GI-03

KRISHNA KANTA HANDIQUI STATE OPEN UNIVERSITY

POST GRADUATE CERTIFICATE PROGRAMME IN GEOINFORMATICS

COURSE 3

SPATIAL ANALYSIS AND APPLICATION IN GEOINFORMATICS

CONTENTS

UNIT-1: Spatial Data Modeling and Database Management

UNIT-2: Data Analysis

UNIT-3: Geo-processing Functions and Tools

UNIT-4: Analytical Modeling and Output

UNIT-5: Basics of Image Processing-I

UNIT-6: Basics of Image Processing-II

UNIT-7: Basics of Image Classification

UNIT-8: Accuracy Assessment in Image Classification

UNIT-9: Geoinformatics for Natural Resources Assessments

UNIT-10: Geoinformatics for Environmental Studies, Planning and Development

Subject Experts

Prof. Anup Saikia, Dept. of Geography, Gauhati University, Guwahati

Dr. Santanu Sarma, Dept. of Geology, Cotton University, Guwahati

Prof. B. S. Mipun, Dept. of Geography, North East Hill University, Meghalaya

Course Coordinators : Antara Mahanta Barua, Assistant Professor, KKHSOU

SLM Preparation Team

UNITS	CONTRIBUTORS
1	Dr. Manoj Hazarika, Associate Professor, Dept. of Geography, Cotton University, Guwahati
2, 4, 7, 8	Dr. Sarat Phukan, Professor, Department of Geological Science, Gauhati University, Guwahati
3	Mr. Abul Fazal Murtaza Ahmed, Guest Faculty, Department of Geography, Cotton University, Guwahati
5, 6	Dr. Suranjana Borah, Senior Specialist, GIS and Remote Sensing, International Rice Research Institute, Guwahati
9, 10	Dr. Swarup Jyoti Baishya, Assistant Professor, Dept. of Earth Science, USTM, Meghalaya

Editorial Team

Content : Dr. Narayan Chetry, Formerly Associate Professor, Department of Geography, Cotton University, Guwahati

Structure, Format & Graphics : Antara Mahanta Barua, KKHSOU

First Edition : November, 2020 **ISBN:** 978-93-89955-78-1

© Krishna Kanta Handiqui State Open University

This Self Learning Material (SLM) of the Krishna Kanta Handiqui State University is made available under a Creative Commons Attribution-Non Commercial-ShareAlike4.0 License (International): http.//creativecommons.org/licenses/by-nc-sa/4.0.

For the avoidance of doubt, by applying this license KKHSOU does not waive any privileges or immunities from claims that it may be entitled to assert, nor does KKHSOU submit to the jurisdiction, courts, legal processes or laws of any jurisdiction.

Printed and published by Registrar on behalf of the Krishna Kanta Handiqui State Open University.

Headquarters: Patgaon, Rani Gate, Guwahati-781017

City Office: Khanapara, Near Bodoland Guest House, Guwahati-781022; Web: www.kkhsou.in

The University acknowledges with strength the financial support provided by the **Distance Education Bureau, UGC** for preparation of this material.

Post Graduate Certificate Programme in Geoinformatics

Course 3
GI-03 SPATIAL ANALYSIS AND APPLICATION IN GEOINFORMATICS

MARKS 100
DETAILED SYLLABUS

Credits - 4

CONTENTS

		Marks	Pages
UNIT 1 :	**Spatial Data Modeling and Database Management** Spatial Data and Its Characteristics; Various Sources of Spatial Data; Spatial Data Models: Raster Data Model, Vector Data Model, Topology; Spatial Data Structures: Raster Data Structure, Vector Data Structure; Database Management Technology in GIS; Methods of Data Input and Data Editing: Georeferencing of Objects, Data Input and Methods, Data Editing.	**10**	**6-26**
UNIT 2 :	**Data Analysis** Introduction to Data Analysis basics; Database Queries; Geospatial Measurements; Neighbourhood Operations, Overlay Operations; Surface Analysis Uses of DEM / DTM / DSM; Network Analysis.	**10**	**27-51**
UNIT 3 :	**Geo-processing Functions and Tools** Vector overlay; Raster overlay; Spatial Buffering; Union and Intersection; Spatial Autocorrelation; Weighted Regression.	**10**	**52-66**
UNIT 4 :	**Analytical Modeling and Output** Analytical Modeling basics; Process Models: Physical and Environmental Process Modeling, Decision Making Process Modelling, Human Process Modelling; Output of GIS: Maps as output; Spatial Decision Support Systems.	**10**	**67-78**

UNIT 5 :	**Basics of Image Processing-I** Introduction to Digital Image Processing (DIP); DIP software Systems: User interface of DIP software, Conversion of Digital Data in DIP Software; Image properties: Spatial resolution, Radiometric resolution, Image histogram; Data preparation: Geometric corrections, Radiometric corrections; Reprojection and resampling of digital data; Subsetting digital images.	**10**	**79-96**
UNIT 6 :	**Basics of Image Processing-II** Geo-referencing Data; Pattern Recognition in Digital Image Processing;Image enhancement: Image filtering, Band ratio, Principal component analysis.	**10**	**97-113**
UNIT 7 :	**Basics of Image Classification** Introduction to Image Classification; Unsupervised Classification, Supervised Classification.	**10**	**114-125**
UNIT 8 :	**Accuracy Assessment in Image Classification** Concept of Accuracy Assessment**:** Need for accuracy assessment; Types of accuracy assessment.	**10**	**126-134**
UNIT 9 :	**Geoinformatics for Natural Resources Assessments** Application of Geoinformatics in land use and land cover classification systems; Mapping and change detection analysis; Geo-environmental studies; Biotic resource assessments.	**10**	**135-144**
UNIT 10:	**Geoinformatics for Environmental Studies, Planning and Development** Geoinformatics for environmental pollution and climate change studies; Floods and drought Prone Area mapping; Early warning systems for disasters and hazards; Rural and urban infrastructure survey; Mapping and planning.	**10**	**145-157**

COURSE INTRODUCTION

"Spatial Analysis and Application in Geoinformatics" is the third course of this programme which is designed to provide the conceptual overview of Geospatial data analysis, image processing, image classification and some applications of Geoinformatic in natural resources and environmental studies.

The whole course is divided into ten units.

The **first unit** covers the concept of spatial data and its characteristics, various sources of spatial data, spatial data models, spatial data structures, database management technology in GIS and methods of data input and data editing.

The **second unit** includes the Introduction to data analysis, database query, geospatial measurements, neighbourhood operations, overlay operations, surface analysis and network analysis.

The **third unit** deals with vector overlay, raster overlay, spatial buffering, union and intersection, spatial autocorrelation and weighted regression.

The **fourth unit** covers the analytical modelling basics, process models, output of GIS and spatial decision support systems.

The **fifth unit** covers the Introduction to digital image processing, dip software systems, image properties, data preparation, reprojection and resampling of digital image data and subsetting of digital image data.

The **sixth unit** deals with the Geo-referencing of raster data set, pattern recognition in digital image processing and image enhancement.

The **seventh unit** includes the Introduction to image classification, unsupervised classification and supervised classification.

The **eighth unit** covers the concept and need of accuracy assessment, types of accuracy and assessment of classification accuracy.

The **ninth unit** covers the application of geoinformatics in land use and land cover classification, mapping and change detection analysis, geo-environmental studies and biotic resource assessments.

The **tenth unit** covers the Geoinformatics for environmental pollution and climate change studies, Floods and drought prone area mapping, Early warning systems for disasters and hazards, Rural and urban infrastructure survey and Mapping and planning.

Each unit of this block has some questions as "Check Your Progress", to be answered by the learners while going through the units. The learners can tally their answers with the answers given at the end of each unit as "Answers to Check Your Progress". This will help them in judging their performance. The units of this block include some along-side boxes to help you know some of the difficult, unfamiliar terms. The learners will come across some "Activity" within the units which will help them in gaining practical experience in the related topics.

UNIT- 1 : SPATIAL DATA MODELLING AND DATABASE MANAGEMENT

UNIT STRUCTURE

1.1 Learning Objectives

1.2 Introduction

1.3 Spatial Data and Its Characteristics

1.4 Various Sources of Spatial Data

1.5 Spatial Data Models

- 1.5.1 Raster Data Model
- 1.5.2 Vector Data Model
- 1.5.3 Topology

1.6 Spatial Data Structures

- 1.6.1 Raster Data Structure
- 1.6.2 Vector Data Structure

1.7 Database Management Technology in GIS

1.8 Methods of Data Input and Data Editing

- 1.8.1 Georeferencing of Objects
- 1.8.2 Data Input and Methods
- 1.8.3 Data Editing

1.9 Let us Sum up

1.10 Further Reading

1.11 Answers to Check Your Progress

1.12 Model Questions

1.1 LEARNING OBJECTIVES

After reading this unit, you will be able to:

- explain about the spatial data
- discuss about the various sources of spatial data
- describe the spatial data models

- outline the spatial data structures
- illustrate database management technology in GIS
- narrate methods of data input and data editing.

1.2 INTRODUCTION

This is the first unit of the third course. This unit focuses on how geographic reality is modelled or simplified in a GIS, with particular emphasis on choosing one particular style of data model over another. In this unit you are going to learn about spatial data and its characteristics. We are also going to discuss the various sources of spatial data, spatial data models and spatial data structures. This unit also deals with the GIS database management technology and its applications. Apart from this, the unit will familiar you to the methods of data input, and data editing along with an overview of Integrated Database Systems.

1.3 SPATIAL DATA AND ITS CHARACTERISTICS

A data ***model*** is an essential component of any operational GIS. It has important implications for the types of operations and results that can be performed, as well as, obtained through GIS. The heart of any GIS is the data model, which is a set of constructs for representing objects and processes in the digital environment of the computer (Figure 1.1). People (GIS users) interact with operational GIS in order to undertake tasks like making maps, querying databases, and performing site suitability analyses.

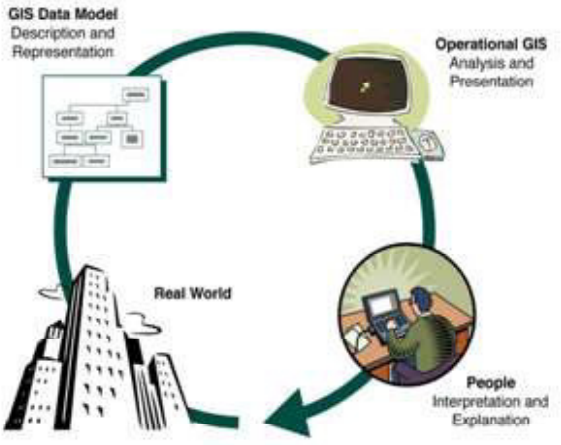


Figure 1.1: GIS Data Model and representation

Model:
A simplified representation explains the workings of a real world system.

A data model is a set of idea or concept for describing and representing selected aspects of the real-world in a computer. Some choices have to be made on the things that are modeled in a GIS and how they are represented. Different types of people use GIS for different purposes, and the phenomena these people study have different aspects. But unfortunately, there is no single type of GIS data model that is suits for all circumstances.

2-D Plane: It is also called Euclidean plane or surface with length and breadth of the space.

Spatial Data is the graphic data on the space that describes the geometry of the spatial features. The important characteristics of spatial data are:

- Spatial data describe the location based (x,y) on a ***2-D plane***.
- Spatial data represents the shape and size of Earth surface features.

Discrete or continuous: An individual building is a discrete object. The slope of an area is continuous.

- Spatial data can either be ***discrete or continuous***.
- Spatial data can be represented graphically by using various colours, symbols or signs.
- Spatial data can be linked to a desired attribute. The attribute may text, number, colour or symbols/ graphic objects.



CHECK YOUR PROGRESS

Q.1: State True or False:

I. Geographic reality can be modeled with the help of GIS.

II. The word ‘users’ refers to the people, who interact with operational GIS in order to undertake tasks.

III. Spatial data can neither be discrete nor continuous.

Q.2: Fill in the blanks:

I. _____ is an essential component of any operational GIS.

II. Geographic reality is _____, but computers are simple.

III. _____ represents information about the shape of Earth surface features on the Earth’s surface.

1.4 VARIOUS SOURCES OF SPATIAL DATA

Data Acquisitions:

The act or process of acquiring data into GIS software.

The power of GIS comes from the ability to relate different information in a spatial context and to reach a conclusion about these relationships. Most of the information we have about our world contains a location reference. A GIS, which can use such information from many different sources in many different forms, can help in various analyses. The data sources for ***data acquisitions*** should be carefully selected for specific purpose and the identified sources are as follows:

Data Acquisitions: The act or process of acquiring data into GIS software.

- **Conventional analog map sources:** Conventional analog maps including all types of Administrative maps, Topographic maps etc. are the most primary sources of spatial data. The locational information in terms of geographic grids should be available along with the scale of the map. These maps can be digitised with the help of a scanner and the data files with .tiff, .tif, .jpg or .bmp formats can be added as raster data in a GIS (described in the following section)
- **Aerial remote sensing:** Data acquired though aerial remote sensing e.g. Radar, Lidar data have been playing a vital role in the field of spatial data.
- **Satellite remote sensing:** Now-a-days remotely sensed satellite imageries provide the most popular source of spatial data for any GIS project.
- **Field data sources:** Data collected from surveying and Global Positioning System (GPS) is an inseparable source of spatial data.
- **Existing digital map sources:** There are a number of digital map sources for spatial data such as; USGS, Earth Explorer, BHUVAN, SRTM etc. where one can easily acquire spatial data for any purpose.



CHECK YOUR PROGRESS

Q. 3: What are various sources of spatial data?

.....

.....

1.5 SPATIAL DATA MODELS

In general, data models are constructed to represent space. These data models define how spatial features are represented in GIS. The two main spatial data models used in GIS are–

- Raster data model
- Vector data model

1.5.1 Raster Data Model

A Raster Data Model is variously called a grid, a raster map, a surface cover or an image in GIS. A raster represents a continuous surface with the help of discrete grids/ cells. But for data storage and analysis, a raster is divided into rows and columns forming cells. Cells are also called ***pixels*** with images. The origin of rows and columns is typically at the upper left corner of the raster. Rows function as y-co-ordinates and columns as x- co-ordinates. Each cell in the raster is explicitly defined by its rows and column position. The pixel will hold the attribute or the property of the object like the value of elevation in a model of terrain.

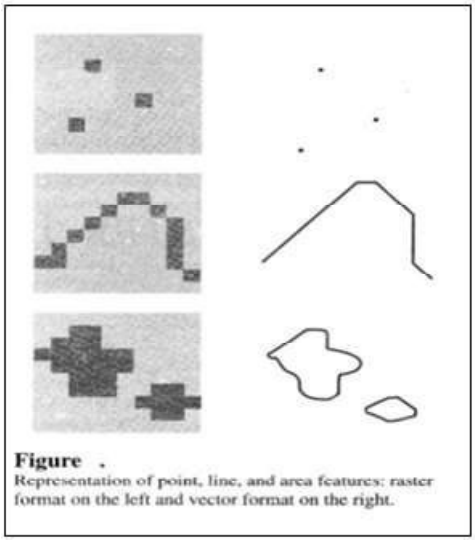


Figure1.2: Representation of point, line and area features.

Raster data represents points of single cells, lines with sequence of neighbouring cells and areas with collection of contiguous cells. In computing algorithms, a raster can be treated as a matrix with rows and columns, and its cell values can be stored in a two dimensional array. All commonly used programming languages can easily handle arrayed variables. Raster data are therefore much easier to manipulate, aggregate and analyse than vector data.

- **Cell Value:** Each cell in a raster carries a value, which represents the characteristics of spatial phenomenon at the location denoted by its rows and columns. Depending on the coding of its cell values, a raster can be either an integer or a floating-point.
- **Cell Size:** The cell size determines the spatial resolution of the raster data model. A cell size of 10 metre means that each cell measures 100 square metre (10 m x10m).
- **Raster bands:** A raster may have a single ***band*** or multiple bands. Each cell in a multiband raster is associated with more than one cell value. An example of a multiband band raster is a satellite image, which may have 4, 7 or more bands at each cell location. Each cell in a single band raster has only one cell value. An example of a single-band raster is an elevation raster, which has one elevation value at each cell location.
- **Spatial Reference:** Raster data must have the spatial reference information so that they can align spatially with other data sets are based on the same coordinate system. A raster that has been processed to match a projected coordinate system is often called a georeferenced raster.

Band: Generally, an image is formed due to variation in the characteristics of electromagnetic radiation in various wave lengths. A very small range in wave length is called a band or spectral band.

1.5.2 Vector Data Model

The Vector data Model uses the Geometric objects of point, line and area to represent simple spatial features. Dimensionality and property distinguish the three types of geometric objects, as well as, the features they represent.

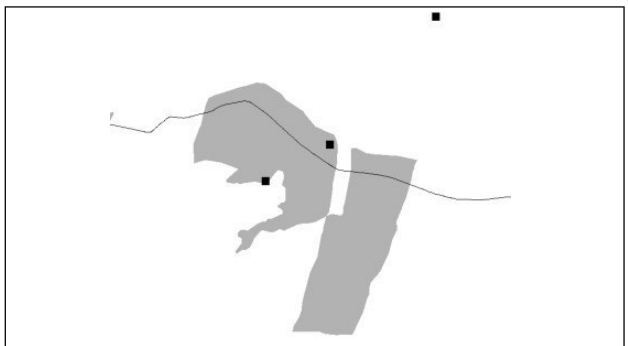


Figure 1.3: Representation of Point, Line, Polygons

- A **Point** has 0 dimension and has only the property of location. A point may also be called a node, vertex or 0-cell/ 0-D object. Wells, benchmarks, gravel pits are example of point features.
- A **Line** is one dimensional and has the property of length. A line is also called an edge, link chain or 1-cell/ 1-D object. Roads, streams and contour lines are examples of line features.
- An **Area** is two dimensional and has the property of area (size) and perimeter. An area is also called a polygon, face, zone or 2-cell/ 2-D object. An area feature is made of polygons. Examples of area features include land parcels, water bodies etc.

1.5.3 Topology

Topology is the study of those properties of geometric objects that remain invariant under certain transformations such as, bending or stretching (Massey, 1967). Topology based data sets require additional data files to store the spatial relationship between features.

Topology has two main advantages-

- The first advantage is the assurance of data quality. The Topological relationships enable us to detect errors such as lines that do not meet correctly or polygons that are not closed properly.
- Secondly, topology can enhance GIS analysis.

The coverage model enforces three topological relationships with its data structure. The rules that apply to features within a feature class are functionally similar to the build-up in topology for the coverage model. Some topology rules applied in data model are:

- **Polygon:** Must not overlap, must not have gaps, must not overlap with, must be covered by feature class of, must cover each other, must be covered by, boundary must be covered by, area boundary must be covered by boundary of, and contains points.
- **Line:** Must not overlap, must not intersect, must not have pseudo-nodes, must not intersect or touch interior, must not overlap with,

must be covered by feature class of, must be covered by boundary of, end point must be covered by, must not self overlap, must not self intersect, and must be single part.

- **Point:** Must be covered by boundary of, must properly be inside polygons, must be covered by end point of, and must be covered by line.

There are mainly three types of topological relationships (ESRI, Inc. 1998):

- **Connectivity:** Arcs connect to each other at nodes.
- **Area definition:** An area is defined by a series of connected arcs.
- **Contiguity:** Arcs have directions as well as, left and right polygons.

Apart from the types, name of some data models used in GIS are–

DATA MODEL	TYPE	EXAMPLES IN APPLICATIONS
CAD	Vector	Engineering design and drafting
Graphical	Vector	Simple mapping
Image	Raster	Image processing and grid analysis
Raster/Grid	Raster	Spatial analysis and modelling
Network	Vector	Network analysis
Geo-relational	Vector	Geoprocessing geometric features
TIN	Vector	Surface / terrain analysis / modelling
Object	Vector	Features with behaviour



CHECK YOUR PROGRESS

Q.4: State True or False:

I. Raster is divided into rows and columns form cells.

II. The origin of rows and columns is typically at the upper right corner of the raster.

III. Vector data uses geometric objects to represent earth surface features.

IV. A line is two dimensional, while an area is a one dimensional feature.

Q.5: Fill in the blanks:

I. Cells are also called _____with images.

II. A raster that has been processed to match a _____ called geo referenced raster.

III. There are mainly _____ types of topological relationship are present.

IV. In topology _____ must not have gaps.

1.6 SPATIAL DATA STRUCTURES

As we know, GIS lets us visualise, store, manipulate, question, analyse and interpret data to understand relationships, patterns and trends. In order to describe geographical features, we can take the help of:

- **Spatial Data,** which describes location(where)
- **Attribute data,** which specifies characteristics on that location.

We have discussed on representing data in GIS. We can represent these data digitally in GIS with the help of following two formats of data structure.

- Raster data structure and
- Vector data structure.

1.6.1 Raster Data Structure

Raster data structure represents geography via grid cells.

I. Cell-by Cell Encoding:

It is the satellite imagery which consists the scene of the same area at different wavelength, of electromagnetic radiation.

The cell-by-cell encoding method provides the simplest raster data structure. A raster is stored as a matrix, and its cell values are written into a file by row and column (as shown in the Figure 1.4). With multiple spectral bands, however, each pixel in a satellite image has more than one value.

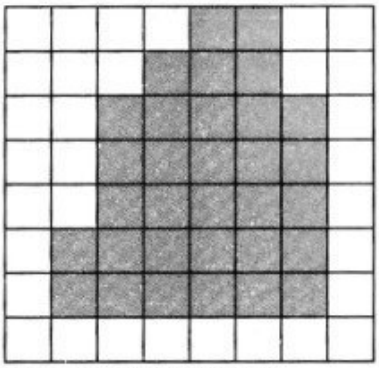


Figure 1.4 Cell by cell data structure records each cell value. The gray cells have the cell value of 1.

Multiband Imagery: It is the satellite imagery which consists the scene of the same area at different wavelength, of electromagnetic radiation.

Multiband imagery is typically stored in the following three formats–

- **BSQ:** Band Sequential method stores the values of an image band as one file.
- **BIL:** Band Interleaved by Line method stores, row by row, the values of all the bands in one file. Therefore the file consists of row 1, band 1; row 1, band 2. . . . . row 2, band 1; row 2, band 2. . . . and so on.
- **BIP:** Band Interleaved by Pixel method stores the values of all the bands by pixel in one file. The file is therefore comprised of pixel (1,1), band 1; pixel(1,1), band2. . . pixel (2,1), band1; pixel(2,1), band 2. . . . and so on.

II. Run length Encoding:

Raster models with many repetitive cell values can be more efficiently stored using Run-Length Encoding (RLE) method, records the cell values by row

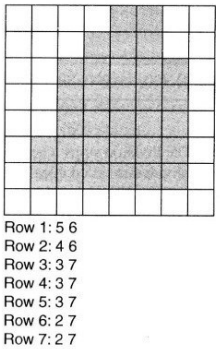


Figure 1.5

and by group. A group refers to adjacent cells with the same cell value. For each row, the starting cell and the end cell denotes the length of the group ("run") that falls within the polygon.

Many GIS packages use RLE in addition to the cell-by-cell encoding method for storing raster data. They include GRASS, IDRISI and Arc GIS.

III. Quad tree:

The Quad tree method divides a geographic area into square cells using the principle of recursive subdivision of a non-homogeneous square array of cells into four equal size quadrants.

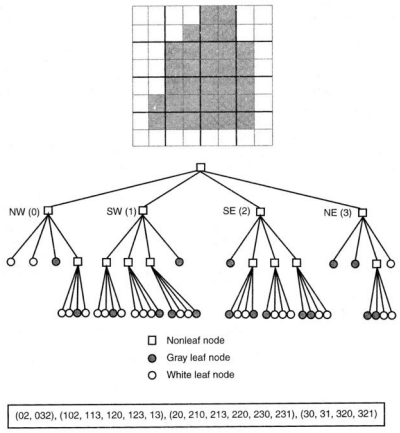


Figure 1.6 Quad tree

The quartering is continued to a suitable level until a square is found to be homogeneous. The underlying structure is a level tree where all non-leaf nodes have exactly four descendants.

IV. Header File:

To import raster data such as a DEM or a satellite image, a GIS package requires information about the raster, such as the data structure, area extent, cell size, number of bands, a value for no data etc. Often this information is contained in a header file. Functionally, a header file is similar to a projection file for a coordinate system.

1.6.2 Vector Data Structure

This data structure represents geography via point, line and polygons:

I. Whole Polygon:

Polygons described by listing coordinate of points in order as you walk around the outside boundary of the polygon.

- all data stored in one file
  - could also store—inefficiently—attribute data for polygon in same file.
- Coordinates/borders for adjacent polygons stored twice.
- no topological information about polygons
  - Which are adjacent and have common boundary?
  - How to relate different geographies?

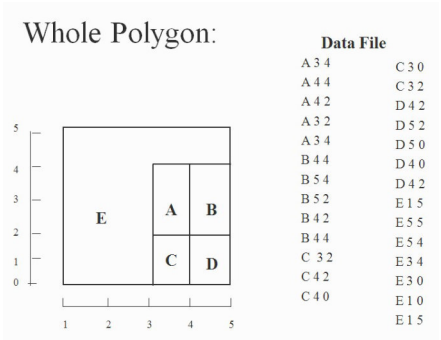


Figure 1.7: Vector Polygon (Whole)

In order to understand this more clearly, refer to the Figure given and you will receive an idea of how data is stored in vector data structure with the help of whole polygon part.

II. Points and Polygons: Polygons described by listing ID numbers of points outside the polygon, a second file lists all points and their coordinates.

- Solves the duplicate coordinate/double border problem
- Lines can be handled similar to polygons (list of IDs), but how to handle networks?
- Still no topological information.

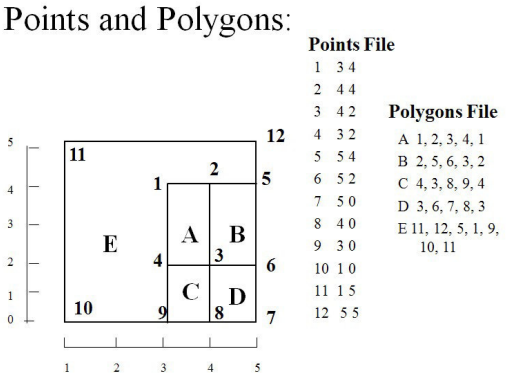


Figure 1.8: Points and polygon

In order to understand this more clearly, check the Figure given and you will receive an idea of how data is stored in vector data structure with the help of whole polygon part.

III. Node/Arc/Polygon Topology:

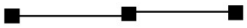
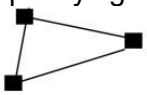
Node:

Start or end point of a line segment.

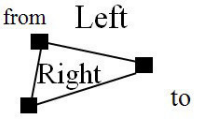
Arc: Same as a line.

Area/ polygon: A closed geometric figure.

