UNIT-III

(Part – 2)
DATA ENTRY AND PREPARATION

5.1 SPATIAL DATA INPUT

Spatial data can be obtained from various sources. It can be collected from scratch, using direct spatial data acquisition techniques, or indirectly, by making use of existing spatial data collected by others.

- **Direct Spatial Data Capture**

One way to obtain spatial data is by direct observation of the relevant geographic phenomena. This can be done through ground-based field surveys, or by using remote sensors in satellites or airplanes. Many Earth sciences have developed their own survey techniques, as ground-based techniques remain the most important source for reliable data in many cases.

Data which is captured directly from the environment is known as primary data. With primary data the core concern in knowing its properties is to know the process, by which it was captured, the parameters of any instruments used and the rigour with which quality requirements were observed.

Remotely sensed imagery is usually not fit for immediate use, as various sources of error and distortion may have been present, and the imagery should first be freed from these. An image refers to raw data produced by an electronic sensor, which are not pictorial, but arrays of digital numbers related to some property of an object or scene, such as the amount of reflected light. For an image, no interpretation of reflectance values as thematic or geographic characteristics has taken place. When the reflectance values have been translated into some 'thematic' variable, called raster

Factors of cost and available time may be a hindrance in using existing remotely sensed images because previous projects sometimes have acquired data that may not fit the current project's purpose.

- **Indirect Spatial Data Capture**

In contrast to direct methods of data capture described above, spatial data can also be sourced indirectly. This includes data derived from existing paper maps through scanning, data digitized from a satellite image, processed data purchased from data capture firms or international agencies, and so on. This type of data is known as

secondary data.

Any data which is not captured directly from the environment is known as secondary data.

Digitizing

A traditional method of obtaining spatial data is through digitizing existing paper maps. This can be done using various techniques. Before adopting this approach, one must be aware that positional errors already in the paper map will further accumulate, and one must be willing to accept these errors.

There are two forms of digitizing: on-tablet and on-screen manual digitizing. In on-tablet digitizing, the original map is fitted on a special surface (the tablet), while in on-screen digitizing, a scanned image of the map is shown on the computer screen. In both of these forms, an operator follows the map's features with a mouse device, thereby tracing the lines, and storing location coordinates relative to a number of previously defined control points. The function of these points is to 'lock' a coordinate system onto the digitized data: the control points on the map have known coordinates, and by digitizing them we tell the system implicitly where all other digitized locations are. At least three control points are needed, but preferably more should be digitized to allow a check on the positional errors made.

Another set of techniques also works from a scanned image of the original map, but uses the GIS to find features in the image. These techniques are known as semi-automatic or automatic digitizing, depending on how much operator interaction is required. If vector data is to be distilled from this procedure, a process known as vectorization follows the scanning process. This procedure is less labour-intensive, but can only be applied on relatively simple sources.

Scanning

A scanner is an input device that illuminates a document and measures the intensity of the reflected light with a CCD array. The result is an image as a matrix of pixels, each of which holds an intensity value. Office scanners have a fixed maximum resolution, expressed as the highest number of pixels they can identify per inch; the unit is dots-per-inch (dpi). For manual on-screen digitizing of a paper map, a resolution of 200–300 dpi is usually sufficient, depending on the thickness of the thinnest lines. For manual on-screen digitizing of aerial photographs, higher resolutions are recommended—typically, at least 800 dpi.

Digitizing requires a resolution that results in scanned lines of at least three pixels wide to enable the computer to trace the centre of the lines and thus avoid displacements. For paper maps, a resolution of 300–600 dpi is usually sufficient. Automatic or semi-automatic tracing from aerial photographs is obtained through visual interpretation.

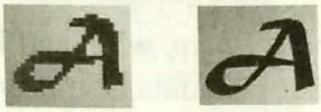
After scanning, the resulting image can be improved with various image processing techniques. It is important to understand that scanning does not result in a structured data set of classified and coded objects. Additional work is required to recognize features and to associate categories and other thematic attributes with them.

Vectorization

The process of distilling points, lines and polygons from a scanned image is called vectorization. As scanned lines may be several pixels wide, they are often first thinned to

retain only the centreline. The remaining centreline pixels are converted to series of (x, y) coordinate pairs, defining a polyline. Subsequently, features are formed and attributes are attached to them. This process may be entirely automated or performed semi-automatically, with the assistance of an operator. Pattern recognition methods—like Optical Character Recognition (OCR) for text—can be used for the automatic detection of graphic symbols and text.

Vectorization causes errors such as small spikes along lines, rounded corners, errors in T- & X-junctions, displaced lines or jagged curves. These errors are corrected in an automatic or interactive post-processing phase. The phases of the vectorization process are illustrated in Figure below



Input Image | After Vectorzation

Selecting a digitizing technique

The choice of digitizing technique depends on the quality, complexity and contents of the input document. Complex images are better manually digitized; simple images are better automatically digitized. Images that are full of detail and symbols—like topographic maps and aerial photographs are therefore better manually digitized.

The optimal choice may be a combination of methods. For example, contour line film separations can be automatically digitized and used to produce a DEM. Existing topographic maps must be digitized manually, but new, geometrically corrected aerial photographs, with vector data from the topographic maps displayed directly over it, can be used for updating existing data files by means of manual on-screen digitizing.

- **Obtaining Spatial Data Elsewhere**

Spatial data has been collected in digital form at increasing rate, stored in various databases by the individual producers for their own use and for commercial purposes. More and more of this data is being shared among GIS users. This is for several reasons. Some of this data is freely available, although other data is only available commercially, as is the case for most satellite imagery. High quality data remain both costly and time-consuming to collect and verify, as well as the fact that more and more GIS applications are looking at not just local, but national or even global processes. New technologies have played a key role in the increasing availability of geospatial data. As a result of this increasing availability, we have to be more careful that the data we have acquired is of sufficient quality to be used in analysis and decision making.

Clearinghouses and web portals

Spatial data can also be acquired from centralized repositories. More often those repositories are embedded in Spatial Data Infrastructures, which make the data available through what is sometimes called a spatial data clearinghouse. This is essentially a marketplace where data users can 'shop'. It will be no surprise that such markets for digital data have an entrance through the internet. The first entrance is typically formed by a web portal which categorizes all available data and provides a local search engine and links to data documentation also called metadata. It often also points to data viewing

and processing services. Standards-based geo-webservices have become the common technology behind such portal services.

