead of each having independent data base results in reduction of data redundancy. They act as a big file, different users may have different needs,but may use the same data.

- Mainteance of data integrity and Quality: Data base should have as minimum redundancy as possible.Redundant is unnecessary duplication. A data is redundant if we can delete it without information being lost. Controlling the redundancy and updating procedures can be implemented, more efficiently using a data base.

- Security restrictions: Data base include security tools to control access to the data.This is important particularly for inserting and deleting data. Only authorized users can have access to the data base.

3.3 Geographic Data Types

There are four primitive types of geometric entities to encode spatial data: Points, lines, polygons and continuous surface.

Point data-

Points are the simplest type of spatial data. Points can be of two kinds, observations relating to discretely distributed phenomena such as oil and water wells and observations relating to continuous distributions such as weather station reporting perception measurements of soil temperature.

Line data:

Line entities are all linear features built up of straight line segments made up of two or more coordinates, Line entities can be static (structural type) or dynamic (flow), An arc, chain, or a string is a set of n x-y coordinate pairs describing a continuous complex line.

Polygons or Area Data:

Polygons constitute the most common data type used in GIS. They are bounded regions. The boundaries may be defined by natural phenomena such a land forms or by man made such as forest stand or land use units.

Continuous surfaces:

Examples of continuous surfaces are elevation (as part of topographic data), rainfall, temperature, etc. Most of the GIS products can not handle true 3-D data although they can handle topographic data, usually as Digital Elevation Model (DEM).

3.4 Spatial data model:

There are two fundamental approaches for representation of the spatial component of geographic information, the vector model and the raster model. Fig. 7 illustrates both models.

3.4.1 Vector Models:-

In the vector model,objects or conditions in the real world are represented by the points and times that define their boundaries such as if they were being drawn on a map.The fundamental primitive in the vector model is the point.Objects
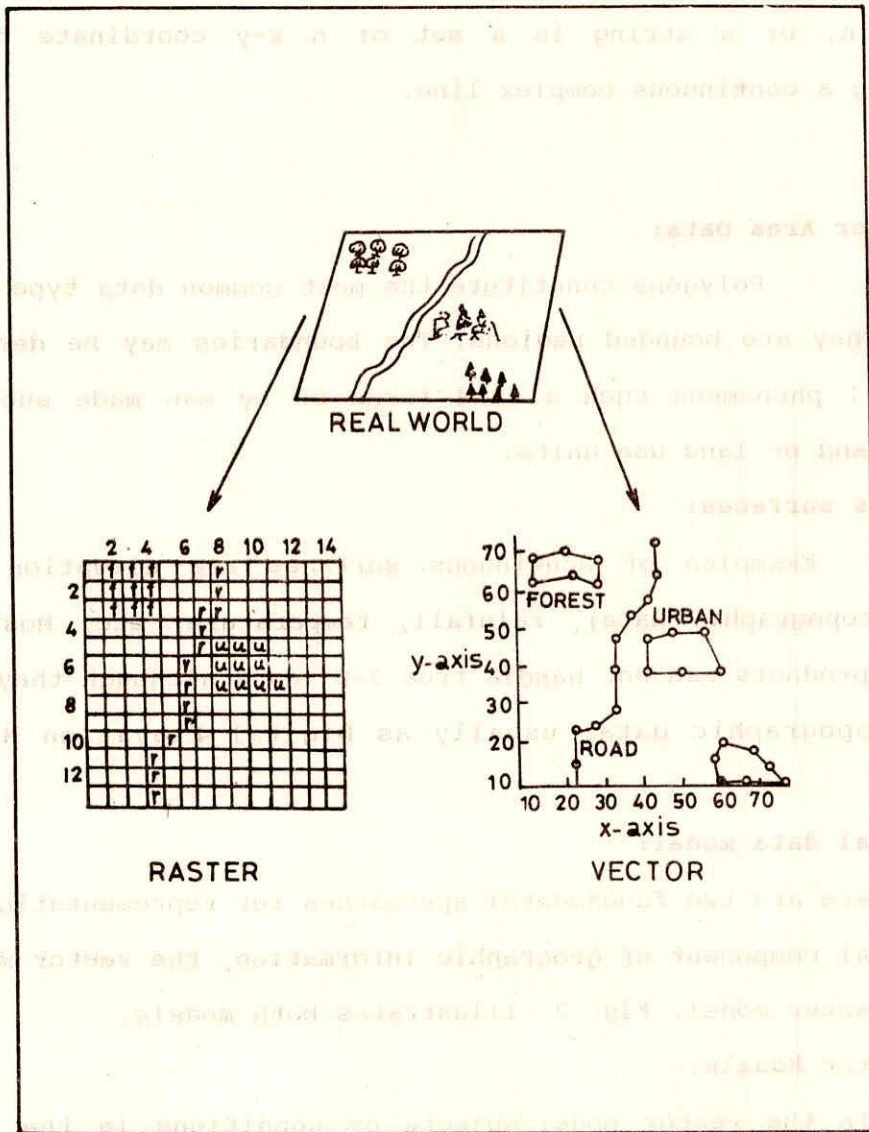
Fig 7 Representation of data in vector and raster models

are created by connecting points with straight lines, areas are
defined by set of lines. Vector models have the lime as the basic
logical unit in a geographic context.

The most common vector models are the whole polygon
structure and the topologic model.

The Whole Polygon Structure:-

In the whole polygon structure or Spaghetti model
each polygon is encoded in the data base as one logical record In
this model (fig.8) the paper map is translated line into a
list of X-Y coordinates, points is encoded as a single X-Y
coordinate pair.

The common boundary between adjacent polygons must be recorded
twice, once for each polygon. The main disadvantage of this model
are:-
- line between adjacent polygons must be digitized and stored
twice.

- editing and updating the data base are carried out comparing
it visually with a graphic content.

- islands are only graphical constructions.

The Topologic Model:

The topological model is the most widely used method of
encoding spatial relationship in a GIS. In this structure basic
logical entity is the arc, a series of points that start and end
at a node. A node is an intersection point where two or more area
met. In addition, the identifiers, or names of the polygons on
either side of the line are recorded, thus the more elementary
spatial relationship are explicitly retained and can be used for
analysis and more over, the basic data are stored in a non-
redundant manner. The data structure of this model is depicted in

21

ORIGIONAL MAP

DATA MODEL

DATA STRUCTURE

FEATURE	NUMBER	LOCATION
POINT	5	xy
LINE	24	xy .... xy
POLYGON	15	xy .... _ xy
	18	xy .... xy
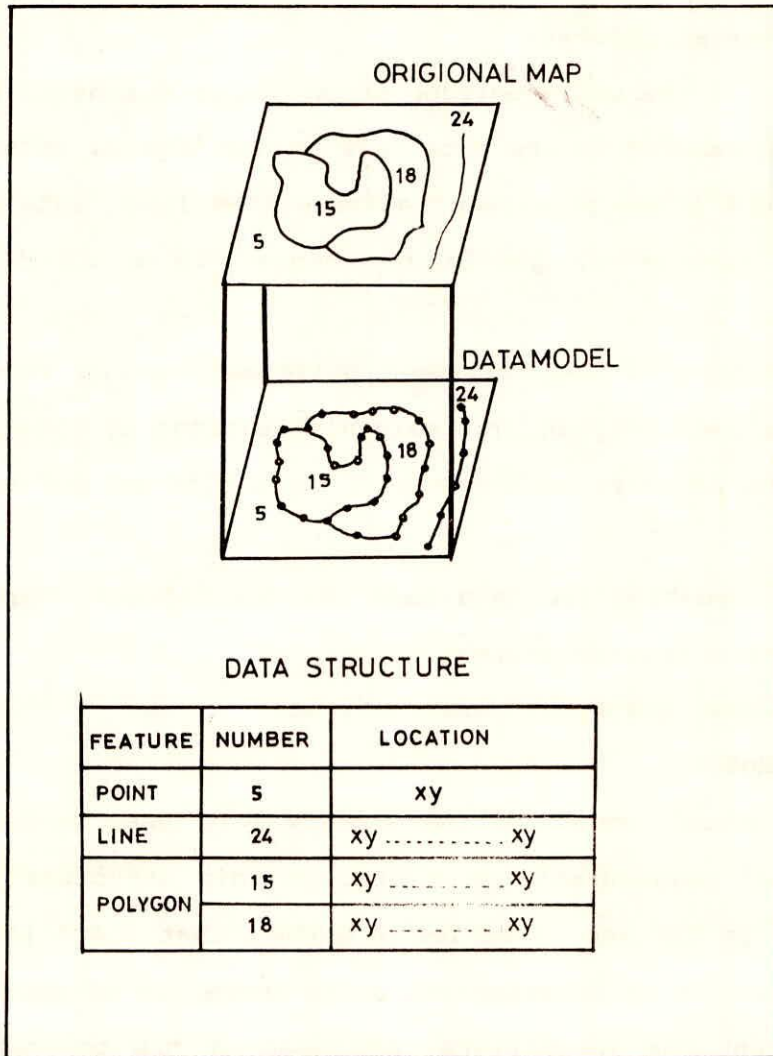
Fig.8 Whole polygon structure   After Dangermond,1982

fig.9.