Comprise 3 topological components which permit relationships between all spatial elements to be defined (note: does not imply inclusion of attribute data)

- **ARC-node topology:**
  - defines relations between points, by specifying which are connected to form arcs. 
  - defines relationships between ***arcs***, by specifying which arcs are connected to form routes and networks.
- **Polygon-Arc Topology:** 
  - defines polygons (areas) by specifying which arcs comprise their boundary. 
- **Left-Right Topology:**
  - defines relationships between polygons (and thus all areas) by
    - defining from-nodes and to-nodes, which permit
    - Left polygon and right polygon to be specified.

    (also left side and right side arc characteristics)





CHECK YOUR PROGRESS

Q. 6: What are the formats that represent data digitally in GIS?

.....

.....

Q.7: What is BSQ?

.....

.....

1.7 DATABASE MANAGEMENT TECHNOLOGY IN GIS

Database is a collection of logically related data designed to meet the information needs of one or more users. These data are often organized in database tables in terms of rows and column.

Generally, the column of a database or database table is called field and the row is called record. A typical database table is shown below.

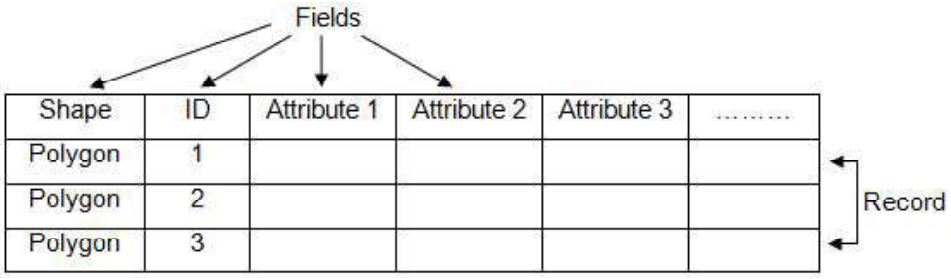


Figure 1.9: A database table with rows and columns

The computer program used to manage, query, link to another database is known as Database Management System (DBMS).

Maintaining a standard DBMS involves the following:

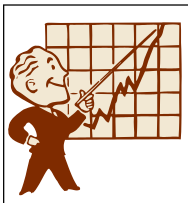
- Data definition and manipulations
- Adding records
- ***Querying***
- Modifying records
- Deleting records etc.
- Data security and integrity
- Data recovery and concurrency
- Data dictionary.
- Query optimization

Querying:
The process by which data are retrieved from a database in order to gain information from it.

Enabling the user to store, retrieve and update data in the database. In addition to this, the system should also allow the user to create relationship between the various tables. Examples of some DBMS include:

- Microsoft Access
- Oracle
- Sybase
- SQL server
- DB2

The presence of feature attribute and non spatial data tables means that a GIS requires a DBMS. However, the object based data model such as the geo-database has actually blurred the boundary between GIS and a DBMS as it stores both spatial and attributes data as tables in a single database. It allows flexibility in keeping various data types and structures for various applications.



CHECK YOUR PROGRESS

Q.8: What do you mean by database?

.....

.....

Q.9: Give some examples of Database Management System.

.....

.....

1.8 METHODS OF DATA INPUT AND DATA EDITING

Methods of Data Input and Data Editing are considered as one of the basic technical modules of the software packages for Geographic Information System. The choice of data input and editing method is governed largely by the application, the available budget, and the type and the complexity of data to be incorporated.

1.8.1 Georeferencing of Objects

Map projection: It is a device to represent an area of the earth surface by preserving the properties of the distance, direction, shape and area.

Digital data in GIS should have locational information in terms of geographic co-ordinate system and the mathematical model of the earth. In the later stage, the spatial data may be transformed into a ***map projection*** system with real world co-ordinates either in meter or feet that may be adopted as linear units of measurement in a 2-D plane, where we create maps.

The data acquired through scanning of map will not give such information until and unless georeferencing procedure is not adopted in the dataset. All the GIS packages provide this basic tool to give input of such information.

For the purpose of georeferencing, we have to adopt a co-ordinate system and the earth model such as WGS84 (World Geodetic Survey, 1984), which is also termed as reference ***ellipsoids***. There are many reference ellipsoids for various regions of the world. For India, the reference ellipsoid is the Everest 1975. The WGS84 reference ellipsoid is used by the GPS system for the purpose of mapping entire globe.

Ellipsoid: It is a mathematical model of the earth. Examples of ellipsoids are WGS84, GRS 1980, Everest 1975 and many more.

Ellipsoid: It is a mathematical model of the earth. Examples of ellipsoids are WGS84, GRS 1980, Everest 1975 and many more.

In the case of image acquired by scanner for a map, the intersection of the lines of latitude and longitude will provide locational information. A minimum of 4 such references are required to georeference a scanned map. The input of co-ordinates of these points is to be provided during the process of georeferencing. These values are stored and the location of any point is shown by the map.

These inputs are the first steps creating geographic data/ spatial data. A large number of map projection systems are available to project the data in geographic system to rectangular co-ordinate system with (x,y) values in desired units of measurement for each location. By performing this kind of transformation, map reading in terms of distance, azimuth, shape and area can be made. Further analysis of spatial data will also require the projected system of co-ordinates.

Based on a georeferenced map only other layers of vector data can be created, manipulated and edited. The scale of the map is dependent on the georeferenced data only, whether the data set is a raster layer or vector layer.

1.8.2 Data Input and Methods

The most expensive part of a GIS project is data acquisition. There is no single method of entering the spatial data into a GIS. Rather, there are several, mutually compatible methods that can be used singly or in combination. These are:

I. Manual digitizing: Manual digitizing has many advantages including:

- Low capital cost, e.g. digitizing tables is cheap;
- Low cost of labor;
- Flexibility and adaptability to different data types and sources;
- Easily taught in a short span of time - an easily mastered skill;
- Generally the quality of data is high;
- Digitizing devices are very reliable and most often offer a greater precision that the data warrants;
- Ability to easily register and update existing data.

II. Automatic scanning: A variety of scanning devices exist for the automatic capture of spatial data. While several different technical approaches exist in scanning technology, all have the advantage of being able to capture spatial features from a map at a rapid rate of speed.

III. Entry of coordinates using coordinate geometry: This involves entering, from survey data, the explicit measurement of features from some known monument.

IV. Conversion of existing digital data: A fourth technique that is becoming increasingly popular for data input is the conversion of existing digital data.

1.8.3 Data Editing

Data Editing or Spatial Data Editing in GIS refers to the process of adding, deleting and modifying features in digital layers. A major part of spatial data editing is removing digitizing errors. Newly digitized maps, no matter how carefully prepared, always have some errors. There are basically two types of digitizing errors:

- **Location Error:** It refers to the geometric inaccuracies of digitized features. We can examine location errors by referring to the data source used for digitizing.
- **Topological Error:** To correct topological errors, we must first learn about the topological relationships and then use a topology-based GIS package to help us make corrections. However, many GIS software allows us to identify and remove the errors automatically. This process includes the major actions shown in the table.

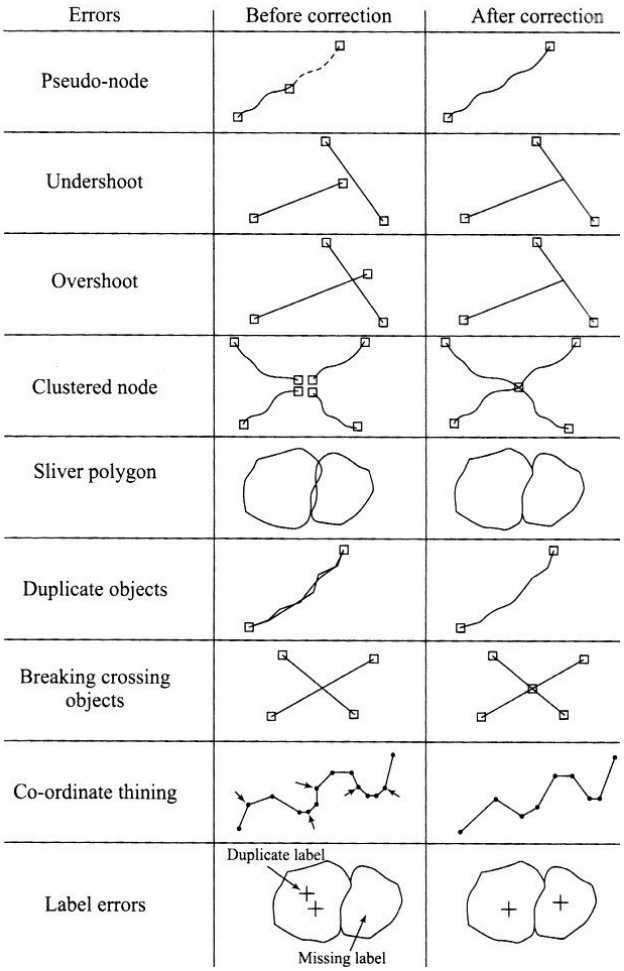


Figure 1.10



CHECK YOUR PROGRESS

Q.10. What is the referencing ellipsoid for India?

.....

.....

.....

.....

Q.11. What are the digitizing errors during data editing in GIS?

.....

.....

.....

.....



1.9 LET US SUM UP

- A data model is a set of constructs for describing and representing selected aspects of the real-world in a computer.
- In GIS we basically use two types of data i.e. Spatial Data and Attribute Data.
- Spatial Data describe the location(x,y) information for an Earth surface feature while, attribute data describe the characteristics of spatial features.
- A raster represents a continuous surface which is also called a grid, a raster map, a surface cover or an image in GIS.
- Vector data model uses the geometric objects of point, line and area to represent earth surface features.
- Connectivity, Area definition and Contiguity are mainly the three types of topological relationship.
- Raster data structure and vector data structure are the basic spatial data structure types in GIS.
- A database is collection of records stored in computer in a systematic way.



1.10 FURTHER READING

1) Basudeb Bhatta, *Remote Sensing and GIS*, published by Oxford University press, 2008.

2) Kang-Tsung Chang, *Introduction to Geographic Information System,* Published by McGraw-Hill, 2002

3) Paul A. Longley, Michael F. Goodchild, David J. Maguire, David W. Rhind, *Geographic Information Systems and Science*, Published by Wiley, 2010.



1.11 ANSWERS TO CHECK YOUR PROGRESS

Q. No. 1:

I. True

II. True

III. False

Q. No. 2:

I. Data Model

II. Complex

III. Spatial data

Q. No. 3: The various data sources for spatial data are:

- Conventional analog map sources
- Aerial remote sensing
- Satellite Remote Sensing
- Field data sources
- Existing digital map sources

Q. No. 4:

I. True

II. False

III. True

IV. False

Q. No. 5:

I. Pixel
II. Projected Coordinate System
III. Three
IV. Polygons

Q. No. 6: The two formats that represent data digitally in GIS are: Raster data structure and vector data structure.

Q. No. 7: Band Sequential (BSQ) method stores the values of an image band as one file.

Q. No. 8: Database is a collection of logically related data designed to meet the information needs of one or more users. In other words, a database is collection of records stored in computer in a systematic way.

Q. No. 9: Some examples of DBMS are: Microsoft Access, Oracle, Sybase, SQL server and DB2.

Q. No. 10: For India, the reference ellipsoid is the Everest 1975.

Q. No. 11: There are basically two types of digitizing errors in GIS namely location error and topological error.



1.12 MODEL QUESTIONS

Q.1: Define Data. Write about some characteristics of spatial data.

Q.2: What do you mean by spatial data source in GIS? Give a few examples of spatial data sources for GIS use.

Q.3: Define spatial data model. Differentiate between Raster Data model and Vector Data model.

Q.4: Explain the term topology. What are the topology rules in data model?

Q.5: Write some Data model names and their applications.

Q.6: Discuss Raster data structure and Vector data structure.

Q.7: Explain the importance of data input and data editing in GIS.

---×---

UNIT- 2: DATA ANALYSIS

UNIT STRUCTURE

2.1 Learning Objectives

2.2 Introduction

2.3 Introduction to Data Analysis

2.4 Database Query

2.5 Geospatial Measurements

2.6 Neighbourhood Operations

2.7 Overlay Operations

2.7.1 Types of Overlays in Raster

2.8 Surface Analysis

2.9 Network Analysis

2.10 Let us Sum up

2.11 Further Reading

2.12 Answers to Check Your Progress

2.13 Model Questions

2.1 LEARNING OBJECTIVES

After reading this unit, you will be able to :

- explain the GIS data analysis technique
- list out the database query techniques in GIS
- discuss the geospatial measurements
- illustrate the neighbourhood operations
- conceptualise the overlay analysis in GIS database
- describe the concept of surface and network analysis.

2.2 INTRODUCTION

In the previous unit, you have received an introduction to the concept data modelling, as well as an idea on the database management technology in GIS. This unit will now discuss the GIS data analysis technique and will provide an overview of how these data analyse techniques can be classified. It will also discuss the database query techniques in GIS. This unit will also provide you the geospatial measurements. Further it comprises the neighbourhood operations and the overlay analysis in GIS database. Apart from these, the unit will help you to familiarise with the concept of surface and network analysis in GIS.

2.3 INTRODUCTION TO DATA ANALYSIS

The heart of GIS is the analytical capabilities of the system. The spatial data analysis functions are the major factors that distinguish the GIS from other information systems. GIS is bundled with many geospatial analysis functions. The analysis functions use the geospatial and non-spatial attributes in the database to answer questions about the real world. Results of geographic analysis can be communicated with the help of maps, reports, charts etc.

The objective of analysis is to find the answer as useful information for decision making. An important use of the analysis is the possibility of predicting events on the basis of the datasets used for the purpose.

Geospatial analysis is undertaken to answer geographical questions about the real world, including the present situation of specific areas and features, the change in situation, the trends, the evaluation of capability, or the possibility of using overlay technique and/or modelling, and prediction of events. Therefore, geospatial analysis ranges from simple arithmetic and logical operation to complicated model analysis. This chapter highlights the basic analysis of geospatial data or spatial data.

The GIS has the capability of carrying out any number of analyses related to any discipline oriented problems. The range of analysis techniques is very large. Accordingly, this section focuses on providing an overview of the fundamental primitive functions that are most often utilised in geospatial analysis.

The GIS analysis techniques can be broadly classified in the following broad categories:

- Database query
- Geospatial measurements
- Overlay operations
- Neighbourhood operations / Proximity analysis
- Network analysis
- Surface analysis

The geospatial data analysis techniques can be classified from different points of view and may significantly differ.

CHECK YOUR PROGRESS



Q. 1: What are the different categories of GIS analysis techniques?

.....

.....

2.4 DATABASE QUERY

Selective display and retrieval of information from a database are among the fundamental function of GIS. Database query just asks to find already stored information. Database of both raster and vector data structure can be queried.

Vector Data Query:

There are three types of query that a GIS generally allows on vector data:

1. Query by attribute data
2. Query by spatial data
3. Hybrid data query

1. Query by Attribute Data:

Query by attribute is to retrieve the attribute data without altering the existing data according to specifications given by the operator. Map features are also simultaneously retrieved on the basis of attributes. For example, showing all the villages (polygon feature) having the population density greater than 500 per square kilometer.

Many GIS software include the Structured Query Language (SQL) to search a GIS database. The attribute data, stored in a table with a unique code, is linked to the spatial data. These attribute data from the database can be searched with some specifications. Simple to more complex queries can be made with the help of SQL.

The specifications include the following three items, given usually in SQL:
SELECT <attribute / field name(s)> FROM <table(s)> WHERE <condition statement(s)>

The conditional statement is represented by the following three types of operators:

Relational operator: The relational operators are >, <, =, >=, <= that measures greater than, less than, equal to, greater than and equal to, less than and equal to respectively.

Arithmetic operators: +, -, x, / are the arithmetic operators.

Logical (Boolean) operators: The logical operators are AND, OR, NOT, and XOR (exclusive OR).

2. Query by Spatial Data:

Spatial data query refers to the process of retrieving data from a map by working with map or spatial features. The GIS can carry out a number of such queries. There are three forms of spatial database query; selection by cursor, selection by graphic objects and selection by spatial relationship.

Selection by Cursor: It is the simplest application of spatial data query by clicking on the object or by defining an area. The attributes of displayed spatial features can be seen by identifying them with a graphical cursor.

Selection by Graphic Objects: This method of spatial data query uses a graphic object (such as a point, line, rectangle, circle, or polygon) to select map features, which fall inside or are intersected by the graphic object. For example, select all schools that fall inside a drawn circle.

Selection by Spatial Relationship: This query is useful to select map features based on their spatial relationship or topological relationship to other features. Map features to be selected may be in the same layer or in different layers. Spatial relationships used in query include the following:

- **Containment:** Selects features that fall completely within features used for selection. Feature is a geographic object of a data layer, and not just a drawn graphic object. For example, selecting all cities within the selected districts.
- **Adjacency:** Selects features that are adjacent to features used for selection. For example, selecting all adjacent districts of a selected state.
- **Intersect:** Selects features that intersect features used for selection. For example, selecting all districts on which a selected highway passes.

3. Hybrid Data Query:

A further complex query is one that uses both geometric and attributes search criteria together. Many GIS software force the separation of the two different types of query. However, some GIS software use databases to store both geometric and attribute data and allow true hybrid spatial queries. For example, selecting all roads which fall entirely inside a drawn circle and having a road width greater than 10m.

Raster Data Query

Raster data may also be queried in GIS by the following methods:

- **Query by pixel values of the raster**
- **Query by attribute**
- **Query by spatial data**

1. Query by Pixel Value:

A common method of query on a raster selecting pixels is based on the values of the pixels; for example, selecting all pixels which have values less than 100. Raster data query can also use Boolean operations like AND, OR, NOT. Relational operators and logical operators may also be combined to execute query by pixel values. Some areas will be shown as true or 1, which will fulfill the condition and the remaining areas as false or 0 in the resulting map.

2. Query by Attribute:

If the raster is a classified image or a thematic image or raster pixels are joined with any attribute table, it can be queried by the attributes. For example, in a land use / land cover map acquiring pixel count of forest class and water body class.

3. Query by Spatial Data:

A raster data may be queried by pointing the cursor on a cell to identify its location and cell value. Graphic objects (like point, line, circle, rectangle, or polygon) can also be used to make a query on raster data. For example, obtaining the pixel values which fall within the graphic object used for the query.



CHECK YOUR PROGRESS

Q. 2: Mention the types of query on vector data.

.....

.....

Q. 3: Mention the methods of raster data queries.

.....

.....

2.5 GEOSPATIAL MEASUREMENTS

Measurement is also a type of query; however, it does not select any object from the GIS database, instead it gives some statistical/ mathematical / geometrical result by measuring the map or resulting in a map. Three types of geospatial measurements are mainly very popular - measurement of

density, measurement of distance, measurement of length, area and perimeter.

1. Measurement of Density

By calculating density, point values are spread out over a surface. The magnitude at each sample location (line or point) is distributed throughout a landscape, and a density value is calculated for each cell in the output raster.

Density surfaces are good for showing where point or line features are concentrated. By calculating density, we can create a surface showing the predicted distribution of the signal throughout the landscape. This is often done by interpolating the 'known values' to predict the 'unknown values'. Some of the popular interpolation functions are Inverse Distance Weightage (IDW), Spline, Krigging and trend surface.

2. Measurement of Distance

Distance is one of the important elements in measuring spatial objects in GIS. Distance measurement refers to measuring distances between points, or between points and the distances from a given point or time. Distances can be used directly in data analysis. Several different concepts of distance are defined as follows:

Euclidean distance: Euclidean distance is defined as the distance measured along a straight line from one point to another in the Cartesian coordinate system.

Manhattan distance: Manhattan distance is defined as the rectilinear route measured along parallels to x and y axes.

Great circle distance or spherical distance: Great circle distance is defined as distance along the great circle of the spherical earth surface from a point (latitude and longitude) to another point.

Time distance: Time distance is defined as the time required to move from one point to another by using specific transportation means.

Measurement of length, area and perimeter: Measurement tools of geometric features are available in most of the standard GIS packages. Measurements are done in map units and stored in attribute table.



CHECK YOUR PROGRESS

Q.4: Name three types of Geospatial measurements.

.....

.....

Q.5: Mention the various concepts of distance measurements.

.....

.....

2.6 NEIGHBOURHOOD OPERATIONS

Proximity or neighborhood functions evaluate the characteristics of an area surrounding a feature's location. Neighborhood functions also include the calculation of a value representing a weighted average, maximum value, minimum value, measure of diversity etc.

Neighborhood operations can be of the following kinds:

1. Buffer: One of the most popular proximity analyses is based on 'buffering'. Buffer can be defined as an influenced area / zone of an object in GIS. There is a range of functions available in GIS that allow a spatial entity to influence its neighbours. The most common example is buffering, the creation of a zone of interest around a point, line, or area with a given distance as shown in Figure 2.1.

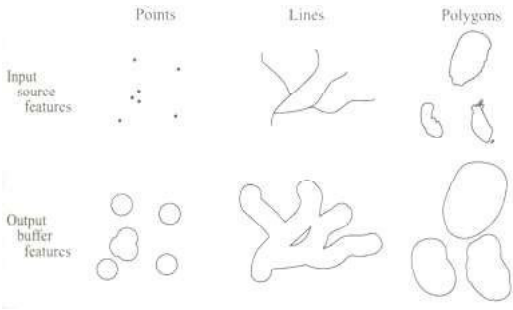


Figure 2.1 Buffering point, line and polygon features

The varying buffering zones for point, line and polygon features are possible. For example, road objects may be buffered by their varying width stored in a field in the linked attribute table of road. An object can be buffered by multiple values to generate multiple influenced zones. For example, influence of a river by flood may be analysed by three different buffered zones - highly affected, moderately affected, and less affected (Figure 2.2). Proximity analysis is not always based on distance but also on time, as it is known that distance can be converted into time. For example, proximity analysis based on access time or travel time provides the distribution of time zones indicating the time to reach a certain point.

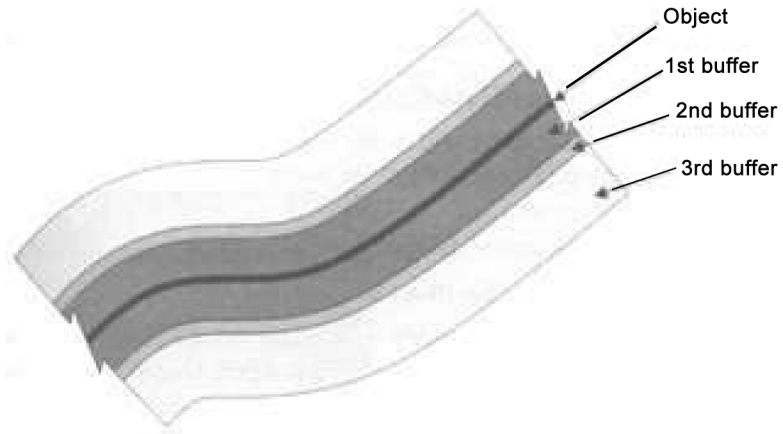


Figure 2.2 Multi-ring buffers around a line feature

2. Measurement of Neighbourhood: Measurement of neighbourhood or neighbourhood analysis is based on Voronoi diagram which is known as **thiessen polygon**. A Voronoi diagram is a special kind of decomposition of a geographic space determined by distances to a specified discrete set of objects, e.g., by a discrete set of points. For example, we are given a set of points; the segments (boundary) of thiessen polygons are equidistant to the nearest two points (Figure 2.3). Thiessen polygons are used in many neighbourhood analyses; for example, to calculate the rainfall of an area, based on a series of point measurements, or to find the object that is closest to a given query point. This tool defines the area of influence around the sampling locations.

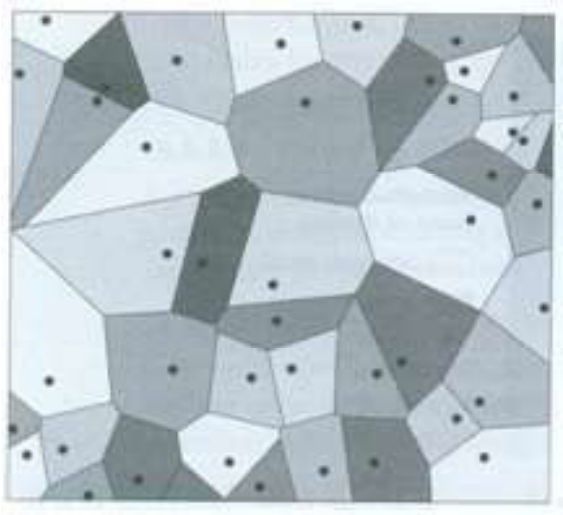


Figure 2.3 Thiessen polygons for set points

3. **Near Distance:** It computes the distance from each point in the input feature class to the nearest point or polyline in the near feature class or layer within the maximum search radius.

4. **Distance Calculation:** It computes the distances between point features in one feature class to all points in a second feature class that are within the specified Search Radius. The results are stored in a table.



CHECK YOUR PROGRESS

Q. 6: Mention the different kinds of neighbourhood operation.

.....

.....

.....

Q. 7: What is multi ring buffer?

.....

.....

.....

Q. 8: How Thiessen polygon is used for proximity analysis?

.....

.....

.....

2.7 OVERLAY OPERATIONS

The hallmark of GIS is overlay operations. The capability to overlay multiple data layers in a vertical fashion is the most required and common technique in geographic data processing.

Topological overlay is predominantly concerned with overlaying polygon data with polygon data, e.g., soils and forest cover. However, there are requirements for overlaying point, linear, and polygon data in selected combinations, e.g., point-in-polygon, line-on- polygon, and polygon-on-polygon are the most common. Vector and raster based software differs considerably in their approach to overlay the data layers.

Raster-based software is oriented towards arithmetic overlay operations, e.g., the addition, subtraction, division, and multiplication of data layers. The nature of the one- attribute map approach, typical of the raster data model, usually provides a more flexible and efficient overlay capability. The raster data model affords a strong numerical modeling (quantitative analysis) capability. Most sophisticated spatial modeling is undertaken within the raster domain.

In vector-based systems, topological overlay is achieved by the creation of a new topological theme from two or more existing themes. This requires the rebuilding of topological tables, e.g., arc, node, polygon, and therefore can be time-consuming and CPU-intensive. In this way, selected queries of the original layer can then be undertaken, e.g., soils and forest cover, to determine where specific situations occur, e.g., deciduous forest cover where drainage is poor.

1. Vector Overlay:

Overlay of vector data is slightly complicated because it must update the topological tables of spatial relationships between points, lines, and polygons. During the process of overlay, the attribute data associated with each feature type are merged. The resulting table contains both the attribute data. The process of overlay depends upon the modelling approach the user needs. Again, different GIS software utilises varying approaches for the display and reporting of overlay results. Some systems require that topological

overlay occur on only two data layers at a time, creating a third layer. One might need to carry out a series of overlay procedures to arrive at the conclusion, which depends upon some criteria.