Metadata

Metadata is defined as background information that describes all necessary information about the data itself. More generally, it is known as 'data about data'.

This includes :

- Identification information : Data source(s), time of acquisition, etc.
- Data quality information : Positional, attribute and temporal accuracy, lineage, etc.
- Entity and attribute information: Related attributes, units of measure, etc.

Metadata answer who, what, when, where, why, and how questions about all facets of the data made available. Maintaining metadata is an key part in maintaining data and information quality in GIS. This is because it can serve different purposes, from description of the data itself through to providing instructions for data handling. Depending on the type and amount of metadata provided, it could be used to determine the data sets that exist for a geographic location, evaluate whether a given data set meets a specified need, or to process and use a data set.

Data formats and standards

An important problem in any environment involved in digital data exchange is that of data formats and data standards. Different formats were implemented by different GIS vendors; different standards came about with different standardization committees. The phrase 'data standard' refers to an agreed upon way of representing data in a system in terms of content, type and format. The good news about both formats and standards is that there are many to choose from; the bad news is that this can lead to a range of conversion problems. Several metadata standards for digital spatial data exist, including the International Organization for Standardization (ISO) and the Open Geospatial Consortium (OGC) standards.

5.2 DATA QUALITY

GIS is being increasingly used for geospatial decision support applications, with increasing reliance on secondary data sourced through data providers or via the internet, through geo-webservices. The implications of using low-quality data in important decisions are potentially severe. There is also a danger that uninformed GIS users introduce errors by incorrectly applying geometric and other transformations to the spatial data held in their database.

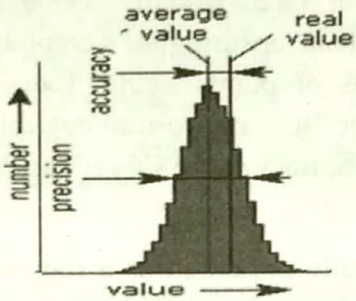
The main issues related to data quality in spatial data are positional, temporal and attribute accuracy, lineage, completeness, and logical consistency.

- **Accuracy and Precision**

Accuracy should not be confused with precision, which is a statement of the smallest unit of measurement to which data can be recorded. In conventional surveying and mapping practice, accuracy and precision are closely related. Instruments with an appropriate precision are employed and surveying methods chosen, to meet specified accuracy tolerances. In GIS, however, the numerical precision of computer processing and storage usually exceeds the accuracy of the data. This can give rise to so-called spurious

accuracy, for example calculating area sizes to the nearest m2 from coordinates obtained by digitizing a 1: 50, 000 map.

Using graphs that display the probability distribution of a measurement against the true value T, the relationship between accuracy and precision can be clarified. An accurate measurement has a mean close to the true value; a precise measurement has a sufficiently small variance.



- **Positional accuracy**

The surveying and mapping profession has a long tradition of determining and minimizing errors. This applies particularly to land surveying and photogrammetry, both of which tend to regard positional and height errors as undesirable. Cartographers also strive to reduce geometric and attribute errors in their products, and, in addition, define quality in specifically cartographic terms, for example quality of linework, layout, and clarity of text. It must be stressed that all measurements made with surveying and photogrammetric instruments are subject to error.

These include:

1. **Human errors in measurement** (e.g. reading errors) generally referred to as gross errors or blunders. These are usually large errors resulting from care lessness which could be avoided through careful observation, although it is never absolutely certain that all blunders have been avoided or eleminated.
2. **Instrumental or systematic errors** (e.g. due to misadjustment of instruments). This leads to errors that vary systematically in sign and/or magnitude, but can go undetected by repeating the measurement with the same instrument. Systematic errors are paticularly dangerous because they tend to accumulate.
3. **Random errors** caused by natural variations in the quantity being measured. These are effectively the errors that remain after blunders and systematic errors have been removed. They are usually small, and dealt with in least–squares adjustment. more general ways of quantifying positional accuracy using root mean square error (RMSE).

Measurement errors are generally described in terms of accuracy. In the case of spatial data, accuracy may relate not only to the determination of coordinates (positional error) but also to the measurement of quantitative attribute data. The accuracy of a single measurement can be defined as:

"The closeness of observations, computations or estimates to the true values or the values perceived to be true".

In Surveying and mapping, the 'truth' is usually taken to be a value obtained from a survey of higher accuracy, for example by comparing photogrammetric measurements with the coordinates and heights of a number of independent check points determined by field survey. Although it is useful for assessing the quality of definite objects, such as cadastral boundaries, this definition clearly has practical difficulties in the case of natural resource mapping where the 'truth' itself is uncertain, or boundaries of phenomena become fuzzy.

Prior to the availability of GPS, resource surveyors working in remote areas sometimes had to be content with ensuring an acceptable degree of relative accuracy among the measured positions of points within the surveyed area. If location and elevation are fixed with reference to a network of control points that are assumed to be free of error, then the absolute accuracy of the survey can be determined.

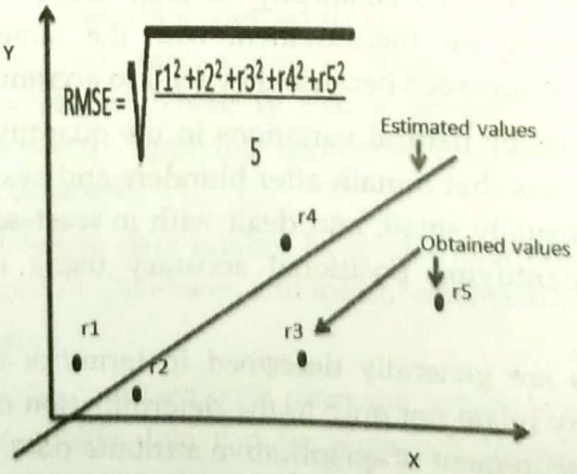
Root mean square error

Location accuracy is normally measured as a root mean square error (RMSE). The RMSE is similar to, but not to be confused with, the standard deviation of a statistical sample. The value of the RMSE is normally calculated from a set of check measurements (coordinate values from an independent source of higher accuracy for identical points). The differences at each point can be plotted as error vectors, as is done in Figure 5.3 for a single measurement. The error vector can be seen as having constituents in the x- and y-directions, which can be recombined by vector addition to give the error vector representing its locational error. For each checkpoint, the error vector has components δx and δy. The observed errors should be checked for a systematic error component, which may indicate a (possibly repairable) lapse in the measurement method. Systematic error has occurred when

$\delta x = 0$ or $\delta y = 0$.

