ISBN: 978-81-968830-3-4

2. Process in GIS

Dr. Jagruti R. Chavan

Assistant professor,
Department of Environmental Science,
KTHM College,
Nashik, Maharashtra.

Abstract:

Geographic Information Systems (GIS), highlighting advancements in spatial data analysis, mapping, and decision-making tools. Emphasizing the integration of diverse data sources, the evolving GIS process facilitates enhanced geospatial insights, contributing to improved resource management, urban planning, and environmental monitoring. You can save a basic map using a geographical reference system, such as longitude or latitude, and then add extra layers of other information using a GIS (geographic or geospatial information system). It is significant that the same geographical referencing is used to identify that information. The GIS then enables linking between the various levels, or themes as they are also known. The statistical and analytical capabilities offered by the GIS may then be used to conduct an analysis of the data. It is feasible to produce compelling, visual representations of data by doing spatial analysis on appropriately coded data. Without the use of GIS technology, these representations frequently make patterns and trends visible that could have gone overlooked otherwise.

Keywords:

GIS technology, GIS data, Process.

2.1 Data Pre-processing and Manipulation:

Under this section, we will discuss a wide range of tasks that a GIS may be required to complete in order to convert digitally mapped data into the format that is required for producing the necessary maps or securely enabling any ensuing data analysis. This basically indicates that the original digital data may need to be modified in some way, such as by being corrected, updated, refined, or transformed in a desired way. Many of the tasks may be accomplished by other kinds of software, such image processing programmes. Because a GIS may do pre-processing, there is a lot of room for the user to "interactively experiment" with the data. Consequently, it is possible to generate the relevant data for the task at hand. The specific algorithms that each GIS uses and the way the data is organised will determine how efficiently they can manipulate the data.

A. Data Validation and Editing
B. Structure Conversion
C. Geometric Conversion
D. Generalization and Classification

E. Integration
F. Map Enhancement
G. Interpolation
H. Buffer Generation
I. Data Searching and Retrieval

A. Data Validation and Editing:

This procedure essentially entails reviewing and editing any previously collected data, with the obvious goal of minimising mistakes. However, it is important to keep in mind that many GIS software programmes allow for the detection and correction of digitising errors as a pre-processing function. Editing is frequently possible and desirable to be performed immediately after data capture, i.e., as a final stage in the digitising process, in the case of digitised data. Programmes for checking the accuracy of all geometric, topological, and attribute data are also included in GIS software. For example, these programmes ensure that all graphical data is properly specified, attribute data does not exceed anticipated ranges, and impossible combinations of attributes do not exist. Data may be added, removed, relocated, joined, changed, and so on. Any of these data altering operations ought to be able to work with both graphical and textual data. Any final GIS output will be invalidated, or at the very least rendered less usable, if data changes at a subsequent processing stage result in error propagation and multiplication due to careless data verification.

B. Structure Conversion:

It may also be desirable to transform data from a raster to a vector structure, or vice versa, for various tasks. The lack of properly integrated GISs that can manage both raster and vector data with equal ease makes this vital. It is crucial to remember that precision will inevitably be lost during the vector to raster conversion (rasterising), a problem that will only get worse as the sinuosity of the lines and the size of the raster cells rise. In the raster to vector conversion process (also known as "vectorising"), the GIS software application "threads" a line across clusters of pixels using a unique "thinning" method. The construction of topological information and the identification of specific characteristics will be required as a result. These latter needs may necessitate significant operator involvement, however there are GIS tools that can calculate new nodes and linkages and create topology tables automatically.

