

**Q.1 Attempt any TWO question of the following :** [10]

**Q.1(a) List and explain components of GIS.** [5]

**Ans.:** Like any other information technology, GIS requires the following four components to work with geospatial data:

- **Computer System** : The computer system includes the computer and the operating system to run GIS. Typically the choices are PCs that use the Windows operating system (e.g., Windows 2000, Windows XP) or workstations that use the UNIX or Linux operating system. Additional equipment may include monitors for display, digitizers and scanners for spatial data input. GPS receivers and mobile devices for fieldwork, and printers and plotters for hard-copy data display.
- **GIS Software** : The GIS software includes the program and the user interface for driving the hardware. Common user interfaces in GIS are menus, graphical icons, command lines, and scripts.
- **People** : People refers to GIS professionals and users who define the purpose and objectives, and provide the reason and justification for using GIS.
- **Data** : Data consist of various kinds of inputs that the system takes to produce information.
- **Infrastructure** : The infrastructure refers to the necessary physical, organizational, administrative, and cultural environments that support GIS operations. The infrastructure includes requisite skills, data standards, data clearinghouses, and general organizational patterns.

**Q.1(b) Explain datum with a suitable example.** [5]

**Ans.:** **Datum** :A datum is a mathematical model of the Earth, which serves as the reference or base for calculating the geographic coordinates of a location (Burkard 1984; Moffitt and Bossler 1998). The definition of a datum consists of an origin, the parameters of the spheroid selected for the computations, and the separation of the spheroid and the Earth at the origin. Many countries have developed their own datums for local surveys. Among these local datums are the European Datum, the Australian Geodetic Datum, the Tokyo Datum, and the Indian Datum (for India and several adjacent countries).

Until the late 1980s, Clarke 1866, a ground-measured spheroid, was the standard spheroid for mapping in the United States.

In 1986 the National Geodetic Survey (NGS) introduced NAD83 (North American Datum of 1983).

**Q.1(c) Explain ESRI's Coverage Model.** [5]

**Ans.:** Besides work by federal agencies, commercial GIS vendors such as ESRI, Inc. have also used topology to develop proprietary vector data formats. ESRI, Inc. introduced the coverage model in the 1980s to separate GIS from CAD (computer-aided design) at the time. AutoCAD by Autodesk was, and still is, the leading CAD package. A data format used by AutoCAD for the transfer of data files is called DXF (drawing exchange format). DXF maintains data in separate layers and allows the user to draw each layer using different line symbols, colors, and text. But DXF files do not support topology.

Coverage is a topology-based vector data format. A coverage can be a point coverage, line coverage, or polygon coverage. The coverage model supports three basic topological relationships (Environmental Systems Research Institute, 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 and left and right polygons.

**Q.1(d) State the topology rules associated with Geospatial Data.**

**[5]**

**Ans.:** The following lists some real-world applications of topology rules :

- Counties must not overlap.
- County must not have gaps.
- County boundary must not have dangles (i.e. must be closed).
- Census tracts and counties must cover each other.
- Voting district must be covered by county.
- Contour lines must not intersect.
- Interstate route must be covered by feature class of reference line.
- Milepost markers must be covered by reference line (i.e., road feature class).
- Label points must be properly inside polygons.

**Q.2 Attempt any TWO question of the following :**

**[10]**

**Q.2(a) Explain Metadata with all its categories.**

**[5]**

**Ans.:** The FGDC has developed the content standards for metadata and provides detailed information about the standards at its website (<http://www.fgdc.gov/>). These standards have been adopted by federal agencies in developing their public data. FGDC metadata standards describe a data set based on the following categories:

- Identification information – basic information about the data set, including title, geographic data covered, and currency.
- Data quality information – information about the quality of the data set, including positional and attribute accuracy, completeness, consistency, sources of information, and methods used to produce the data.
- Spatial data organization information – information about the data representation in the data set, such as method for data representation (e.g., raster or vector) and number of spatial objects.
- Spatial reference information – description of the reference frame for and means of encoding coordinates in the data set, such as the parameters for map projections or coordinate systems, horizontal and vertical datums, and the coordinate system resolution.
- Entity and attribute information – information about the content of the data set, such as the entity types and their attributes and the domains from which attribute values may be assigned.
- Distribution information – information about obtaining the data set.
- Metadata reference information – information on the currency of the metadata information and the responsible party.

**Q.2(b) Write short notes on :**

**[5]**

**(i) US Geological Survey                      (ii) Natural Resources Conservation Service**

**Ans.:** **U.S. Geological Survey :**

Through its National Map program, the U.S. Geological Survey (USGS) is the major provider of GIS data in the United States. Its website (<http://geography.usgs.gov/>) offers pathways to USGS national mapping and remotely sensed data and to thematic data clearinghouses on biological, geologic, and water resources data. Public data available from the USGS include both vector and raster data.

**Natural Resources Conservation Service :**

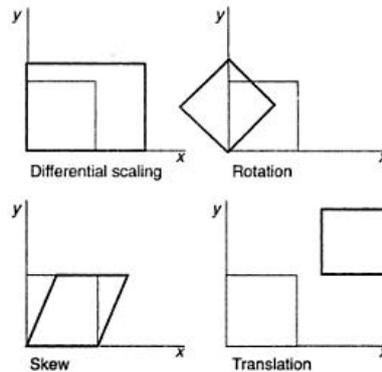
The Natural Resources Conservation Service (NRCS) of the U.S. Department of Agriculture distributes soils data nationwide through its website (<http://soils.usda.gov/>). There are two soil databases: STATSGO and SSURGO. Compiled at 1:250,000 scales, the STATSGO (State Soil Geographic) database is suitable for broad planning and management uses. Compiled from field mapping at scales ranging from 1:12,000 to 1:63,360, the SSURGO (Soil Survey Geographic) database is designed for uses at the farm, township, and county levels.

**Q.2(c) Explain Affine Transformations with example.**

**[5]**

**Ans.: Affine Transformation :**

The affine transformation allows rotation, translation, skew, and differential scaling on a rectangular object, while preserving line parallelism (Pet-tofrezzo 1978; Loudon et al. 1980; Chen et al. 2003). Rotation rotates the object's x- and y-axes from the origin. Translation shifts its origin to a new location. Skew allows a non-perpendicularity (or affinity) between the axes, thus changing its shape to a parallelogram with a slanted direction. And differential scaling changes the scale by expanding or reducing in the x and/or y direction. Figure shows these four transformations graphically.