3.4.2 The Raster Data Model:

In its simplest form, the raster data model consist of a regular grid of square or rectangular cells. The location of each cell or pixel is defined by its row and column numbers. Each cell in a raster file is assigned only one value, different attributes are stored as separate data files. (Fig. 11) illustrates the structure of the raster model. As shown in Fig. 7, a point (the house) is represented by a single pixel, a line (river) by several pixel with the same value forming a linear grouping and an area (the forest stand) by a clump of cells all having the same value. The main raster models are the grid and other regular tessellation. Fig. 10 shows the rasterization process. In the raster data model, each cell represents an area of the land surface. Since the attribute of each cell is stored as a unique value, the total numbers of values to be stored is the product of the number of columns times the number of rows. The smaller the area of land that each cell represents, the higher the regulation of the data and larger the files needed to store the data. So it is common for a raster file to be comprised of several million cells. However many of the cells may contain the same value as neighbouring cell.Therefore reductions in the size of the raster file can be achieved by using various data compression methods such as ,chain codes,run length,block codes etc.The Quadtree data model provides a more compact raster representation by using a variable size grid cell. Instead of dividing an area into one size cell, finer sub-divisions are performed in areas with finer

ORIGIONAL MAP

DATA MODEL

DATA ENCODIND

Node	Area
n1	$a_1 a_2 a_3$
n2	$a_1 a_2 a_3$
n3	a_6
n4	a_4
n5	a_4
n6	a_6

Area	From Node	To Node	Right Polygon	Left Polygon
a_1	n1	n2	12	15
a_2	n2	n1	18	15
a_3	n1	n2	18	15
a_4	n4	n5	12	12
a_5	n3	n3	12	12
a_6	n4	n6	0	12

Polygon	Arcs
12	$a_1 a_3, a_4 a_5 a_6$
15	a_1, a_2
18	a_2, a_3

Are	Start	Cordinate			End
a_1	xy	xy	xy	xy	xy
a_2	xy	xy	xy	xy	xy
a_3	xy	xy	xy	xy	xy
a_4	xy	xy	xy	xy	xy
a_5	xy	xy	xy	xy	xy
a_6	xy	xy	xy	xy	xy