The systematic error δx in x is then defined as the average deviation from the true value:

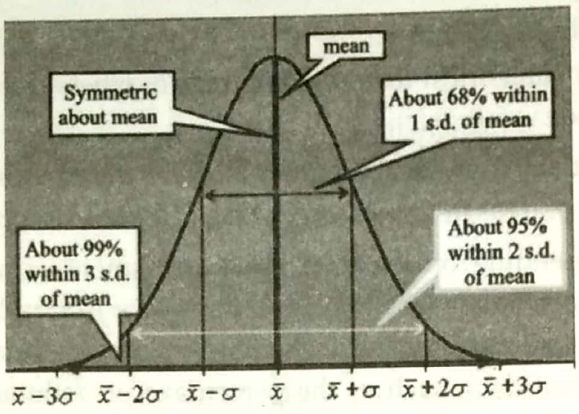
$$\delta \overline{x} = \frac{1}{n} \sum_{i=1}^{n} \delta x_i.$$



- **Accuracy tolerances**

Many kinds of measurement can be naturally represented by a bell-shaped probability density function p. This function is known as the normal (or Gaussian)

distribution of a continuous, random variable, in the figure indicated as Y. It shape is determined by two parameters: μ, which is the mean expected value for Y , and σ which is the standard deviation of Y . A small σ leads to a more attenuated bell shape.

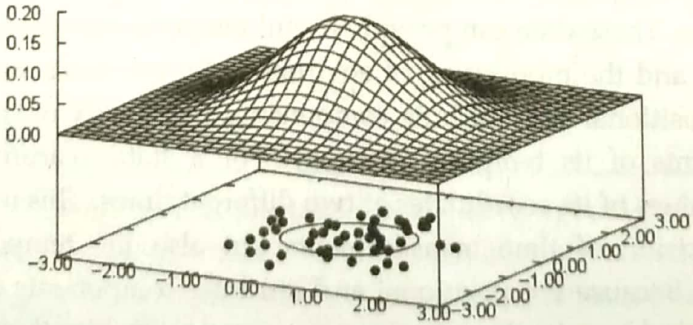


Any probability density function p has the characteristic that the area between its curve and the horizontal axis has size 1. Probabilities P can be inferred from p as the size of an area under p's curve. Figure above, for instance, depicts P (x - σ ≤ Y ≤ x - σ), i.e. the probability that the value for Y is within distance σ from μ.

In a normal distribution this specific probability for Y is always 0.6826.

The RMSE can be used to assess the probability that a particular set of measurements does not deviate too much from, i.e. is within a certain range of, the 'true' value. In the case of coordinates, the probability density function often is considered to be that of a two-dimensional normally distributed variable. The three standard probability values associated with this distribution are:

- 0.50 for a circle with a radius of 1.1774 mx around the mean (known as the circular error probable, CEP);
- 0.6321 for a circle with a radius of 1.412 mx around the mean (known as the root mean square error, RMSE);
- 0.90 for a circle with a radius of 2.146 mx around the mean (known as the circular map accuracy standard, CMAS).



The RMSE provides an estimate of the spread of a series of measurements around their (assumed) 'true' values. It is therefore commonly used to assess the quality of transformations such as the absolute orientation of photogrammetric models or the spatial referencing of satellite imagery. The RMSE also forms the basis of various statements for reporting and verifying compliance with defined map accuracy tolerances. An example is the American National Map Accuracy Standard, which states that:

"No more than 10% of well-defined points on maps of 1:20,000 scale or greater may be in error by more than 1/30 inch."

- **Attribute Accuracy**

Two types of attribute accuracies, related to the type of data it is dealing with:

- For nominal or categorical data, the accuracy of labeling (for example the type of land cover, road surface, etc).
- For numerical data, numerical accuracy (such as the concentration of pollutants in the soil, height of trees in forests, etc).

It follows that depending on the data type, assessment of attribute accuracy may range from a simple check on the labelling of features—for example, is a road classified as a metalled road actually surfaced or not?—to complex statistical procedures for assessing the accuracy of numerical data, such as the percentage of pollutants present in the soil. When spatial data are collected in the field, it is relatively easy to check on the appropriate feature labels. In the case of remotely sensed data, however, considerable effort may be required to assess the accuracy of the classification procedures. This is usually done by means of checks at a number of sample points. The field data are then used to construct an error matrix (also known as a confusion or misclassification matrix) that can be used to evaluate the accuracy of the classification. An example is provided in Table below,

	Urban	Agriculture	Range	Forest	Water	Total
Urban	**310**	9	18	23	18	378
Agriculture	61	**1051**	92	147	12	1363
Range	12	32	**561**	86	17	708
Forest	23	87	218	**1202**	8	1538
Water	11	7	12	27	**270**	327
Total	417	1186	901	1485	325	**3394**

- **Temporal Accuracy**

Spatial data sets captured through remotely sensed data has increased enormously over the last decade. These data can provide useful temporal information such as changes in land ownership and the monitoring of environmental processes such as deforestation. Analogous to its positional and attribute components, the quality of spatial data may also be assessed in terms of its temporal accuracy. For a static feature this refers to the difference in the values of its coordinates at two different times. This includes not only the accuracy and precision of time measurements but also the temporal consistency of different data sets. Because the positional and attribute components of spatial data may change together or independently, it is also necessary to consider their temporal validity. For example, the boundaries of a land parcel may remain fixed over a period of many years whereas the ownership attribute may change more frequently.

- **Lineage**

Lineage describes the history of a data set. In the case of published maps, some lineage information may be provided as part of the metadata, in the form of a note on the data sources and procedures used in the compilation of the data.

Examples include the date and scale of aerial photography, and the date of field verification. For digital data sets, however, lineage may be defined as: "that part of the data quality statement that contains information that describes the source of observations or materials, data acquisition and compilation methods, conversions, transformations, analyses and derivations that the data has been subjected to, and the assumptions and criteria applied at any stage of its life."

All of these aspects affect other aspects of quality, such as positional accuracy. If no lineage information is available, it is not possible to adequately evaluate the quality of a data set in terms of 'fitness for use'.

- **Completeness**

Completeness refers to whether there are data lacking in the database compared to what exists in the real world. Essentially, it is important to be able to assess what does and what does not belong to a complete dataset as intended by its producer. It might be incomplete (i.e. it is 'missing' features which exist in the real world), or overcomplete (i.e. it contains 'extra' features which do not belong within the scope of the data set as it is defined).