**Fig. :** Differential scaling, rotation, skew, and translation in the affine transformation.

Mathematically, the affine transformation is expressed as a pair of first-order polynomial equations:

$$X = Ax + By + C \quad \dots (1)$$

$$Y = Dx + Ey + F \quad \dots (2)$$

where x and y are the input coordinates that are given, X and Y are the output coordinates to be determined, and A, B, C, D, E, and F are the transformation coefficients.

**Q.2(d) List and explain various resampling methods with suitable example.**

**[5]**

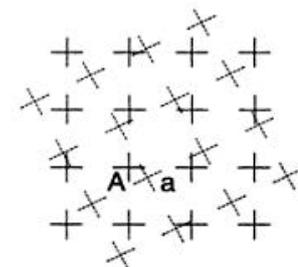
**Ans.: Resampling of Pixel Values**

The result of geometric transformation of a satellite image is a new image based on a projected coordinate system. But the new image has no pixel values. The pixel values must be filled through resampling. Resampling in this case means filling each pixel of the new image with a value or a derived value from the original image.

**Resampling Methods :** Three common resampling methods are nearest neighbor, bilinear interpolation, and cubic convolution. The nearest neighbor resampling method fills each pixel of the new image with the nearest pixel value from the original image. For example, Figure 1 shows that pixel A in the new image will take the value of pixel a in the original image because it is the closest neighbor. The nearest neighbor method is computationally efficient.

Both bilinear interpolation and cubic convolution fill the new image with distance-weighted averages of the pixel values from the original image.

**Fig. 1 :** Because a in the original image is closest to pixel A in the new image, the pixel value at a is assigned to be the pixel value at A using the nearest neighbor technique.



The bilinear interpolation method uses the average of the four nearest pixel values from three linear interpolations, whereas the cubic convolution method uses the average of the 16 nearest pixel values from five cubic polynomial interpolations (Richards and Jia 1999). Cubic convolution tends to produce a smoother output than bilinear interpolation but requires a longer processing time.

**Q.3 Attempt any TWO question of the following :** [10]

**Q.3(a) Explain with example what is meant by Normalization.** [5]

**Ans.:** Preparing a relational database such as SSURGO involves following certain rules. An important rule is called normalization. **Normalization** is a process of decomposition, taking a table with all the attribute data and breaking it down into small tables while maintaining the necessary linkages between them (Vetter 1987). Normalization is designed to achieve the following objectives:

- To avoid redundant data in tables that waste space in the database and may cause data integrity problems.
- To ensure that attribute data in separate tables can be maintained and updated separately and can be linked whenever necessary.
- To facilitate a distributed database.

An example of normalization is offered here.

**Table 1 : An Unnormalized Table**

PIN	Owner	Owner address	Sale date	Aeres	Zone code	Zoning
P101	Wang	101 Oak St	1-10-98	1.0	1	Residential
	Chang	200 Maple St				
P102	Smith	300 Spruce Rd	10-6-68	3.0	2	Commercial
	Jones	105 Ash St				
P103	Costello	206 Elm St	3-7-97	2.5	2	Commercial
P104	Smith	300 Spruce Rd	7-30-78	1.0	1	Residential

**Table 2 : First Step in Normalization**

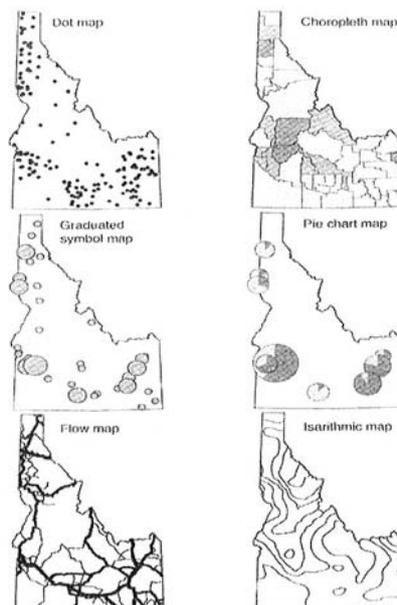
PIN	Owner	Owner address	Sale date	Aeres	Zone code	Zoning
P101	Wang	101 Oak St	1-10-98	1.0	1	Residential
P101	Chang	200 Maple St	1-10-98	1.0	1	Residential
P102	Smith	300 Spruce Rd	10-6-68	3.0	2	Commercial
P102	Jones	105 Ash St	10-6-68	3.0	2	Commercial
P103	Costello	206 Elm St	3-7-97	2.5	2	Commercial
P104	Smith	300 Spruce Rd	7-30-78	1.0	1	Residential

**Q.3(b) Explain the different types of maps in GIS.** [5]

**Ans.:** Types of Maps

Cartographers classify maps by function and by symbolization

The following describes several common types of quantitative maps (Figure 1)



**Fig 1 : Six common types of quantitative maps**

The dot map uses uniform point symbols to show spatial data, with each symbol representing a unit value.

The choropleth map symbolizes, with shading, derived data based on administrative units.

The **graduated symbol map** uses differentsized symbols such as circles, squares, of triangles to represent different ranges of values.

The **chart map** uses either pie charts or bar charts. A variation of the graduated circle, the pie chart can display two sets of quantitative data.

The **flow map** displays different quantities of flow data such as traffic volume and stream flow by varying the line symbol width.

The **isarithmetic map** uses a system of isolines to represent a surface. Each isoline connects points of equal value.

**Q.3(c) Explain the role of Typography in GIS.**

[5]

**Ans.:** Type Variations

Type can vary in typeface and form. **Typeface** refers to the design character of the type. Two main groups of typefaces are **serif** (with serif) and **sans serif** (without serif). Serifs are small, finishing touches at the ends of line strokes, which tend to make running text in newspapers and books easier to read. Compared to serif types, sans serif types appear simpler and bolder. Although rarely used in books or other text-intensive materials, a sans serif type stands up well on maps with complex map symbols and remains legible even in small sizes. Sans serif types have an additional advantage in mapmaking because many of them come in a wide range of type variations.