Most GIS software makes use of a consistent logic for the overlay of multiple data layers. The rules of Boolean logic are used to operate on the attributes and spatial properties of geographic features. Boolean algebra uses the operators AND, OR, XOR, and NOT for overlay operations. Boolean logic represents all possible combinations of spatial interaction between different features. The implementation of Boolean operators is often transparent to the user. A **union overlay**, based on OR operator, combines the geographic features and attribute tables of both inputs into a single new output. An **intersect overlay**, based on AND operator, defines the area where both inputs overlap and retains a set of attribute fields for each. A **symmetric difference** *overlay*, based on XOR operator, defines an output area that includes the total area of both inputs except for the overlapping area. A **delete overlay**, based on NOT operator, defines an area of one input after excluding the area of another input. Using these operations, new spatial elements are created by the overlaying of maps as shown in Figure 2.4.

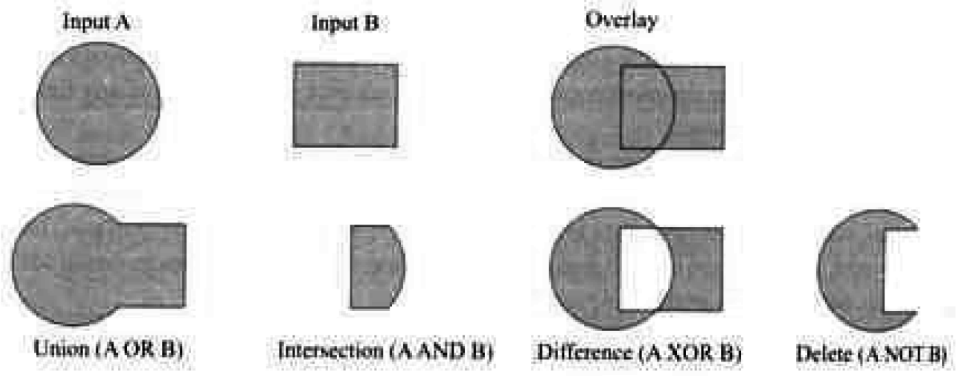


Figure 2.4 Overlay operations based on Boolean logic

There are three types of vector overlay - point-in-polygon, line-on-polygon, and polygon- on-polygon.

- **Point-in-Polygon Overlay:**

It overlays a point layer on a polygon layer as shown in Figure 2.5. Topology of point in polygon computes the 'contained in' relationship. The point features of the resulting point layer assume the polygon attributes they lie within. For example, combining wells (point feature class) and planning districts (polygon feature class) will help in queries such as how many wells are there in each district.

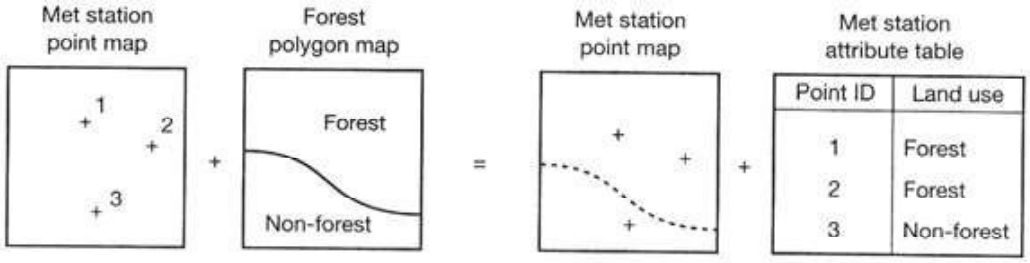


Figure 2.5 Point-in-Polygon Overlay

- **Line-on-Polygon Overlay:**

It overlays line layer onto a polygon layer as shown in Figure 2.6. The line features in the output layer assumes the attribute of the polygon they lie within. The lines are broken at each area object boundary.

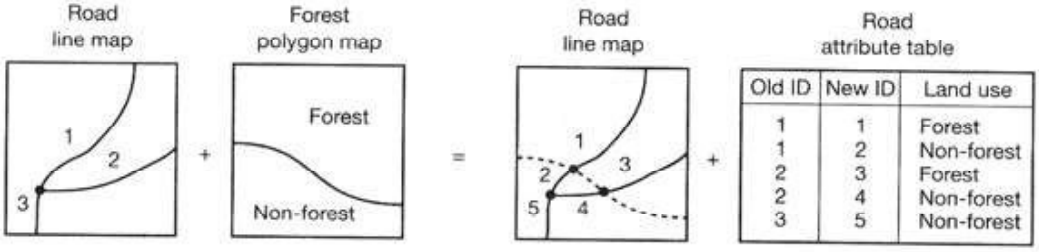


Figure 2.6 Line-in-Polygon Overlay

- **Polygon-on-Polygon Overlay:**

Polygon overlay is a spatial operation which overlays one polygon layer onto another to create a new polygon layer as shown in Figure 2.7. The spatial location of each set of polygons and their polygon attributes is joined to derive new data relationships. The output layer of such overlay operation is a polygon layer.

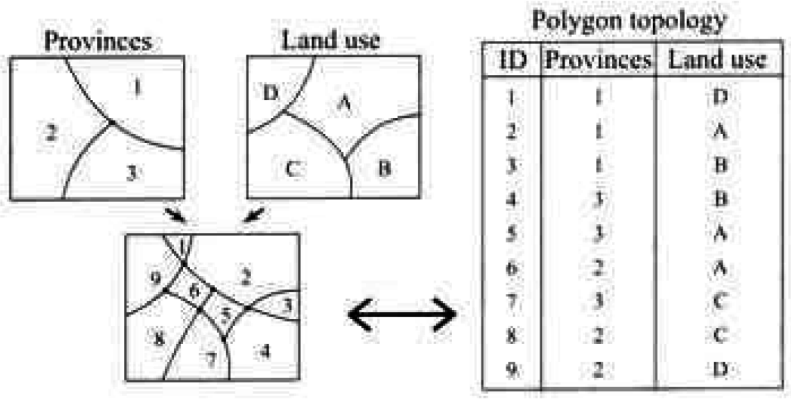


Figure 2.7 Polygon-in-Polygon Overlay

2. Raster Overlay:

Overlay of raster data with more than two layers is faster than overlay of vector data. It does not include any topological operation but it uses pixel-by-pixel operations. In raster data analysis, the overlay of datasets is accomplished through a process known as 'local operation on multiple raster' or 'map algebra', through a function that combines the values of each raster's matrix. The raster data model performs overlay functions more efficiently than the vector model. Raster cells have a one-to-one relationship between layers. Map algebra is used to identify relationship between layers, or to derive indices that describe phenomena.

This function may weigh some inputs more than others, for example, through use of an 'index model' that reflects the influence of various factors upon a geographic phenomenon.

In raster overlay, the pixel or grid cell values in each map are combined using arithmetic and Boolean operators to produce a new value in the composite raster map. The maps can be treated as arithmetical variables and we can perform complex algebraic functions. The method is often described as map algebra. This is particularly important for the modelling in which various maps are combined using various mathematical functions. Conditional operators are the basic mathematical functions that are also supported in such cases.

2.7.1 Types of Overlays in Raster

There is a wide range of operations and functions that are used to analyse raster maps with domain type values. In raster method, overlay analysis manipulates spatial data organised in several layers to create combined spatial features.

There are three types of overlay in raster:

- **Arithmetical overlays**
- **Logical Overlays**
- **Conditional Overlays**

- **Arithmetical Overlays:** This overlay includes operations like subtraction, addition, division and multiplications of each value in a layer by the value in corresponding location in the second layer. It is obvious that arithmetical overlays can only be used on value maps.

  For example, Add the pixel values of Map A and Map B and store the result in Map C i.e.

  Map C = Map A + Map B

- **Logical Overlays:** It works with logical operations AND, OR, NOT and XOR based on Boolean logic. Boolean logic operates on the attributes and spatial properties. The operations are used to see whether a particular condition is true or false. Simple Boolean logic is often portrayed visually in the form of Venn diagrams.

If set A is the set of items 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 and A NOT B will return items having attribute combinations covered by the shaded portions of the respective Venn diagrams (Figure 2.8).

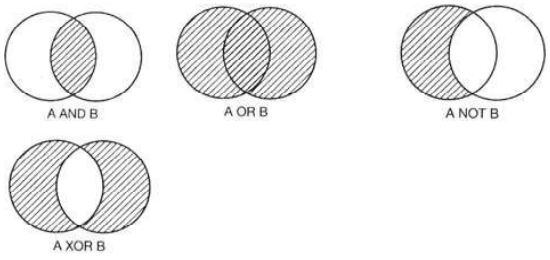


Figure 2.8 Venn diagram showing logical overlay

They should not be regarded as applicable to the attributes of the geographical elements alone, because they can be also related to the spatial elements.

For example, find all mapping units with clay loam texture in combination with a pH >7 where the area > 5 hectares.

Conditional Overlays: These overlay operations comprise selection of an area based on a certain set of preset conditions that are met. In practice we use the conditional operators mostly with the so-called conditional IFF function. The general syntax for the conditional IFF function is:

Outmap = IFF (Condition, Then statement, Else statement)

Conditional overlays are of two types:

1. Matrix Overlay: This overlay operations involve combining two map layers at a time. This is possible using a so-called 2-D table. A 2-D table is a matrix where each row represents a class or an ID in the first input map and each column represents a class or an ID in the second input map.

The Figure 2.9 shows the fields in the 2-D table using a class domain with two classes, **suitable and unsuitable**. The records and columns in the table are defined by two class domains (geology and landuse).

Input Maps

F	F	F
G	F	L
G	G	L

Landuse

A	A	A
S	A	A
S	S	S

Geology

(F=Forest, G=Grass, L=Lake)

(A=Alluvium, S=Shale)

2-D table

	Alluvial	Shale
Forest	Suitable	Unsuitable
Grass	Unsuitable	Suitable
Lake	Unsuitable	Unsuitable

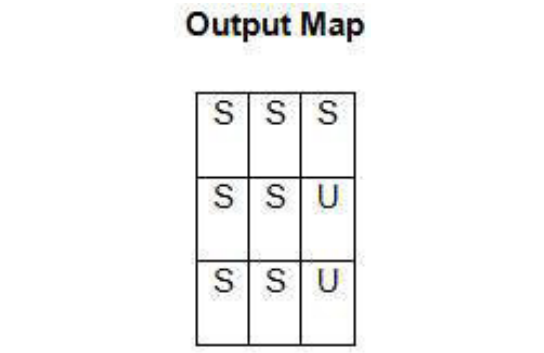


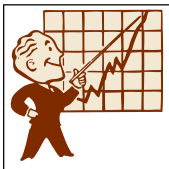
Figure 2.9 Matrix overlay

2. Index Overlay: By this operation, 15 – 20 layers can be overlaid in a single operation, at a time. Ranks and weightages can also be assigned to individual map as well as their attributes. The indexing approach may be best used for suitability or probability mapping. The individual layers can be weighted depending on their influence.

Likewise, classes on each map also influence the selection of suitable areas. Each class on a single map may be ranked on a predefined scale depending upon its contribution to the analysis. The higher the rank, the more suitable the area is. Classes which have negative influence may be ranked with a negative value. All areas with a negative value will be excluded from the analysis. For each cell, the index is applied using the following equation:

Cell value = {(Weight of map 1 x Rank of class) + (Weight of map 2 x Rank of class) + …..} / Total Weight

The output map contains values, which range from the lowest negative value assigned to the highest positive ranking.



CHECK YOUR PROGRESS

Q. 9: What are the operators used in Boolean algebra.

.....

.....

Q. 10: State the three types of Vector Overlay.

.....

.....

Q. 11: Mention the operators used in Arithmetic Overlay.

.....

.....

2.8 SURFACE ANALYSIS

The modelling of surface entities such as height, rainfall, humidity etc. poses interesting problems in GIS. Surface modelling is explored by looking in detail at Digital Terrain Models (DTM) or Digital Elevation Model (DEM).

The abbreviation DTM / DEM is used to describe a digital data set, which is used to model a topographic surface (a surface representing height data). To model a surface accurately, it would be necessary to store an infinite number of observations. Since this is impossible, a surface model approximates a continuous surface using a finite number of observations. Thus, an appropriate number of observations must be selected, along with their geographical location. The 'resolution' of a DTM is determined by the frequency of observations used. DTMs are created from a series of either regularly or irregularly spaced (x, y, z) data points (where x and y are the horizontal co-ordinates and z is the vertical or height co-ordinate). DTMs may be derived from a number of data sources. These include contour and spot height information found on topographic maps, stereoscopic aerial photography, satellite images and field surveys.

There are differences between the way in which DTMs are constructed in raster-based and vector based GIS.

The Raster Approach to Digital Terrain Modelling: In raster GIS a DTM is simply a grid of height values in which each cell contains a single value representative of the height of the terrain that is covered by that cell. How

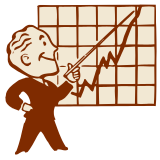
accurately the terrain can be modelled using this method depends on the complexity of the terrain surface and the spacing (resolution) of the grid. If the terrain is complex, for example in a mountainous area, a fine grid will be required to represent it with any degree of accuracy. If the terrain is relatively flat, fewer points will be needed to model it to the same degree of accuracy. Thus, a coarser resolution DTM may be used. A suitable method of interpolation is also required to generate DTM / DEM. Some of them are- Inverse Distance Weighted method, Spline and Krigging. Vector dataset of point or line may be used to create raster grids for the elevation.

From the elevation raster the two important components the slope and aspects can be computed. DEM / DTMs are used to find the areas of visibility from a given location. They are used to align the road/ railway routes to find the strategic defence locations and to estimate the volume of earth filling of cutting. 3-D visualisation is possible for DEM. In hydrological modelling DTMs / DEMs are prerequisite dataset.

The Vector Approach to Digital Terrain Modelling: In its simplest form, a vector DTM mimics the raster version by using a regularly spaced set of spot heights to represent the terrain surface. A more advanced, more complex and more common form of vector DTM is the triangulated irregular network (TIN) (Figure 2.10). In vector GIS a TIN is used to create a DTM from either regular or irregular height data. The TIN method joins the height observations together with straight lines to create a mosaic of irregular triangles. In the TIN model of a surface, the vertices of the triangles produced represent terrain features such as peaks, depressions and passes, and the edges represent ridges and valleys. The surfaces of individual triangles provide area, gradient (slope) and orientation (aspect). These values can be stored as TIN attributes or can be quickly calculated when the TIN is used in further analysis. TIN data can be converted into grid data for 2-D viewing. TN data allows 3-D viewing also.

The main advantage of the TIN model is efficiency of data storage, since only a minimum number of significant points need to be stored to reproduce a surface. In the previous section, we considered the difficulties of choosing an appropriate resolution for a raster DTM in an area of mixed terrain. Since

TINs are created from irregularly spaced points, more points are used to represent the mountaineous areas and fewer to represent the flatter areas. To achieve this efficiency in storage, TIN models use only 'surface significant' points to reproduce a terrain surface. These points are selected by the TIN model from the input data on the basis of their spatial relationship with their neighbours. Those points that cannot be closely interpolated from their neighbours are considered to be 'surface significant' and are used as TIN vertices. Two methods are generally used to achieve the linking of TIN vertices.



CHECK YOUR PROGRESS

Q. 12: What is known as the TIN Model?

.....

.....

Q. 13: What does 'Surface Significant points' refer to?

.....

.....

2.9 NETWORK ANALYSIS

A network is a set of interconnected linear features through which materials, goods and people are transported or along which communication of information is achieved. The vector network model is made up of the same arc (line segments) and node elements as any other vector data model but with the addition of special attributes as illustrated in the Figure 2.12.

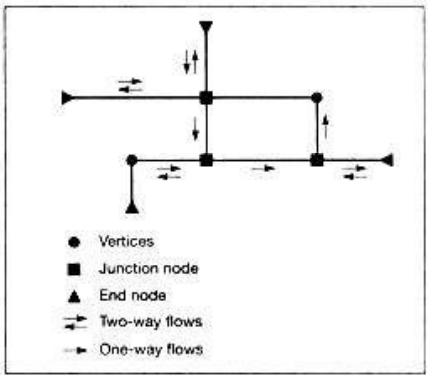


Figure 2.12 Network Model

In the network model, the arcs become network links representing the roads, railways and air routes of transport networks; the power lines, cables and pipelines of the utilities networks; or the rivers and streams of hydrological systems. The nodes in turn become network nodes, stops and centres. Network nodes are simply the endpoints of network links and as such represent junctions in transport networks, confluences in stream networks, and switches and valves in utility networks. Stops are locations on the network that may be visited during a journey. They may be stops on a bus route, pick-up and drop-off points on a delivery system, or sediment sources in a stream network. They are points where goods, people or resources are transferred to and from some form of transport system. Centres are discrete locations on a network at which there exists a resource supply or some form of attraction. Examples include shopping centres, airports, schools and hospitals. At a larger scale centres may be a whole city if the transport, resource or information networks for an entire country are being considered. Turns represent the transition from network link to network link at a network node. Therefore turns represent the relationships between network links and greatly affect movement through the network system. For example, turns across oncoming traffic on a road network take longer than turns down slipways, whereas turns that go against the flow of traffic on one-way streets are prohibited altogether.

All the data regarding the characteristics of network links, nodes, stops, centres and turns are stored as attribute information in the vector model database. Two key characteristics of network features are **impedance** and **supply or demand**.

Impedance sometimes referred to as constraints is the cost associated with traversing a network link, stopping, turning or visiting a centre. For example, link impedance may be the 'time' it takes to travel from one node to another along a network link. If we use the example of a delivery van travelling along a city street then the impedance value represents time, fuel used and the driver's pay. Impedance values for turn and stop are, therefore,

very important in determining the outcome of route finding, allocation and spatial interaction operations.

Supply and demand are equally important concepts in network analysis. Supply is the quantity of a resource available at a centre that is available to satisfy the demand associated with the links of a network. If we take the example of a hospital, then supply would be represented by the number of beds available. Demand, on the other hand, is represented by the number of people requiring treatment living within the hospital's catchment area. Correct topology and connectivity are extremely important for network analysis. Digital networks should be good topological representations of the real-world network they mimic. Correct geographical representation in network analysis is not so important, so long as the key attributes of impedance and distance are preserved. Finding the shortest path, least cost route, allocated route in between the nodes is the essence of network analysis to model geographic reality.



CHECK YOUR PROGRESS

Q. 14: Fill in the blanks:

a) Network analysis is essentially adaptation of the _____ data model.

b) In the network model the nodes become network nodes, _____ and _____.

c) Centres are _____ locations on a network at which there exists a resource supply or some form of attraction.

d) _____ and _____ are equally important concepts in network analysis.



2.10 LET US SUM UP

- The objective of analysis is to transform data into useful information to satisfy the requirements.
- The GIS analysis techniques can be classified as Database Query, Geospatial Measurements, Overlay Operations, Neighbourhood Operations, Network Analysis, and Surface Analysis.

- The vector database queries are Query by attribute data, Spatial data query and Hybrid data query.
- The conditional statement is represented by the relational operators, arithmetic operators and logical operators.
- The three forms of spatial database query are selection by cursor, selection by graphic objects and selection by spatial relationship.
- Raster data query method is classified into query by pixel values of the raster, query by attribute and query by spatial data.
- Three types of geospatial measurements are measurements of density, measurement of distance and measurement of length, area and perimeter.
- Different kinds of neighbourhood operations are Buffer, Measurement of Neighbourhood, Near Distance and Distance calculation
- The three types of Vector overlay are point-in-polygon, line-on-polygon, and polygon- on-polygon.
- Digital Terrain Models (DTM) is used to describe a digital data set, which is used to model a topographic surface.
- A common form of vector DTM is the triangulated irregular network (TIN).
- Points that cannot be closely interpolated from their neighbours are considered to be 'surface significant' and are used as TIN vertices.
- Network models are essentially adaptations of the vector data model.



2.11 FURTHER READING

1) Bernhardsen, *Geographic Information Systems: An Introduction,* (3rd Ed), Published by Wiley India.
2) Kang-Tsung Chang, *Introduction to Geographic Information Systems;* Published by Mc-Graw Hill.

2.12 ANSWERS TO CHECK YOUR PROGRESS

Q. No. 1: The GIS analysis techniques can be broadly classified in the following broad categories:

- Database query
- Geospatial measurements
- Overlay operations
- Neighbourhood operations
- Network analysis
- Surface analysis

Q. No.2: There are three types of vector data queries. They are query by attribute data, spatial data query and hybrid data query.

Q.No.3: Raster data query methods in GIS are: query by pixel values, attribute and spatial data.

Q.No.4: Measurement of density, distance and length, area and perimeter.

Q.No.5: Euclidean distance, manhattan distance, great circle distance or spherical distance and time distance.

Q.No.6: Buffer, measurement of neighbourhood, near distance and distance calculation

Q.No.7: When an object is buffered by multiple values to generate multiple influenced zones it is known as multi ring buffer.

Q.No.8: Thiessen polygons are used in many neighbourhood analyses; as a tool define the area of influence around a point location.

Q.No.9: Boolean algebra uses the operators AND, OR, XOR, and NOT for overlay operations.

Q.No.10: There are three types of vector overlay - point-in-polygon, line-on-polygon, and polygon- on-polygon.

Q.No.11: Operations used in arithmetic overlay like subtraction, addition, division and multiplication.

Q.No.12: In vector GIS a triangulated irregular network (TIN) is used to create a DTM from either regular or irregular height data.

Q.No.13: Points that cannot be closely interpolated from their neighbours are considered to be 'surface significant' and are used as TIN vertices.

Q. No. 14: a) vector b) stops, centres c) discrete d) supply, demand.



2.13 MODEL QUESTIONS

Q.1: Discuss the vector overlay analysis in GIS.

Q.2: Explain the two methods used to create the TIN model by linking the vertices.

Q.3: Define a network. Mention the various components and their functions in a network model.

Q.4: Discuss the two types of Conditional Overlay in Raster data structure.

Q.5: Discuss the vector data query in GIS.

Q.6: Discuss the methods that are used to achieve the linking of TIN vertices.

---×---

UNIT- 3: GEOPROCESSING FUNCTIONS AND TOOLS

UNIT STRUCTURE

3.1 Learning Objectives

3.2 Introduction

3.3 Vector Overlay

3.4 Raster Overlay

3.5 Spatial Buffering

3.6 Union and Intersection

3.7 Spatial Autocorrelation

3.8 Weighted Regression

3.9 Let us Sum up

3.10 Further Reading

3.11 Answers to Check Your Progress

3.12 Model Questions

3.1 LEARNING OBJECTIVES

After reading this unit, you will be able to:

- perform overlay operations of vector and raster data
- create buffer zones
- perform union and intersection of geographic data
- grasp the concepts of spatial autocorrelation and weighted regression.

3.2 INTRODUCTION

In the previous unit we have learnt about the GIS data analysis. This unit focuses on different methods of processing the geospatial data. In this unit you are going to learn the overlaying of different data formats- vector and

raster as well as their outcomes. You will also learn about buffer zones. This unit also deals with the concepts of spatial autocorrelation and weighted regression.

3.3 VECTOR OVERLAY

We have already discussed about the vector data in unit 1. One of the most basic questions asked with regard to GIS is “what’s on top of what?” For example:

- What land use is on top of what soil type?
- What roads are within what counties?

Before the advent of GIS, in order to answer such questions, cartographers would create maps on clear plastic sheets and overlay these sheets on a light table in order to create a new map of the overlaid data. Because overlay yields such valuable information, it was paramount to the development of GIS. An overlay operation is much more than a simple merging of different layers; all the attributes of the features taking part in the overlay are carried through. For example, watershed boundaries and a vegetation layer can be overlaid to calculate the amount of each vegetation type in each watershed.

A vector overlay involves combining point, line, or polygon geometry and their associated attributes. All overlay operations create new geometry and a new output geospatial data set. However, with certain overlay operations, very large attribute tables may result if the overly operations combine many layers, and each layer has a very large attribute table. Additionally, it might be possible that the combined attribute tables would cause duplicate attribute fields to exist. In these cases, one should consider reducing the number of transferred attributes to the minimum required, and renaming duplicate fields so that there is no ambiguity.

Overlay of vector data is slightly complicated because it must update the topological tables of spatial relationships between points, lines and polygons. During the overlay, the attribute data associated with each feature type are merged. The resulting table contains both the attribute data. The most common methods of vector overlay are Union Overlay, Intersection Overlay

and Difference Overlay. There are three types of vector overlay: point-in-polygon overlay, line-on-polygon overlay and polygon-on-polygon overlay.

Point-in-Polygon Overlay

Points are overlaid on polygon map and hence point topology in the new data layer is a new attribute of polygon for each point (Fig 3.1).

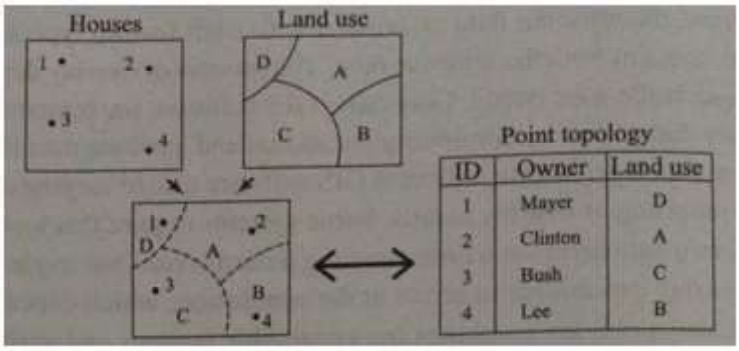


Fig 3.1: Point-in-Polygon overlay

Line-on-Polygon Overlay:

Lines are overlaid polygon map and line topology in the new data layer is the attribute of old line ID and containing area ID (Fig 3.2).

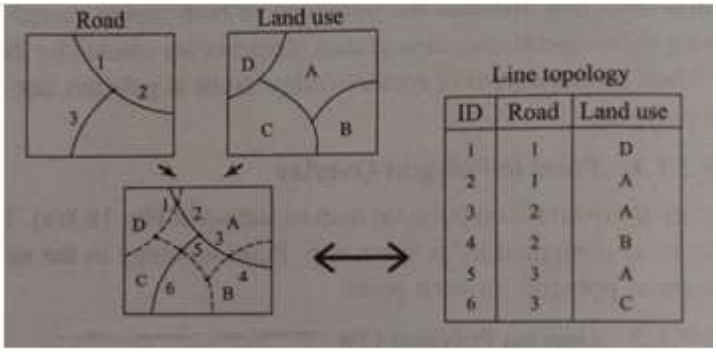


Fig 3.2: Line-on-Polygon overlay

Polygon-on-Polygon Overlay:

Two layers of area objectives are overlaid resulting in new polygons and intersections. The number of new polygons is larger than that of the original polygons. Polygon topology in the new data layer is a list of original polygon IDs (Fig 3.3).