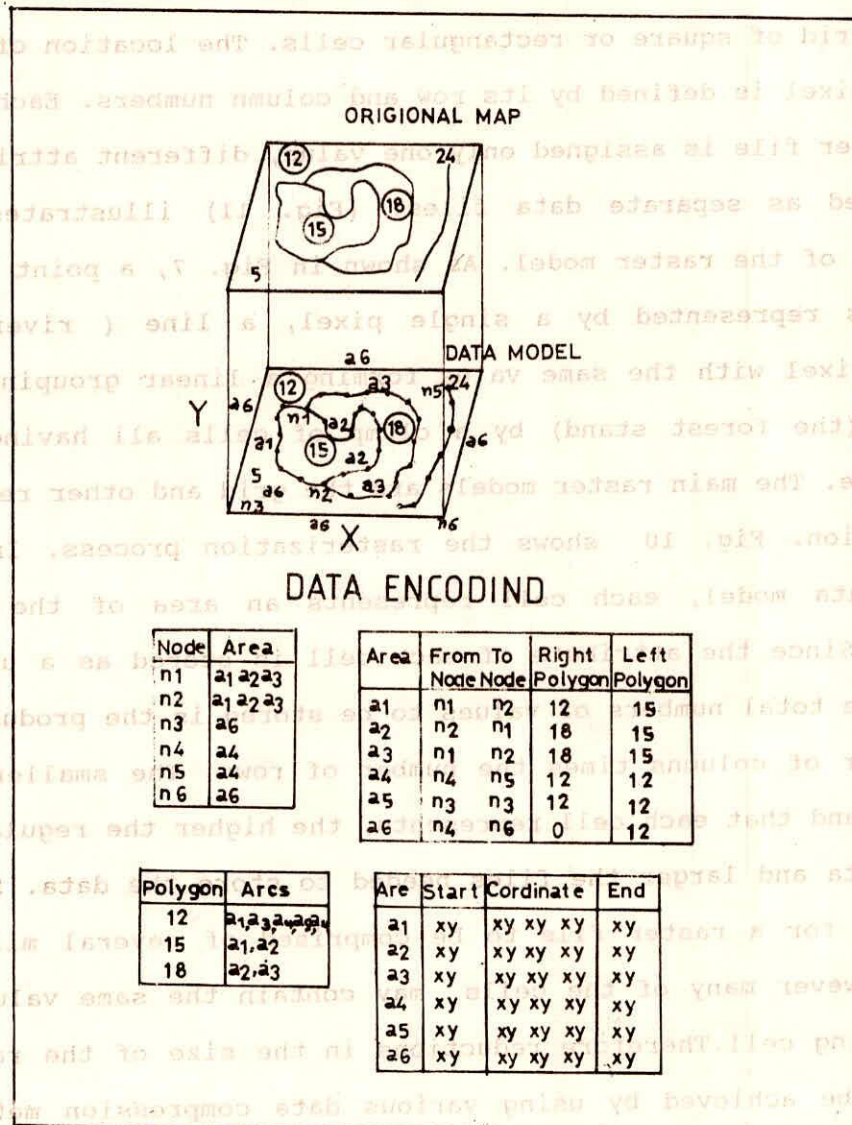Fig. 9 The topographic data model

24

ORIGINAL MAP

GRID OVERLAY
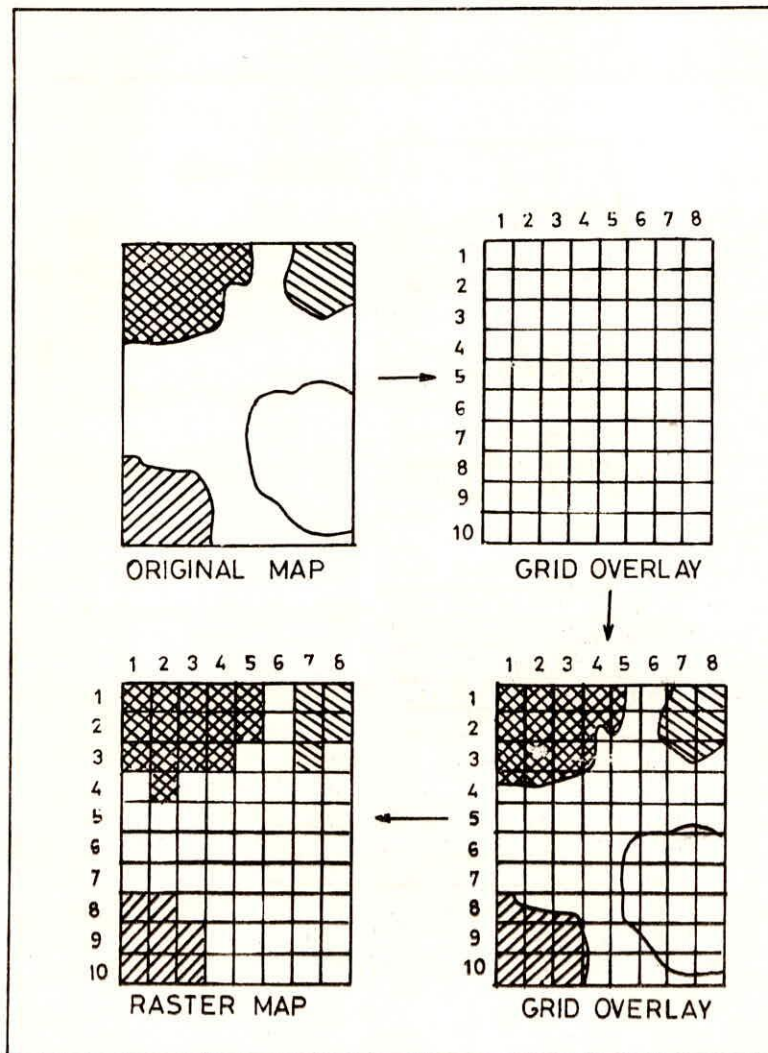
RASTER MAP

GRID OVERLAY

Fig. 10 Rasterization process

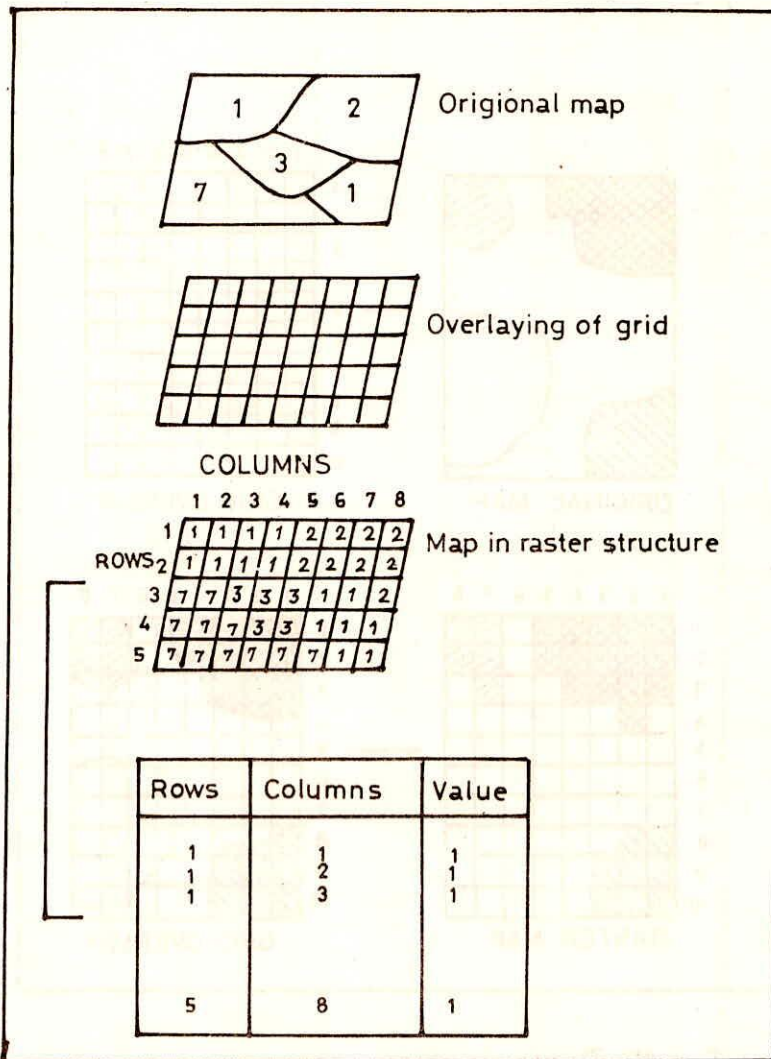Fig. 11 Structure of the raster model

details as shown in Fig.12   . This provides a greater resolution only where it is needed.The use of quadtree model allows the use of large cells (coarse resolution) to encode large homogenous areas. Higher level of resolution is used for areas that have higher spatial variability.

3.5 Comparison between data models

Vector and raster models have both, advantages and disadvantages.

3.5.1 Raster Model

Advantages:

-   It is a simple data structure.

-   Overlay and combination of mapped data with remotely sensed data is easy.

-   The raster format is more or less required for efficient manipulation and enhancement of digital images.

-   Technology is cheap and it is being energetically developed.

-   Simulation is easy because each spatial unit has the same shape and size.

Disadvantages:-

-   Wasteful use of computer storage for spatially spaces data.

-   Network images are difficult to establish because of non neighborhood connection.

-   Errors in estimating parameter and shape.

-   Boundaries of a object can not represented perfectly, they appear as step staircase.

3.5.2 Vector Model

THE QUADTREE REPRESENTATION

A.   Land-Use Map

INDUSTRIAL	
Reside Detache	
RESIDENTIAL MULTI-UNIT	RURAL
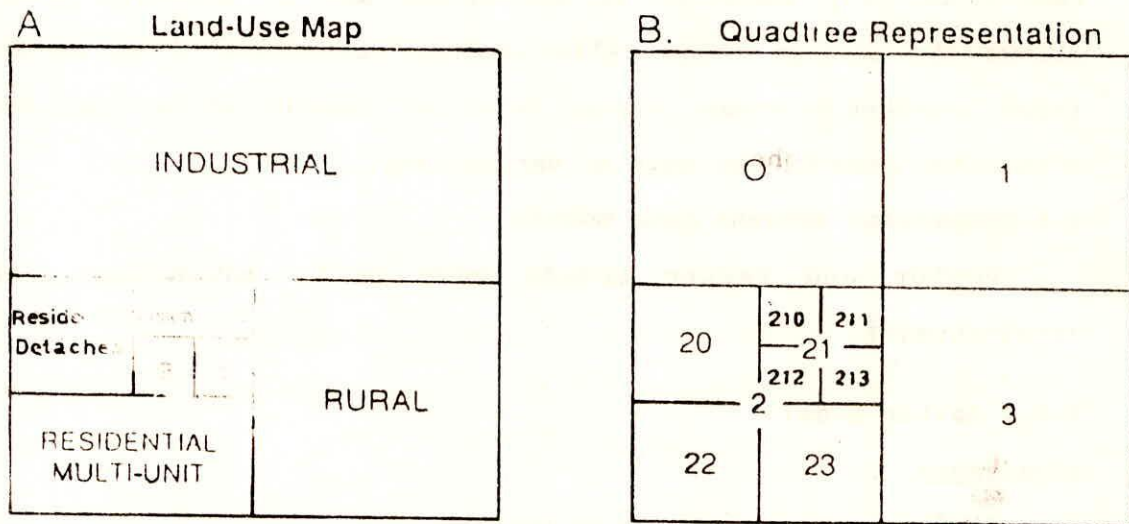
a. Community Services
b. Recreation

B.   Quadtree Representation

O	1
20 210 211 -21- 212 213 -2- 22 23	3

C.   Schematic Representation of the Quadtree

MAP
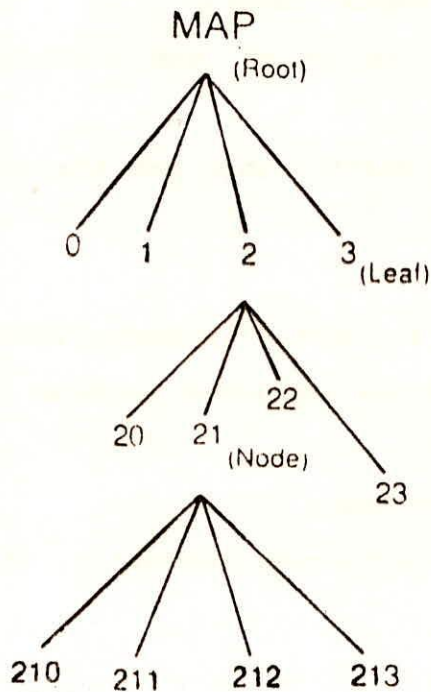
D.   Table of Attributes

Quadtree Levels			ATTRIBUTES
1	2	3	
0			Industrial
1			Industrial
2			Residential
	20		Residential. Detached
	21		Services
		210	Commercial
		211	Commercial
		212	Community Services
		213	Recreation
	22		Residential. Multi-Unit
	23		Residential. Multi-Unit
3			Rural

Fig. 12 The Organization of Data Using the Quadtree Model

Advantages-

- It provides a more compact data structure

- Retrieval, updating and generalization of graphics and attributes are possible.

- Widely used to describe administrative zones.

<u>Disadvantages</u>

- It is a more complex data structure.

- Overlay operations are more difficult to implement.

- Simulation is difficult because each unit has different topological form.

- Manipulation and enhancement of digital images can not be effectively done in the vector domain.

- Technology is expensive.

- Error correction in the data file is also difficult.

- During overlay operations vector data sets create large number of polygons which may create problems while interpretating output data set/map etc.

the ground resolution.