Helvetica Normal

*Helvetica Italic*

**Helvetica Bold**

***Helvetica Bold-Italic***

Times Roman Normal

*Times Roman Italic*

**Times Roman Bold**

***Times Roman Bold-Italic***

**Fig 1 :** Type variations in weight and roman versus italic.

Type form variations include **type weight** (bold, regular, or light), **type width** (condensed or extended), upright versus slanted (or roman versus italic), and uppercase versus lowercase (Figure 1).

**Q.3(d) Explain the various color schemes used in GIS.**

[5]

**Ans.:** The following is a summary of color schemes :

- The single hue scheme. This color scheme uses a single hue but varies the combination of value and chroma to produce a sequential color scheme such as from light red to dark red. It is a simple but effective option for displaying quantitative data.
- The hue and value scheme. This color scheme progresses from a light value of one hue to a darker value of a different hue. Examples are yellow to dark red and yellow to dark blue. color sequences incorporating both regular hue and value variations outperform other color schemes on the recall or recognition of general map information.
- The diverging or double-ended scheme. This color scheme uses graduated colors between two dominant colors.
- The part spectral scheme. This color scheme uses adjacent colors of the visible spectrum to show variations in magnitude. Examples of this color scheme include yellow to orange to red and yellow to green to blue.

- The full spectral scheme. This color scheme uses all colors in the visible spectrum. A conventional use of the full spectral scheme is found in elevation maps. Cartographers usually do not recommend this option for mapping other quantitative data because there is no logical sequence between hues.

**Q.4 Attempt any TWO question of the following :** [10]

**Q.4(a) Explain the various descriptive statistics used in GIS.** [5]

**Ans.:** **Descriptive Statistics :** Descriptive statistics summarize the values of a data set. They include the following :

- Range: the difference between the minimum and maximum values.
- Median: the midpoint value, or the 50<sup>th</sup> percentile.
- First quartile: the 25<sup>th</sup> percentile.
- Third quartile: the 75<sup>th</sup> percentile.
- Mean: the average of data values. The mean can be calculated by  $\sum_{i=1}^n x_i / n$ , where  $x_i$  is the  $i^{\text{th}}$  value and  $n$  is the number of values.
- Variance: the average of the squared deviations of each data value about the mean. The variance can be calculated by  $\sum_{i=1}^n (x_i - \text{mean})^2 / n$
- Standard deviation: the square root of the variance.
- Z score: a standardized score that can be computed by  $(x - \text{mean})/s$ , where  $s$  is the standard deviation.

**Q.4(b) Explain the concept of Attribute Data Query.** [5]

**Ans.:** **Attribute Data Query**

Attribute data query retrieves a data subset by working with attribute data. The selected data subset can be simultaneously examined in the table, displayed in charts, and linked to the highlighted features in the map. The selected data subset can also be saved for further processing.

Attribute data query requires the use of expressions, which must be interpretable by a GIS of a database management system. The structure of these expressions varies from one system to another, although the general concept is the same. ArcGIS, for example, uses SQL (Structured Query Language) for query expressions

#### **SQL (Structured Query Language)**

SQL is a data query language designed for relational databases. IBM developed SQL in the 1970s, and many commercial database management systems such as Oracle, Informix, DB2, Access, and Microsoft SQL Server have since adopted the query language. A new development is to extend SQL to object-oriented database management systems and to spatial data (Shekhar and Chawla 2003).

To use SQL to access a database, we must follow the structure (i.e., syntax) of the query language. The basic syntax of SQL, with the keywords in italic, is

```
select < attribute list >
from < relation >
where < condition >
```

**Q.4(c) Explain the concept of Spatial Data Query.** [5]

**Ans.:** **Spatial Data Query**

Spatial data query refers to the process of retrieving a data subset from a layer by working directly with features. We may select features using a cursor, a graphic, or the spatial relationship between features. As the geographic interface to the database, spatial data query complements attribute data query in data exploration.

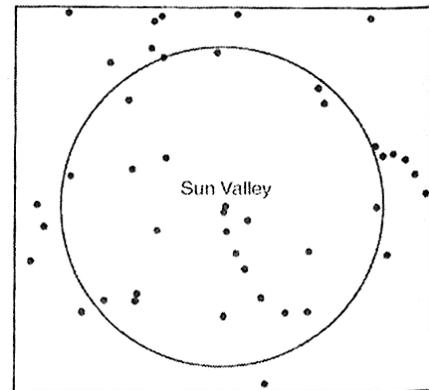
Similar to attribute data query, the results spatial data query can be simultaneously inspected in the map, linked to the highlighted records in the table, and displayed in charts. They can also be saved as a new data set for further processing.

#### Feature Selection by Cursor

The simplest spatial data query is to select a feature by pointing at it or to select features by dragging a box around them.

#### Feature Selection by Graphic

This query method uses a graphic such as a circle, a box, a line, or a polygon to select features that fall inside or are intersected by the graphic object (Figure 1).



**Fig. 1:** Select features by a circle centered at Sun Valley

#### Feature Selection by Spatial Relationship

This query method selects features based on their spatial relationships to other features. Features to be selected may be in the same layer as features for selection.

Spatial relationships used for query include the following :

- **Containment** : selects features that fall completely within features for selection.
- **Intersect** : selects features that intersect features for selection. Examples include selecting land parcels that intersect a proposed road, and finding urban areas that intersect an active fault line.
- **Proximity** : selects features that are within a specified distance of features for selection. Examples include finding state parks within 10 miles of an interstate highway, and finding pet shops within 1 mile of selected streets.

**Q.4(d) How map comparison can be used for data exploration?**

**[5]**

**Ans.:** **Map Comparison :**

Map comparison can help a GIS user sort out the relationship between different maps. For example, the display of wildlife locations on a vegetation layer may reveal the association between the wildlife species and the distribution of vegetation covers.

If the maps to be compared consist of only point or line features, they can be coded in different colors and superimposed on one another in a single view. But this process becomes difficult if they include polygon features or raster data. One option is to use transparency as a visual variable. A semitransparent layer allows another layer to show through. For example, to compare two raster layers, we can display one layer in a color scheme and the other in semitransparent shades of gray. The gray shades simply darken the color symbols and do not produce confusing color mixtures.