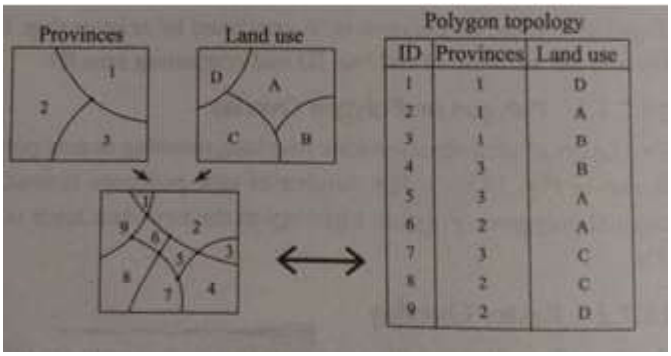


Fig 3.3: Polygon-on-Polygon overlay



CHECK YOUR PROGRESS

Q.1: State True or False:

I. During vector overlay, the attribute data associated with each feature type are deleted.

II. The combined attribute tables in vector overlay would cause duplicate attribute fields to exist.

Q.2: Fill in the blanks:

I. Vector overlay operations create _____ and _____ data set.

II. The resulting table in a vector overlay contains _____

3.4 RASTER OVERLAY

Overlay of raster data with more than two layers is easier than overlay of vector data because it does not include any topological operations but only pixel by pixel operations. In raster data analysis, the overlay of datasets is accomplished through a process known as 'local operation on multiple rasters' or 'map algebra', through a process that combines the values of each raster's matrix for mathematical calculations. This function may weigh some inputs more than others through use of an 'index model' that reflects the influence of various factors upon a geographic phenomenon.

In raster overlay, the pixel or grid cell values in each map are combined using arithmetic and Boolean operators to produce a new value in the composite raster map. The maps are treated as arithmetical variables and

one can perform complex algebraic functions. The raster GIS provides the ability to perform operations on map layers mathematically. This is particularly important for modelling in which various maps combined using mathematical functions as in Fig 3.4.

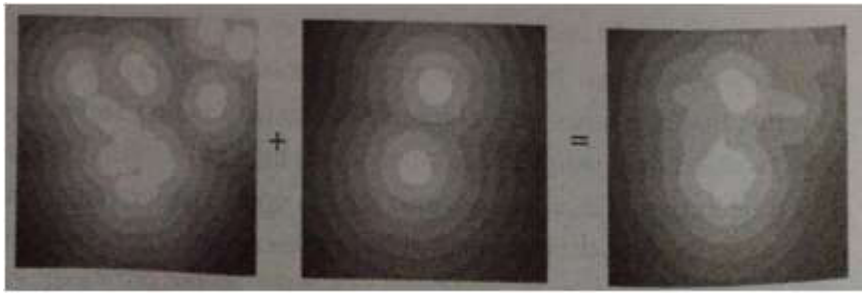


Fig 3.4: Raster overlay



CHECK YOUR PROGRESS

Q.3: Fill in the blanks:

I. In raster overlay the overlay of datasets is accomplished through a process known as _____.

II. In raster overlay the _____ values in each map are combined.

3.5 SPATIAL BUFFERING

Buffer analysis is an important pre- analysis technique used to constrain space around individual land features. It combines spatial data query techniques and cartographic modeling. Buffering usually creates two areas: one area that is within a specified distance to selected real world features and the other area that is beyond. The area that is within the specified distance is called the buffer zone.

A buffer zone is any area that serves the purpose of keeping real world features distant from one another. Buffer zones are often set up to protect the environment, protect residential and commercial zones from industrial accidents or natural disasters, or to prevent violence. Common types of buffer zones may be greenbelts between residential and commercial areas,

border zones between countries, noise protection zones around airports, or pollution protection zones along rivers.

A buffer is, thus, a reclassification based on distance: classification of within/ without a given proximity. Buffering involves measuring distance outward in directions from an object. Buffering can be done on all three types of vector data: point, line, area as in Fig 3.5. The resulting buffer is a polygon file.

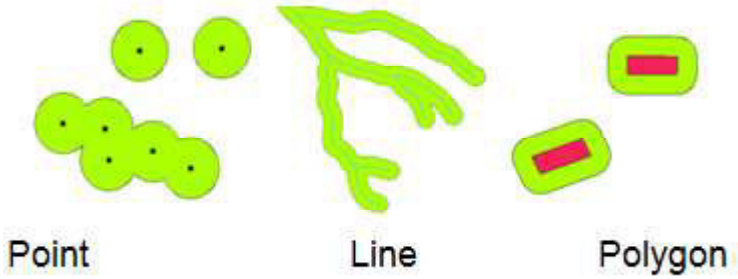


Fig 3.5: Buffer around Point, Line and Polygon

Multiple Buffer Zones:

A feature can also have more than one buffer zone. A nuclear power plant may be buffered with distances of 10, 15, 25 and 30 km, thus forming multiple rings around the plant as part of an evacuation plan (Fig 3.6)

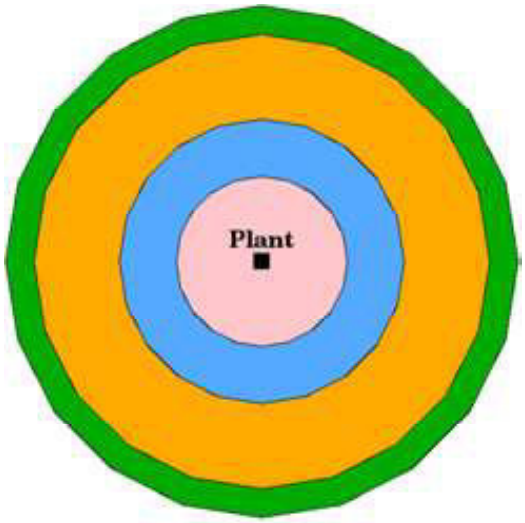
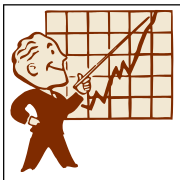


Fig 3.6: Buffering a point feature with distances of 10, 15, 25 and 30 km.



CHECK YOUR PROGRESS

Q.4: Define the term buffer.

.....

.....

Q.5: What is the purpose of buffer zones?

.....

.....

Q.6: Explain multiple ring buffer.

.....

.....

3.6 UNION AND INTERSECTION

Union:

In GIS, a Union is an analytical process in which the features from two or more map layers are combined into a single, composite layer. Union includes the data from all the included layers, meaning that overlapping and non-overlapping areas are included in a new polygon. Union is an overlay feature and analysis tool found in the overlay toolbox with several other useful tools like Identity, Intersect, etc. Union is a tool that can only be used with polygon features. It takes the two polygon features and combines both the spatial polygon and the associated attribute data into a single, composite layer. This allows for analysis of two datasets information together in one layer.

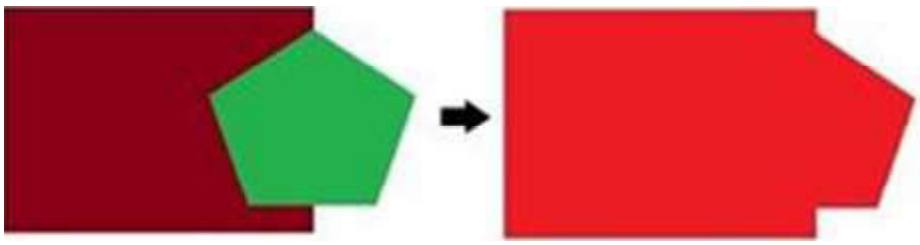


Fig 3.7: In a union, the data from both layers is used to create a new layer

In simple terms, Union in GIS terminology means the same as it does in any other terminology; 'to put together'. In map making, two different sets of information or layers can be put together into a single set using a union operation. This can be done with most mapping software systems in an

internal algorithm that takes the two layer inputs and outputs a composite of the two. An example of when this can be used is in analysing the habitat of many different animals. For example, if we have an area that is inhabited by elk, deer, and antelope, we can take the three sets of habitat information, use a union operation on them, and have an output of total animal habitat for that area. Hence, Union is particularly valuable in map algebra in allowing the cartographer to do suitability analysis.

Intersection:

In GIS, Intersection is an analytical operation that can be used to select any part of a feature that intersects with one or more other features. The areas of the map where all the input features intersect will create a feature as the intersect output. The Intersect is a spatial relationship operation in which a query method is used. This query method selects features based on spatial relationships to other features. Features for selection can come from the same or different layers. Besides Intersect, other operations that use query to define spatial relationships are Containment and Proximity.

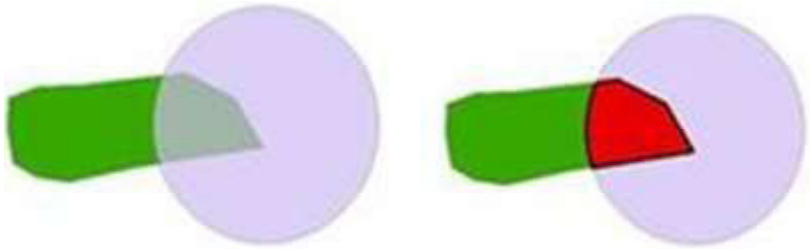
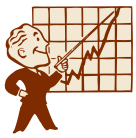


Fig 3.8: When two polygons are intersected, a third polygon is produced that only covers the area where the two original polygons overlap (The red portion).

Input features must be simple features that can include any combination of point, line, and polygon features. The inputs cannot be complex features like annotation features, dimension features, or network features. If the features used have different geometry types, such as a point/line or line/ polygon relationship, then the output feature will be the same as the input with the dimension geometry. For example the output of a point/line intersect operation will come out as a point shape file. Examples of Intersection include

selecting land parcels that intersect a proposed road or finding urban areas that intersect an active fault line. The Intersect operation can be very useful with Geocoding. Also another example of the intersect tool would be if the composite of several distribution areas of wildlife species is needed to determine a buffer zone for a proposed highway. The intersect tool could produce an output polygon with all the common areas of the species distribution polygons, subsequently excluding all areas of the distribution polygons that do not share common areas with other polygons.



CHECK YOUR PROGRESS

Q.7: What is the difference between union and intersection operation in GIS?

.....

.....

3.7 SPATIAL AUTOCORRELATION

Spatial autocorrelation in GIS helps understand the degree to which one object is similar to other nearby objects. Moran's I (Index) is the technique used to measures spatial autocorrelation. It measures how much close objects are in comparison with other close objects. Moran's I can be classified as positive, negative and no spatial auto-correlation.

- Positive spatial autocorrelation is when similar values cluster together in a map.
- Negative spatial autocorrelation is when dissimilar values cluster together in a map.

Spatial autocorrelation, thus, indicates if there is clustering or dispersion in a map. While a positive Moran's I hints at data is clustered, a negative Moran's I implies data is dispersed. One of the main reasons why spatial autocorrelation is important is because statistics relies on observations being independent from one another. If autocorrelation exists in a map, then this violates the fact that observations are independent from one another.

Positive Spatial Autocorrelation:

Positive spatial auto-correlation occurs when Moran's Index is close to +1. This means values cluster together. For example, elevation datasets have similar elevation values close to each other.

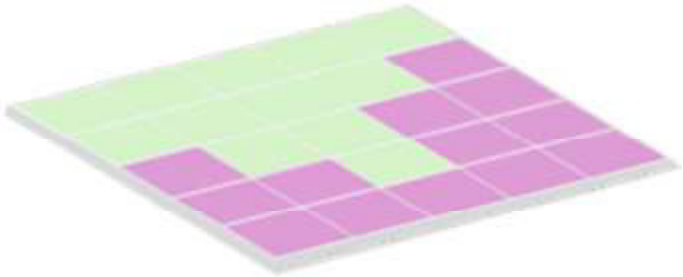


Fig 3.9: Clustered Image Spatial Autocorrelation

There is clustering in the land cover image above. This clustered pattern generates a Moran's I of 0.60. The z-score of 4.95 indicates there is a less than 1% likelihood that this clustered pattern could be the result of random choice.

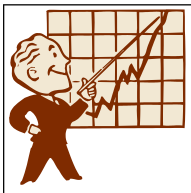
Negative Spatial Autocorrelation:

Negative spatial auto correlation occurs when Moran's Index is near -1. A checkerboard is an example where Moran's Index is -1 because dissimilar values are next to each other. A value of 0 for Moran's I typically indicates no autocorrelation.



Fig 3.10: Checkerboard Pattern: Spatial Autocorrelation

The checkerboard pattern generates a Moran's index of -1.00. This checkerboard pattern has a less than 1% likelihood that it is the result of random choice.



CHECK YOUR PROGRESS

Q.8: What do you mean by spatial autocorrelation?

.....

.....

Q.9: What is Moran's Index?

.....

.....

3.8 WEIGHTED REGRESSION

Geographically Weighted Regression (GWR) is a local spatial statistical technique used to analyse spatial non stationarity when the input variable differs from location to location. It provides a local model to predict an independent variable or process (e.g., plant growth, yield) by fitting a regression equation to the available datasets of dependent variables. It enables identification of the yield stability of a region as well as the association of the independent environmental factors to the yield. GWR is, thus, prevalent in the comprehension of the process related with the interaction amongst variables. In geographically weighted regression the observations taken into consideration for the formation of the model are weighted with regard to their location. The direct effect is that multiple linear regression involves analysis in the global level whereas geographically weighted regression involves analysis at local level.

Geographically weighted regression analysis responds to the vital question of spatial differentiation of parameters. Indeed the cartographic representation of results contributes to the comprehension of these differentiations as well as in their interpretation.

GWR is a useful and practical tool for evaluating the spatial heterogeneity of a dependent variable. Instead of calibrating a single equation, GWR generates a separate regression equation for each observation and thus allows parameter values to vary continuously in geographical space. Each equation is calibrated using a different weighting of the observations contained in the dataset. Because the regression equation is calibrated

independently for each observation, a separate parameter prediction (*z* value) and R^2 value are calculated for each observation. The complexity of a GWR model depends not only on the number of variables in the model but also on the bandwidth.



CHECK YOUR PROGRESS

Q.10: What do you mean by GWR?

.....

.....



3.9 LET US SUM UP

- A vector overlay involves combining point, line, or polygon geometry and their associated attributes.
- All vector overlay operations create new geometry and a new output geospatial data set.
- In vector overlay, the attribute data associated with each feature type are merged.
- The most common methods of Vector Overlay are Union Overlay, Intersection Overlay and Difference Overlay.
- There are three types of Vector Overlay: point-in-polygon Overlay, Line-on-polygon Overlay and Polygon-on-polygon Overlay.
- In raster data analysis, the overlay of datasets is accomplished through a process known as 'local operation on multiple rasters' or 'map algebra'.
- In raster overlay the pixel or grid cell values in each map are combined using arithmetic and Boolean operators
- A buffer is a reclassification based on distance: classification of within/without a given proximity.
- Buffering can be done on all three types of vector data: point, line, area. The resulting buffer is a polygon file.

- A Union is an analytical process in which the features from two or more map layers are combined into a single, composite layer.
- Intersection is an analytical operation that can be used to select any part of a feature that intersects with one or more other features
- Spatial autocorrelation in GIS helps understand the degree to which one object is similar to other nearby objects.
- Moran's I (Index) is the technique used to measures spatial autocorrelation.
- Moran's I can be classified as positive, negative and no spatial auto-correlation.
- Geographically Weighted Regression (GWR) is a local spatial statistical technique used to analyse spatial non stationarity when the input variable differs from location to location.



3.10 FURTHER READING

1) Basudeb Bhatta, *Remote Sensing and GIS*, published by Oxford University press, 2008.
2) Kang-Tsung Chang, *Introduction to Geographic Information System,* Published by McGraw-Hill, 2002.
3) Paul A. Longley, Michael F. Goodchild, David J. Maguire, David W. Rhind, *Geographic Information Systems and Science*, Published by Wiley, 2010.



3.11 ANSWERS TO CHECK YOUR PROGRESS

Q. No. 1:

I. False

II. True

Q. No. 2:

I. New geometry, a new output geospatial

II. Both the attribute data

Q. No. 3:

I. Map algebra

II. Pixel

Q. No. 4:

A buffer is a reclassification based on distance- classification of within/ without a given proximity. Buffering involves measuring distance outward in directions from an object.

Q. No. 5:

Buffer zones are set up to protect the environment, protect residential and commercial zones from industrial accidents or natural disasters, or to prevent violence. Common types of buffer zones may be greenbelts between residential and commercial areas, border zones between countries, noise protection zones around airports, or pollution protection zones along rivers.

Q. No. 6:

A particular feature can have more than one buffer zone. For example, a nuclear power plant may be buffered with distances of 10, 15, 25 and 30 km, thus forming multiple rings around the plant.

Q. No. 7:

A Union is an analytical process in which the features from two or more map layers are combined into a single, composite layer.

Whereas Intersection is an analytical operation that can be used to select any part of a feature that intersects with one or more other features.

Q. No. 8:

Spatial autocorrelation in GIS helps understand the degree to which one object is similar to other nearby objects. Spatial autocorrelation measures how much close objects are in comparison with other close objects. Spatial autocorrelation, thus, indicates if there is clustering or dispersion in a map.

Q. No. 9:

Moran's I (Index) is the technique used to measures spatial autocorrelation. Moran's I can be classified as positive, negative and no spatial auto-correlation.

- Positive spatial autocorrelation is when similar values cluster together in a map.
- Negative spatial autocorrelation is when dissimilar values cluster together in a map.

Q. No. 10:

Geographically Weighted Regression (GWR) is a local spatial statistical technique used to analyse spatial non stationarity when the input variable differs from location to location.



3.12 MODEL QUESTIONS

Q.1: Explain the term Vector Overlay.

Q.2: What are the different types of vector overlay?

Q.3: Explain raster overlay.

Q.4: What is a 'map algebra'?

Q.5: Explain the application of spatial buffering in GIS.

Q.6: Explain the concept of spatial autocorrelation.

Q.7: Explain the application of weighted regression in GIS analysis.

---×---

UNIT- 4: ANALYTICAL MODELLING AND OUTPUT

UNIT STRUCTURE

4.1 Learning Objectives

4.2 Introduction

4.3 Analytical Modelling Basics

4.4 Process Models

4.4.1 Physical and Environmental Process Modelling

4.4.2 Decision Making Process Modelling

4.4.3 Human Process Modelling

4.5 Output of GIS

4.6 Spatial Decision Support Systems

4.7 Let us Sum up

4.8 Further Reading

4.9 Answers to Check Your Progress

4.10 Model Questions

4.1 LEARNING OBJECTIVES

After reading this unit, you will be able to:

- grasp the concept of analytical modelling in GIS
- list out the types of process models
- conceptualise the various nature of process models
- explain the map generation as GIS output
- define the concept of Spatial decision support systems.

4.2 INTRODUCTION

In the previous unit, you have learnt about the GIS data analysis techniques. In this unit, we are going to discuss the concept of analytical modelling and output of GIS. An important challenge of GIS development is to transfer the

static GIS system into solving spatio-temporal processes application. Linking a GIS to a spatially distributed, physically-based process model offers many advantages.

This unit will provide a discussion on the basic concept of analytical modelling. The integration of GIS with modelling has enabled GIS users in thematic mapping to conduct sophisticated analysis and simulation for both scientific research and policy analysis. In this unit, you will also learn about various types of process models. This unit also deals with the output of GIS. Apart from this, the unit will also focus on the spatial decision support systems and the various functional applications of GIS.

4.3 ANALYTICAL MODELLING BASICS

Analytical modelling is the technique used to predict or simulate the real world phenomena or processes. It is the art and science of predictive analysis of spatial forms and processes. For example to predict future urban landscape we need to understand the present form and the processes of the growth of urban areas to predict the change in future. The most dominant factors of change may be taken into account to model the future urban landscape. One such tool in the GIS is the model builder. Analytical modelling can be done using several layers and GIS functions to have a single predictive layer.

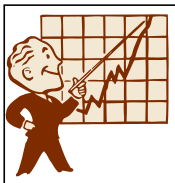
It helps to make decisions on what to do or what actions one should take. It also helps to evaluate our understanding of behaviours and relationships. Spatial form means how the world looks and the spatial process refers to the world works. The factors acts to change spatial form are the processes. Examples are population change, consumer spending patterns, soil erosion, Climate change etc.

Various Problems has been identified while using GIS to model spatial processes. They include :

1. Quality of source data for model calibration,

2. Availability of real world data for validation,
3. Implementation within a GIS using various functions.

The analytical model creates complexity depending on the nature of available data to run it. Because at times data tends to be large and problematic and a long term data may not exist. In certain cases, one has to wait a long time to prove or disprove the model. One should have object orientented programming skills to use spatial macro language for analytical modelling in GIS.



CHECK YOUR PROGRESS

Q. 1: Define the term analytical modelling.

.....

.....

Q. 2: What are the problems associated with analytical modelling in GIS?

.....

.....

4.4 PROCESS MODELS

Process models describe the interlinked processes to have an expected response or a kind of probable response. Process models explain the results that can be anticipated under existing or given conditions.

Various nature of process models are:

1. Physical and environmental process models
2. Human process models of various socio-economic behaviour.
3. Decision making models (single criterion or multi-criteria type)
4. Any model simulates real world processes

Process models aid in understanding the real world processes. It is cheaper and safer to simulate rather than do the real thing. Process models aims to describe, prescribe and explain the reality or phenomena under the existing or given conditions. A set of processes may be linked to explain the situation or to receive response.

There are two types of process models namely:

1. **A priori model:** This model is used where a body of theory is yet to be established. This means that the model is built first followed by a theory. This model is also used to investigate whether a phenomenon is actually happening or not. Example: Global Warming
2. **A posteriori model:** This model is designed to explore an established theory. Example: Applying avalanche or erosion model to new areas.

In GIS, the process model uses analogue, conceptual or mathematical model concept.

- **Natural and scale analogue models:** It uses actual events or real world objects as a basis. Example: Slope stability model, scale models, maps, images etc.
- **Conceptual models:** Usually expressed in verbal or graphical form.
- **Mathematical models:** It uses various statistical techniques.

4.4.1 Physical and Environmental Process Modelling

One of the greatest challenges in GIS is to build reliable model of physical or environmental processes. If the functional relationships among the variables are understood properly, such models can be built in GIS environment very effectively.

Forecasting future conditions is very difficult but very important. Forecast models often linked to GIS. Physical and environmental process model designed to handle variables such as weather conditions, flooding scenario, sea level raising or surface conditions of land and sea etc.

Various types of exhibiting models can be designed for:

- Trends
- Conditions
- Changes
- Predictions or forecasts etc.

4.4.2 Decision Making Process Modelling

In decision making modelling, large number of data layers may be used. Data preparation, setting the objectives and functional associations are important requirement in such analysis.

A sequence of events and processes are involved in decision making.

Steps follow to execute the decision making process modelling includes:

a) Define problem
b) Selection of data layers and criteria
c) Standardisation of criterion scores
d) Allocation of weights or specification of weights to data layers
e) Linking events and processes with data layers by choosing specific functions/ algorithms.
f) Running the model.

A skill in object oriented programming is required to design multi-criteria based decision making models. An easy Graphic User Interface (GUI) may be designed for a specific datasheet and the goals to be achieved.

4.4.3 Human Process Modelling

Generally, human processes are very complex than the natural process. Therefore, the modelling approach in this case differs widely than the physical processes. However, significant components of the system can be modelled by generating spatial data sets for many attributes which are functionally interrelated.

Spatial interaction models are used to “understand and predict” the 1) Location of activities and 2) Movement of materials, people, information. Generally, human process model tries to answer the questions of the following nature:

How do people move through space?
How do they make decisions such as …?
Where to live?
Which shops to buy from?

Where to go on holiday?

Spatial interaction models

Predict flow

Supply, demand and distance

Distance decay functions

The data layers needed to obtain specific result may vary according to the aim of modelling human response. Data selection preparation, designing criteria applying the specific algorithm or function govern the modelling process. The user is to be trained to use various functions and tools of modelling in a particular package in which he is accustomed to work. Any kind of modelling is based on data, events, functions or operations/ processes leading finally to a result. These components are to be well organised before any kind of modelling attempt.



CHECK YOUR PROGRESS

Q. 3: What is the meaning of spatial form and process?

.....

.....

Q. 4: Mention the various nature of process model.

.....

.....

Q. 5: Which type of process model is used for flood forecasting?

.....

.....

Q. 6: In which process model the map overlay technique is extensively applied in GIS?

.....

.....

Q. 7: Which process model is used to model the supply and demand?

.....

.....

4.5 OUTPUT OF GIS

All the activities in GIS are map driven. Mapping a single event or process is based on its attribute but the essence of GIS mapping is to handle many attributes either to combine or interrelate and derive map outputs. Mapping is the most convenient way to convey the results of analytical model executed through the spatial analysis technique in GIS. It provides several tools to place data on a map and display it in an effective manner.

For each map creation process, one need to be ascertain its purpose by answering the following questions:

- What will be shown in the map? Or what is the theme?
- Does the map need to stand on its own, or will it be part of a larger presentation?
- Who are the targeted map users?
- What information is to be communicated?
- What may be the combined results of few data layers put together to answer a geographic question?

Answering these questions will help in determining the organisation and presentation of information on the map.

Irrespective of the kind of map to make and the kind of GIS software used, the process always begins with creating a new map document. Either an empty map is created with nothing on it or uses a map template as a starting point. Map templates typically contain a predefined page layout that arranges map elements such as North arrows, scale bars, neatlines and logos. Here just the data is to be added and then immediately the map can be printed. Otherwise, one has to follow the entire map creation stage.

It is possible to create great looking publication-quality maps and create interactive displays that link charts, tables, drawings, photographs and other elements to the data. The results presented in various formats can be embedded in reports, published on the Web, exported to standard formats, or printed to hang on the wall.

Once a map is created, it is necessary to export it from a map document to another file type. Map can be exported in the form of several types of file

formats. These include EMF (Enhanced Windows Metafiles), BMP (Windows Bitmap), EPS (Encapsulated PostScript), PDF (Portable Document Format), JPEG (Joint Photographic Experts Group) etc. for the purpose of easy access by the relevant software other than GIS software.



CHECK YOUR PROGRESS

Q. 8: What is contained in map templates?

.....

.....

Q. 9: Mention some types of file formats where maps can be exported.

.....

.....

4.6 SPATIAL DECISION SUPPORT SYSTEMS

Many spatial problems are complex and require the use of analysis tools / algorithm and the skills of modelling. They are semi-structured or ill-defined because all of their aspects cannot be measured or modelled

Spatial Decision Support Systems (SDSS) are designed to help decision-makers in solving complex spatial problems. SDSS provide a framework for integrating:

1. Analytical modelling capabilities
2. Database Management Systems (DBMS)
3. Graphical display capabilities
4. Tabular reporting capabilities
5. The decision-maker's expert knowledge

GIS is equipped with a DBMS, graphical display and tabular display capabilities. So, the addition of analytical modelling capabilities and the decision-maker's expert knowledge, the defined structure of information flow only allow to create SDSS.

Four key modules are needed in a SDSS.

1. A Database Management System (DBMS):

GIS database management systems are designed to support cartographic display and spatial query. Database of an SDSS must support cartographic display, spatial query and analytical modelling by integrating locational, topological and thematic data. Database must permit the user to construct and exploit complex spatial relations between all three types of data at a variety of scales, degrees of resolution and levels of aggregation. Database management systems, found in many GIS, use the relational data model.