In the above paragraphs a number of sources of errors in GIS has discussed. So while using GIS technology, we must be aware of the errors which can occur and where possible take appropriate steps to keep such errors to an acceptable minimum.

4.0 DATA ANALYSIS AND MODELLING

4.1 Introduction

The most important characteristics of GIS is the provision of capabilities for data analysis and spatial modelling. These functions use the spatial and nonspatial attribute data in the GIS data base to answer questions about the real world. The data base in a geographic information system is model of the real world that can be used to simulate certain aspects of reality.A model may be represented in words ', in mathematical equations or as a set of spatial relationships/displayed on a map. The general problem in data analysis is stated below:

User's query- Database- Link - output

The user has a particular problem or query. The database contains information in the form of maps that can be used to answer the users query. All that is necessary is to establish a link between data base and output that will provide the answer in the form of a map, tables or figures. The link is any function that can be used to convert data from one or more input maps into an output.

4.2 Analysis Function-

The power of a GIS lies in its ability to analyse spatial and attribute data together. A large range of analysis procedures of functions is there but they have been subdivided into four categories, (a) retrieval/classification and measurement,(b) overlay,(c) distance and connectivity and (d) neighborhood.

4.2.1 Retrieval, Classification and Measurement Functions

In these functions spatial analysis and attribute data

are retrieved, but only the attribute data are modified. No changes are made to the location of spatial elements and no new spatial elements are created.

Retrieval Operations:-

Retrieval operations on the spatial and attribute data involve the selective search and manipulation, and output of data with out the need to modify the geographic location of features or to create new special entities. Retrieval operations include-

- Retrieval of data using geometric classification.
- Retrieval of data using symbolic specifications.
- Retrieve data using a name of code of an attribute.
- Retrieval of data using conditional and logical statements.

By using the rules of Boolean logic to operate on the attributes and spatial properties, the data can also be retrieved. Boolean algebra uses the operations AND, OR XOR, NOT to see whether a particular condition is true or false as shown in table given below:-

A	B	NOT A	A AND B	A OR B	A XOR B
1	1	0	1	1	0
1	0	0	0	1	1
0	1	1	0	1	1
0	0	1	0	0	0

N.B. 1 implies 'true' and 0 implies 'false'

Simple Boolean logic is often portrayed visually in the form of Venn diagrams (Fig.13).If set A is the set of items

VENN'S DIAGRAM
BOOLEAN LOGIC
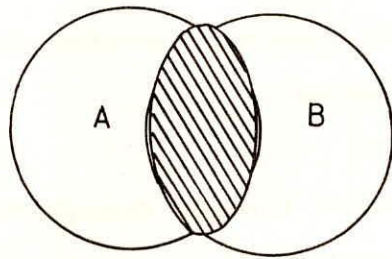
A AND B

A NOT B

A OR B

A XOR B
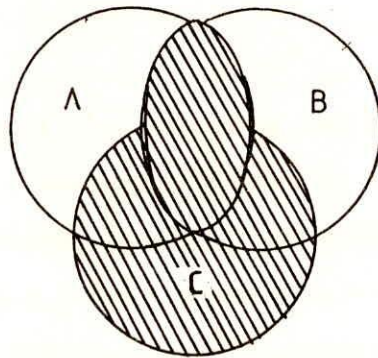
(A AND B) OR C
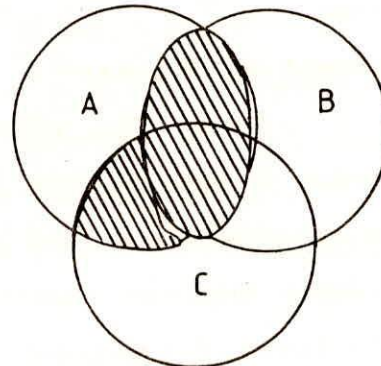
A AND (B OR C)

(From Burrough, 1986)

Fig.13

having attribute a and set B is the set of items having attribute b, then the statements A AND B, A OR B, A XOR B, A NOT B will return items that have attribute combinations covered by the shaded portions of the respective Venn diagrams.Boolean operations should not be thought of as being only applicable to the attributes of the geographical elements, because they can also be applied to the spatial properties.
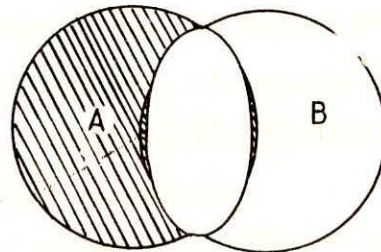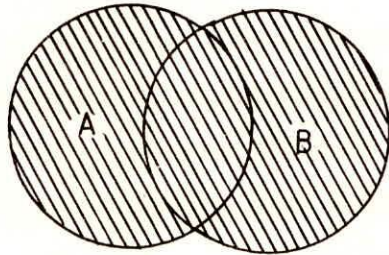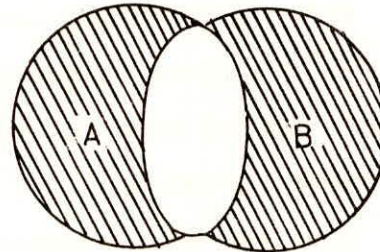
Reclassification Procedures-

Reclassifying procedures involve operations that reassign thematic values to the categories of an existing map as a function of the initial value the position size or shape of the spatial configuration associated with each category, for instance a soil map, reclassified into a permeability map. In a raster based GIS, numerical values are often used to indicate classes. A cell might be assigned the value to indicate classes.A cell might be assigned the value 1 to indicate a agriculture land, 2 for forest land and so on.Classification is done using single data layers an well as with multiple data layers as part of an overlay operations.

Measurement Functions:-

Every GIS provides some measurement functions, Measurement of spatial data involve the calculation of distance, lengths of lines, area and perimeter of polygons, and volumes, Measurements involving points include distances from a point to a point, a lines or a polygon, enumeration of total number as well as the enumeration of points falling within polygon.

4.2.2 Overlay Operations

Overlaying of maps result in the creative on a map where the values assigned to every location on that map are computed as a function of independent values associated with that location on two or more existing maps. Overlaying operation create a new data set containing new polygons formed from the intersection of the boundaries of the two or more sets of separate polygon layers. Arithmetical and logical overlay operations are common in all GIS software packages. Arithmetical overlay includes operations such as addition, substraction, division and multiplication of each value in a data layer by the value in the corresponding location in the second data layer. Logical overlay involves the selection of an area where a set of conditions are satisfied.

Fig.14 depict the overlaying concept in a vector structure (topologic overlay)

4.2.3 Neighbourhood Operations:

Neighbourhood operations involve the creation of new data based on the consideration of roving window of neighbouring points about selected target locations. They evaluate characteristics of an area surrounding a specified location. In all Neighbourhood operations it is necessary to indicate one or more target locations, the neighbourhood considered around each target and the type of function to be executed on the attributes within the neighbourhood. The typical neighbourhood operations in most GIS are Search, Topographic functions and Interpolation.

Search Functions-

The search function is one of the most common of the

35

SOIL MAP LAND USE MAP
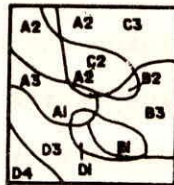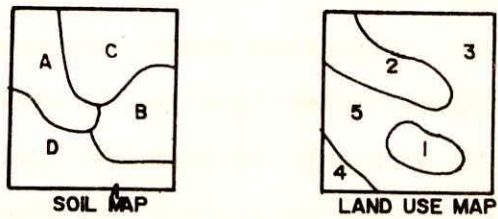
SOILS AND LAND USE MAP

LAND USE

USE CO	USE NA
1	FOREST
2	AGRICULTURE
3	PASTURE
4	SHRUBS

SOILS

SOILS CO	DEPTH
A	150
B	200
C	160
D	190

LAND USE / SOILS

SOIL-USE	USE NA	DEPTH
A1	FOREST	150
A2	AGRICULTURE	160
A3	PASTURE	150
B1	FOREST	200
B2	AGRICULTURE	200
B3	PASTURE	200
C2	AGRICULTURE	160
C3	PASTURE	160
D1	FOREST	190
D3	PASTURE	190
D4	SHRUBS	190

FIG 14 : OVERLAY PROCEDURE IN AVECTOR SYSTEM

neighbourhood operations. This function assigns a value to each target feature according to some characteristics of its neighbourhood. The three basic parameters to be defined in a neighbourhood search are targets, the neighbourhood, and the function to be applied to the neighbourhood to generate the neighbourhood value. The search area is commonly square, rectangular or circular with a size selected by the analyst.