**Q.5 Attempt any TWO question of the following :**

**[10]**

**Q.5(a) Explain the variations in buffering.**

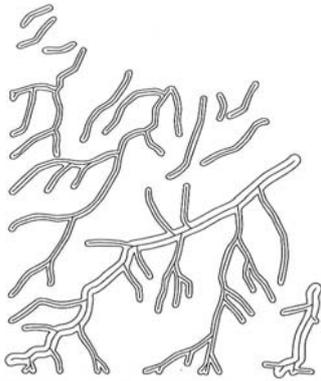
**[5]**

**Ans.:** **Variations in Buffering**

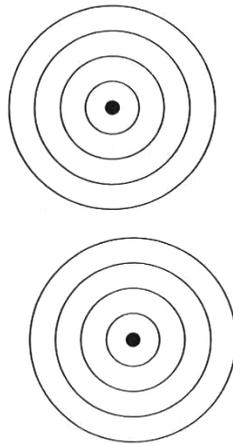
The buffer distance or buffer size does not have to be constant; it can vary according to the values of a given field. A feature may have more than one buffer zone. As an example, a nuclear power plant may be buffered with distances of 5, 10, 15, and 20 miles, thus forming multiple rings around the plant.

Buffering around line features does not have to be on both sides of the lines, it can be on either the left side or the right side of the line feature. Likewise, buffer zones around polygons can be extended either outward or inward from the polygon boundaries. Boundaries

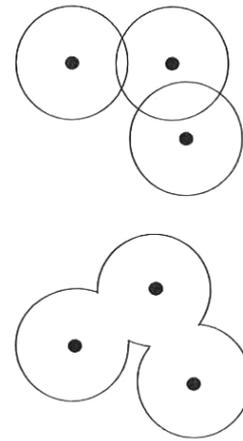
of buffer zones may remain intact so that each buffer zone is a separate polygon. Or these boundaries may be dissolved.



**Fig. 1 :** Buffering with different buffer distances.



**Fig. 2 :** Buffering with four rings



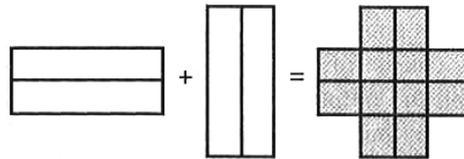
**Fig. 3 :** Buffer zones not dissolved (top) or dissolved (bottom)

**Q.5(b) List and explain various overlay operations.**

**[5]**

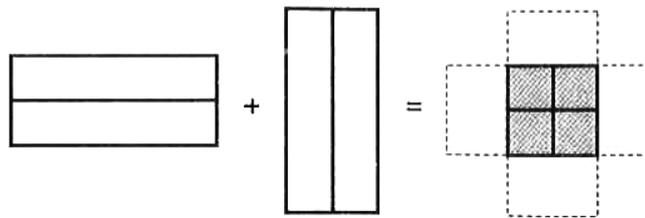
**Ans. :** All overlay methods are based on the Boolean connectors AND, OR, and XOR.

Union preserves all features from the inputs (Figure 1).



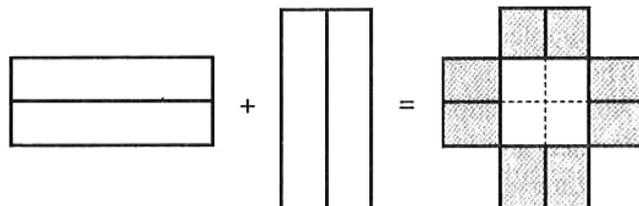
**Fig. 1 :** The Union method keeps all areas of the two input layers in the output.

Intersect preserves only those features that fall within the area extent common to the inputs (Figure 2).



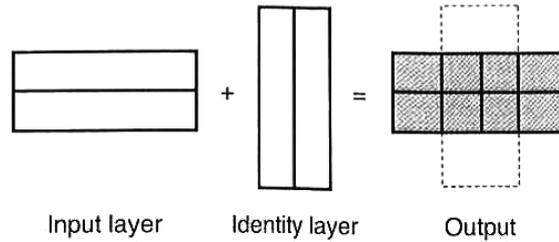
**Fig. 2 :** The Intersect method preserves only the area common to the two input layers in the output.

Symmetrical Difference preserves features that fall within the area extent that is common to only one of the inputs (Figure 3).



**Fig. 3 :** The Symmetrical Difference method preserves areas common to only one of the input layers in the output.

Identity preserves only features that fall within the area extent of the layer defined as the input layer (Figure 4).



**Fig. 4 :** The Identity method produces an output that has the same extent as the input layer. But the output includes the geometry and attribute data from the identity layer.

**Q.5(c) Explain local operations in Raster Data Management.**

**[5]**

**Ans.:** Local Operations

Constituting the core of raster data analysis, local operations are cell-by-cell operations. A local operation can create a new raster from either a single input raster or multiple input rasters.

**Local Operations with a Single Raster**

Given a single raster as the input, a local operation computes each cell value in the output raster as a mathematical function of the cell value in the input raster. As shown in Figure 1, a large number of mathematical functions are available in a GIS package.

Arithmetic	+, -, /, *, absolute, integer, floating-point
Logarithmic	exponentials, logarithms
Trigonometric	sin, cos, tan, arcsin, arccos, arctan
Power	square, square root, power

**Fig. 1 :** Arithmetic, logarithmic, trigonometric and power functions for local operations.

15.2	16.0	18.5	8.64	9.09	10.48
17.8	18.3	19.6	10.09	10.37	11.09
18.0	19.1	20.2	10.20	10.81	11.42

(a)

(b)

**Fig. 2 :** A local operation can convert a slope raster from percent (a) to degrees (b)

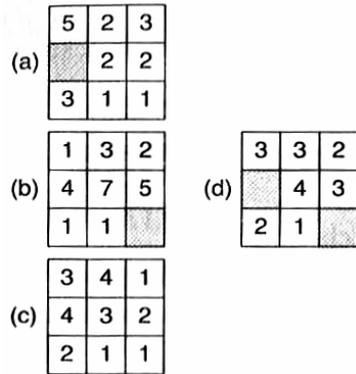
**Reclassification :**

A local operation, reclassification creates a new raster by classification. Reclassification is also referred to as recoding, or transforming, through lookup tables. The first method is a one-to-one change, meaning that a cell value in the input raster is assigned a new value in the output raster. The second method assigns a new value to a range of cell values in the input raster. For example, cells with population densities between 0 and 25 persons per square mile in a population density raster are assigned a value of 1 in the output raster and so on.