2. Analysis Procedures in a Model Base Management System (MBMS):

A Model Base Management System (MBMS) consists of small pieces of code, each of which solves a step in an algorithm. As many of these steps are common to several algorithms, this approach saves large amounts of code. The system developer only has to modify one piece of code to update a step in several algorithms. The MBMS also contains information about how steps are sequenced to execute a given algorithm. Using an MBMS facilitates rapid development and testing of new algorithms. Implementation may be achieved simply by adding a new formula to the MBMS. In other cases, new code for additional steps also may be added to the model-base.

3. A Graphical and Tabular Report Generator:

The following capabilities should be provided in a SDSS.

- A high-resolution cartographic displays
- General-purpose statistical graphics, including two and three-dimensional scatter plots and graphs
- Specialised graphics for depicting the results from analytical models and sophisticated statistical techniques
- The full range of tabular reports normally associated with each of the above.



CHECK YOUR PROGRESS

Q. 10: What is SDSS?

.....

.....

Q. 11: What type of Database Management System is normally used in most GIS?

.....

.....



4.7 LET US SUM UP

- A model is a representation of reality in many ways including maps capable of being displayed and manipulated by using various tools available in GIS.
- To help decision-makers in solving complex spatial problems Spatial Decision Support Systems (SDSS) provide a framework for integrating various capabilities of GIS.
- A problem specific SDSS may be designed for the planner decision makers.
- Mapping is the most convenient way to convey the results of analytical model executed through the spatial analysis technique in GIS.
- Using various tools of map composition in GIS, results can be displayed in pictorial and tabular form.
- Wherever there is a geographical elements in area of resource or utility management, GIS can be used.
- GIS can be used in different streams of planning, natural resource based applications, land parcel based applications, facilities management, street network based applications etc.
- For all location based services this technology is useful.



4.8 FURTHER READING

1) Paul A. Longley & Michael Batty. *Spatial Analysis: Modelling in a GIS Environment*. Published by Wiley, 1997.

2) Paul A. Longley, Michael F. Good child, David J. Maguire & David W. Rhind, *Geographic Information Science and Systems* (4th Edition), *Published by Wiley,* February 2015.

3) Maguire David J, Michael F Goodchild, and Michael Batty (Editor), *GIS. Spatial Analysis and Modeling*, Jenson Books Inc.



4.9 ANSWERS TO CHECK YOUR PROGRESS

Q.No.1: Analytical modelling is the technique used to predict or simulate the real world phenomena or processes. It is art and science of predictive analysis of spatial forms and processes.

Q. No. 2: The problems associated with analytical modelling of GIS are:

- Quality of source data for model calibration,
- Availability of real world data for validation,
- Implementation within a GIS using various functions.

Q. No. 3: Spatial form means how the world looks and the spatial process refers how the world works.

Q. No.4: Various nature of process models are physical and environmental, human, decision making and any model simulates real world processes

Q. No.5: Physical and environmental process model.

Q. No.6: Decision making process model.

Q. No.7: Human process model.

Q. No.8: Map templates typically contain a predefined page layout that arranges all the map elements.

Q. No.9: BMP (Windows Bitmap), PDF (Portable Document Format), JPEG (Joint Photographic Experts Group) etc.

Q. No.10: Spatial decision support systems (SDSS) are designed to help decision-makers in solving complex spatial problems in GIS by providing a framework for integrating all capabilities of GIS.

Q. No.11: Relational Database Management System (RDBMS).



4.10 MODEL QUESTIONS

Q. 1: Explain the map creation process in GIS.

Q. 2: Discuss the key modules in SDSS.

Q. 3: What is the meaning of priori and posteriori models?

Q. 4: List the required capabilities of GIS for the SDSS.

Q. 5: Discuss the types of mathematical process model.

---×---

UNIT 5: BASICS OF IMAGE PROCESSING - I

UNIT STRUCTURE

5.1 Learning Objectives

5.2 Introduction

5.3 Introduction to Digital Image Processing

5.4 DIP Software Systems

- 5.4.1 User Interface of DIP software
- 5.4.2 Loading of Digital Data into DIP Software
- 5.4.3 Conversion of Digital Data in DIP Software

5.5 Image Properties

- 5.5.1 Spatial resolution
- 5.5.2 Radiometric resolution
- 5.5.3 Image histogram

5.6 Data preparation

- 5.6.1 Geometric corrections
- 5.6.2 Radiometric corrections

5.7 Reprojection and resampling of Digital Image data

5.8 Subsetting of Digital Image Data

5.9 Let us Sum Up

5.10 Further Reading

5.11 Answers to Check Your Progress

5.12 Model Questions

5.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- describe the concept of image processing
- discuss the system configuration of digital signal processing
- explain the general architecture of a digital image processing software
- get an idea of different properties of an image
- illustrate about data preparation methods implemented on an image

- explain processing steps like reprojection and subsettting of Digital Image data.

5.2 INTRODUCTION

In the previous unit you have learnt about the analytical modelling and output of GIS from various data sources. In this unit, we are going to discuss the concept of image processing, a mjaor source of data for GIS output. This unit provides a discussion on the image processing and system configuration of digital image processing system. The unit will also deals with the digital image processing software. In this unit, you will also learn various properties of an image. Apart from this, the unit will also focus on the digital data pre-processing and reprojection and subsettting of Digital Image data.

5.3 INTRODUCTION TO DIGITAL IMAGE PROCESSING

The word image comes from the Latin word *imago* which represents a visual observation in 2D (e.g., a picture) or 3D (e.g., a hologram) space of a physical object. It may be real or imaginary. In the present context, an image usually means a picture of the surface of the earth, acquired from a space borne or airborne camera utilising one or some bands of electromagnetic spectrum originating from the sun or artificial sources.

Before the advent of digital technologies, an image represented a colour or B&W photograph printed on a piece of paper. In the digital context, an image is a matrix or array of pixels (picture elements) arranged in rows and columns also called as raster image (Figure 5.1 and 5.2). Every pixel in a raster dataset is identified by the row and column numbers.



Figure 5.1: A black & white picture



Figure 5.2: A pixelated portion of the Figure 5.1

The main objective of image processing is to enhance the quality of a raw image or extract some information from it. In imaging science, image processing is a technique by which we try to recognize the features of the earth's surfaces. In doing so, some statistical or mathematical principles are applied to regroup the pixels that give characteristics to the features of earth's surface. The output of the processing is an image with a kind of information derived from the image. In the last five to six decades, various techniques have developed with regard to image processing. In Digital Image Processing (DIP), images are considered as two dimensional arrays of pixels. DIP is computer-based manipulation and interpretation of digital images. The images are processed in computer utilising Digital Image Processing software.



CHECK YOUR PROGRESS

Q.1: Fill in the blanks:

a) The word image comes from the Latin word _____.

b) An image is a matrix or array of _____ arranged in rows and columns.

c) In digital image processing, images are considered as _____ signals.

d) The images are processed in computer utilising _____ software.

e) The main objective of an image processing is to _____ the quality of a raw image.

5.4 DIP SOFTWARE SYSTEMS

The fundamental requirement of an image processing system is a computer with fast, modern and image processing software. High resolution satellite images consume a lot of memory spaces. Therefore, a computer with large module of RAM and large capacity hard disk is more preferable for image processing operations dealing with high resolution and large images. A dedicated graphics card rather than inbuilt graphics processor makes the operations smoother and faster. Usually, a workstation computer is more suitable than a desktop or laptop PC for image processing operations.

Among the input devices, a good quality mouse is essential, because image processing operations may frequently require precise positioning. Some years before, a digitizing table often accompanied an image processing system. Now-a-days, it is replaced by on screen digitizing utilizing the mouse. For photogrammetric operations a 3D mouse is often required. A large format scanner is also required for digitalization of old hardcopy images, topographic maps or manuscript maps.

Modern image processing laboratories are equipped with large format (A_0) printer / plotter and scanning systems. Photogrammetric software is an integral part of image processing.

There are numerous software packages available for digital image processing. The most frequently used softwares are- ERDAS Imagine, ENVI, Geomatica etc. Each of these software systems is bundled with image correction, rectification, map projections, classifications, image catalogue and modelling functions along with readily available analytical tools. These can be customised as per the user's need. Besides these some spatial macro language (SML) tools are also available for advanced users, who are trained in programming. They can use SML tools for ***batch processing*** of the image from image catalogue.

Batch processing: It is the process of classifying the images from an image catalogue/ image repository by issuing pre-defined commands through SML tools.

These softwares are used to perform mathematical or artistic processing on digital images. The input data are of two types. One is analog data e.g. topographical sheet, aerial photo etc. and the other is digital data

e.g. satellite data. To process these data digital image processing system requires computer hardware and image processing software.

5.4.1 User Interface of DIP Software

The digital image processing software (DIP) usually have graphical user interface. The images can be viewed in viewer windows, where it can be enlarged or reduced. In ***ENVI*** (versions earlier than 5.0), several viewer windows are utilised for viewing the same image in different enlargements. However, in later versions, all these windows are consolidated into one like the other image processing software. The algorithms of manipulations are usually menu driven in most of the software. However, they were command line driven in the past. There are other DIP softwares like **ERDAS Imagine** and **Geomatica**. Some tools / modules of image processing are available in Arc GIS version 10 and above.

5.4.2 Loading of Digital Data into DIP Software

Digital images are often supplied by acquiring agency (e.g. NRSC for IRS LISS 3 Data) in CD or DVD in their own image formats. Many other digital image data can be downloaded from the online archives of the acquisition agencies.

The images that are supplied in CD or DVD ROMs by the acquisition agencies are usually stored in their own formats. These digital data are generally loaded into the DIP software using 'Import' or similar functions. While importing, suitable image formats must be selected to convert them into the default image format of the respective DIP software. For example, if we have to load an IRS LISS 3 1D satellite image from the CD ROM supplied by the NRSC into ERDAS Imagine software, we have to import it choosing 'IRS-1C/1D(super structure format)' as the input file type. After successful implementation of the import function of the ERDAS Imagine software, the supplied image data would be converted to the default imagine file format with *.img extension. For NOAA's AVHRR data, we will require to choose AVHRR (NOAA) as the import file type.

Many online archived image data are stored in compressed formats like ZIP or RAR in separate bands. We need to download the required bands or all the bands from the online archives. The compressed data must be decompressed after downloading and required bands are to be stacked into one image file. In the ERDAS Imagine software there is a special command, called, '*Layer Stack*' to accomplish this function.

Multispectral satellite images usually contain more than 3 bands. When we load a multispectral image in DIP software as colour image, we can utilise only 3 of the bands and project them through the blue, green and red electron guns of the display device of the image processing computer to visualise in the form of a colour image. If we project the true blue, green and red bands of the multispectral image through the blue, green and red electron guns, it gives a natural colour image. If we project different spectral bands through these electron guns, it gives a False Colour Composite (FCC). Different combinations of spectral bands of multispectral satellite images give different FCCs. The most common FCC is created when we project infrared band through the red electron gun, red band through green electron gun and green through blue electron gun. This is also called the standard FCC. This FCC is very useful in studying the vegetation cover of the surface of the earth.

5.4.3 Conversion of Digital Data in DIP Software

There are many formats of digital image data with different advantages and disadvantages. Different DIP software utilises their own proprietary data formats for processing of images. However, when we work with multiple image processing software we may require to convert one format into another for the ease of processing. The image processing software usually has export and import functions to convert the formats of the digital image data. Nowadays, interoperability functions are highly implemented by the DIP software.

The image processing software also supports many other proprietary or non-proprietary raster data formats, i.e., geotiff, MrSID, jp2 etc. These raster data format utilise different kinds of lossy or non-lossy compressions. We may also require using these data formats in different situations while implementing an image processing project.



CHECK YOUR PROGRESS

Q. 2: State true or false:

(a) A digital image is an array or matrix of pixels arranged in columns and rows.

(b) A digital image is usually printed on a piece of paper.

(c) For acquisition of digital images, electromagnetic spectrum may or may not be utilised.

(d) Digital image processing is usually done in computers and but can also be done by utilising optical instruments.

(e) The image processing software usually has export and import functions.

Q.3: Fill up the blanks:

(a) DIP stands for _____.

(b) FCC stands for _____.

(c) IRS LISS 3 satellite data are acquired by the _____.

(d) The default image file format of the ERDAS Imagine DIP software is _____.

(e) In the ERDAS Imagine software there is a special command, called _____ to accomplish this function.

5.5 IMAGE PROPERTIES

Knowing the properties of a digital image before analysis is very important to understand its usability. It can be done in every DIP software by looking at the image ***metadata***. The two most important image properties that are important to understand before trying to digitally process them are spatial resolution and radiometric resolution.

Metadata: It is data based on data. For an image the rows, columns, pixel size, co-ordinate system, size of the image, map extents etc. are metadata.

5.5.1 Spatial Resolution

Spatial resolution of an image is the unit area on the ground that the individual pixel of the image represent. It defines the size of the smallest possible feature that can be detected. It is usually expressed in metres. For example, an IRS LISS 3 multispectral image has spatial resolution of 23.5m. It means that each pixel of this particular type of satellite image represents an area of 23.5x23.5 m^2 of the surface of the earth. The spatial resolution of an image is determined primarily by the instantaneous field of view (IFOV) of the sensor, which is a measure of the ground area viewed by a single detector element in a given instant. A high resolution image refers to one with pixels representing small unit area on the ground and thus, finer details can be identified. On the other hand, pixels represent large unit area of the ground in a low resolution image and thus, finer details cannot be observed. It is important to note that lower resolution leads to data reduction. An example of different spatial resolutions of the same image is shown in the Figure 5.3. Satellite images acquired by some modern sensors may have very high spatial resolution. For example, the Quick Bird satellite images acquired by the earth observation satellite, owned by Digital Globe have 61cm spatial resolution in its panchromatic band and 2.44m in its multispectral bands. Satellite images with very high spatial resolution are important for urban planning, while those with medium resolution can be used for regional level planning or analyses.



Figure 5.3 Example of different spatial resolutions of an image from high to low (L to R)

5.5.2 Radiometric Resolution

The quality of a satellite image that represents or distinguishes differences of intensity is usually defined by the radiometric resolution. It defines the number of shades or gray levels in the image or picture. It is usually expressed as number of bits. For example, an 8-bit image has 2^8=256 gray levels or shades, while a 16-bit image has 2^{16}= 65536 gray levels or shades. An image with higher radiometric resolution would resolve finer details of the surface of the earth. For example, the radiometric resolution of a LANDSAT-7 satellite image is 8 bit, while that of a modern LANSAT-8 OLI image is 16-bit. The LANDSAT-8 image would, therefore, show more shades of colour and thus, more details of the surface of the earth. An example of different radiometric resolution of the same image is shown in the Figure 5.4. The gray level is also known as brightness value or intensity value or digital number (DN) of pixels.



Figure 5.4 Example of different radiometric resolution of an image

5.5.3 Image Histogram

An image histogram is a graphical representation of the pixel intensity values and their frequency. It is constructed by plotting each gray level or intensity level value of the image against the total number of pixels that show that particular gray level. The image histogram is usually used to detect image defects, for example, saturation, spikes, gaps etc. The good understanding of the image histogram is very important in image processing. The Figure 5.5 (a), (b) & (c) shows the histograms of 3 versions of the same picture. The histogram of the first image shows normal distribution (error free)

of the intensity range, while the second image shows gap (absence of some shades) and the third image shows spikes (abnormally high counts of some shades). DIP software generally has tools to remove such kind of defects from the images..

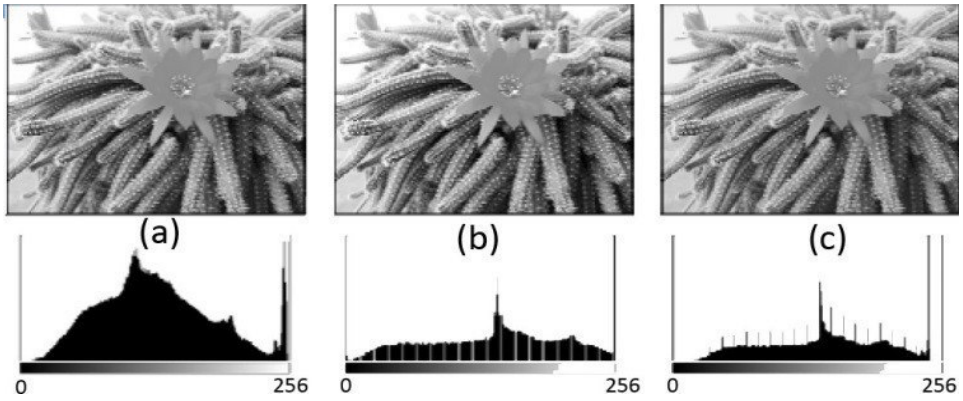


Figure 5.5 The figure shows image histograms of 3 versions of the same picture



CHECK YOUR PROGRESS

Q. 4: What do you mean by image histogram?

.....

.....

5.6 DATA PREPARATION

Data or image pre-processing is done to correct distorted or degraded image data to create more faithful representation of the original scene. It is the first step in digital image processing and is performed to correct image data for distortions or degradations stemming from the image acquisition process. Data pre-processing processing or image corrections consists of two primary types: geometric and radiometric corrections. The details of each type of correction is discussed in the following sections.

5.6.1 Geometric Corrections

Raw images contain geometric distortions or errors in geometry due to various factors such as variations in altitude, attitude,

and velocity of sensor platform, panoramic distortions, distortions due to the earth's curvature, atmospheric refraction, relief displacement and non-linearities in the sweep of a sensor's IFOV. Geometric corrections are performed by the process of rectification and registration. **Rectification** is performed on images by finding Ground Control Points (GCPs) on the image which correspond to the actual locations on the ground through field surveys or ground maps. **Registration** is the linking of an image with a geometrically correct image to obtain locations from it. Mathematical transformations are done to obtain geometrically corrected images using different functions and the output is saved to a new raster file.

5.6.2 Radiometric Corrections

Radiometric corrections which includes atmospheric and topographic corrections is concerned with improving the accuracy of surface spectral reflectance, emittance, or back-scattered measurements obtained using a remote sensing system. Radiance measured by a sensor may be influenced by changes in scene illumination, atmospheric conditions, viewing geometry and instrument response characteristics. Some of the processes involved in image enhancement include sun elevation correction, haze compensation, de-noising and de-striping. Sun elevation correction is done by using the metadata of the image and using the time and day of the year in order to normalise illumination in images from different dates with the sun in the zenith. Haze compensation is usually performed by identifying dark objects in the image and assigning them the lowest value and stretching the image histogram accordingly. Random noise due to unwanted signals from the sensor can be removed by threshold of difference in pixel values. Striping present in images due to sensor errors can be corrected by averaging neighbour pixel values.



CHECK YOUR PROGRESS

Q. 5: What are the two process of Geometric corrections?

.....

.....

Q. 6: What is image registration?

.....

.....

Q. 7: What are the factors that might influence radiance of image?

.....

.....

Q. 8: How are differences in scene illumination due to sun elevation corrected?

.....

.....

5.7 REPROJECTION AND RESAMPLING OF DIGITAL IMAGE DATA

Re-projection is another important technique of pre-processing of satellite images before any image analysis job. Re-projection involves transforming and rectifying the projection, ellipsoid and datum of an image. Satellite images and other spatial data obtained from different sources may have different projection, ellipsoid and datum. Without bringing them to a common projection, ellipsoid and datum system, these data cannot be superimposed and analysed correctly in the DIP software. The DIP software usually supports hundreds of projection and datum which are used in different parts or countries of the world. The DIP software also provides re-projection algorithms to change one projection and datum system to another.

For the entire coverage of the world, WGS84 ellipsoid and datum with geographic co-ordinate system is employed by the GPS system. For the very small areas, the UTM (Universal Transverse Mercator) projection

system is employed. In this system, the entire globe is divided into 60 zones of 6^0 longitudinal extents. The geometric properties of area, distance and direction can be maintained for small areas using the UTM system of projection. Moreover, it employs Cartesian system of co-ordinates for the mapping space.

Resampling is the process of determining new values for pixels in an output raster that results from applying some geometric transformation to an input raster after geometric corrections are applied to it. There are three common methods of resampling:- (i) Nearest neighbor, (ii) bilinear interpolation and (iii) cubic convolution.

(i) *Nearest neighbour* resampling uses the digital value from the pixel in the original image which is nearest to the new pixel location in the corrected image.

(ii) *Bilinear interpolation* resampling takes a weighted average of four pixels in the original image nearest to the new pixel location.

(iii) *Cubic convolution* resampling goes even further to calculate a distance weighted average of a block of sixteen pixels from the original image which surround the new output pixel location.

5.8 SUBSETTING OF DIGITAL IMAGE DATA

In many applications, the acquired image size or coverage is larger than the area of interest; in that case the image can be resized with a vector polygon or the required map extent in terms of co-ordinates of extreme corners of the image. This process is called subsetting the image. When the size of the image is reduced it allows faster processing also.

This operation of subsetting is also called **clipping or extraction** in different platforms. Subsetting of a satellite image is usually done with the help of a mask which may be represented by a vector polygon file or an *aoi* (area of interest) file (in ERDAS Imagine) or a raster file which marks the exact boundary of the required extent of the satellite image. This mask file may have any shape. The input image file and the mask file should have the same projection and datum system. In certain softwares (e.g. ArcGIS) the required extent of the area can also be defined as a square or rectangle

by inserting the coordinates of the corner points during the sub-setting operation. In this case, only the east-west and north-south extent of the area is extracted with the help of co-ordinates.



CHECK YOUR PROGRESS

Q. 9: Fill in the blanks:

a) Re-projection involves transforming and rectifying the _____, _____ and _____ of an image.

b) UTM stands for _____.

c) The algorithms for reprojection in the DIP software allow us to change _____ to another.

d) For UTM the whole globe is divided into _____ zones and _____ longitudinal extents.

Q. 10: State whether true or false:

a) Subsetting or clipping operation slices a portion of a large scene.

b) In subsetting operation the area of interest over the input image is generally masked by vector, raster or aoi files.

c) AOI stands for Area of Image.



5.9 LET US SUM UP

- An image is a visual perception in 2D or 3D space of a physical object, which may be real or imaginary.
- In satellite image processing, an image usually means a picture of the surface of the earth, acquired by a space borne or airborne camera utilizing some bands of electromagnetic spectrum.
- Some popular image processing software are ERDAS Imagine, ENVI, PCI Geomatica etc.
- The image processing software usually have viewer windows to display images and the image processing algorithms are generally menu driven.

- Spatial resolution of an image is the unit area on the on the ground that the individual pixels of the image represent.
- The spectral resolution describes the ability of a sensor to define spectral bands with finer wavelength intervals.
- An image histogram is a graphical representation of the pixel intensity values of a digital image and is constructed by plotting the number of pixels against each gray level or intensity level value of the image.
- Data pre-processing includes two major types of corrections: geometric and radiometric.
- Geometric corrections are performed by the process of rectification and registration to correct errors in geometry.
- Radiometric corrections which includes atmospheric and topographic corrections is concerned with improving the scene radiometry using various techniques.
- Re-projection involves transforming and rectifying the projection, ellipsoid and datum of an image.
- Resampling is the process of determining new values for pixels in an output raster that results from applying some geometric transformation to an input raster.
- There are three common methods of resampling: (i) Nearest neighbor, (ii) bilinear interpolation and (iii) cubic convulation.
- Subsetting is the extraction or clipping of the area of interest or study area from an image covering a large area.



5.10 FURTHER READING

1) James B. Campbell and Randolph H. Wynne, *Introduction to Remote Sensing,* Published by the Guilford Press.
2) Rafael C. Gonzalez and Richard E. Woods**,** *Digital image processing,* Published by Prentice Hall.
3) Thomas Lilesand, Ralph W. Kiefer and Jonathan Chipman, *Remote sensing and image* interpretation, Published by *Willey.*



5.11 ANSWERS TO CHECK YOUR PROGRESS

Q. No. 1:

(a) imago

(b) pixels

(c) two dimensional

(d) digital image processing software

(e) enhance

Q. No. 2:

(a) True

(b) False

(c) False

(d) False

(e) True

Q. No. 3:

(a) Digital Image Processing

(b) False Colour Composite

(c) NRSC, India

(d) img

(e) Layer Stack

Q. No. 4: An image histogram is a graphical representation of the pixel intensity values of a digital image. It is constructed by plotting each gray level or intensity level value of the image against the total number of pixels that show that particular gray level. It is usually used to detect image defects, for example, saturation, spikes, gaps etc.

Q. No. 5 Geometric corrections are performed by the process of rectification and registration.

Q. No. 6: Registration is the linking of an image with a geometrically correct image to obtain locations from it.

Q. No. 7: Radiance may be influenced by changes in scene illumination, atmospheric conditions, viewing geometry and instrument response characteristics.

Q. No. 8: Sun elevation correction is done by using the metadata of the image and using the time and day of the year in order to normalise illumination in images from different dates with the sun in the zenith.

Q. No. 9:

a) projection, ellipsoid, datum

b) Universal Transverse Mercator

c) one projection and datum

d) 60, 6^0

Q. No. 10:

a) True

b) True

c) False



5.12 MODEL QUESTIONS

Q. 1: Define the term 'digital image'. Why and how is image processing done?

Q. 2: Write shortly about different requirements for a computer system to be utilized in image processing.

Q. 3: Describe the utility of 'import' and 'export' commands of DIP software.

Q.4: Write about spatial resolution of satellite image.

Q. 5: What is radiometric resolution? How is the radiometric resolution expressed?

Q. 6: What is image histogram? How is an image histogram constructed? What are the utilities of an image histogram?

Q. 7: What are the processes involved in geometric and radiometric corrections?

Q. 8: What is reprojection? What are the universal projections used?

Q. 9: What is resampling? What are the methods of resampling digital data?

Q. 10: Why is subsetting or clipping required in image processing? Write shortly about the procedure of subsetting in image processing.

---×---

UNIT 6: BASICS OF IMAGE PROCESSING- II

UNIT STRUCTURE

6.1 Learning Objectives

6.2 Introduction

6.3 Geo-referencing of Raster Data Set

6.4 Pattern Recognition in Digital Image Processing

6.5 Image Enhancement

- 6.5.1 Image Filtering
- 6.5.2 Band Ratio
- 6.5.3 Principal Component Analysis

6.6 Let us Sum up

6.7 Further Reading

6.8 Answers to Check Your Progress

6.9 Model Questions

6.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- identify a few advanced image processing techniques.
- perform geo-referencing of raster data set.
- describe pattern recognition in digital image processing.
- discuss the different processes of image filtering.
- explain band ratios.
- describe principal component analysis of various bands in image.

6.2 INTRODUCTION

In the previous unit you have got an idea about the basics of digital image processing. In this unit also, we shall discuss digital image processing but in a slightly advanced manner. The unit focuses on the process of geo-referencing and pattern recognition in a digital image. This unit also provides a discussion on the image enhancement techniques such as filtering and spectral operations of band ratioing and principal component analysis.