Topographic functions:-

Topography refers to the surface characteristics i.e. the slope, relief and form of an area. The topography of a surface can be represented in a digital elevation Model (DEM). A DEM represents a topographic surface in terms of a set of elevation values measured at a finite number of points and contains terrain features of geomorphological importance such as valleys and ridges, peaks etc. Topographic functions are used to calculate values that describe the topography of an area. The most common transformations working with elevation data are the slope, which is defined as the rate of change of elevation and the aspect which is the direction that a slope faces. The maximum slope is called the gradient.

Interpolation:

Interpolation is the procedure of predicting unknown values using the known values at neighboring locations.
The neighboring points may be regularly or irregularly spaced. Interpolation programs employ a range of methods to predict unknown values including polynomial regression, fourier series moving averages and kriging etc. The quality of the interpolation results is a function of the precision accuracy, number and

37

function to calculate proximity (straight line distance) and area to be analysed. A buffer zone may be the result of a proximity analysis.

Network:-

Network functions are commonly used in analysis that require moving resources from one location to another. Examples of networks includes stream of a drainage basin, a city's streets, a grid of power transmission lines etc. GIS is used to perform network analysis such as prediction of network loading like for instance transport of water and sediments in a fluvial system, route optimization such as airline scheduling municipal garbage collection, and resource allocation e.g. subdivision of municipal district into zones that can be efficiently serviced by hospital and schools.In network analysis four component are usually considered: a set of resources (e.g. sediments transported by water), one or more locations where the resources are located (e.g. a fluvial system), a destination (e.g.outlet of the watershed);and a set of constraints (e.g. only permanent streams of higher order).

4.3 Modelling

In the last paragraphs an over view of the analysis functions available in GIS is given. A characteristic of modelling is the use of attribute data i.e. each map has one or several tables that include a specific single datum (attribute) of the pertinent map. For example in Fig. 15 universal soil loss equation (USLE) is used to depict soil loses estimation. The TMU map are reclassified with the KLS factor (erosion susceptibility) and
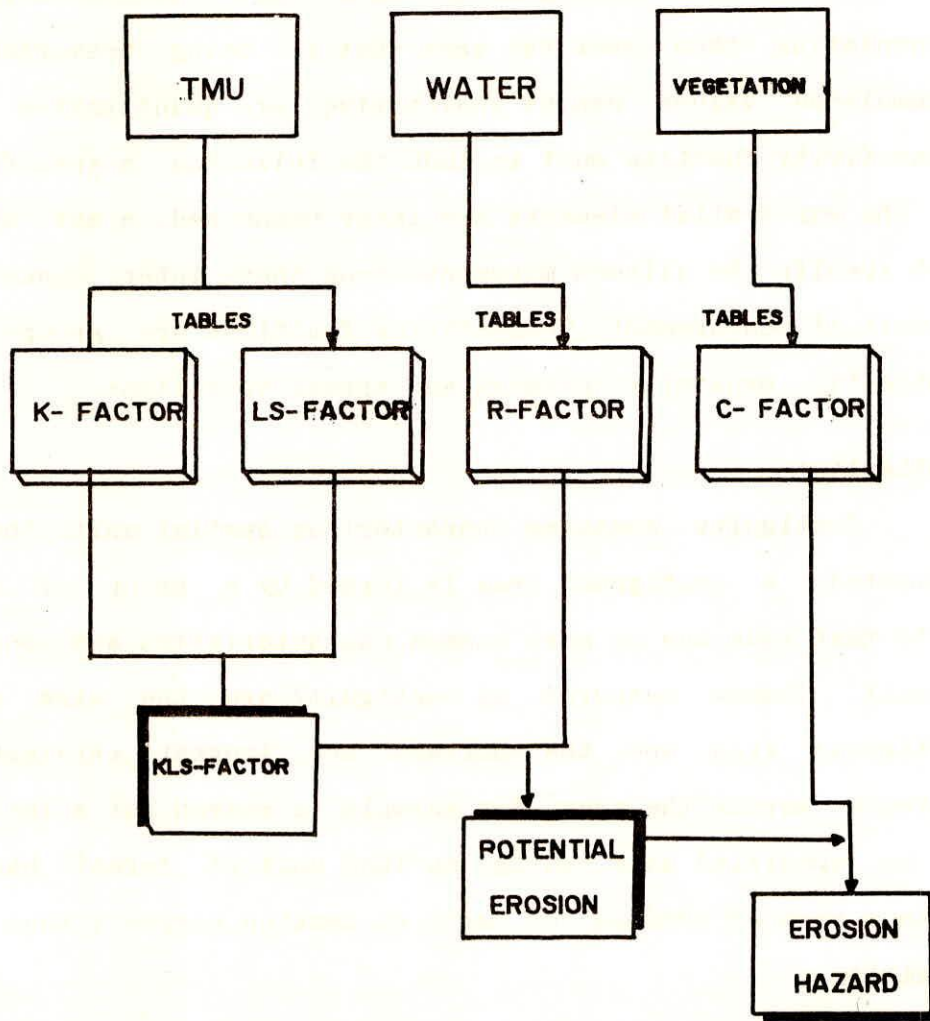
38

FIG. 15 : SCHEMATIC REPRESENTATION OF AN EROSION MODEL

distribution of the known points used in the calculation and the manner in which the mathematical function models reality. The unknown values are then calculated according to this function.

4.2.4 Connectivity Functions

Connectivity operations are those that estimate values by accumulating them over the area that is being traversed. The accumulated values can be qualitative or quantitative. Every connectivity function must include the following: a specification of the way spatial elements are inter connected, a set of rule that specify the allowed movement along these inter connections, a unit of measurement. Connectivity functions are grouped into contiguity, proximity, network and spread operations.

Contiguity:-

Contiguity measures characterizes spatial units that are connected. A contiguous area is formed by a group of spatial units that have one or more common characteristics and constitute a unit. Common measures of contiguity are the size of the **contiguous area** and the **shortest** and **longest** straight link distance across the area. For example, a search for a land unit to be specified as a contiguous land unit of forest having a minimum area of 1000 sq. km. with no section narrower than 20 km.

Proximity :

Proximity involves the measurement of the distances between features. The measurement unit can be distance in length, travel distance in time or other units. Four parameters must be specified to measure proximity: the features or objects (roads, wells, houses), the units of measure (meters, seconds), a

multiplied with the R- factor (rain erosivity) to obtain the potential soil erosion. This map is then multiplied by vegetation map reclassified. with the C-factor (vegetative cover) to generate the soil erosion hazard map (soil losses) under the current conditions.

5.Ø ERRORS IN GEOGRAPHIC INFORMATION SYSTEMS

5.1 Introduction

Geographic Information Systems are used to edit, store, manipulate and display the data as maps and colour images for use in a variety of resource management applications. The data used in GIS should be of good quality and without error. Two sources of error, inherent and operational ,contribute reduction in accuracy of the products that are generated by GIS. Inherent error is the error present in source documents. Operational error is produced through the data capture and manipulation functions of a GIS. Although it is not possible to remove errors from the system, they can at least be managed and kept to an acceptable minimum.

5.2 Sources of Error -

There is error associated with all geographic information. Error can arise at every stage of using a GIS, from the collection of the original data to the output and use of the resulting information. The objective in dealing with error should not be to eliminate is but to manage it. The level of error in a GIS needs to be managed that data errors will not invalidate the information that the system is used to provide common sources of error encountered in using a GIS areas follows:-

Errors in the source data-

Error exists in the original source materials that are entered into the GIS. These errors may be a result of inaccuracies in field measurements, inaccurate equipment ,or incorrect recording procedures common errors are-

- errors in the analysis of remote sensed data
- errors in field data collection
- errors in· the compilation of source maps

Errors in data input-

The data input devices used to enter geographic data all introduce opositional errors. These are digitizing errors due to operator mistakes and limited precision of the digitizer and errors in attribute data entry (typing error).

Errors in Data Storage-

When data are stored in digital form, they must be stored with a finite level of precision. In data storage errors are arised due to the limited precision with which coordinates and other numerical data are stored. Errors also occur while converting data from vector to raster format.

Errors in data analysis and manipulation

Many operations in GIS involve the procedure of overlaying one or more spatial networks. As the number of overlays used in an analysis increase, the number of possible opportunities for error increase. Errors are also arised from interpolation e.g. for the determination of terrain slope from digitized coutours. The errors are also due to incorrect use of formula.