Completeness can relate to spatial, temporal, or thematic aspects of a data set. For example, a data set of property boundaries might be spatially incomplete because it contains only 10 out of 12 suburbs; it might be temporally incomplete because it does not include recently subdivided properties; and it might be thematically over complete because it also includes building footprints.

- **Logical consistency**

For any particular application, (predefined) logical rules concern:

- The compatibility of data with other data in a data set (e.g. in terms of data format),
- The absence of any contradictions within a data set,
- The topological consistency of the data set, and
- The allowed attribute value ranges, as well as combinations of attributes.

For example, attribute values for population, area, and population density must agree for all entities in the database. The absence of any inconsistencies does not necessarily imply that the data are accurate.

5.3 DATA PREPARATION

Spatial data preparation aims to make the acquired spatial data fit for use. Images may require enhancements and corrections of the classification scheme of the data. Vector data also may require editing, such as the trimming of overshoots of lines at intersections, deleting duplicate lines, closing gaps in lines, and generating polygons. Data may require conversion to either vector format or raster format to match other data sets which will be used in the analysis. Additionally, the data preparation process includes associating attribute data with the spatial features through either manual input or reading digital attribute files into the GIS/DBMS.

The intended use of the acquired spatial data may require only a subset of the original data set, as only some of the features are relevant for subsequent analysis or

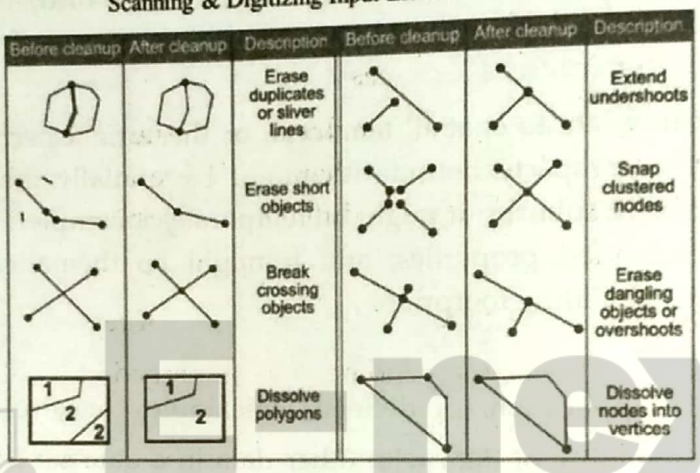
subsequent map production. In these cases, data and/or cartographic generalization can performed on the original data set.

- **Data Checks and Repairs**

Acquired data sets must be checked for quality in terms of the accuracy, consistency and completeness parameters discussed above. Often, errors can be identified automatically, after which manual editing methods can be applied to correct the errors. Alternatively, some software may identify and automatically correct certain types of errors. Below, we focus on the geometric, topological, and attribute components of spatial data.

'Clean-up' operations are often performed in a standard sequence. For example, crossing lines are split before dangling lines are erased, and nodes are created at intersections before polygons are generated. These are illustrated in Table below.

Scanning & Digitizing Input Errors



With polygon data, one usually starts with many polylines, in an unwieldy format known as spaghetti data that are combined in the first step. This results in fewer polylines with more internal vertices. Then, polygons can be identified. Sometimes, polylines that should connect to form closed boundaries do not, and therefore must be connected (either manually or automatically).

Rasterization or vectorization

Vectorization produces a vector data set from a raster. We have looked at this in some sense already: namely in the production of a vector set from a scanned image. Another form of vectorization takes place when we want to identify features or patterns in remotely sensed imagery. The keywords here are feature extraction and pattern recognition, which are dealt with in Principles of Remote Sensing.

If much or all of the subsequent spatial data analysis is to be carried out on raster data, one may want to convert vector data sets to raster data. This process is known as rasterization. It involves assigning point, line and polygon attribute values to raster cells that overlap with the respective point, line or polygon. To avoid information loss, the raster resolution should be carefully chosen on the basis of the geometric resolution. A cell size which is too large may result in cells that cover parts of multiple vector features, and then ambiguity arises as to what value to assign to the cell. If, on the other hand, the cell size is too small, the file size of the raster may increase significantly.

Rasterization itself could be seen as a 'backwards step': firstly, raster boundaries are

only an approximation of the objects' original boundary. Secondly, the original 'objects' can no longer be treated as such, as they have lost their topological properties. Often the reason for rasterisation is because it facilitates easier combination with other data sources also in raster formats, and/or because there are several analytical techniques which are easier to perform upon raster data. An alternative to rasterization is to not perform it during the data preparation phase, but to use GIS rasterization functions on the-fly, that is when the computations call for it. This allows keeping the vector data and generating raster data from them when needed.

Topology generation

Topological relations may sometimes be needed, for instance in networks, e.g. the questions of line connectivity, flow direction, and which lines have over- and underpasses. For polygons, questions that may arise involve polygon inclusion: Is a polygon inside another one, or is the outer polygon simply around the inner polygon? Many of these questions are mostly questions of data semantics, and can therefore usually only be answered by a human operator.

- **Combining data from multiple sources**

A GIS project usually involves multiple data sets, so the next step addresses the issue of how these multiple sets relate to each other. There are four fundamental cases to be considered in the combination of data from different sources:

1. They may be about the same area, but differ in accuracy,
2. They may be about the same area, but differ in choice of representation,
3. They may be about adjacent areas, and have to be merged into a single data set.
4. They may be about the same or adjacent areas, but referenced in different coordinate systems.

The following may be the situation :

- ❖ Differences in accuracy
- ❖ Differences in representation
- ❖ Merging data sets of adjacent areas
- ❖ Differences in coordinate systems

Differences in accuracy

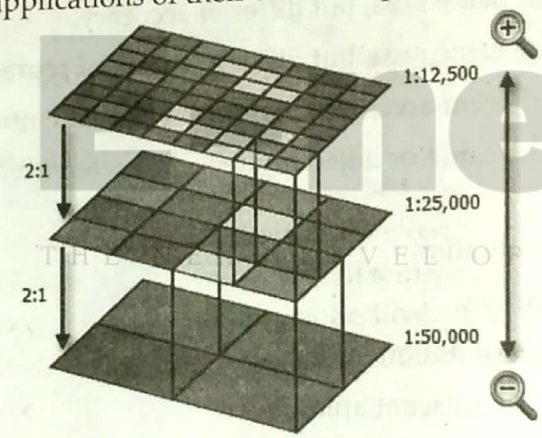
These are clearly relevant in any combination of data sets which may themselves have varying levels of accuracy. Images come at a certain resolution, and paper maps at a certain scale. This typically results in differences of resolution of acquired data sets, all the more since map features are sometimes intentionally displaced to improve readability of the map. For instance, the course of a river will only be approximated roughly on a small-scale map, and a village on its northern bank should be depicted north of the river, even if this means it has to be displaced on the map a little bit. The small scale causes an accuracy error. If we want to combine a digitized version of that map, with a digitized version of a large-scale map, we must be aware that features may not be where they seem to be. Analogous examples can be given for images at different resolutions.