6.3 GEO-REFERENCING OF RASTER DATA SET

Geo-referencing is the process of assigning location in terms of geographic co-ordinate systems to image with the help of a reference ellipsoid and datum.

Spatial analysis of a raster data set requires its every pixel to represent a location of definite portion of the surface of the earth. Modern high resolution multispectral images or LiDAR or RADAR data usually contain spatial reference information embedded in them. But when we work with medium resolution or old satellite images, scanned maps or aerial photos, we require to assign them the earth's coordinate system. This operation is also known as image geo-referencing or image geo-rectification or image registration. A map coordinate system is defined using a map projection (a method by which the curved surface of the earth is portrayed on a flat surface) and an ellipsoid along with associated datum.

Once the geo-referencing is done, the locations on the scanned map or satellite image are defined using earth's coordinate system (latitude and longitude for an ellipsoidal surface) or a projection system (metre grid for a flat surface). Geo-referencing raster data allows it to be viewed, queried, and analysed with other geographic data.

The geo-referencing process involves identifying a series of ground control points (GCPs) with known x, y coordinates that link locations on the raster data set with actual locations on the surface of the earth. The GCPs can also be taken from an already geo-referenced satellite image or map (source). Both the source and target images should represent the same area on the surface of the earth. The GCPs are generally defined by some permanent features on the surface of the earth such as road junctions that can viewed in both target satellite images and maps. The coordinates of the ground control points can also be collected from the field using a survey grade GPS. Topographic maps may be utilised to geo-reference a satellite image.

Utilising the GCPs, a polynomial transformation can be made that shift the raster data set from its current location to the spatially correct

location. The number of GCPs required for this depends on complexity of transformation or warping of the raster data set to map coordinates. The GCPs should uniformly spread over the entire data set rather than concentrating in some areas. Excluding polynomial transformations, we can use several methods of transformation like spline, projective, adjust transformation, etc. Usually polynomial transformation is preferred over the others.

The polynomial transformation can be achieved in several orders and number of the GCPs required for it must be 3 for a first order, 6 for a second order and 10 for a third order (Figure 6.1).

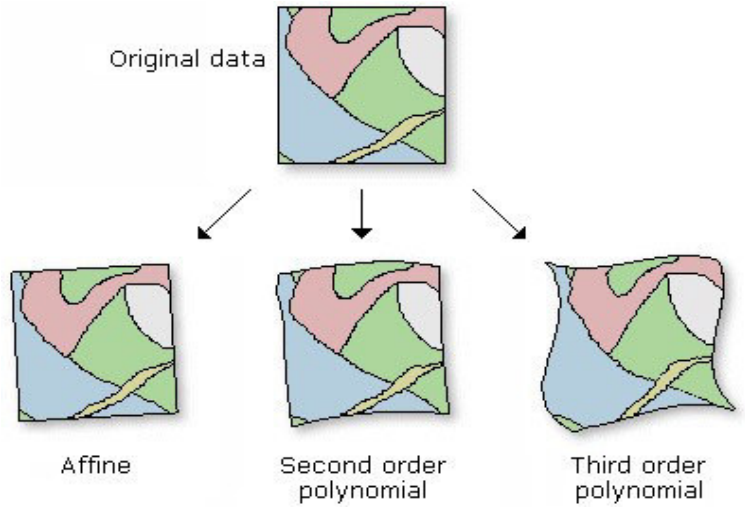


Figure 6.1: The first, second and third order polynomial transformation (Courtsey: ArcGIS 9.3 Desktop help)

The first-order or affine transformation shifts, scales and rotates a raster data set (Figure 6.1). Straight lines remain as straight lines and rectangles change into parallelograms with arbitrary angle of rotation and scale.

The distortion becomes more complex if we choose higher order polynomial transformation. However, for geo-rectification of the satellite image, transformations higher than third order are rarely needed.

The error in geo-referencing is calculated by a measure of the residual error given by the general formula. The total error is computed by taking root mean square (RMS) sum of all the residuals that describes the

consistency of the transformation between different control points. If the error is large, it can be reduced by (i) removing GCPs with very high residual error; (ii) slightly displacing the control points with high residual errors, and (iii) by adding additional control points to adjust the errors. However, a low RMS error may not mean an accurate registration.

The image geo-referencing requires an earth model (ellipsoid) and datum. The projection system represents a mathematical model to convert the spherical surface of the earth into a flat surface, in accordance with an ellipsoid and its datum. There are various types of projections and these can be associated with various ellipsoids and datums. A projection is not required when we consider the earth's surface as a curved surface and can use earth's latitude and longitude as coordinates. In that case the georeferenced image remains as unprojected (geographic), but, an ellipsoid or datum still would be required. However, when we require doing any kind of measurements (distance and area) over the georeferenced image, it must be converted to projected image. In the limited space of this document it will be difficult to discuss about projection and datum in detail. But the learners are encouraged to consult relevant books for better understanding of these concepts.

6.4 PATTERN RECOGNITION IN DIGITAL IMAGE PROCESSING

As we noted in the previous units, analysis of remotely sensed data involves an identification of various targets in that particular image. Those targets may be environmental or artificial with features which consist of points, lines, or areas. Pattern Recognition is the key for interpretation and information extraction from Remote Sensing images. Observing the differences between targets and their backgrounds involve comparing different targets based on any, or all, of the visual elements of tone, shape, size, pattern, texture, shadow, height, site and association, different features may be interpreted in the satellite imagery. The following nine elements of image interpretation which are regarded to have general significance, irrespective of the precise nature of the imagery and the features it portrays are–

- **Tone**: Tone is a measure of the intensity of EMR reflected (emitted) by the earth surface features. Areas of lower reflectance appear as dark gray tones, while higher reflectance areas as light tones in a black and white image. In simple words, tone refers to the relative brightness or colour of objects in an image.
- **Shape**: Another important element for image interpretation is the Shape. In simple terms, shape refers to the general form, structure, or outline of individual objects. In satellite images, large number of components whether natural or man-made can be identified with reasonable certainty merely by their shape.
- **Size**: Size of an object in an image is a function of scale. It is important to assess the size of a target relative to other objects in a scene, as well as the absolute size, to aid in the interpretation of that target. In many cases, the length, breath, height area and/or volume of an object can be significant in determining whether these are surface features or atmospheric phenomena.
- **Texture**: Texture is the frequency of tonal change on an image. Texture gives rough or smooth appearance images. Rough textures would consist of a mottled tone where the gray levels change abruptly in a small area, whereas smooth textures would have very little tonal variation. In fact, in case of radar imagery texture is considered to be one of the most important elements for distinguishing features.
- **Pattern**: Pattern refers to the spatial arrangements of objects. Typically, an orderly repetition of similar tones and textures will produce a distinctive and ultimately recognisable pattern. Natural occurrences like lithological patterns, drainage, man-made objects like road networks have specific patterns, which enables their identification in satellite images.
- **Shadow**: Shadow is also helpful in interpretation as it may provide an idea of the profile and relative height of a target or targets which may make identification easier. This element is especially useful in geomorphologic studies where micro relief features may be easier

to detect under conditions of low angle solar illumination then when the sun is high in the sky.

- **Site and Association**: At an advanced stage in image interpretation, the location of an object with respect to terrain features of other objects is helpful in refining the identification and classification of certain picture contents. In other words, association takes into account the relationship between other recognisable objects or features in proximity to the target of interest.

In such a way, these elements assure a variety of ranks of importance in interpreting an image. In fact, sometimes they can lead to assessment of conditions that are not directly visible in the images, in addition to the identification of features or conditions that are explicitly revealed.



CHECK YOUR PROGRESS

Q. 1: What is image geo-referencing?

.....

.....

Q. 2: Define the Ground Control Points (GCPs).

.....

.....

Q. 3: How can the error in geo-referencing be reduced?

.....

.....

Q. 4: What is pattern reorganisation?

.....

.....

Q. 5: What are the nine elements of image interpretation?

.....

.....

6.5 IMAGE ENHANCEMENT

Image enhancement is the process of adjusting digital images so that the results are more suitable for display or for further image analysis. Image enhancement is largely a subjective process and depends solely on the judgement person making the adjustments. The major processes involved in image enhancement can be categorised under contrast manipulation techniques, spatial feature manipulation and multiple band enhancement. Contrast manipulation techniques include gray level thresholding, contrast stretching and level slicing. Spatial feature manipulation techniques include spatial filtering, convulations, edge enhancement and fourier analysis. Enhancement involving multiple bands of an image include spectral ratioing and Principal Component Analysis (PCA).

All image enhancement techniques can be categorised into three different types based on the type of operation. These are:

i. **Point operations** change the value of each individual pixel independent of all other pixel.
ii. **Local operations** change the value of individual pixels in the context of the values of neighbouring pixels.
iii. **Spectral operations** change the values of individual pixels based on values obtained in different bands.

6.5.1 Image Filtering

Image or spatial filtering is the technique to highlight or suppress the specific features in an image based on their spatial frequency. Filtering helps prepare images for segmentation and / or analysis. Filtering accomplishes noise removal, correction of uneven background and elimination of image details other than the objects of interest. No image filter will work perfectly on all images.

Spatial filtering essentially requires the use of a mask known as a **kernel**. The kernel is a group of pixels of different sizes so that there is a specific central pixel (Fig. 6.2). This mask is moved on

the image such that the centre of the mask traverses all image pixels. The process is also known as convolution. In this process the elements of the kernel are multiplied with the DN number of the matching pixel and the average of them replaces the original value of the central pixel. This is done over the entire image to modify the original values of all pixels. In this process the output image would lose the boundary columns as rows at the end of the operation.

There are a wide variety of filters available to extract information from the images based on the requirement. Image filters can be classified into two types based on their function:

(i) Smoothing filters and (ii) Sharpening filters

(i) *Smoothing filters:* Smoothing filters are used for blurring and noise reduction in the image. Low pass filters are mostly used for smoothing. Low pass filters emphasise low frequency features; suppresses high frequencies by replacing each pixel intensity with the local average intensity (local operation). Mean, median and mode filters are mostly used for smoothing.

(ii) *Sharpening filters:* Sharpening filters enhance details in images by applying a high pass filter (edge enhancement/edge detection). It is also known as derivative filter. The purpose of the sharpening spatial filter is just the opposite of the smoothing spatial filter. Its main focus in on the removal of blurring and highlighting of the edges. It is based on the first and second order derivative.

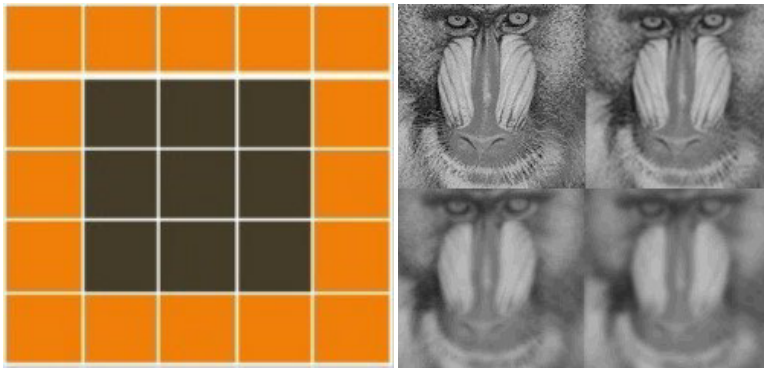


Fig. 6.2: (a) A 3X3 kernel on a 5X5 image, (b) A smoothing filter of different kernel sizes applied on the same image. Larger kernel gives more blur.

6.5.2 Band ratio

Band ratioing or spectral ratioing is the most common transforms applied to images which is a spectral operation involving two or more bands of data. The ratio images are created by dividing the DN numbers or intensity values of pixels of one spectral band by the same values of corresponding pixels of another spectral band of the same scene. This normalisation operation effectively removes shadows from the output ratio image since, the effect of shadow is approximately similar for all bands in the same scene.

By ratioing the data from two different spectral bands, the resultant image enhances variations in the slopes of the spectral reflectance curves between the two different spectral ranges that may otherwise be masked by the pixel brightness variations in each of the bands.

The ratio images are extensively used in studying vegetation, soils, minerals and snow covers on the surface of the earth. Few of important ratio images are described herewith.

(i) TM4/TM3:

It is the ratio between the NIR and the red bands. This ratio image is useful in vegetation mapping. For healthy vegetation the ratio is high and low for stressed or yellow vegetation and non-vegetated areas.

(ii) TM3/TM1:

It is the ratio between the red and the blue bands which highlights reddish coloured iron oxide minerals. It is useful in exploration of super egenesulphide enrichment deposits.

(iii) TM5/TM7:

This is the ratio between shorter wavelength MIR (TM5) and longer wavelength MIR bands (TM7) of the LANDSAT images. In the TM5 band, nearly all minerals are highly reflective. However, in the TM7 band clay minerals show high absorption. This ratio, therefore, highlights the clay minerals found in alteration zones.

(iv) Normal Difference Vegetation Index (NDVI):

This is possibly the most widely used ratio images. It is a ratio between the NIR and the red bands, with the formulae: (NIR-Red)/ (NIR+Red). The result of the subtraction of the corresponding intensity values in the red and the NIR bands is then divided by the sum of the two. This effectively removes sensor noise as well as the topographic effects. In the resultant ratio image the pixel values range from -1 to +1 symmetrically around 0.

(v) Normal Difference Snow Index (NDSI):

The NDSI is useful in discriminating snow from rocks and cloud cover especially in rough topography. This index also provides good contrast between bare ice and its surroundings in glaciers. The NDSI is calculated with the following ratioing:

NDSI = [TM2 (green) – TM5 (shortwave MIR)] / [TM2(green) + TM5 (shortwave MIR)].

Similar indices are designed for ascertaining moisture content in the soil or water bodies.



Fig. 6.3: NDVI of an agricultural area

6.5.3 Principal Component Analysis

Vibrant colours of the earth's landscapes can be easily captured with terrestrial cameras. But while taken from the space, the colours of landscapes are surprisingly limited. In satellite images, vegetated areas are seen to appear faint green, the desert belts are

dominated by grey and brown while the ice covered areas are seen in white. It is because of the fact that the spectral components in the satellite images are highly correlated. Therefore, they have very little contrast. On the other hand, the spectral components of terrestrial photographs are highly uncorrelated and they retain beautiful colours of landscapes. This is evident in the Fig. 6.4, where the scatter plots of brightness values of red and green bands for a terrestrial photograph and a satellite image are shown alongside them.

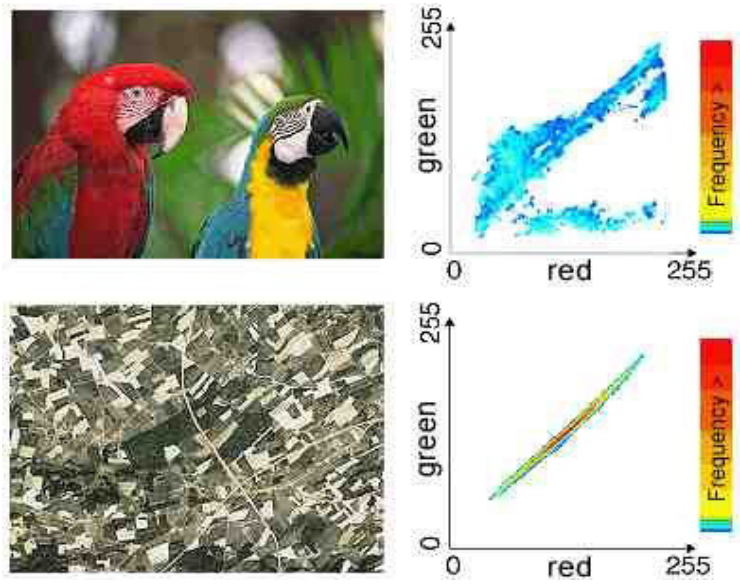


Fig. 6.4: Scatter plot of red vs green bands of terrestrial photograph (top) and satellite image (bottom)

Because of the highly correlated nature, the brightness values of the blue band of the satellite image increases with the brightness value of the red band (Fig. 6.4). This means that various spectral bands of a multispectral satellite image actually carry redundant information and the same information can also be practically retrieved even from a panchromatic satellite image.

The **Principal Component Analysis (PCA)** has its importance in satellite image processing for this reason. It is performed on the spectral bands of a satellite image to reorganise the data so that

they are no longer correlated. If the multispectral satellite image has several spectral bands, the PCA will concentrate almost all the information in first two or three of its components. Other components actually carry noise. This is elaborately explained by the Fig. 6.5.

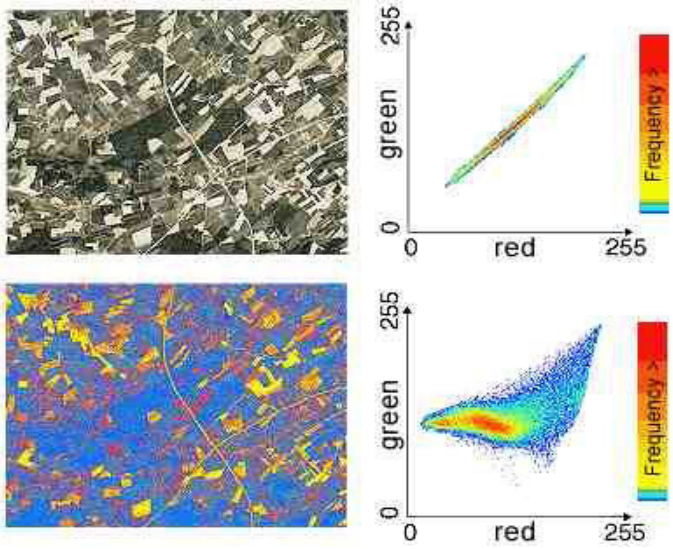


Fig. 6.5: A satellite image on the top where the bands are highly correlated. The bottom image is the FCC of the first three principal components of the same image. The scatter plot of two components projected as red and green bands show that they are highly uncorrelated.

To perform the principal component analysis, two bands of the input satellite image are first plotted in a scatter plot and an ellipse is constructed that binds the points (Fig. 6.6). The major axis of the ellipse is rotated then to make it the new x- axis and it becomes the first principal component (PC1). The PC1 shows the greatest variation in the data since it is the largest transect across the ellipse. The perpendicular to the PC1 is the second principal component (PC2) and the new axis for the original y- axis. It shows the greatest variance is not shown by the PC1 (Fig. 6.7). Any transact across the ellipse would form the other principal components but would carry decreasing amount of non-redundant information.

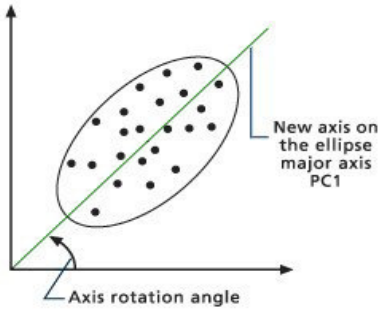


Fig. 6.6: The rotated axis of the scatter plot which is the Principal Component 1.

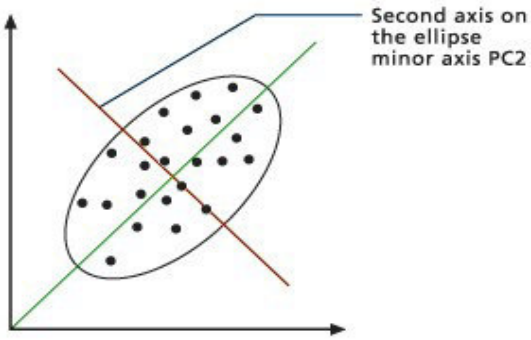


Fig. 6.7: PC2 of the scatter plot



CHECK YOUR PROGRESS

Q. 6: Fill in the blanks:

a) _____ helps prepare images for segmentation and/or analysis.

b) Spatial filtering essentially requires the use of a mask of different sizes known as a _____.

c) Smoothing filters are used for _____ and noise reduction in the image.

d) A sharpening filter is also known as a _____ filter.

e) Low pass filters _____ high frequencies.

f) NDVI stands for _____.

g) The equation for NDVI ratioing is _____.

h) Clay minerals show high absorption in the LANDSAT band no. _____.

i) Image filtering is a _____ operation.

j) The array of a high pass filter contains _____ value at the centre.

Q. 7: State true or false:

(a) Spectral ratio is a spectral operation involving values in a single band.

(b) Variations in scene illumination as a result of topographic effects are reduced after band ratio.

(c) NDVI uses pixel values from Red and Infrared bands for highlighting vegetation.

(d) PCA is used to increase correlation between bands of an image.

(e) PCA results in fewer number of bands.



6.6 LET US SUM UP

- The GCPs are generally defined by some permanent features on the surface of the earth such as road junctions that can viewed in both target satellite images and maps.
- The error in geo-referencing is calculated by a measure of the residual error given by the general formula applied to the control point locations.
- The image geo-referencing requires a co-ordinate system, a reference ellipsoid and datum.
- An image after geo-referencing with the help of reference ellipsoid and datum can be transformed into a map projection system of the choice of user.
- Pattern Recognition is the key for interpretation and information extraction from Remote Sensing images.
- Image enhancement is the process of adjusting digital images so that the results are more suitable for display or for further image analysis.

- The techniques involved can be categorised into point, local and spectral operations.
- Filtering is the technique to highlight or suppress the specific features in an image based on their spatial frequency.
- Spatial filtering essentially requires the use of a mask known as a kernel. This mask is moved on the image such that the centre of the mask traverses all image pixels.
- Image filters can be classified into two types based on their function: Smoothing filters and Sharpening filters.
- Smoothing filters are used for blurring and noise reduction in the image. Low pass filters are mostly used for smoothing.
- Sharpening filters enhance details in images by applying a high pass filter (edge enhancement/edge detection).
- Band ratioing or spectral ratioing is the most common transforms applied to images which is a spectral operation involving two or more bands of data.



6.7 FURTHER READING

1) James B. Campbell and Randolph H. Wynne, *Introduction to Remote Sensing,* Published by the Guilford Press.
2) Paul M. Mather*, Computer Processing of Remotely-Sensed Images: An Introduction,* Published by Wiley-Blackwell.
3) Rafael C. Gonzalez and Richard E. Woods**,** *Digital image processing,* Published by Prentice Hall.
4) Thomas Lilesand, Ralph W. Kiefer and Jonathan Chipman, *Remote sensing and image* interpretation, Published by *Willey.*



6.8 ANSWERS TO CHECK YOUR PROGRESS

Q. No. 1: When we work with medium resolution or old satellite images, scanned maps or aerial photos were required to assign them the earth's

coordinate system. This operation is also known as image geo-referencing.

Q. No. 2: The ground control points are generally defined by some permanent features on the surface of the earth such as road junctions that can viewed in both target satellite images and maps.

Q. No. 3: The error in geo-referencing can be reduced by (i) removing GCPs with very high residual error; (ii) slightly displacing the control points with high residual errors, and (iii) by adding additional control points to adjust the errors.

Q. No. 4: Pattern Recognition is the key for interpretation and information extraction from Remote Sensing images.

Q. No. 5: The nine elements for image interpretation are tone, shape, size, texture, pattern, shadow, height, site and association.

Q. No. 6:

(a) Filtering
(b) kernel
(c) blurring
(d) derivative
(e) suppress
(f) Normal Difference Vegetation Index
(g) (Nir-Red)/(NIR+Red)
(h) Band 7
(i) neighbourhood
(j) positive or negative

Q. No. 7:

(a) False
(b) True
(c) True
(d) False
(e) True



6.9 MODEL QUESTIONS

Q. 1: What do you mean by image geo-referencing? How is it done?

Q. 2: What are the nine elements of image interpretation? Define each of them.

Q. 3: What is image enhancement? Briefly describe the different techniques involved in image enhancement.

Q. 4: What is mean by image filtering? What are the major types of image filters?

Q. 5: What do you mean by band ratio? Give examples of common band ratios used in remote sensing.

Q. 6: What is Principal Component Analysis? For what reason principal component analysis (PCA) is done on multispectral satellite images?

Q. 7: Write briefly on the working principle of Principal Component Analysis (PCA).

---×---

UNIT 7: BASICS OF IMAGE CLASSIFICATION

UNIT STRUCTURE

7.1 Learning Objectives

7.2 Introduction

7.3 Introduction to Image Classification

7.4 Unsupervised Classification

7.5 Supervised Classification

7.6 Let us Sum up

7.7 Further Reading

7.8 Answers to Check Your Progress

7.9 Model Questions

7.1 LEARNING OBJECTIVES

After going through this unit, you will be able to –

- explain the basics of image classification
- perform unsupervised classification of digital image
- describe supervised classification of digital image

7.2 INTRODUCTION

In the previous unit, you were introduced to the concept of data processing and image enhancement techniques. In this unit you will gain an idea on image classification. This unit shall hence provide you an introductory explation of image classification covering both the areas of unsupervised classification and supervised classification of digital images.

7.3 INTRODUCTION TO IMAGE CLASSIFICATION

The image classification methods are some of the most important techniques of image processing and analysis. In the context of image classification, an image consists of pixels, each of which has a definite spectral characteristic. The pixels that represent similar features of the surface of the earth usually have similar spectral characteristics, based on

which they can be categorised into information classes. The output of the image classification is a thematic map where the information classes represent land use or land cover themes. According to Lillesand and Keifer (1994), image classification is defined as the process of categorising all pixels in an image or raw remotely sensed satellite data to obtain a given set of labels or land cover themes. Digital image classification utilises spectral information of spectral bands of multispectral images. The spectral information is represented by digital numbers or gray level or tone of the pixels. Therefore, the digital image classification is also known as spectral pattern recognition.

It is to be noted that image classification process should be attempted after meeting the needs of correction of the image on the following aspects:

a) Radiometric corrections
b) Geometric corrections
c) Enhancement of the image.

These are the preparation steps for the image.

Land cover refers to the physical and biological cover over the surface of the earth, which includes vegetation, water, land, snow field, wasteland, etc. The term 'land use' is rather complicated and it includes human intervention on the surface of the earth. Typical land use classes include agricultural land, plantation, urban settlements, rural settlements, etc. However, in practice both the terms are quite amalgamated, and a holistic land use/land cover (LULC) mapping would include identification of both natural and man-made covers of the surface of the earth for the same thematic map.

Image classification techniques can be manual or computer based. **Manual or visual image classification** can be done either over hard copy images or in image processing software, by identifying the land use or land cover classes with eyes.

Some tools such as magnifier, colour viewer or analyser may be utilised to recognise the pattern exhibited by a group of pixel for a class of land use/ land cover. While using Digital Image Processing (DIP) tools for visual interpretation, a group of pixels indicating a particular land use/ land

cover class, may be mapped by vector polygons. This process is time consuming and very tedious.