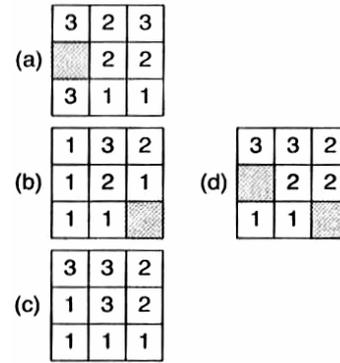
**Local Operations with Multiple Rasters**

Local operations with multiple rasters are also referred to as compositing, overlaying, or superimposing maps. Another common term for local operations with multiple input rasters is map algebra.

Summary statistics, including maximum, minimum, range, sum, mean, median, and standard deviation, are measures that apply to rasters with numeric data. Figure 3, for example, shows a local operation that calculates the mean from three input rasters.



**Fig. 3 :** The cell value in (d) is the mean calculated from three input rasters (a, b and c) in a local operation. The shaded cells have no data.

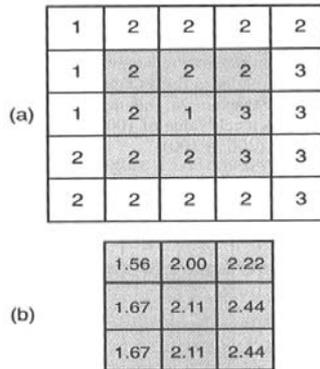


**Fig. 4 :** The cell value in (d) is the majority statistic derived from three input rasters (a, b and c) in a local operation. The shaded cells have no data.

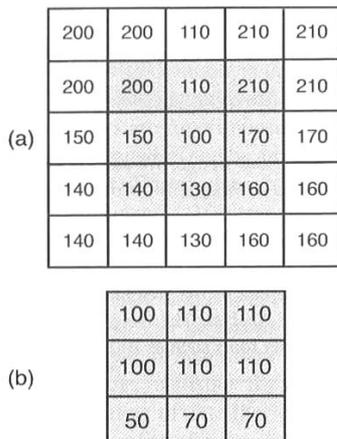
**Q.5(d) Explain applications of Neighborhood operations.**

**[5]**

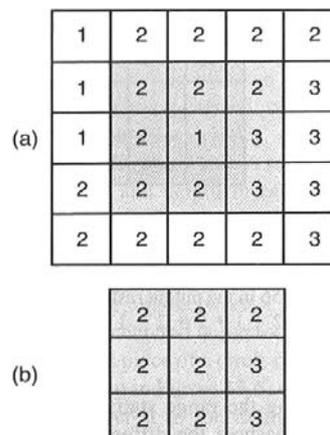
Ans. :



**Fig. 1 :** The cell values in (b) are the neighborhood means of the shaded cells in (a) using a 3-by-3 neighborhood. For example, 1.56 in the output raster is calculated from  $(1 + 2 + 2 + 1 + 2 + 2 + 1 + 2 + 1)/9$



**Fig. 2 :** The cell values in (b) are the neighborhood range statistics of the shaded cells in (a) using a 3-by-3 neighborhood. For example, the upper-left cell in the output raster has a cell value of 100, which is calculated from  $(200 - 100)$ .



**Fig. 3 :** The cell values in (b) are the neighborhood majority statistics of the shaded cells in (a) using a 3-by-3 neighborhood. For example, the upper-left cell in the output raster has a cell value of 2 because there are five 2s and four 1s in its neighborhood.

**Q.6 Attempt any TWO question of the following :** [10]

**Q.6(a) What is spatial interpolation? List the types of spatial interpolations and explain any one.** [5]

**Ans.:** Spatial interpolation is the process of using points with known values at other points. Through spatial interpolation, we can estimate the precipitation value at a location with no recorded data by using known precipitation readings at nearby weather stations.

### Elements of Spatial Interpolation

#### 1) Control Points

Control points are points with known values. Also called known points, sample points, or observations, control points provide the data necessary for the development of an interpolator (e.g. a mathematical equation) for spatial interpolation. The number and distribution of control points can greatly influence the accuracy of spatial interpolation.

**Q.6(b) Explain Regression Models in Spatial interpolation.** [5]

**Ans.:** Regression models

A regression models relates a dependent variable to a number of independent variables in a linear equation (an interpolator), which can then be used for prediction or estimation. Many regression models use nonspatial attributes and are not considered methods for spatial interpolation. But exceptions can be made for regression models that use spatial variables such as distance to a river or location-specific elevation.

A watershed-level regression model for snow accumulation developed by Chang and Li uses snow water equivalent (SWE) as the dependent variable and location and topographic variables as the independent variables. One of their watershed models takes the form of :

$$SWE = b_0 + b_1EASTING + b_2 SOUTHING + b_3ELEV + b_4PLAN1000 \quad \dots (1)$$

where EASTING and SOUTHING corresponds to the column number and the row number in an elevation raster, ELEV is the elevation value, and PLAN1000 is a surface curvature measure.

**Q.6(c) Write a short note on density estimation.** [5]

**Ans.:** Density estimation measures cell densities in a raster by using a sample of known points. There are simple and kernel density estimation methods. To use the simple density estimation method, we can place a raster on a point distribution, tabulate points that fall within each cell, sum the point values, and estimate the cell's density by dividing the total point value by the cell size.

Kernel density estimation associates each known point with a kernel function for the purpose of estimation .

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{1}{h}(x - x_i)\right) \quad \dots (1)$$

where  $K()$  is the kernel function,  $h$  is the bandwidth,  $n$  is the number of known points within the bandwidth, and  $d$  is the data dimensionality.

**Q.6(d) Define :** [5]

(i) **Nugget**    (ii) **Range**    (iii) **Sill**    (iv) **Partial Sill**    (v) **Anisotropy**

**Ans.:** (i) **Nugget** : Nugget is the semivariance at distance 0.

(ii) **Range** : Range is the distance at which the semivariance starts to level off.

(iii) **Sill** : The semivariance at which the leveling takes place is known as sill.

(iv) **Partial Sill** : The difference between the sill and the nugget is known as partial sill.

(v) **Anisotropy** : It is the term describing the existence of directional differences in spatial dependence.