Errors in data output and application

At the output stage, errors can be introduced in the plotting of maps by the output device and by the shrinkage and swelling of map material. As paper shrinks and swells, measurements taken from the map will be changed. On a small scale map, the millimeter changes can represent several meters at

6.0 A CASE STUDY

Evaluation of the Ground Water Recharge to the Bandung basin (West Jawa Indonesia)

6.1 Introduction:-

The city of Bandung (West Java, Indonesia) is a rapidly expanding city (fig-16). The city faces great problems regarding inadequate supply of water for both human consumption agriculture and Industry. Therefore more and more groundwater is used in the recent decades. Potentially the Cikapundung catchment located north of Bandung, is a well protected area with recharge possibilities of groundwater of good quality. However, the lembang fault, situated between the catchment and the Bandung basin may or may not inhibit the flow of ground water. If this flow is restricted the importance of the Cikapundung catchment for groundwater recharge would be minor.

The objective of this study is to asses the importance of the Cikapundung catchment as a recharge area for the bandung basin using a simple water balance model within a GIS environment.

6.2 Data availability-

To make a reliable assessment the groundwater recharge normally requires a adequate amount of good vegetation, topography and climate. The data made available to us were in the form of image file and GIS files.

The file are:-

Topographic Map   : Lembang.Seg

Lembang.Crd

FIG. 16 : LOCATION MAP

```
Catchment area Map          : Catchment. Seg

                              Catchment. Crd

                              Catchment. Pol

                              Catchment. Top

Soil Map                    : Soil. Crd

                              Soil. Seg

                              Soil. Pol

                              Soil. Top

Detention for March         : Detmar. mpd

                              Detmar. mpc
```
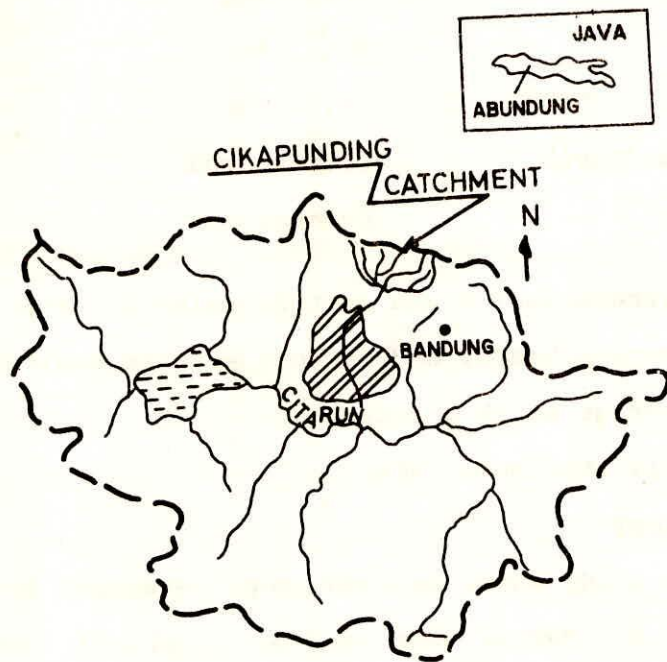
Table 1   Average Yearly and monthly rainfall (mm)

Table 2   Average Yearly and monthly evapotranspiration (mm)

Table 3   Average monthly discharge

Table 4   Soil attribute data.

6.3 Methodology-

In this study water balance model developed by C.W. Thornth waite and J.R. Mather was applied. Fig.17 shows a graphic representation of this model.

As input for the water balance calculation data from different types, formats and data source are needed. This is shown in Table given below

Table - Data needed for the water balance calculations

```
-----------------------------------------------------
Data types                Format
-----------------------------------------------------

Monthly rainfall          Tables

Monthly Pan evaporation   Tables

Soil profile descript     Tables
```

ET   RAIN

SURFACE STORE $SR = C_1 * RAIN$

                       $SPECH = (RAIN-SR)-ET$

EA

WHC

SOIL STORE   PREDICTED CATCHMENT
                                             OUTFLOW

                       $GRECH = SRECH-(WHC-CH)$

WHC $SM = WH(C*EXP^{(A7WI WHC)}$

SOIL MOISTURE DEPLETION CURVE

GROUND WATER STORE $GRO = (DET + GRE) * C_2$
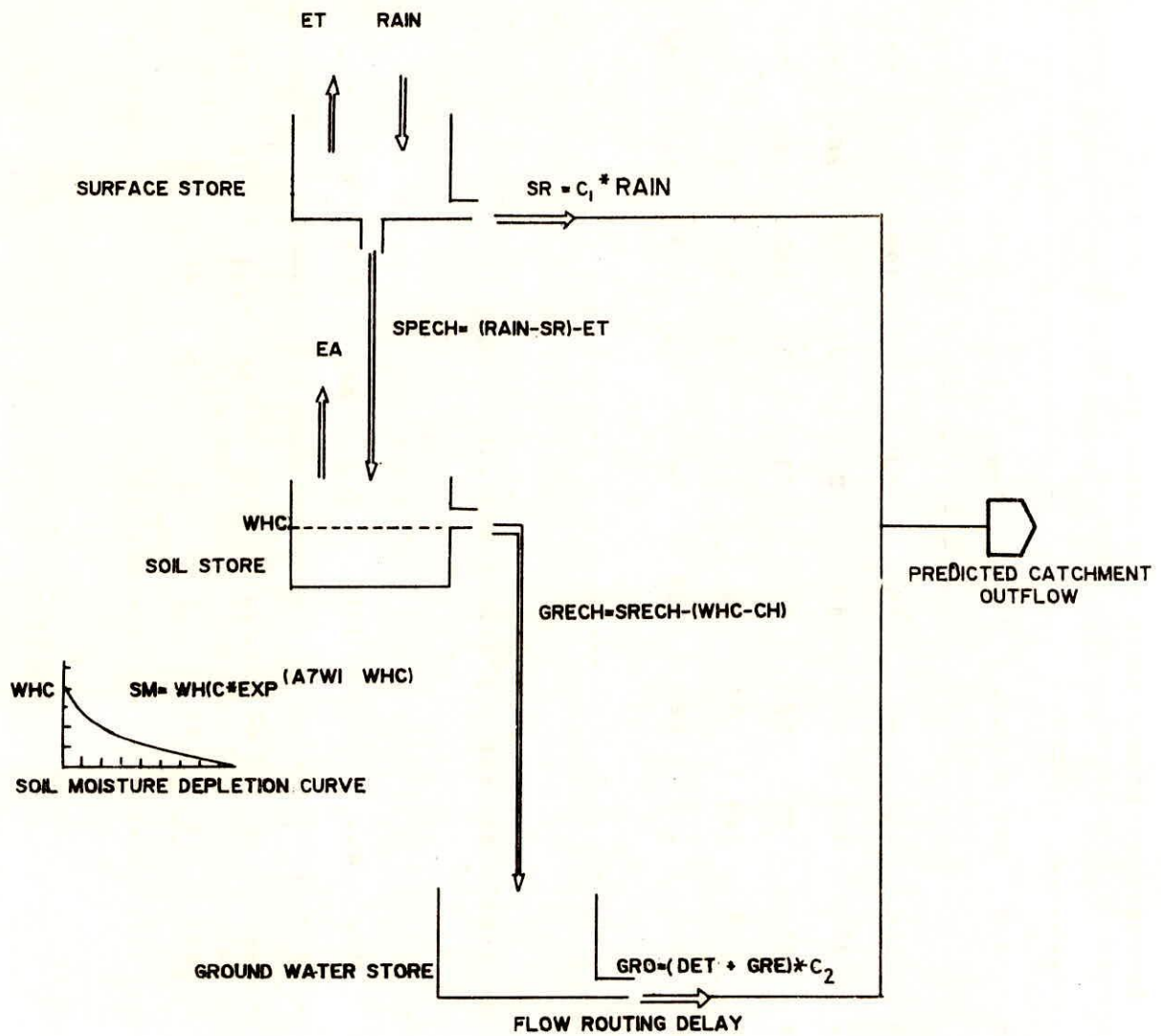
              FLOW ROUTING DELAY

FIG. 17 : GRAPHIC REPRESENTATION OF THE WATER
BALANCE MODEL

Table: 1 Average yearly and monthly rainfall (mm)