There can be good reasons for having data sets at different scales. A good example is found in mapping organizations; European organizations maintain a single source database that contains the base data. This database is essentially scale-less and contains all

data required for even the largest scale map to be produced. For each map scale that the mapping organization produces, they derive a separate database from the foundation data. Such a derived database may be called a cartographic database as the data stored are elements to be printed on a map, including, for instance, data on where to place name tags, and what color to give them. This may mean the organization has one database for the larger scale ranges (1:5,000–1:10,000) and other databases for the smaller scale ranges. They maintain a multi-scale data environment.

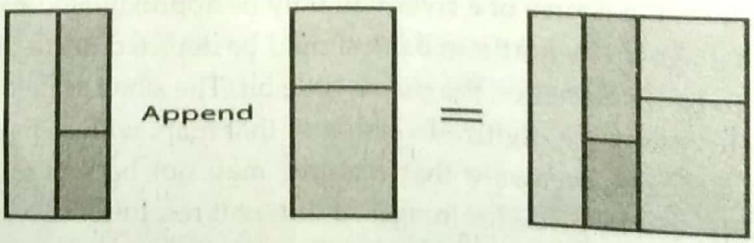
Differences in representation

Some advanced GIS applications require the possibility of representing the same geographic phenomenon in different ways. These are called multi representation systems. The production of maps at various scales is an example, but there are numerous others. The commonality is that phenomena must sometimes be viewed as points, and at other times as polygons. For example, a small-scale national road network analysis may represent villages as point objects, but a nation-wide urban population density study should regard all municipalities as represented by polygons. The complexity that this requirement entails is that the GIS or the DBMS must keep track of links between different representations for the same phenomenon, and must also provide support for decisions as to which representations to use in which situation. The links between various representations for the same object maintained by the system allows switching between them, and many fancy applications of their use seem possible.



Merging data sets of adjacent areas

When individual data sets have been prepared as described above, they some- times have to be matched into a single 'seamless' data set, whilst ensuring thatthe appearance of the integrated geometry is as homogeneous as possible. Edge matching is the process of joining two or more map sheets, for instance, after they have separately been digitized.



Merging adjacent data sets can be a major problem. Some GIS functions, such as line smoothing and data clean-up (removing duplicate lines) may have to be performed. Some

GIS have merge or edge-matching functions to solve the problem arising from merging adjacent data. At the map sheet edges, feature representations have to be matched in order for them to be combined. Coordinates of the objects along shared borders are adjusted to match those in the neighboring data sets. Mismatches may still occur, so a visual check, and interactive editing is likely to be required.

Differences in coordinate systems

Map projections provide means to map geographic coordinates onto a flat surface (for map production), and vice versa. It may be the case that data layers which are to be combined or merged in some way are referenced in different coordinate systems, or are based upon different datums. As a result, data may need coordinate transformation, or both a coordinate transformation and datum transformation. It may also be the case that data has been digitized from an existing map or data layer. In this ase, geometric transformations help to transform device coordinates (coordinates from digitizing tablets or screen coordinates) into world coordinates (geographic coordinates, meters, etc.).

Other data preparation functions

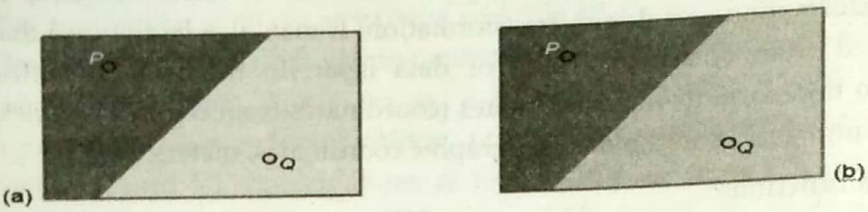
A range of other data preparation functions exist that support conversion or adjustment of the acquired data to format requirements that have been defined for data storage purposes. These include:

- Format transformation functions. These convert between data formats of different systems or representations, e.g. reading a DXF file into a GIS. Although we will not focus on the technicalities here, the user should be warned that conversions from one format to another may cause problems. The reason is that not all formats can capture the same information, and therefore conversions often mean loss of information. If one obtains a spatial data set in format F , but needs it in format G (for instance because the locally preferred GIS package requires it), then usually a conversion function can be found, often within the same GIS software package. The key to successful conversion is to also find an inverse conversion, back from G to F , and to ascertain whether the double conversion back to F results in the same data set as the original. If this is the case, both conversions are not causing information loss, and can safely be applied.
- Graphic element editing. Manual editing of digitized features so as to correct errors, and to prepare a clean data set for topology building.
- Coordinate thinning. A process that is often applied to remove redundant or excess vertices from line representations, as obtained from digitizing.

5.4 POINT DATA TRANSFORMATION

We may have captured a sample of points (or acquired a dataset of such points), but wish to derive a value for the phenomenon at another location or for the whole extent of our study area. We may want to transform our points into other representations in order to facilitate interpretation and/or integration with other data. Examples include defining homogeneous areas (polygons) from our point data, or deriving contour lines. This is generally referred to as interpolation, i.e. the calculation of a value from 'surrounding' observations. The principle of spatial autocorrelation plays a central part in the process of interpolation.

In order to predict the value of a point for a given (x, y) location, we could simply find the 'nearest' known value to the point, and assign that value. This is the simplest form of interpolation, known as nearest-neighbour interpolation. We might instead choose to use the distance that points are away from (x, y) to weight their importance in our calculation. In some instances we may be dealing with a data type that limits the type of interpolation. A fundamental issue in this respect is what kind of phenomena we are considering: is it a discrete field—such as geological units, for instance—in which the values are of a qualitative nature and the data is categorical, or is it a continuous field—like elevation, temperature, or salinity— in which the values are of quantitative nature, and represented as continuous measurements?