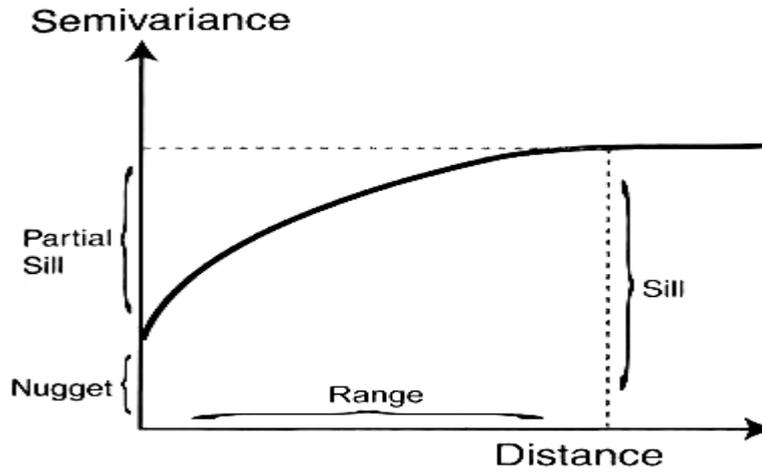


Fig.: Nugget, Range, Sill, and Partial sill

Q.7 Attempt any THREE question of the following : [15]

Q.7(a) Explain Quad Tree with a suitable example. [4]

Ans.: Quad tree :

Instead of working along one row at a time, quad tree uses recursive decomposition to divide a raster into a hierarchy of quadrants (Samet 1990). Recursive decomposition refers to a process of continuous subdivision until every quadrant in a quad tree contains only one cell value.

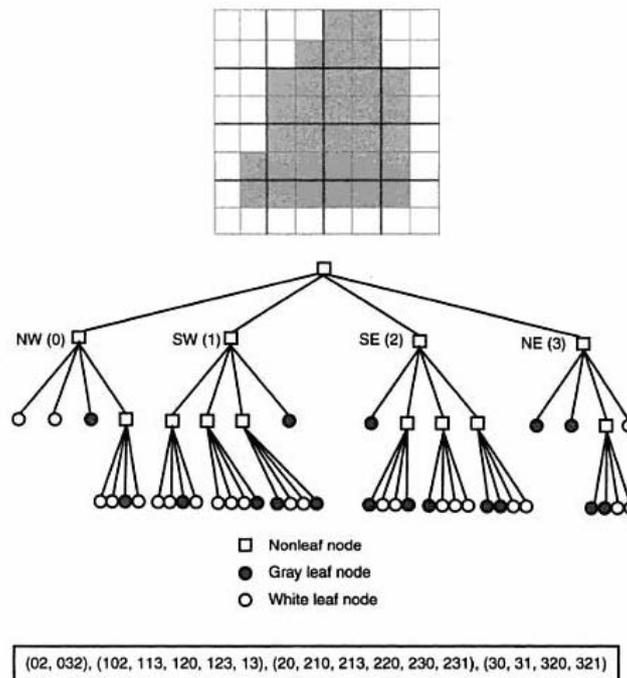


Fig.1: The regional quad tree method divides a raster into a hierarchy of quadrants. The division stops when a quadrant is made of cells of the same value (gray or white).

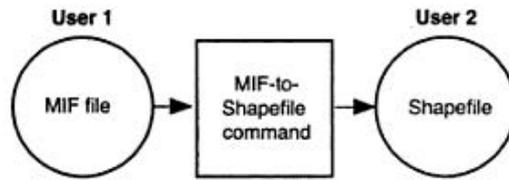
A quadrant that cannot be subdivided is called a leaf node. In the diagram, the quadrants are indexed spatially: 0 for NW, 1 for SW, 2 for SE, and 3 for NE.

Using the spatial indexing method and the hierarchical quad tree structure, the gray cells can be coded as 02,032, and so on.

Q.7(b) Explain Conversion of Existing Data in GIS. [4]

Ans.: Direct Translation :

Direct translation uses a translator in a GIS package to directly convert geospatial data from one format to another (Figure1). Direct translation used to be the only method for data conversion before the development of data standards and open GIS. Many users will prefer direct translation because it is easier to use than other methods.

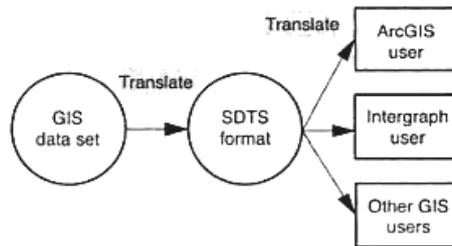


**Fig. 1 :** The MIF to Shapefile tool in ArcGIS converts a MapInfo file to a shapefile.

**Neutral Format :**

A neutral format is a public or de facto format for data exchange.

The Spatial Data Transfer Standard (SDTS) is a neutral format approved by the Federal Information Processing Standards (FIPS) Program in 1992 (<http://mcmcweb.er.usgs.gov/sdts/>). Several federal agencies have converted some of their data to SDTS format. They include the USGS, U.S. Army, U.S. Army Corps of Engineers, Census Bureau, and U.S. National Oceanic and Atmospheric Administration.



**Fig. 2 :** To accommodate users of different GIS packages, a government agency can translate public data into a neutral format such as SDTS format. Using the translator in the GIS package, the user can convert the public data into the format used in the GIS.

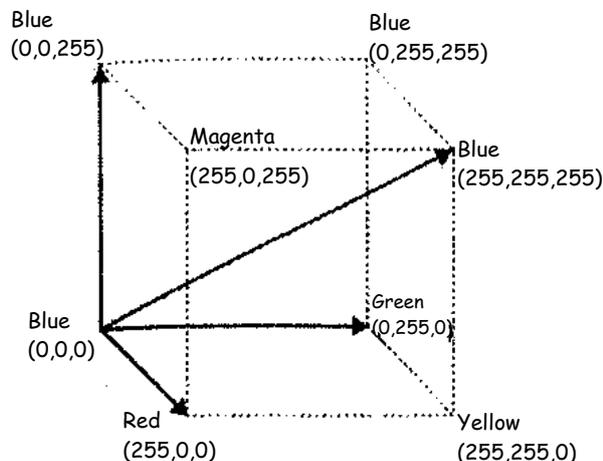
**Q.7(c) Explain the process of Map Production.**

**[4]**

**Ans.:** GIS users design and make maps on the computer screen. These soft-copy maps can be used in a variety of ways. They can be printed, exported for use on the Internet, used in overhead computer projection systems, exported to other software packages, or further processed for publishing.