Name	Height	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	Year
Pakar Dago	900	245	201	256	216	146	67	56	59	37	175	253	267	2028
Cikapundung	1400	307	270	351	257	197	62	43	23	79	139	268	336	2332
Lembang	1241	237	202	216	258	188	85	51	63	106	155	266	259	2086
Marghayu	1250	233	214	249	268	184	102	94	94	118	207	299	301	2363
Tangkuban Prahu	1800	353	349	428	396	278	227	128	149	162	353	421	380	3624

Table 2 : Average yearly and monthly evapotranpiration (mm)

Name	Height	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	Year
Bandung	670	51	65	69	76	75	77	82	90	91	94	78	71	920
Margahayu	1250	63	62	63	69	77	74	82	82	79	91	77	65	881

Table 3 : Average monthly discharge (m^3/s)

Location   : Cikapundung - Maribaya
Main river   : Citarum
Station no.   : 02-016-04-05
Number of years : 37

Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	Year
4.15	4.25	4.43	4.97	4.36	3.30	2.73	2.36	2.16	2.47	3.59	4.12	3.57

Table 4 : Soil attribute data

Soil Unit	Available water mm/m	Rooting depth
Unit 1	125	1.10
Unit 2	75	0.60
Unit 3	100	0.80
Unit 4	150	1.50

Run-off coefficients	Tables
contours	Topographical
	Map 1:80,000
Soil Map	Soil Map 1,50,000
Landuse	Digital raster image

The Procedure for integrating the different Data sets is shown in a flow chart in fig.18 and flow chart of the water balance calculations are shown in fig. 19.

For the processing and integration of the data they integrated land and watershed management information system (ILWIS) is used. Rainfall point data of 5 different stations located in and around the catchment area are stored in the ILWIS tabular data base. Using the capabilities of the tabular data base for calculating linear regression, the altitude rainfall relationship for every month is established. y (rainfall) = a +a * x(altitude) . So for each month we get a function say CJan, C Feb ----and so on.

Pan evaporation data from one regional meteorological stations was stored in ILWIS tabular data base.

A file Lembang. Seg was provided in which contours, roads, towns were shown. From this we get a contour map by masking other features. By interpolation from isoline we get elevation of each point in the map Height (say). Now classifying this map with function of respective month we get rainfall at each point. For January

P Jan (Say) = CJan (Height) & so on.

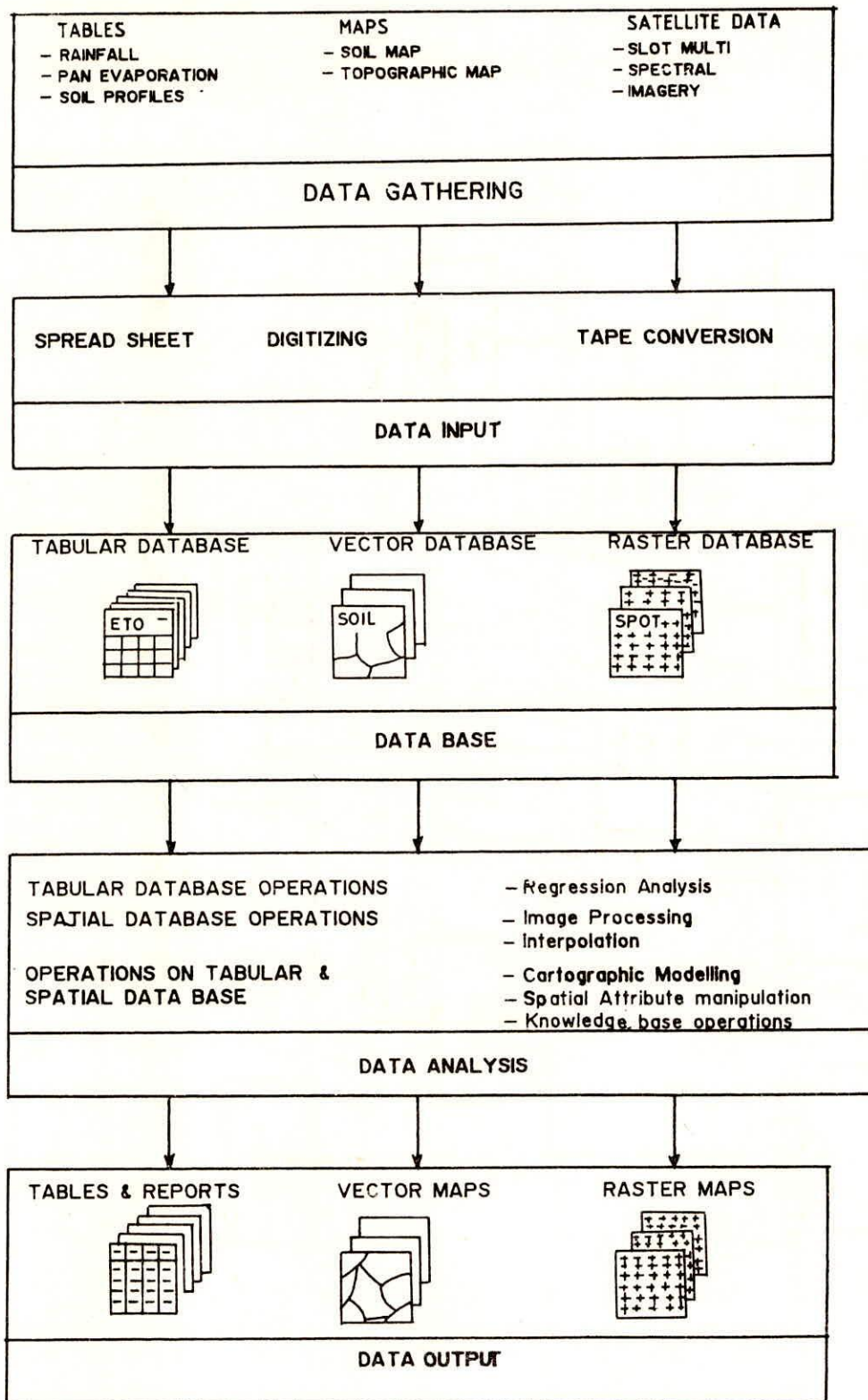From the total amount of effective rainfall that reaches

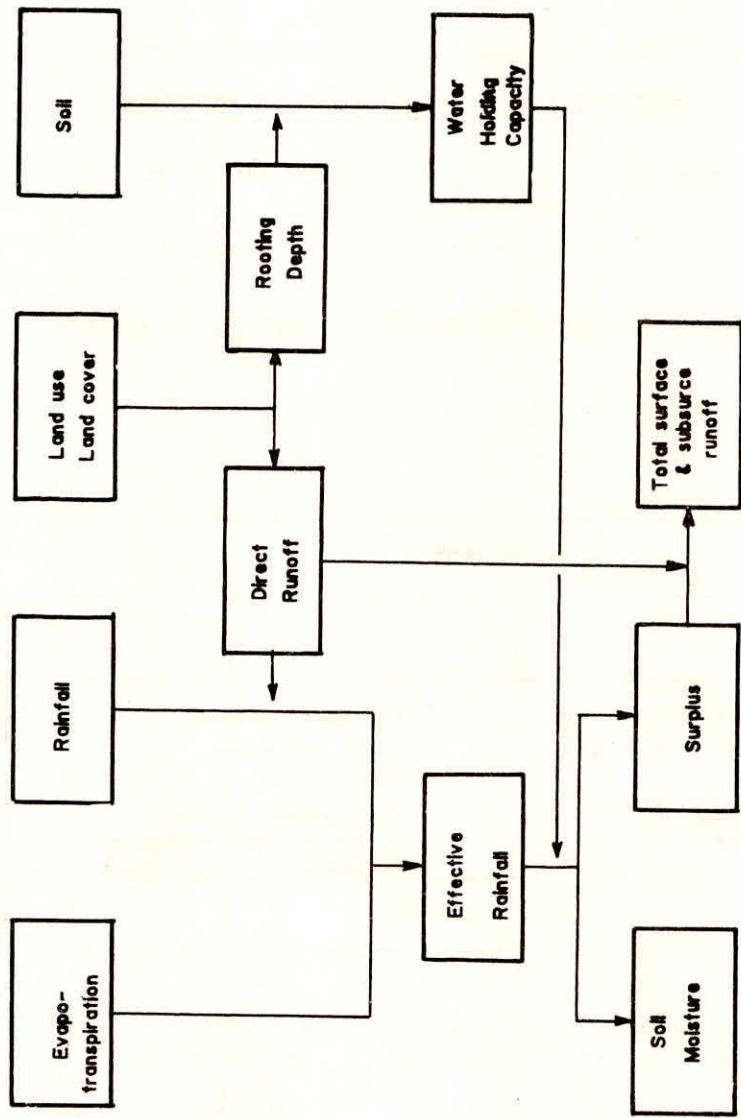FIG 18 : FLOW CHART INTEGRATING DIFFERENT DATA SETS

FIG. 19 : FLOW CHART OF THE WATER BALANCE CALCULATIONS

the earth surface, part is returned to the atmosphere by evapotranspiration (ET). The remaining part (SRECH) IS available for infiltration into the soil.