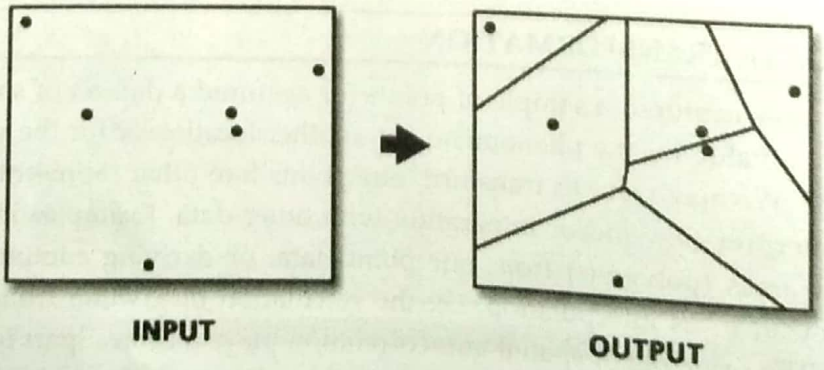
field representation obtained from two point measurements: (a) for qualitative (categorical), and (b) for quantitative (continuous) point measurements. The value measured at *P* is represented as dark green, that at *Q* as light green.

How we represent a field constructed from point measurements in the GIS also depends on the above distinction. A discrete field can either be represented as a classified raster or as a polygon data layer, in which each polygon has been assigned a (constant) field value. A continuous field can be represented as an unclassified raster, as an isoline (thus, vector) data layer, or perhaps as a TIN. Some GIS software only provide the option of generating raster output, requiring an intermediate step of raster to vector conversion. The choice of representation depends on what will be done with the data in the analysis phase.

- **Interpolating Discrete Data**

If we are dealing with discrete (nominal, categorical or ordinal) data, we are effectively restricted to using nearest-neighbour interpolation. This is the situation shown in Figure below, though usually we would have many more points. In a nearest-neighbour interpolation, each location is assigned the value of the closest measured point. Effectively, this technique will construct 'zones' around the points of measurement, with each point belonging to a zone assigned the same value. Effectively, this represents an assignment of an existing value (or category) to a location.

If the desired output was a polygon layer, we could construct Thiessen polygons around the points of measurement. The boundaries of such polygons, by definition, are the locations for which more than one point of measurement is the closest point. If the desired output was in the form of a raster layer, we could rasterize the Thiessen polygons.



- **Interpolating Continuous Data**

Interpolation of values from continuous measurements is significantly more complex. Since the data are continuous, we can make use of measured values for interpolation. There are many continuous geographic fields—elevation, temperature and ground water salinity are just a few examples. Continuous fields are represented as rasters, and we will almost by default assume that they are.

The main alternative for continuous field representation is a polyline vector layer, in which the lines are isolines. We will also address these issues of representation below. The aim is to use measurements to obtain a representation of the entire field using point samples. In this section we outline four techniques to do so:

1. Trend surface fitting using regression,
2. Triangulation,
3. Spatial moving averages using inverse distance weighting,
4. Kriging.

Trend surface fitting

In trend surface fitting, the assumption is that the entire study area can be represented by a formula f(x, y) that for a given location with coordinates (x, y) will give us the approximated value of the field in that location.

The key objective in trend surface fitting is to derive a formula that best describes the field. Various classes of formulæ exist, with the simplest being the one that describes a flat, but tilted plane:

$$f(x, y) = c_1 \cdot x + c_2 \cdot y + c_3.$$

The field under consideration can be best approximated by a tilted plane, then the problem of finding the best plane is the problem of determining best values for the coefficients c_1, c_2 and c_3. This is where the point measurements earlier obtained become important. Statistical techniques known as regression techniques can be used to determine values for these coefficients ci that best fit with the measurements. A plane will be fitted through the measurement that makes the smallest overall error with respect to the original measurements. We have used the same set of point measurements, with four different approximation functions. Part (a) has been determined under the assumption that the field can be approximated by a tilted plane, in this case with a downward slope to the southeast. The values found by regression techniques were:

$c_1 = -1.83934$, $c_2 = 1.61645$ and $c_3 = 70.8782$,

giving $f(x, y) = -1.83934 \cdot x + 1.61645 \cdot y + 70.8782$.

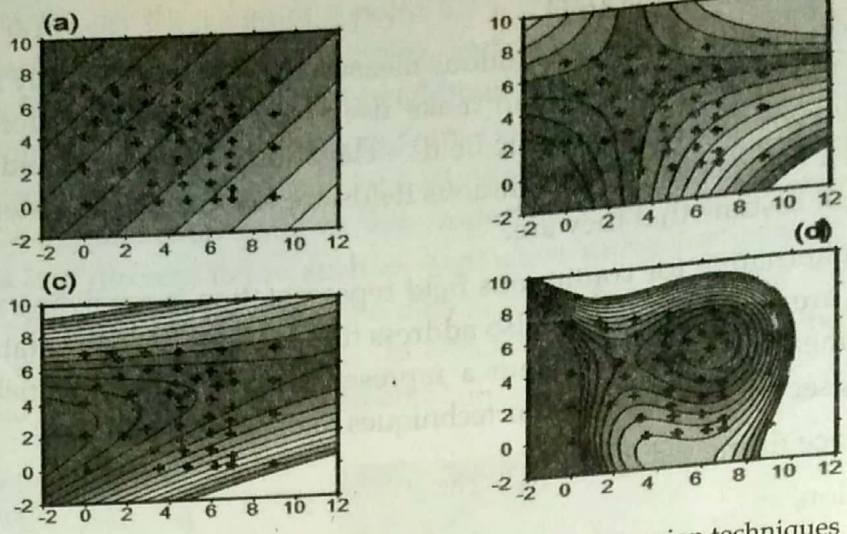


Fig. : Various global trend surfaces obtained from regression techniques :
a) simple titled pland; b) bilinear saddle; c) quadratic surface; cubic surface. Values range from white (low), via black and light grey to dark grey (high)

Not all fields are representable as simple, tilted planes. Sometimes, the theory of the application domain will dictate that the best approximation of the field is a more complicated, higher-order polynomial function. The simplest extension from a tilted plane, that of bilinear saddle, expresses some dependency between the x and y dimensions:

$$f(x, y) = c_1 \cdot x + c_2 \cdot y + c_3 \cdot xy + c_4.$$

This is illustrated in part (b). A further step up the ladder of complexity is to consider quadratic surfaces, described by:

$$f(x, y) = c_1 \cdot x_2 + c_2 \cdot x + c_3 \cdot y_2 + c_4 \cdot y + c_5 \cdot x_y + c_6.$$

The objective is to find six values for our coefficients that best match with the measurements. A bilinear saddle and a quadratic surface have been fitted through our measurements in (b) and (c), respectively.