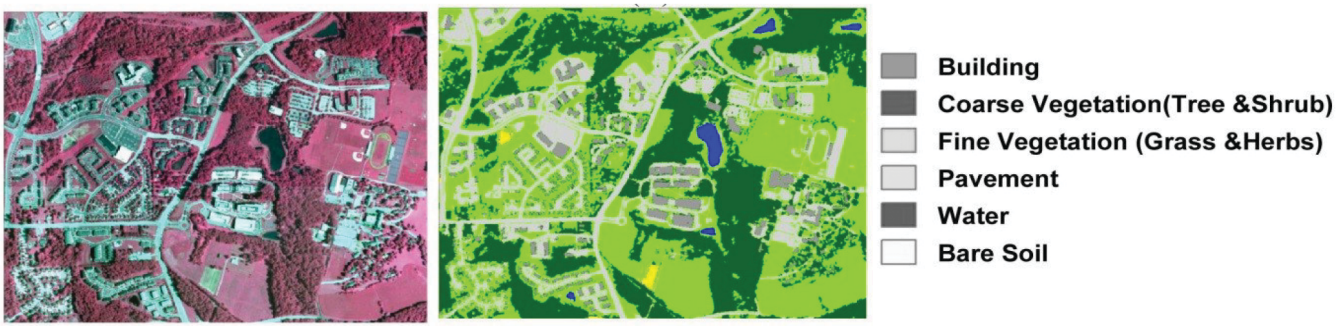
The identification of land use or land cover classes in manual method is based on different image elements such as tone, texture, shape, size, pattern, shadow, etc.

In **computer based image classification** methods, the spectral classes are identified by the algorithms of image processing software and they are essentially based on DN values or brightness values. Common classification methods are grouped into two broad subdivisions, i.e., supervised and unsupervised image classifications.

In the unsupervised image classification, the spectral classes are clustered first based solely on spectral characteristics of the pixels and then matched by the analyst with the real world land use/land cover classes. The analyst usually specifies how many spectral classes or clusters are to be identified in the beginning of the procedure. In some instances, a large number of land use/land cover classes may be specified by the user. Some clusters may have the same characteristics with another cluster. In such cases those spectral classes have to be merged into one spectral class. Each spectral classes are coded with new values and class-names can be assigned at a later stage.

In the supervised classification, the analyst first identifies the representative samples of information classes with definite spectral characteristics which are also known as training classes/sets/samples. The identification of training classes requires prior knowledge about the land use or land cover pattern within the geographical area of the area of interest. Thus, the analyst 'supervises' the procedure of this particular image classification. After proper identification of the training classes, the image classification software categorises all the pixels of the image into spectrally distinct classes which resemble spectral characteristics of the training classes in the rest of the image also.

Fig. 7.1 shows a raw image (a) and the classified image (b) with the index of land use/land cover classes.



(a) Satellite image (b) Classified image

Fig. 7.1: An example of land use and land cover classification (Zhou et al, 2014)



CHECK YOUR PROGRESS

Q. 1: What is image classification?

.....

.....

Q. 2: What are the two common image classification methods?

.....

.....

7.4 UNSUPERVISED CLASSIFICATION

Unsupervised image classification is a method in which the image interpreting software separates a large number of unknown pixels in an image based on their reflectance values into classes or clusters with no direction from the analyst (Tou, Gonzalez 1974). This method is purely based on spectral characteristics of pixel and no prior knowledge of the characteristics of the land use and land cover classes of the study is required. There are two clustering algorithms utilised for unsupervised classification, which rely essentially on spectral characteristics of pixels. These two algorithms are known as K-means and ISODATA (Iterative Self-organizing Data Analysis technique).

Both of these algorithms are iterative procedures. At first both of the methods first assign arbitrary cluster properties as initial values which are evenly distributed in the data space. In the second step, all the pixels of the image are classified to the closest cluster. In the third step, the mean cluster properties are selected based on pixel properties of each cluster. After this the second and third steps are iterated again and again until the change in clustering of pixels is minor. This is measured either from (i) the distances of the mean cluster properties from the counter part of the previous iteration or by (ii) the difference in percentages of total pixels in each cluster between iterations.

The ISODATA algorithm adds some refinement by splitting and merging the clusters. Merging of clusters is based on two thresholds which are defined by minimum number of pixels in each cluster and minimum distance between the mean values of the clusters. A cluster is split into two when the standard deviation of the cluster exceeds a predefined value or total number of member pixels in that particular cluster is twice the threshold for the minimum number of pixels.

The K-means algorithm differs from the ISODATA in the fact that the number of clusters remains the same throughout the iteration.

The output clusters of the unsupervised classification are then visually correlated with the real world land use/land cover classes by ground truthing. Sometimes, a single land use/land cover class may be represented by more than one spectral cluster in the output thematic map. In that case, these spectral clusters are grouped into a single one.

The chief advantages of the unsupervised classification are:

(i) No detailed prior knowledge about the area of interest is required.

(ii) Clustering is done by computer based algorithms, thus, chances of human error are less.

(iii) In case of supervised classification, classification is based on identification of training classes. Therefore, minor land use/land cover classes may be left unrecognised. In unsupervised image classification distinct spectral classes, be it minor, never left unrecognised.

Despite these advantages of unsupervised classification over supervised classification, it also bears some disadvantages, notably:

(i) Unsupervised image classification algorithms may recognise distinct spectral classes within the same information class, because of spatial variation in spectral characters of the same information class caused by illumination differences.

(ii) The relation between spectral classes and information classes is not constant and change over space (location) and time (seasonal variation). Therefore, this relationship varies from image to image.

(iii) Identification of spectral classes in the output of the unsupervised classification may be very time consuming.

The major steps involved in the unsupervised image classification method is shown in Fig. 7.2 below.

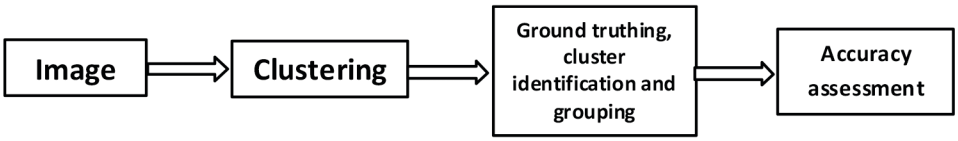


Fig. 7.2 Major steps involved in unsupervised image classification.



CHECK YOUR PROGRESS

Q.3: State true or false.

(a) The main purpose of digital image classification is to identify land use/land cover classes of the surface of the earth.

(b) For digital image classification all image elements are usually taken into consideration.

(c) The prior knowledge of the analyst about the land use/land cover pattern of the area of interest is prior requirement for unsupervised classification.

(d) The algorithms of unsupervised classification are ISODATA and K-means.

(e) Image classification is usually done with multispectral satellite images.

Q. 4: What are the two clustering algorithms used in unsupervised classification?

.....

.....

7.5 SUPERVISED CLASSIFICATION

In supervised image classification, the user first identifies representative samples of the land use/land cover classes over the digital image. These representative classes are known as 'training classes'. To select the training classes over the image, ground truth data with positional accuracy is a prime requisite in this classification. The image processing software performs a statistical characterisation of spectral reflectance of the each training class over the spectral bands of the digital image. This stage is known as **signature analysis** and the statistical characterisation of the reflectance values of the training classes involves determination of mean, variance and covariance over the spectral bands of the digital image. The image is then classified by examining the spectral reflectance of each pixel and aligning the pixel to the training class to which it resembles the most. Since, the user's supervision is necessary in the initial stage of the classification scheme in identifying the training classes, this classification is known as **supervised classification**.

Sometimes, spectral reflectance of some pixels of the digital image may not resemble those of the training classes. The analyst can leave an option not to classify such pixels and these pixels remain as unclassified pixels in the output image.

The classification scheme and steps of the supervised image classification is shown in Fig. 7.3.

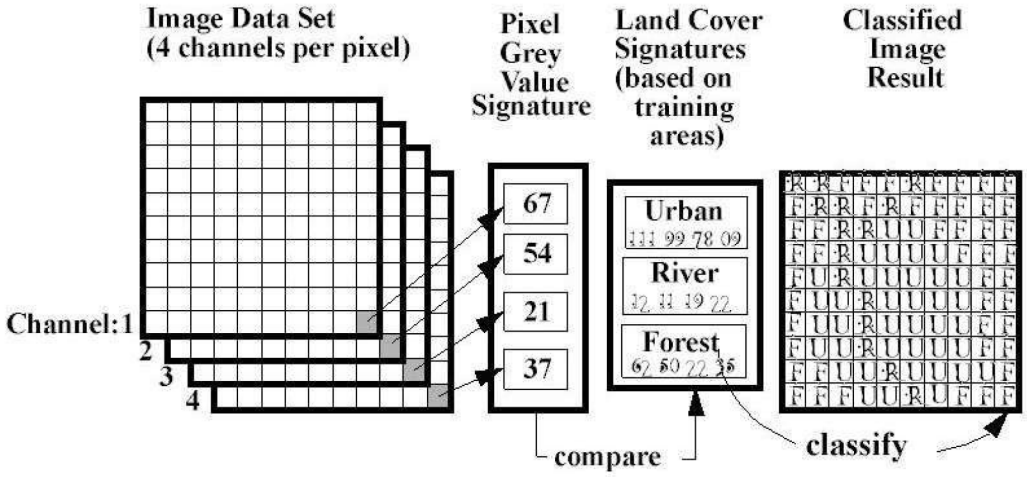


Fig. 7.3: Steps in supervised classification (Courtesy: http:// www.sc.chula.ac.th /courseware/ 2309507/ Lecture/ remote18.htm)

There are three statistical methods used for statistical characterisation of the spectral reflectance of the training classes. These methods are minimum distance, parallelepiped and maximum likelihood methods.

Minimum Distance Method:

This is the simplest statistical method of supervised image classification. The minimum distance method determines the mean DN value (spectral reflectance or pixel brightness) of each training class over each spectral band of the digital image. In the second step, it assigns the unclassified pixels of the image to the training class for which it has the minimum Euclidean distance in terms of spectral reflectance. Thus, all pixels of the image are assigned to the nearest training classes. The analyst can designate a minimum distance threshold for the classification. Pixels which fall beyond the threshold minimum distance to the training classes would be left as unclassified. This classification method does not require high processing power of the computer and was very popular few decades back when computers were not very efficient.

This method is still being a good option when analyst requires classifying a large and high resolution digital image within a short time.

Parallelopiped Method:

This parallelopiped classifier is little more complex than the minimum distance classifier. In this classification, the unclassified pixels of the image are assigned to a training class, when their brightness values fall within a range of the training mean brightness value. A standard deviation from the mean selects the minimum and maximum value of the class range of each training class. However, sometimes the parallelopiped classifier creates training class overlap, which causes misclassification of the some pixels of the image.

Maximum Likelihood Method:

It is the most modern and popular classification algorithm of supervised classification, but computationally complex. The classifier first determines the variance and covariance about the mean of the training classes. Based on variance and covariance, the classifier, then, determines the Gaussian probability of the each unknown pixel and assigns it to the class for which it has the highest probability.



CHECK YOUR PROGRESS

Q. 5: What is meant by training classes?

.....

.....

Q.6: What are the statistical methods used for statistical characterisation of the spectral reflectance of the training classes?

.....

.....



7.6 LET US SUM UP

- An image consists of pixels, each of which has a definite spectral characteristic which is represented by DN value.

- Image classification is defined as the process of categorising all pixels in an image or raw remotely sensed satellite data to obtain a given set of labels or land cover themes.
- Land cover refers to the physical and biological cover over the surface of the earth, which includes vegetation, water, land, snow field, wasteland, etc.
- The term 'land use' is rather complicated and it includes human intervention on the surface of the earth.
- The two common digital image classification methods are supervised and unsupervised image classifications.
- In the unsupervised image classification, the analyst usually specifies how many spectral classes or clusters are to be identified in the beginning of the procedure.
- In the supervised classification, the analyst first identifies the representative samples of information classes with definite spectral characteristics which are also known as training classes.
- Two clustering algorithms used for unsupervised classification are K-means and ISODATA.
- Three statistical methods used for statistical characterisation of the spectral reflectance of the training classes are minimum distance, parallelopiped and maximum likelihood methods.



7.7 FURTHER READING

1) James B. Campbell and Randolph H. Wynne, *Introduction to Remote Sensing,* The Guilford Press.
2) Paul M. Mather, *Computer Processing of Remotely-Sensed Images An Introduction,* Wiley-Blackwell.
3) Rafael C. Gonzalez and Richard E. Woods, *Digital image processing,* Prentice Hall.
4) Thomas Lilesand, Ralph W. Kiefer and Jonathan Chipman, *Remote sensing and image interpretation*, Willey

5) US Army Corps of Engineers, *Remote Sensing: Engineer Manual,* EM-1110-2-2907.



7.8 ANSWERS TO CHECK YOUR PROGRESS

Q. No. 1: Image classification is defined as the process of categorising all pixels in an image or raw remotely sensed satellite data to obtain a given set of labels or land cover themes.

Q. No. 2: Common image classification methods are supervised classification and unsupervised image classifications.

Ans. to Q. No. 3: (a) True (b) False (c) False (d) True (e) True

Q. No. 4: There are two clustering algorithms utilised for unsupervised classification which are known as K-means and ISODATA (Iterative self-organising data analysis technique).

Q. No. 5: In supervised image classification, the user first identifies representative samples of the land use or land cover classes over the digital image. These representative classes are known as 'training classes'.

Q. No. 6: The three statistical methods used for statistical characterisation of the spectral reflectance of the training classes are minimum distance, parallelopiped and maximum likelihood methods.



7.9 MODEL QUESTIONS

Q.1: What are meant by the terms 'land use' and 'land cover'? Discuss the differences between them.

Q.2: What is digital image classification? What are the bases of classification of digital images?

Q.3: Discuss the methodological differences between unsupervised and supervised classification.

Q.4: Discuss the K-means and ISODATA methods of unsupervised classification.

Q.5: Discuss the advantages and disadvantages of the unsupervised image classification.

Q.6: What are the classification algorithms of supervised classification? Describe briefly.

Q.7: Discuss the advantages and disadvantages of supervised classification.

Q.8: Discuss the role of ground truthing in digital image classification.

---×---

UNIT 8: ACCURACY ASSESSMENT IN IMAGE CLASSIFICATION

UNIT STRUCTURE

8.1 Learning Objectives

8.2 Introduction

8.3 Concept and need of Accuracy Assessment

8.4 Types of Accuracy

8.5 Assessment of Classification Accuracy

8.6 Let us Sum up

8.7 Further Reading

8.8 Answers to Check Your Progress

8.9 Model Questions

8.1 LEARNING OBJECTIVES

After reading this unit, you will be able to-

- explain the concept of Accuracy Assessment
- describe the types of accuracies
- discuss the concept of classification accuracy

8.2 INTRODUCTION

In the previous unit, we have discussed the concepts of Image classification techniques (both supervised & unsupervised classification techniques) which help us to create thematic maps. This unit deals with the accuracy assessment in image classification. It also provides the types of accuracy assessment. Different methods to assess classification accuracy are also discussed in this unit. Assessment of classification and how the accuracy assessment expressed in terms of the accuracies of the User and Producer is obtained from the Error matrix is discussed at the end of this unit.

8.3 CONCEPT AND NEED OF ACCURACY ASSESSMENT

During the image classification using Digital Image Processing techniques, if we can classify imagery that is familiar to us with the area it represents is quite easier than the area which is not familiar to us. Suppose we are going to classify imagery that is not familiar with us than in that case how would we classify the image of that area? If we somehow try to interpret the satellite imagery, it would little difficult to ensure that the information derived from the imagery is correct and truly represent the features of the surface. In such cases, we require an accurate technique to judge our interpretation or classification result which is called Accuracy Assessment.

Definition:

Accuracy assessment is the end and final step in the image analysis technique which determines the quality of information derived from remotely sensed data. According to Campbell, 2007 Accuracy "measures the agreement between a standard assumed to be correct and a classified image of unknown quality." It is carried out once the image interpretation/ image classification procedure has been completed. The accuracy is concerned with the correspondence between class labels (putting during image classification) and 'true' class or information that exist in the remotely sensed data.

The Accuracy Assessment can be either qualitative or quantitative. The qualitative is usually a quick comparison technique that determines if a remotely sensed data or map 'looks right' by comparing what we see in the image or map with what we see on the ground. However, in quantitative assessments efforts are made to identify and quantify remote sensing-based map errors. In such assessments, we compare map data with the reference or ground truth data collected from the field, which is assumed to be 100% correct.

In order to perform accuracy assessment correctly, we need to compare two sources of information which include:

- The classified image derived from the satellite imagery (remote sensing data) and
- Ground truth data collected from the field (in the form of GPS coordinates). Ground truth can be collected only in the field; however, the process is time-consuming and expensive. Ground truth data can also be derived from interpreting high-resolution imagery, existing classified imagery or GIS data layers.

8.4 TYPE OF ACCURACY

There are different ways to represent the accuracy assessment. One basic accuracy measurement is the overall accuracy assessment technique which is calculated by dividing the correctly classified pixels by the total number of pixels checked.

Over all accuracy (%) = (Correctly classified pixels / Total number pixels)…… (1)

Example:-

Classification	Reference				
		Very Dense Forest	Dense Forest	Open Forest	Total
	Very Dense Forest	28	14	15	57
	Dense Forest	1	15	5	21
	Open Forest	1	1	20	22
	Total	30	30	30	100

Overall Accuracy = the sum of the diagonal value (28+15+20 = 63) divided by the grand total. (Expressed as a percentage)

Overall accuracy = 63/100 = 63%.

The overall accuracy assessment does not represent how the accuracy is distributed across the individual categories. Every individual category/class could and frequently exhibit drastically differing accuracies. But this method considers that all the categories having equivalent or similar accuracies.

Besides the overall accuracy assessment technique, classification accuracy assessment of individual classes of the particular LULC (Landuse / landcover) classification can be measured in a similar technique. Two numbers of techniques/approaches are widely applied.

One is the User's accuracy and the other one is the Producer's accuracy.

(1) User's accuracy is a map-based accuracy which is computed by looking at the reference data for a particular category/class and determining the percentage of correct predictions for that samples. For example, if I select any water pixel on the classified map, what is the probability that we will be standing in water during field verification? It computed using the number of correctly classified pixels to the total number of pixels assigned to a particular category. It takes errors of the commission into account by telling the user's that, for all areas identified as category A, a certain percentage are actually correct.

User's accuracy (%) = (Correctly classified pixels / Classified total pixels)…… (2)

Example:

Classification	Reference				
		Very Dense Forest	Dense Forest	Open Forest	Total
	Very Dense Forest	28	14	15	57
	Dense Forest	1	15	5	21
	Open Forest	1	1	20	22
	Total	30	30	30	100

User's Accuracy = the diagonal value divided by the row total.

For example, the Very dense Forest Category:

Very Dense Forest = 28/57 = 49%.

(2) The producer's accuracy is a reference-based accuracy which is computed by looking at the predictions produced for a particular category /class and determining the percentage of correct

predictions. For example, if I know that a particular area is a water, what is the probability of a classified map that will correctly identify that pixel as water? It is computed using the number of correctly classified pixels to the reference total pixels for a particular category. The producer's accuracy measures errors of omission.

Producer's accuracy (%) = (Correctly classified pixels / Reference total pixels)... (3)

Example:

Classification	Reference			
	Very Dense Forest	Dense Forest	Open Forest	Total
Very Dense Forest	28	14	15	57
Dense Forest	1	15	5	21
Open Forest	1	1	20	22
Total	30	30	30	100

Producer's Accuracy = the diagonal value divided by the column total.

For example, the Dense Forest Category:

Dense Forest = 15/30 = 50%.



CHECK YOUR PROGRESS

Q. 1: What is meant by accuracy assessment?

.....

.....

Q. 2: What are the various types of accuracies in image classification?

.....

.....

Q. 3: What is confusion matrix?

.....

.....

Q. 4: State 'True' or 'False regarding the following statements:

a. Accuracy assessment is a pre-classification technique.

b. For accuracy assessment analysis two data sets are required.
c. The confusion matrix is also known as error matrix.
d. Higher values of Kappa shows good agreement between the classified data and reference data.
e. The highest value of Kappa is 100.

8.5 ASSESSMENT OF CLASSIFICATION ACCURACY

The assessment of accuracy in image classification is done by several techniques. The most popular ones are:-

1. Error / Confusion matrix
2. Kappa Analysis

The efficiency of matching the classified result with the ground truth information is done by computing error matrix and Kappa coefficient.

Error Matrix:

Once the image classification exercise has been carried out, the classifier needs to determine the degree of errors that exist in the final classification product. Errors are the result of incorrect classification of the pixels for a particular category. Error matrix (similar name include confusion matrix, evaluation matrix, or a contingency table) is one of the most common means of expressing classification accuracy. Error matrix simply compare the relationship between known reference data (ground truth) and the corresponding results of the classification. Such matrix are a square array of rows and columns in which each row and column represent one category/ class in the classified map. Error matrix is useful for the calculation of overall errors for each category and misclassifications by category, as a result, it is also known as a *confusion matrix*. The advantage of a confusion matrix is that it not only recognizes the nature of the classification errors but also their quantities. The table below represents an example of an error/confusion matrix based on a classification result.

	Ground Reference Information							
Classification	**Class**	**Water**	**Forest**	**Agriculture**	**Built-up**	**Bare Soil**	**Row Total**	**User Accuracy (%)**
	Water	15	0	0	0	0	15	**100**
	Forest	1	33	8	0	2	44	**75**
	Agriculture	0	3	14	0	12	29	**48.8**
	Built-up	0	0	4	9	3	16	**56.25**
	Bare Soil	0	0	11	0	10	21	**47.62**
	Column Total	16	36	37	9	27	125	
	Producer Accuracy (%)	**93.75**	**93.75**	**93.75**	**93.75**	**93.75**		

Kappa Analysis:

Kappa analysis is a widely used image classification accuracy assessment method. It is a discrete multivariate statistical technique for accuracy assessment. Kappa coefficient is also known as Kappa hat or K-hat ($\hat{K}$). The Kappa coefficient value represents how well remotely sensed classification agrees or is accurate to the reference data. The Kappa coefficient value represents three groups of acceptance; a value greater than 0.80 represents strong agreement or good classification performance; a value between 0.40 to 0.80 represents moderate agreement or moderate classification performance, and a value less than 0.40 represents poor agreement or poor classification performance. It is defined as follows:

$$\textbf{K-hat} = (\text{Obs} - \text{exp})/(1 - \text{Exp})$$

Where,

Obs = Observed correct, it represents accuracy reported in the error matrix (overall accuracy).

Exp = Expected correct, it represents correct classification.



8.6 LET US SUM UP

- Assessing accuracy for each class or category derived from remote sensing data as well as for the whole image is essential to check the quality and reliability of the results obtained using various classification techniques.
- Accuracy of the result derived from image classification is affected because of errors of commission and errors of omission.
- Error matrix can be used for accuracy and reliability assessments.
- Error matrix expresses several characteristics about classification performance.
- Overall accuracy is a measure of accuracy for the whole image across all categories.
- Kappa analysis is a discrete multivariate statistical technique for accuracy assessment.



8.7 FURTHER READING

1 Paul M. Mather, *Computer Processing of Remotely-Sensed Images An Introduction,* Wiley-Blackwell.

2 Rafael C. Gonzalez and Richard E. Woods, *Digital image processing,* Prentice Hall.

3 US Army Corps of Engineers, *Remote Sensing: Engineer Manual,* EM-1110-2-2907.



8.8 ANSWERS TO CHECK YOUR PROGRESS

Q. No. 1: Accuracy assessment is the final part of image analysis by which we are in a position to tell how much we are correct in interpreting an object from the image. In such assessments, we compare map data with the reference or ground truth data collected from the field, which is assumed to be 100% correct.

Q. No. 2: There are three types of accuracies involved in image classification, whish are:-

1. Over all accuracy
2. User's accuracy
3. Producer's accuracy

Q. No. 3: Confusion matrix simply compares the relationship between known reference data (ground truth) and the corresponding results of the classification. The errors in classification for each class with reference to ground truth data is determined by this matrix.

Q. No. 4:

a. False
b. True
c. True
d. True
e. False



8.9 MODEL QUESTIONS

Q. 1: Why do we need accuracy assessment in image classification?

Q. 2: What is producer's accuracy?

Q. 3: Define over all accuracy.

Q. 4: What is the use of Kappa statistic?

---×---

UNIT- 9: GEOINFORMATICS FOR NATURAL RESOURCE ASSESSMENTS

UNIT STRUCTURE

9.1 Learning Objectives

9.2 Introduction

9.3 Application of Geoinformatics in Land Use and Land Cover Classification

9.4 Mapping and Change Detection Analysis

9.5 Geo-environmental Studies

9.6 Biotic Resource Assessments

9.7 Let us Sum up

9.8 Further Reading

9.9 Answers to Check Your Progress

9.10 Model Questions

9.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- explain the applications of geoinformatics in land use and land cover classification
- discuss the mapping and change detection analyse in GIS
- illustrate the applications of geoinformatics in geo-environmental studies
- describe applications of geoinformatics in biotic resources

9.2 INTRODUCTION

In the previous unit, you got an idea on accuracy assessment in image classification. This unit provides you the knowledge of applications of geoinformatics techniques in land use and land cover classifications. Change detection in watersheds and basic principles of change detection are also discussed in this unit. Applications of geoinformatics in geo-environmental studies like RS/GIS in forestry, assessment of biomass/

carbon, water resource management are also discussed in this unit. Biotic resource assessments are also a part of discussion in this unit.

9.3 APPLICATION OF GEOINFORMATICS IN LAND USE AND LAND COVER CLASSIFICATION

The earth is constantly under observations from dozens of satellites orbiting the planet and collecting image data of the earth's surface and its environment. The term 'land cover' and 'land use' (LULC) are often used interchangeably, their actual meanings are quite distinct. Land cover mapping is one among the foremost important and typical applications of remote sensing data. Land cover corresponds to the ground surface, for instance forest, grassland, open water etc., while land use reflects human activities like the use of the land for industrial zones, residential zones, agriculture fields, etc.

In more detail, land use refers to the purpose the land serves such as recreation, wildlife habitat or agriculture. It does not describe the surface cover on the ground. For example, a recreational land use could occur in a forest, shrub land, grasslands or on manicured lawns. Whereas land cover refers to the surface cover on the ground, whether vegetation, urban infrastructure, water, bare soil or other. It does not describe the use of land. The use of land may be different for lands with the same cover type. For example, a land cover type of forest may be used for timber production, wildlife management or recreation; it might be private land, a protected watershed or a popular state park.

High-resolution images have often been employed in transportation studies and might be able to identify vehicle types, estimate traffic flows, identify parking problems on city streets, estimate car parking zone usage and even to measure the speed of vehicles on a highway. New construction can be identified on an aerial photograph or on a very high resolution satellite image and permit records can be checked to verify that a building permit was issued for the project. This type of application requires large - scale imagery like 1:5,000.

Remote sensing images have been used to aid in locating businesses or public facilities such as schools, fire stations or libraries. By specifying a collection of criteria that represent desirable locations for the business or public facility, image interpretation are often used to identify sites that satisfy project requirements. Remote sensing methods are often employed to classify styles of land use over large areas in an exceedingly practical, economical and repetitive fashion.