Map production is a complex topic. We are often surprised that color symbols from the color printers do not exactly match those on the computer screen.

Data display on the computer screen uses either **CRT (cathode ray tube)** or **LCD (liquid crystal display)**. It used to be that desktop computers used CRTs and laptop or portable computers used LCDs.



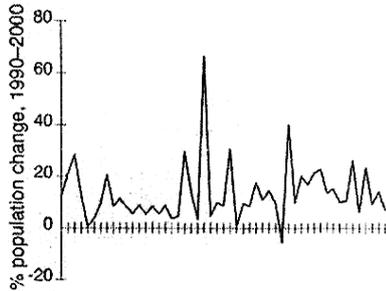
**Fig. 1:** The RGB (red, green and blue) color model

With either a CRT or an LCD, a color symbol we see on a screen is made of pixels, and the color of each pixel is a mixture of **RGB** (red, green, and blue). Typically, the intensity of each primary color can range over 256 shades. Combining the three primary colors produces a possible palette of 16.8 million colors (256 x 256 X 256).

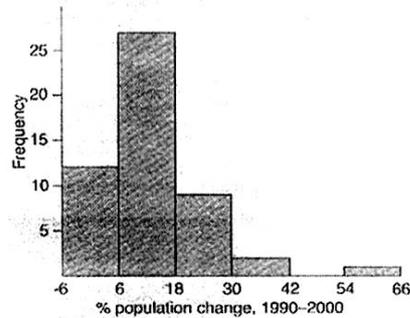
**Q.7(d) List the various types of graphs used in GIS and explain any two. [4]**

**Ans.: Graphs**

Different types of graphs are used for data exploration. A line graph displays data as a line.



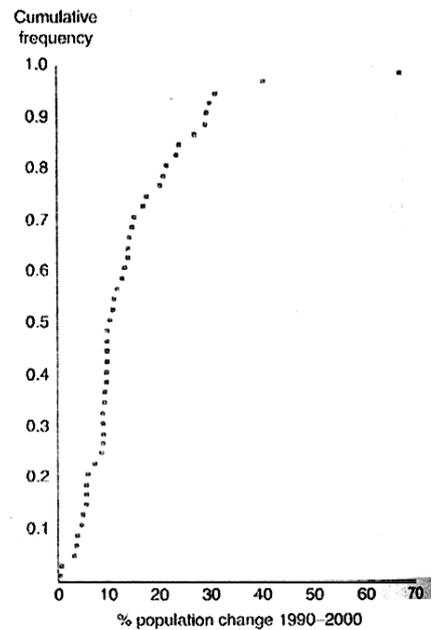
**Fig. 1 : A line graph**



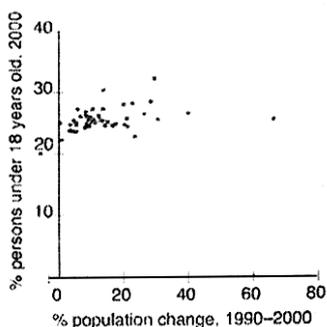
**Fig. 2 : A histogram (bar chart)**

A bar chart, also called a histogram, groups data into equal intervals and uses bars to show the number of frequency of values falling within each class.

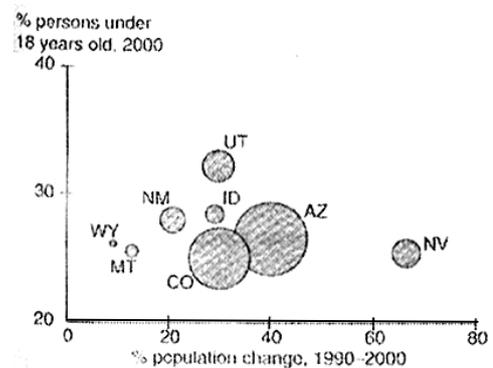
A cumulative distribution graph is one type of line graph that plots the ordered data values against the cumulative distribution values.



**Fig. 3 : A cumulative distribution**



**Fig. 4 : A scatterplot plotting percent persons 18 years old in 2000 against percent population change, 1990-2000. A weak-positive relationship is present between the two variables.**

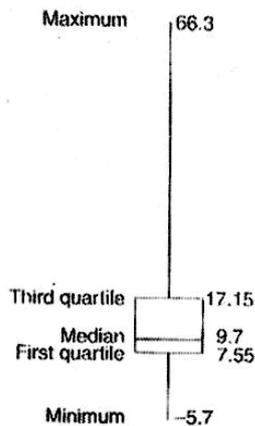


**Fig. 5 : A bubble plot showing percent population change 1990-2000, percent persons under 18 years old in 2000, and state population in 2000.**

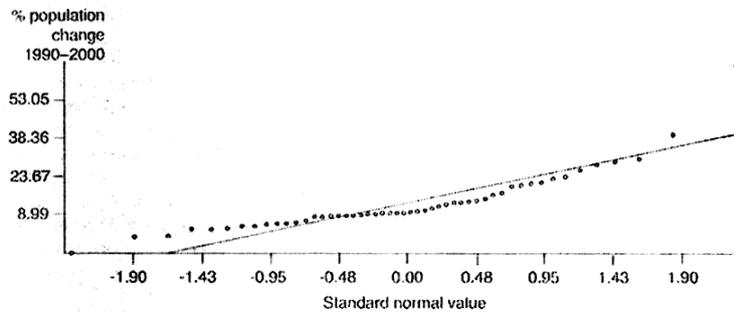
A scatterplot uses markings to plot the values of two variables along the x- and y-axes.

Bubble plots are a variation of scatterplots. Instead of using constant symbols as in a scatterplot, a bubble plot has varying sized bubbles that are made proportional to the value of a third variable.

Boxplots, also called the "box and whisker" plots, summarized the distribution of five statistics from a data set the minimum, first quartile, median, third quartile, and maximum.