Part (d) of the figure illustrates the most complex formula of the surfaces in Figure, the cubic surface. It is characterized by the following formula:

$$f(x, y) = c_1 \cdot x_3 + c_2 \cdot x_2 + c_3 \cdot x +$$
$$c_4 \cdot y_3 + c_5 \cdot y_2 + c_6 \cdot y +$$
$$c_7 \cdot x_2y + c_8 \cdot xy_2 + c_9 \cdot xy + c_{10}.$$

Trend surface fitting is a useful technique of continuous field approximation, though determining the 'best fit' values for the coefficients ci is a time-consuming operation, especially with many point measurements. Once these best values have been determined, we know the formula, making it possible to compute an approximated value for any location in the study area. It is possible to use trend surfaces for both global and local trends. Global trend surface fitting is based on the assumption that the entire study area can be approximated by the same mathematical surface. However in many cases, the assumption that a single formula can describe the field for the entire study area is an unrealistic one. Capturing all the fluctuation of a natural geographic field in a reasonably sized study area, demands polynomials of extreme orders, and these quickly become computationally impossible to decipher. It should also be noted that the spatial distribution of sample measures have significant effect on the shape of the fitting function.

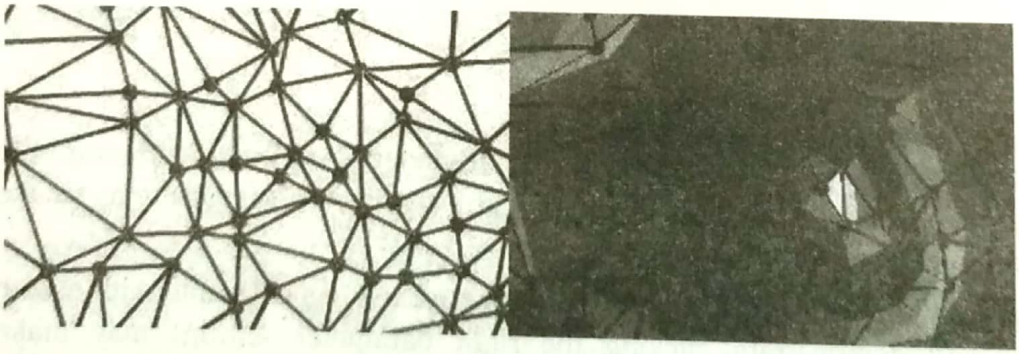
This is especially true for locations that are within the study area, but outside of the area within which the measurements fall. These may be subject to a so-called edge effect, meaning that the values obtained from the approximation function for edge locations may be rather nonsensical. The reader is asked to judge whether such edge effects have taken place in. For these reasons, it is often useful to partition the study area into parts that may actually be polynomially approximated. The decision of how to partition the study area must be taken with care, and must be guided by domain expertise. Once we have identified the parts, we may apply the trend surface fitting technique, and obtain an approximation polynomial for each part.

Local trend surface fitting is not a popular technique in practical applications, because they are relatively difficult to implement, and other techniques such as moving windows are better for the representation and identification of local trends.

If we know the polynomial, it is relatively simple to generate a raster layer, given an appropriate cell resolution and an approximation function for the cell's value. In some cases it is more accurate to assign the average of the computed values for all of the cell's corner points. In order to generate a vector layer representing this data, isolines can be derived, for a given set of intervals.

Triangulation

Another way of interpolating point measurements is by triangulation. Triangulated Irregular Networks (TINs) technique constructs a triangulation of the study area from the known measurement points. Preferably, the triangulation should be a Delaunay triangulation. After having obtained it, we may define for which values of the field we want to construct isolines. For instance, for elevation, we might want to have the 100 m-isoline, the 200 m-isoline, and so on. For each edge of a triangle, a geometric computation can be performed that indicates which isolines intersect it, and at what positions they do so. A list of computed locations, all at the same field value, is used by the GIS to construct the isoline. This is illustrated in Figure below



Moving averages using inverse distance weighting (IDW)

Moving window averaging attempts to directly derive a raster dataset from a set of sample points. This is why it is sometimes also called 'gridding'. The principle behind this technique is illustrated in Figure below. The cell values for the output raster are computed one by one. To achieve this, a 'window' (also known as a kernel) is defined, and initially placed over the top left raster cell. Measurement points falling inside the window contribute to the averaging computation, those outside the window do not. This is why moving window averaging is said to be a local interpolation method. After the cell value

is computed and assigned to the cell, the window is moved one cell to the right, and the computations are performed for that cell. Successively, all cells of the raster are visited in this way.

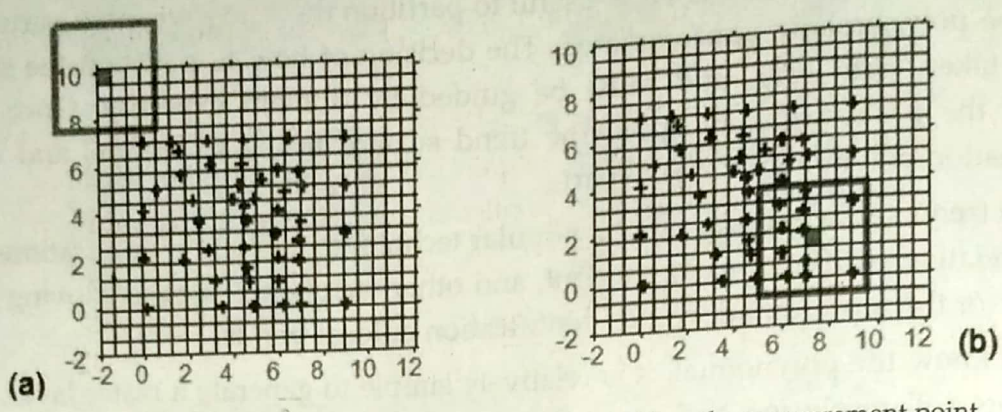


Fig. : The principle of moving window averaging. In blue, the measurement point. A virtual window is moved over the raster cells one by one, and some averaging function computes a field value for the cell, using measurements within the window.