SRECH = (Rain - SR) - ET

C1 = 0.15 (given)

For Jan SRECH (Jan) = 0.85 * P Jan - ET

ET for each month is given in Table 2. Similarly for each month we can calculate SRECH Say CJan and so on

Now next step is to calculate GRECH.

For this first water holding capacity (WHC) is to be calculated using the Tabcalc capability of ILWIS

WHC = A W * RD (Table No. 4)

Now soil.pol map is provided. After rasterising this map and with the help of Table No. 4 (say SOIL)

Map WHC = SOIL WHC [SOIL]

When the soil is not yet at its water holding capacity (WHC) and SRECH is positive SRECH will be used to replenish the soil Moisture (SM):

$$SM_i = (SM)_{i-1} + SRECH_i$$

As soon as the soil moisture reach WHC, the remaining part (GRECH) will percolate to the Ground water.

When SRECH is negative (i.e. the effective rainfall < ET) water will be with drawn from the soil moisture. The high atmospheric demand for water (ET) can not be met by the effective rainfall . This is the reason why additional water is withdrawn from the soil moisture However, the soil moisture depletion curve is not linear but has exponential shape, the drier the soil gets, the more difficult is become to extract

water. Therefore the actual evapotranspiration (EA) is less than ET. The soil moisture depletion curve is described by the following formula :

SM = WHC * EXP (-APWL\WHC)

where APWL = Accumulated potential water loss

The APWL is available which describes the dryness of the soil. In months with a deficit of water (SRECH <0) the

APWL(i) = APWL(i-1) + SRECH

where i = month number

In months with a surplus of water (SRECH >0) the APWL equals 0 .

If month (i-1) , with a surplus of water, is followed by month (i),with a deficit a starting APWL value has to be calculated using the formula

APWL (i-1) = - WHC * LN (SM (i-1)/WHC)

April is one the last month of rainy season and the the soil will be at field capacity, so starting calculation from the month of April.

SMAPR = WHC (in april the soil is saturated)

WLAPR = 0 (Hence no accumulated potential water loss)

Now for, May WL May & SM May are as follows

WL MAY = IF (SMAY.LT.0, WLAPR + SMAY, if(SM APR +SMAY .GT. WHC,0, log (SM APR)/WHC * WHC))

SM MAY = IF (SMAY.LT.0,WHC * Exp (WLMAY/WHC),

 if((SM APR + SMAY) GT. WHC,WHC,SM APR + SMAY))

and similarly for subsequent months we can calculate water
loss & soil Moisture.

Now   Ground water recharge

GRMAY = if (SM   MAY .LT. WHC,0, SM APR + SMAY- WHC) and so on.

Now last step is to calculate Ground water run-off.

When   GRECH in month (i) >0 water is added to the water spill present in the ground water store (DET (i-1)).

The ground water store acts as a buffer and causes a delay in the Ground water run-off. Therefore not all the water in the store will become part of the GW flow. Only a certain fixed percentage will run-off in the same month; the rest is detained till next month.

(GRO) i = C2 * (DET (i-1) + GRECH (i))

The new detention value will be

(DET) i = (1-C2) * DET (i-1)+ GRECH (i))

For MAY, C2= 0.2 given

DET MAY = (DET APR + SUR MAY)* 0.8

GRO MAY = (DET APR + SUR MAY)* 0.2

Now DET MAR is given so we can calculate DET APR and so on.

Now the direct storm run-off (SR) and the ground water run-off (GRO) together will form the total predicted catchment out flow '(both as surface and subsurface run-off).By comparing the predicted out flow with the measured outflow an assessment can be made of the groundwater recharge. so outflow is given by

OJAN = 0.15 * PJAN + GRO JAN and so on for other months. Sum of surface and subsurface run-off is shown in figure 20. By comparing the combined run-off (surface and subsurface run-off) and the measured stream flow at the gauging site (table 5), we find that both values are almost equal. So the results

55

suggests that negligible quantity of water leaves the catchment area through fault and all water is discharge as surface water. However the method is too rough to be used as a prove of ground water flow.
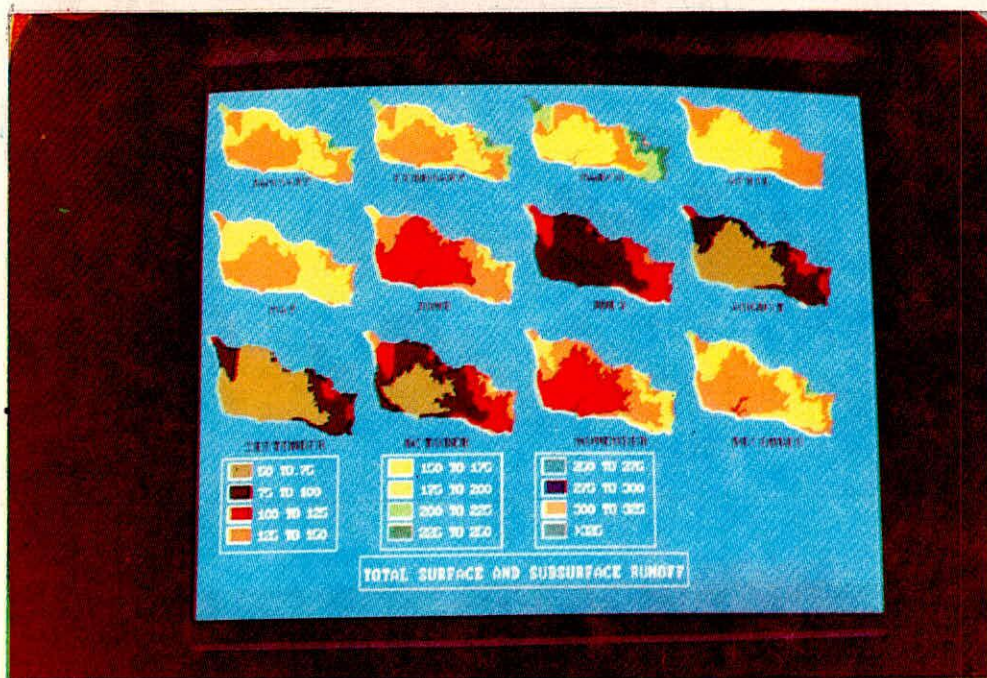
FIGURE 20 TOTAL SURFACE AND SUBSURFACE RUNOFF

Table 5:- Comparison between calculated run-off and measured stream flow at gauging site

Month	Calculated run-off(mm)	Measured Streamflow(mm)
Jan.	164.75	148.20
Feb.	167.26	141.98
March	189.58	158.20
April	176.94	171.76
May	157.63	155.70
June	123.82	114.05
July	94.51	97.49
August	78.83	84.28
Sept.	72.61	74.65
Oct.	93.41	88.21
Nov.	132.54	124.07
Dec.	156.53	147.13

REFERENCES

1.   Aronoff, S. (1989), "Geographic Information System: A management perspective, Ottawa, WDL publications.

2.   Burrough, P.A. (1986), "Principles of geographic information systems for land resources assessment, clarendon press, Oxford, U.K.

3.   Marble, D.F. and D.J. Peuquet (1983), "Geographic information system and remote sensing, In: R.N. Colwell (ed.) Manual of Remote sensing 923-958 ASPRS, Falls Church, Va.

4.   Meijerink A.M.J., C.R. Valenzuela and A.Stewart (1988), "ILWIS, The integrated land and watershed management information system, ITC publication number 7, ITC, Enschede the Netherlands.

5.   Valenzuela C.R., 1988, ILWIS Overview, ITC Journal 1988-1, Enschede, The Netherland.

6.   Valenzuela, C.R. (ed.) 1990), "Introduction Geographic Information System, ITC, Enschede, The Netherlands.