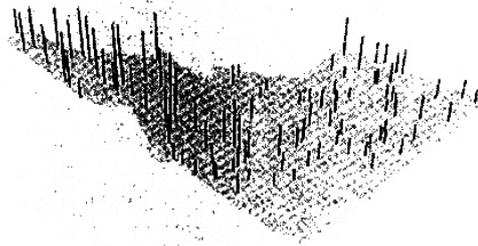
**Fig. 6 :** A boxplot based on the percent population change 1990-2000 data set.



**Fig. 7 :** A QQ plot plotting percent population change, 1990-2000 against the standardized value from a normal distribution.

Some graphs are more specialized. Quantile-quantile plots, also called QQ plots, compare the cumulative distribution of a data set with that of some theoretical distribution such as the normal distribution, a bell-shaped frequency distribution.

Some graphs are designed for spatial data. Figure 8, for example, shows a plot of spatial data values by raising a bar at each point location so that the height of the bar is proportionate to its value.



**Fig.8 :** A 3-D plot showing annual precipitation at 105 weather stations in Idaho. A north-to-south decreasing trend is apparent in the plot.

**Q.7(e) List the various map manipulation techniques and explain any one.**

**[4]**

**Ans.:** **Map Manipulation**

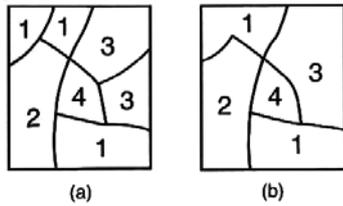
Tools are available in a GIS package for manipulating and managing maps in a database.

Dissolve aggregates features that have the same attribute value or values.

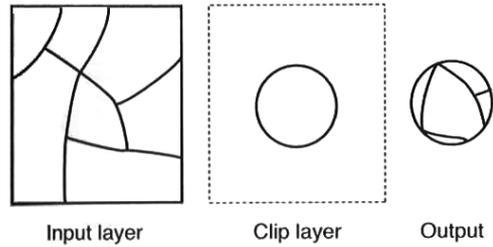
Clip creates a new layer that includes only those features of the input layer that fall within the area extent of the clip layer.

Append creates a new layer by piecing together two or more layers.

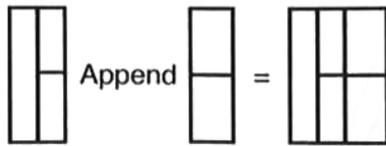
Select creates a new layer that contains features selected from a user-defined query expression.



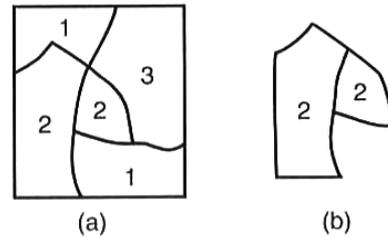
**Fig. 1 :** Dissolve removes boundaries of polygons that have the same attribute value in (a) and creates a simplified layer (b).



**Fig. 2 :** Clip creates an output that contains only those features of the input layer that fall within the area extent of the clip layer.



**Fig. 3 :** Append pieces together two adjacent layers into a single layer but does not remove the shared boundary between the layers.



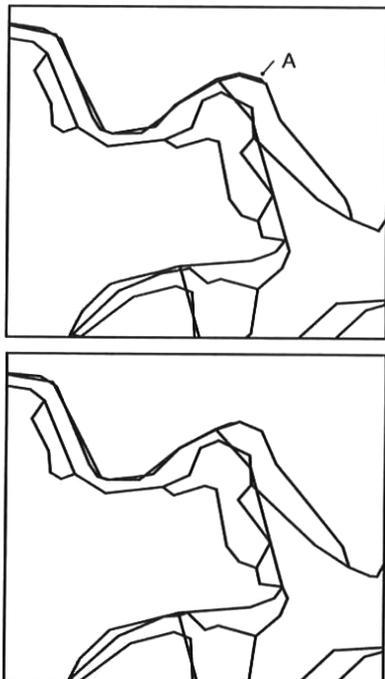
**Fig. 4 :** Select creates a new layer (b) with selected features from the input layer (a)

Eliminate creates a new layer by removing features that meet a user-defined query expression.

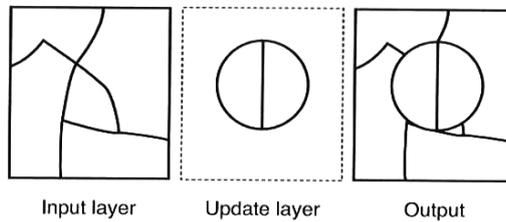
Update uses a "cut and paste" operation to replace the input layer with the update layer and its features.

Erase removes from the input layer those features that fall within the area extent of the erase layer.

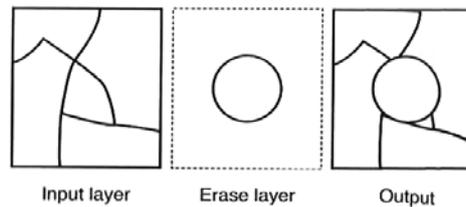
Split divides the input layer into two or more layers (Figure 8).



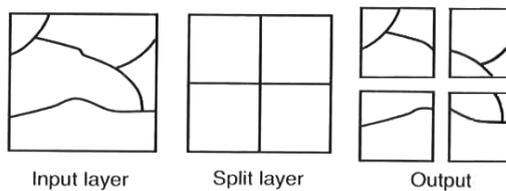
**Fig.5 :** Eliminate removes some small slivers along the top boundary(A)



**Fig. 6 :** Update replaces the input layer with the update layer and its features.



**Fig. 7 :** Erase removes features from the input layer that fall within the area extent of the erase layer.



**Fig. 8 :** Split uses the geometry of the split layer to divide the input layer into four separate layers.

**Q.7(f) Explain the Inverse distance Weighted Interpolation Method.**

**[4]**

**Ans.: Inverse Distance Weighted Interpolation**

Inverse distance weighted (IDW) interpolation is an exact method that enforces the condition that the estimated value of a point is influenced more by nearby known points than by those farther away. The general equation for the IDW method is :

$$z_0 = \frac{\sum_{i=1}^s z_i \frac{1}{d_i^k}}{\sum_{i=1}^s \frac{1}{d_i^k}} \quad \dots (1)$$

where  $z_0$  is the estimated value at point O,  $z_i$  is the z value at known point i,  $d_i$  is the distance between point i and point O, s is the number of known points used in estimation, and k is the specified power.