In part (b) of the figure, the 295th cell value out of the 418 in total, is being computed. This computation is based on eleven measurements, while that of the first cell had no measurements available. Where this is the case, the cell should be assigned a value that signals this 'non-availability of measurements'. The principle of spatial autocorrelation suggests that measurements closer to the cell centre should have greater influence on the predicted value than those further away. In order to account for this, a distance factor can be brought into the averaging function. Functions that do this are called inverse distance weighting functions (IDW). This is one of the most commonly used functions in interpolating spatial data.

Let us assume that the distance from measurement point i to the cell centre is denoted by di. General case the formula is:

$$\sum_{i=1}^{n} \frac{mi}{d_i^p} \Big/ \sum_{i=1}^{n} \frac{1}{d_i^p}$$

Moving window averaging has many parameters. As experimentation with any GIS package will demonstrate, picking the right parameter settings may make quite a difference for the resulting raster.

Moving window averaging has many parameters. As experimentation with any GIS package will demonstrate, picking the right parameter settings may make quite a difference for the resulting raster. We discuss some key parameters below.

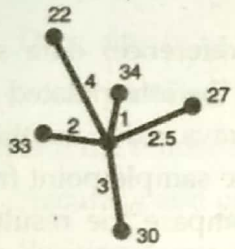
- **Raster resolution :** Too large a cell size will smooth the function too much, removing local variations; too small a cell size will result in large clusters of equally valued cells, with little added value.

- **Shape/size of window :** Most procedures use square windows, but rectangular, circular or elliptical windows are also possible. These can be useful in cases where the measurement points are distributed regularly at fixed distance over the study area, and the window shape must be chosen to ensure that each raster cell will have its window include the same number of measurement points. The size of the window is another

important matter. Small windows tend to exaggerate local extreme values, while large windows have a smoothing effect on the predicted field values.

❖ **Selection criteria :** Not necessarily all measurements within the window need to be used in averaging. We may choose to select use at most five, (nearest) measurements, or we may choose to only generate a field value if more than three measurements are in the window.

❖ **Averaging function :** A final choice is which function is applied to the selected measurements within the window. It is possible to use different distance–weighting functions, each of which will influence the calculation of the resulting value.



$$Z(x) = \frac{\Sigma w_i z_i}{\Sigma w_i} = \frac{\frac{34}{1^2} + \frac{33}{2^2} + \frac{27}{2.5^2} + \frac{30}{3^2} + \frac{22}{4^2}}{\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{2.5^2} + \frac{1}{3^2} + \frac{1}{4^2}} = 32.38$$

Kriging

Kriging was originally developed my mining geologists attempting to derive accurate estimates of mineral deposits in a given area from limited sample measurements. It is an advanced interpolation technique belonging to the field of geostatistics , which can deliver good results if applied properly and with enough sample points. Kriging is usually used when the variation of an attribute and/or the density of sample points is such that simple methods of interpolation may give unreliable predictions.

Kriging is based on the notion that the spatial change of a variable can be described as a function of the distance between points. It is similar to IDW interpolation, in that it the surrounding values are weighted to derive a value for an unmeasured location. However, the kriging method also looks at the overall spatial arrangement of the measured points and the spatial correlation between their values, to derive values for an unmeasured location.

The first step in the kriging procedure is to compare successive pairs of point measurements to generate a semi-variogram.

In the second step, the semi-variogram is used to calculate the weights used in interpolation. Although kriging is a powerful technique, it should not be applied without a good understanding of geostatistics, including the principle of spatial autocorrelation. It should be noted that there is no single best interpolation method, since each method has advantages and disadvantages in particular contexts.

As a general guide, the following questions should be considered in selecting an appropriate method of interpolation:

- For what type of application will the results be used?
- What data type is being interpolated (e.g. categorical or continuous)?
- What is the nature of the surface (for example, is it a 'simple' or complex surface)?
- What is the scale and resolution of the data (for example, the distance between sample points)?

It is important to carry out an evaluation of the data set before interpolation takes place. In such an evaluation, one of the main goals is to establish whether there are any existing trends in the data set that may influence interpolation. Trend surfaces can be fitted to the existing data, followed by an examination of the differences between the existing data and the resulting trend surface. It is also important to assess the spatial variability of the existing data. This can be achieved with simple moving window techniques or some other kind of linear interpolation.

Finally, in order to establish the effect of the interpolation parameters on the result, different sets of interpolation parameters can be employed, and the results of these compared.

One way to evaluate results is to use an independent (reference) data set and calculate the difference between the value from this data set and the interpolated surface at each location. However, 'independent' datasets do not always exist. In this case, another option is to run a series of interpolations, leaving out one sample point from the original data for each run. Again, this makes it possible to compare the results from interpolation with a known value. If the differences ('errors') found using this method are unacceptable, either there are not enough sample points for an accurate result, or one or more of the parameters used for the interpolation is incorrect.

QUESTIONS

1. Write a note on ellipsoid and Geoid reference model of the Earth surface.
2. Write a note on
   a) Ellipsoid
   b) Local horizontal datum
   c) Global horizontal datum
   d) Map Projections
   e) 2D Geometric Coordinate
   f) 3D Geometric Coordinate
   g) 2D Geographic Coordinate
   h) 3D Geographic Coordinate
   i) 2D Cartesian coordinate
   j) 2D Polar Coordinate
   k) Coordinate Transformation
   l) Datum Transformation
3. What are the different classifications of Map Projections? Explain.
4. Write a short note on satellite based positioning.
5. What is absolute positioning? Explain using suitable diagram.
6. Discuss about the errors in absolute positioning related to space segment.
7. Discuss about the errors in absolute positioning related to medium.
8. Discuss about the errors in absolute positioning related to receiver's environment.

9. Discuss about the errors in absolute positioning related to relative geometry of satellite.
10. Explain the following positioning technologies
    a) GPS
    b) GLONASS
    c) Galileo
    d) IRNSS and GAGAN
11. Write a note on direct spatial data capture.
12. Explain the following indirect spatial data capture technique:
    a) Digitizing
    b) vectorization
13. Explain the following data qualities
    a) Accuracy and precision
    b) Positional accuracy
    c) attribute accuracy
    d) Temporal accuracy
14. Why there is a need for combining data from different sources in GIS? Explain each using suitable diagram.
15. Explain the following interpolating technique for continuous data
    a) Trend surface fitting
    b) Triangulation
    c) Krigging