CHECK YOUR PROGRESS

Q. 1: What is the difference between land use and land cover?

.....

.....

9.4 MAPPING AND CHANGE DETECTION ANALYSIS

The visual and digital remote sensing techniques have been extensively used by several organisations in India for surveying, classification and mapping of land use in conjunction with ground truth. Colour infra-red photography (CIR) and multi-band aerial photography are proved most significant tools for detailed land use mapping. The land use mapping, by using remote sensing techniques acquired a ***quasi-operational*** status in the country has now achieved operational status due to high resolution products available from IRS-LISS II, III, PAN and Landsat TM sensors.

Change detection in GIS is a method of understanding how a given area has changed between two or more time periods. Change detection is helpful for understanding the change in forest coverage, ice sheets, and land use. Change detection often involves comparing aerial photographs or satellite imagery of a particular area taken at different times. The change is usually detected by comparison between two multi-date images or sometimes between an old map and an updated remote sensing image. Change detection has been widely used to assess shifting cultivation, deforestation, urban growth, impact of natural disasters like

tsunamis, earthquakes, and use/land cover changes, etc. The method of change detection is divided into the following two categories:

1. Comparison between two land-cover maps which are independently produced.
2. Change enhancement by integrating two images into a colour composite or principal component image.



CHECK YOUR PROGRESS

Q. 2: What is meant by change detection?

.....

.....

Sensors operating in the visible and infrared portion of the spectrum are the most useful data for change detection. Many urban features can be detected on radar and other imagery. Visible and near-infrared data at a high resolution permits fine distinction among land-cover/use classes. Optical imagery acquired during the months of winter is also useful for roughly delineating urban areas versus non-urban areas. Cities appear in dramatic contrast to smooth-texture snow-covered fields. Radar sensors even have some use for all urban/rural delineation applications because of the flexibility of the imaging geometry to boost anthropogenic features like buildings within the manner of corner reflectors.

9.5 GEO-ENVIRONMENTAL STUDIES

Application of RS/GIS in Forestry

Forests are an important global resource that human populations depend on for wood, fuel, air quality, recreation and many other uses. They also serve as habitats for ample plant and animal species. Monitoring the health of forests is crucial for sustainability and conservation issues, because of versatility and scale of remote sensing, it's invaluable in all stages of forest management. Remote sensing may be a great tool for

assessment of environmental conditions either in an existing forest or before planning a replacement one. Remote sensing brings together a large number of tools to higher analyse the scope and deforestation problem. Multi-temporal data provides for change in detection analyses. Images of earlier years are compared to recent scenes to measure the differences within the sizes and extents of the clear-cut or loss of forest. Data from different sources are used to provide complementary information. Radar images, merged with optical data will be accustomed efficiently to monitor the status of existing clear cuts or emergence of recent ones and even assesses regeneration condition. High- resolution data provide an in-depth view of forest depletion, while radar can provide a view which will obscure by clouds.

All remote sensing devices provide a view of remote and inaccessible areas, where illegal cutting or damage could continue unnoticed for long periods of time if aerial surveillances were not possible.

Role of Remote Sensing for Assessment of Biomass/Carbon

Remote sensing techniques allow scientists to examine properties and processes of ecosystems and their inter-annual variability at multiple scales because satellite observations can be obtained over large areas of interest with high revalidation frequencies.

Many studies have demonstrated that indices like Spectral Vegetation index (SVI), Simple Ratio (SR), Normalized Difference Vegetation Index (NDVI), and Corrected Normalized Difference Vegetation Index (NDVI) obtained from satellite data are useful predictors of Leaf Area Index (LAI), biomass and productivity in grasslands and forests. Stand level biomass is frequently calculated from linear and non linear regression models established by species with field measurements.

Water resource management

Remote Sensing, Geographic Information System (GIS) and Global Positioning System (GPS) can play important role in water resources management, such as surface water, groundwater investigation, dynamic monitoring of ecology and estimation of water amount necessary for keeping and recovering ecological environment, existing irrigation area

investigation and irrigation planning, soil moisture and drought monitoring, investigation of soil salinization, planning, monitoring and effect evaluation of returning cultivated and to forest or grassland, dynamic monitoring of desertification and soil erosion, variation of river course and sedimentation in lakes and reservoirs, site selection of water project and its planning, design, construction and management. Water resources investigation, discharge in river channel are often accurately controlled by hydrological measurement, while the world of reservoir and lake are often determined by remote sensing. On the basic of this idea, water storage in lake and reservoir could also be determined by means of stage area volume curve.

This kind of curve can also be worked out on the basis of multitemporal (flood, middle and dry periods) remote sensing images. Groundwater is the most important reproduced natural resources, especially for livelihood, animal husbandry and agriculture in arid regions. Remote sensing can provides the information about geology, hydrogeology geomorphology and urban environment analysis. They are helpful for searching groundwater, provides clue for field investigation and improve successful possibility. For finding groundwater, the penetration of radar is useful to directly find shallow layer groundwater within the places with ancient river channel and also the plain area ahead of mountains. Remote sensing has the flexibility for the observation in these aspects. With the advantage of high temporal resolution of meteorological satellites, it is possible to differentiate cloud and snow cover because of the movement of cloud.

Geoinformatics in environmental studies

In order to plan and monitor the environmental problems, the assessment of hazards and risks becomes important for planning, decisions and for mitigation activities. GIS supports activities in environmental assessment, monitoring, and mitigation and may even be used for generating Environmental models. Below are a number of the applicable areas where GIS are often implemented for effective planning and management.

GIS applicable Areas:

- Wild Land Analysis
- Emergency Services like Fire Prevention
- Hazard Mitigation and Future planning
- Air pollution and control
- Disaster Management
- Forest Fires Management
- Managing Natural Resources
- Waste Water Management
- Oil Spills and its remedial actions
- Sea Water – water interface Studies
- Coal Mine Fires

Apart from data analysis in laboratories GIS also help the environmental data analysts within the field. The GIS tool is flexible enough to figure within the field to administer the precise location of devastation and amount of devastation.

Some examples in field where GIS is applicable are:

1. Using GIS within the field, an environmental inspector can rapidly map waste storage sites; describe the degree, content, and state of waste containers.
2. Retrieve previous inspection records to match with the prevailing environmental conditions.
3. View environmental data in relevancy adjacent geographic features like waterways, neighbour hoods, or other sensitive areas like high-risk zones for landslides, pollution, etc.
4. Integrated with a world positioning receiver, a field crew can use GIS to accurately ground truth satellite imagery in oil spill mapping and its affects on surrounding ecosystem.

9.6 BIOTIC RESOURCE ASSESSMENTS

Biotic components are the living things that shape an ecosystem. Examples of biotic components include animals, plants, fungi, and bacteria. Abiotic components are non-living components that influence an

ecosystem. Examples of abiotic factors are temperature, air currents, and minerals.

Many ecological research projects inherently have a spatial component that is worth exploring. This will range from the distribution of sampling sites to survey tracks, capture locations, movements of individual animals and knowledge about the distribution of specific habitat types. For example, plotting the spatial distribution of sampling sites may help to work out whether sites which are closer to every other are more similar than those which are further apart. Similarly, by plotting the locations where a selected species of animal has been recorded over different habitat variables, we are able to start to develop ideas about what's important for determining the distribution of that species.



CHECK YOUR PROGRESS

Q. 3: Differentiate between the biotic and abiotic components.

.....

.....



9.7 LET US SUM UP

- Land cover corresponds to the physical condition of the ground surface for example forest, grassland etc., while land use reflects human activities such as the use of the land, for example industrial zones, residential zones, agriculture fields etc.
- Biomass estimates at the landscape level are necessary for understanding processes of the target landscapes and provide baseline data for future studies. Models derived from remote sensing need further calibration with ground data before they can be used appropriately to predict AGB for a given landscape.
- Geoinformatics supports activities in environmental assessment, monitoring, and mitigation and can also be used for generating Environmental models.

- Geoinformatics is suited as a tool for the collection, collation, analysis and visualisation of data derived from diverse sources including those from distributed measurement stations (e.g. field-based water quality sensors).



9.8 FURTHER READING

- Agarwal, C. S and Garg, P. K. 2000: *Text book on Remote Sensing in Natural Resource Monitoring and Management* – A. H. Wheeler, New Delhi-1.
- Gautam, N. C., Raghavswamy, V. And Nagaraja, R. (Chief Editor) 1994: *Space Technology and Geography*, NRSC, Hyderabad.
- Guha, P.K. 2013: *Remote Sensing for the Beginner,* Affiliate East West Publishers, New Delhi.
- Heywood, Ian, Cornelius, S. 2011 : *An Introduction to GIS 4th Edition,* Pearson, New Delhi
- Jensen, J. R. 2011: *Remote Sensing of the Environment – An Earth Resource Perspective*: Pearson Education., New Delhi.
- Joseph, G. 2005*: Introduction to Remote Sensing,* Universities Press (India) Pvt. Ltd, Hyderabad.



9.9 ANSWERS TO CHECK YOUR PROGRESS

Q. No. 1: *Land use* refers to the purpose the land serves, for example, recreation, wildlife habitat or agriculture; it does not describe the surface cover on the ground. For example, a recreational land use could occur in a forest, shrub land, grasslands or on manicured lawns.

Land cover refers to the surface cover on the ground, whether vegetation, urban infrastructure, water, bare soil or other; it does not describe the use of land, and the use of land may be different for lands with the same cover type. For example, a land cover type of forest may be used for timber

production, wildlife management or recreation; it might be private land, a protected watershed or a popular state park.

Q. No. 2: Change detection in GIS is a method of understanding how a given area has changed between two or more time periods. Change detection is helpful for understanding the change in forest coverage, ice sheets, and land use.

Q. No. 3: Biotic components are the living things that shape an ecosystem. Examples of biotic components include animals, plants, fungi, and bacteria. Abiotic components are non-living components that influence an ecosystem. Examples of abiotic factors are temperature, air currents, and minerals.



9.10 MODEL QUESTIONS

Q. 1: Explain the role of remote sensing to monitor land-use change.

Q. 2: Explain the application of remote sensing in hydrological sciences.

Q. 3: What are the utilities of remote sensing in mapping science?

Q. 4: Explain the use of remote sensing data for geomorphological and soil mapping in the context of geological mapping.

Q.5: Discuss the application of geoinformatics in biotic resource assessment.

---×---

UNIT-10: GEOINFORMATICS FOR ENVIRONMENTAL STUDIES, PLANNING AND DEVELOPMENT

UNIT STRUCTURE

10.1 Learning Objectives

10.2 Introduction

10.3 Geoinformatics for environmental pollution and climate change studies

10.4 Floods and drought prone area mapping

10.5 Early warning systems for disasters and hazards

10.6 Rural and urban infrastructure survey

10.7 Mapping and planning

10.8 Let us Sum up

10.9 Further Reading

10.10 Answers to Check Your Progress

10.11 Model Questions

10.1 LEARNING OBJECTIVES

After going through this unit, you will be able to –

- explain the application of geoinformatics in mitigating environmental pollution and climate change.
- describe the mapping of floods and drought prone areas.
- explain the application of early warning systems for disasters and hazards.
- describe application of geoinformatics in rural and urban infrastructure survey, mapping and planning.

10.2 INTRODUCTION

In previous units you know about application of geoinformatics in natural resource assessments. This unit provides you the knowledge of application of geoinformatics for environmental pollution and climate change studies. Floods and drought prone area mapping are also discussed in this

unit. Application of geo-informatics in early warning systems for disasters and hazards, rural and urban infrastructures survey are also discussed in this unit. Mapping and planning are also a part of discussion in this unit.

10.3 GEOINFORMATICS FOR ENVIRONMENTAL POLLUTION AND CLIMATE CHANGE STUDIES

Environmental Pollution could be an important issue which is engulfing almost each corner of the planet. Human activities like industrialization, vehicular activity induced by enormous amount of fuel consumption, urbanization, construction, power generation and all other anthropogenic activities contributing to the environmental pollution. Geospatial Technologies operate and manage spatial and statistical data and supply us with the results giving us the correlation between the environmental pollutants and human health. So, GIS and Geospatial Technologies will be widely used for monitoring environmental pollution. Air pollution levels will be monitored by GIS techniques spatially distributed over town. The properties of an interpolated surface are often managed by limiting the calculation of values of the output cell. When the utmost numbers of points are specified for the sampling, the closest output cell points position returns unless the most number is reached. GIS will be utilised by City Planners for the Pollution mitigation and Control programs. Developments in GIS help researchers to gather and analyze information about the globe. The attitude offered by GIS technology and methods will be particularly important in studying global temperature change.

Some of the uses of GIS on temperature change includes:

- Locating areas where temperatures are particularly high or especially erratic compared with the worldwide average.
- Discovering how natural atmospheric processes might affect warming.
- Creating models to indicate how a warming climate might impact the ecology of varied regions.

- Examining the relevance of shifts in land cover, like the removal of trees, to global climate change.
- Visualizing multiple factors with the potential to affect crop growth, industry, wildlife and far more.

Geospatial tools offer perspective on evolving weather patterns, rising sea levels and growing risks to human health. Hazard maps support remote sensing data and satellite imagery informs about current conditions and also helps to identify the areas that require urgent attention. Response teams provide updates and photos from the site, resulting in efficient and effective crisis management.



CHECK YOUR PROGRESS

Q. 1: What are the main factors of air pollution?

.....

.....

Q. 2: How geospatial tools help in climate studies?

.....

.....

10.4 FLOODS AND DROUGHT PRONE AREA MAPPING

Droughts and Floods are water-related natural disasters which affect a wide range of environmental factors and activities related to agriculture, vegetation, human and wild life and local economies. Remote sensing techniques make it possible to obtain and distribute information rapidly over large areas by means of sensors operating in several spectral bands, mounted on aircraft or satellites.

Monitoring and assessment of drought through remote sensing and GIS depend on the factors that cause drought and the factors of drought impact. Based on the causative factors, drought can be classified into Meteorological, Hydrological and Agricultural droughts. Drought is a normal, recurrent feature of climate and occurs in all climatic zones, although its characteristics vary significantly from one region to another. Reduced crop,

rangeland, and forest productivity; increased fire hazard; reduced water levels; increased livestock and wildlife mortality rates; and damage to wildlife and fish habitat are a few examples of direct impacts of drought. The consequences of these impacts illustrate indirect impacts. The remote sensing and GIS technology significantly contributes to all the activities of drought management.

The remote-sensing based method for identification of drought prone areas uses historical vegetation index data derived from National Oceanic and Atmospheric Administration (NOAA) satellite series. It provides spatial information on drought prone area depending on the trend in vegetation development, frequency of low development and their standard deviations. The remote sensing use for drought prediction can benefit from climate variability predictions. The new assimilation techniques have produced a stronger impact of space data on the quality of weather and seasonal climate predictions.

Geographic Information Systems (GIS) successfully visualize the extent of flooding and also to prepare the flood maps to provide flood damage estimation maps and flood risk map. The GIS must be used together with a hydraulic method to estimate flood profile with a given return period. After 1970, Hydrologic Engineering Centers River Analysis System (HEC-RAS) software developed by u. s. Army Corps of Engineers (USACE) is widely utilized in Europe and America. Due to flash flood high levels of water within the streets and roads, causing many problems like bridge collapse, building damage and traffic problems. it's impossible to avoid risks of floods or prevent their occurrence, however it's plausible to work on the reduction of their effects and to chop back the losses which they will cause. Mapping flash flood are helpful to planners, risk managers and disaster response or emergency services during extreme and intense rainfall events. we use SPOT and SRTM DEMs data that accuracy assessment is achieved by using check points, obtained by GPS observations. Analytical Hierarchical Process (AHP) helps to outline the relative impact weight of flood contributing factors to induce a composite flood hazard index (FHI). The contributing factors during this study are runoff, soil type, surface slope, surface

roughness, drainage density, distance to main channel and land use. The areas in high risk flood zones are acquire by superimpose the flood hazard index map with the zone boundaries layer. The affected population number and expanse are determined and compared.



CHECK YOUR PROGRESS

Q. 3: Classify drought based on causative factors.

.....

.....

Q. 4: Define Analytical Hierarchical Process.

.....

.....

Q. 5: What are the contributing factors to induce a composite flood hazard index?

.....

.....

10.5 EARLY WARNING SYSTEMS FOR DISASTERS AND HAZARDS

An Early Warning System (EWS) could be a disaster prediction information tool utilized in several fields to gather and communicate timely and useful warning information for predicted severe events or disasters. this could be as a results of natural factors, geophysical or biological hazard, sociopolitical factors, industrial hazard, and private health risk factors, among other related disasters or hazards. A geospatial tool is also a decision-making support instrument for planning and management of resources, disasters, and hazards through its capability to input, retrieve, process, analyze, and produce georeferenced data. A geospatial tool also can be noted as a broad toolset which is capable of and embedded with the flexibility to store, edit, query, analyze, and display geospatial information.

Geospatial technology plays major roles in disaster management by providing risk modeling and vulnerability analysis, which enable effective early warning and damage assessment. For effective disaster management, there's a desire to include technologies that are capable of handling and contributing to each stage independently. Hence, the requirement for geospatial tool and early warning system for disaster management is very important.



CHECK YOUR PROGRESS

Q. 6: Define early warning system.

.....

.....

10.6 RURAL AND URBAN INFRASTRUCTURE SURVEY

The capabilities within the advance of high-resolution satellites and extensive network of associated infrastructure have contributed to the growing interest within the appliance of GIS for a spread of development needs. The areas which are receiving main attention include natural resources information assessment, monitoring and management, water shed development, environmental planning, urban services and land use planning.

GIS provide a superior service to the user organizations to assist in their own decision-making process for developmental planning. A better Graphical User Interface (GUI) could also be built using GIS, because it provides very widespread and fast access to information, both graphically and non-graphically.

This makes the system more healthy in terms of its communication with a variety of users. Geographical information systems are open and might still evolve to match with our ever changing needs. GIS database with indicators are specifically relevant for analyzing and projecting divergences between rural and urban-industrial areas in India. This GIS should assist decision makers and planners in formulating rural development

policies at the regional and national levels. GIS technology initiate new visions for rural development in India by identifying innovative initiatives in lagging rural regions and by developing alternative development scenarios, which give special emphasis to options outside the quality sectors of agriculture and forestry. GIS assist decision makers and planners in formulating rural development policies at the regional and national levels.

The population of the planet is increasing exponentially, the land available to accommodate the identical remains limited. Urbanization is that the new threat the environment is fighting with. Humans are actively engaging in encroachment of the lands within the forest areas.The need of the hour isn't just urbanization but smart urbanization; to plan urban development smartly so on accommodate more in less, which too nicely and efficiently.

GIS has become ubiquitous. It's pervading almost every sphere of human life today. Using GIS in urban planning, planners can take urbanization to new extent.

GIS provides planners, surveyors, and engineers with the tools they have to style and map their neighborhoods and cities. Visualization, spatial analysis, and spatial modeling are the foremost frequently used GIS functions in plan making. GIS can help to store, manipulate, and analyze physical, social, and economic data of a city. Planners can then use the spatial query and mapping functions of GIS to examine the prevailing situation within town. Through map overlay analysis, GIS can help to spot areas of conflict of exploitation with the environment by overlaying existing exploitation ashore suitability maps.

Using the multilayered mapping feature of GIS, a municipal planning committee can visualize a spread of things, as an example, major agricultural land, surface water, high flood frequency, and highly erodible land. This information used by planners in developing urban infrastructures. GIS can significantly helps in examining component or conducting a feasibility study of a location for a particular purpose, as an example to determine the suitability of a location for the event of a bridge or dam. Feasibility studies of smaller structures like schools and hospitals can even be selected effectively

using GIS. It should be used to ascertain the feasibility of a part for waste disposal and treatment. Planners make use of GIS to smooth the progress of citizen participation and community input as they develop a vision for the community that improve the standard of life for all citizens.



CHECK YOUR PROGRESS

Q. 7: How GIS technology initiate new visions for rural development in India?

.....

.....

10.7 MAPPING AND PLANNING

Mapping could be a fundamental function of Geographic Information Systems (GIS), which offer a visible interpretation of information. GIS-based mapping helps the governments achieve healthier governance as they're able to solve problems more quickly and efficiently. Proper maps helps to achieve better land-use planning, road and utility maintenance, emergency management, infrastructure evaluation and development and property management. Mapping is often of great help when it involves organize urban growth and solving housing issues related to the increasing population.

GIS is getting increasingly employed in urban planning within the developed and developing countries. GIS tools help planners to examine problems more rapidly and systematically, prepare solutions, and supervise improvement toward long-standing objectives for the community. Many planning departments that had obtained mapping systems within the past have shifted to GIS. With the increase in user-friendliness and also the quantity of functions of GIS software, and also the marked decrease within the costs of GIS hardware.

Identification of the patterns of geographic region and analyses of spatial and temporal changes would help immensely within the design for proper infrastructure facilities. Patterns of sprawl and analyses of spatial and temporal changes can be done cheaply and efficiently with the help of spatial and temporal technologies like Geographic Information System (GIS) and Remote Sensing (RS). The assorted stages within the formulation and implementation of a sustainable regional development strategy are often generalized as determination of objectives, resource inventory, analyses of this situation, modeling and projection, development of designing options, selection of designing options, plan implementation, and plan evaluation, monitoring and feedback. The spatial patterns of geographical region over different time periods, are often systematically mapped, monitored and accurately assessed from satellite data (remotely sensed data) along with conventional ground data. Mapping conurbation provides a "picture" of where this type of growth goes on, helps to identify the environmental and natural resources threatened by such sprawls, and to suggest the likely future directions and patterns of sprawling growth. Remote sensing and GIS is also used separately or together for application in studies of geographical area. GIS analyses, database manipulation and models into one analyses system. Remote sensing applications are increasing very rapidly with the availability of high-resolution data from the state of the art satellites like IRS-1C/1D/P4 and LANDSAT. The advancement in constituent and software within the realm of remote sensing also enhances the remote sensing applications. IRS-1C/1D/P4 provides data with good spectral resolution (LISS data) and so the spatial resolution of 5.6 m in panchromatic mode. The remote sensing satellites with high-resolution sensors and broad coverage potential provides data with better resolution, coverage and revisit to meet the growing applications needs.



CHECK YOUR PROGRESS

Q. 8: Define urban sprawl.

.....

.....

Q. 9: How does high resolution satellite data help in planning and mapping?

.....

.....



10.8 LET US SUM UP

- Droughts and Floods are water-related natural disasters which affect a wide range of environmental factors and activities related to agriculture, vegetation, human and wild life and local economies.
- Remote sensing techniques make it possible to obtain and distribute information rapidly over large areas by means of sensors operating in several spectral bands, mounted on aircraft or satellites.
- Space technology has made substantial contribution in all the three phases such as preparedness, prevention and relief phases of drought and flood disaster management.
- Satellite remote sensing with repetitive and synoptic viewing capabilities, as well as multispectral capabilities, is a powerful tool for mapping and monitoring the ecological changes in the urban core and in the peripheral land use planning,.
- GIS mapping helps you to visualize and identify patterns that are difficult to see if the data elements are in table format.



10.9 FURTHER READING

1. Alexander, W.J.R. (2003).Early Warning Systems for the Detection and Response to Severe Floods, in *Early Warning Systems for Natural Disaster Reduction,* eds. Zschau, J and A. N. Kuppers, Springer, New York; pp.311-318
2. Ali. S. (2001). The Design & Development of an Automated Hazard Emergency Management Systems, in *Proceedings of the URISA Caribbean GIS Conference*;pp.125-127

3. Bouloucos and Brown (2001). *ITC Courses in Remote Sensing, GIS and Photogrammetry*; pp.12-24
4. Kidder, S.Q. and Vonder Harr, T.H. (1995). *Satellite Meteorology: An Introduction*, Academic Press; pp. 466
5. Ehlers, M. (2008). Geoinformatics and digital earth initiatives: A German perspective, *International Journal of Digital Earth*, 1:1, pp.17-30, DOI: 10.1080/17538940701781975
6. Vicente-Serrano S. M. (2012). Challenges for drought mitigation in Africa : The potential use of geospatial data and drought information systems. *Applied Geography*, 34, pp. 471-486.

10.10 ANSWERS TO CHECK YOUR PROGRESS

Q. No. 1: Air pollutants like ozone, particulate matter, carbon monoxide, nitrogen oxide, sulfur dioxide, lead cause atmospheric pollution which regards to ozone depletion, GH effect, climate change and human health effects. If anyone say the main factors of air pollution beyond of industry, means of transportation, solid waste generation and urbanization.

Q. No. 2: Geospatial tools like Remote Sensing and GIS helps in investigating evolving weather patterns, temperature change and rising sea levels.

Q. No. 3: Drought is a normal, recurrent feature of climate and occurs in all climatic zones, Based on the causative factors, drought can be classified into Meteorological, Hydrological and Agricultural droughts.

Q. No. 4: The analytic hierarchy process (AHP) is a structured technique for organizing and analyzing complex decisions, based on mathematics and psychology.

Q. No.5: The contributing factors during this study are runoff, soil type, surface slope, surface roughness, drainage density, distance to main channel and land use.

Q. No.6: Early Warning Systems (EWS) are complex processes aimed at reducing the impact of natural hazards by providing timely and relevant information in a systematic way.

Q. No.7: GIS technology initiate new visions for rural development in India by identifying innovative initiatives in lagging rural regions and by developing alternative development scenarios, which give special emphasis to options outside the quality sectors of agriculture and forestry.

Q. No.8: Urban sprawl, also called sprawl or suburban sprawl, the rapid expansion of the geographic extent of cities and towns, often characterized by low-density residential housing, single-use zoning, and increased reliance on the private automobile for transportation.

Q. No.9: IRS-1C/1D/P4 provides data with good spectral resolution (LISS data) and so the spatial resolution of 5.6 m in panchromatic mode. The remote sensing satellites with high-resolution sensors and broad coverage potential provides data with better resolution, coverage and revisit to meet the growing applications needs.



10.11 MODEL QUESTIONS

Q.1: List various applications of geoinformatics in agriculture, water resource and urban planning.

Q.2: Explain the applications of geoinformatics in disaster management.

Q.3: Explain the importance of conservation and planning for the development of water resources.

Q.4: “Unprecedented growth of cities is leading to unprecedented environmental degradation”. Discuss.

Q.5: Explain the role of Remote Sensing in resource appraisal and management.

Q.6: Discuss methodology and techniques of regional planning using geoinformatics.

Q.7: Give an importance and use of aerial photographs and satellite images in geographical research.

Q.8: Explain the effects of global warming and ozone depletion.

Q.9: Use of GIS leads to better decision making in government. Elaborate the statement.

Q.10: Explain the benefits and applications of satellite images and aerial photographs.

Q.11: Write short notes on-

- Importance of survey
- Use of GIS in spatial analysis
- Geometry of aerial photograph.

Q.12: How GIS has helped in the flood hazard mapping of river basin?

---×---