C. Geometric Conversion:

When manipulating mapped digital data, it's crucial that everything adhere to the same geometric reference system if the data is to be combined in any way. Latitude and longitude coordinates are widely employed in small-scale mapping, but the Universal Transverse Mercator (UTM) system is the most used co-ordinate system in GIS. The ability to convert the utilised map reference system to a variety of different map projections or from one co-ordinate system to another is offered by almost all GIS applications. These procedures are sometimes referred to as corrections or transformations. The mathematical links between the various map projections, such as those related to angles, areas, directions, and distances, constitute the foundation for transformations. Registration is a simpler kind of geometric conversion. This is just a simple repositioning of one mapped view to match another, independent of any reference scheme. Maps may be simply rotated to different orientations and scale variations are easily handled with a straightforward multiplication function. The correction for distortions (rectification) is a more complicated operation that can be accomplished by the majority of GISs.

D. Generalization and Classification:

A wide range of operations fall under this broad category, all of which aim to modify the data in some manner so that it may be utilized for a certain purpose more readily, such as:

- Adding to data or removing unwanted data.
- Combining or separating data based on numerical or attribute values.
- Sorting or rearranging data according to user-defined or GIS-recommended classes. Choosing attribute value classes or making modifications to current classes are typically involved in this.
- Generalizing or smoothing linear data using data reduction methods, such as by thinning out coordinates in digitized lines (to significantly reduce the quantity of data stored).
- To make mapped surfaces simpler, lines can be removed or "dissolved".
- Spatial points, lines, and polygons may be given new properties.
- Labels, text, legends, and cartographic symbols can all be used to add annotations on maps.

E. Integration:

These adjustments entail combining two or more previously specified maps in one way or another to create new or altered mapped surfaces. The merging of two or more map layers by overlaying them, or the addition or subtraction of any number of raster or vector map layers to create a desired map, may be the most common operation under this subject. Existing data on a particular topic might be combined; for instance, a water qualitative map could be gradually constructed by combining maps of the water's salinity, pH, dissolved oxygen, temperature, etc.

These layers would be combined to form a new map with several polygons, each of which may contain a unique mix of water quality variables. Overlapping maps can result in serious issues. The creation of several new polygons, some of whose bounds should match those of the two maps being superimposed, is one example. The resultant map will inevitably include many so-called "slivers," or polygons, resulting from a poor match between borders or other lines, because the original data sources will be different, or the digitizing will be wrong.

These layers would be combined to form a new map with several polygons, each of which may contain a unique mix of water quality variables. Overlapping maps can result in serious issues. The creation of several new polygons, some of whose bounds should match those of the two maps being superimposed, is one example. The resultant map will inevitably include many so-called "slivers," or polygons, resulting from a poor match between borders or other lines, because the original data sources will be different, or the digitizing will be wrong.

F. Map Enhancement:

These GIS functions are only a set of activities that enable the final map's cartography to be refined; that is, at the manipulation stage, aspects related to map presentation may be enhanced. This might involve changing the map's layout or the placement of text elements like the key or title, adding an appropriate border, changing the width of the map's lines, changing the selected colours, changing the typefaces or font size, or all of the above. At this point, if a 3-D picture has been constructed, a land use classification could be applied to the image.

G. Interpolation:

This process is used to calculate values for any continuous (as opposed to discrete) "properties" at unsampled sites along a line or inside of a region. This must be based on current point observation data from the region (or along the line) that were obtained using reliable measuring and sampling methods. The challenge with interpolation is selecting the model that is most likely to result in accurate interpolations, i.e., a model that fits the data array and how true variability manifests itself. Numerous straightforward GIS interpolation techniques rely on the usage of weighting functions, which take into account the logical rule that close points should be given more weight in an interpolation than far ones. There are more complicated interpolation models available here, such as Thiessen polygons, kriging, Fourier series, or models that may be used to interpolate linear paths, such as the fitting of spline functions, which span 2-D arrays of data points. Centroid determination is a unique instance of interpolation. The GIS can determine the coordinate position of a polygon's centre using this function. Extrapolation is nothing more than the process of applying interpolation techniques to extend estimated trends outside of the scope of a particular research or region of interest.

H. Buffer Generation:

Since spatial distances are a central focus of GIS, it is usually helpful to identify "buffer zones" (zones of equal distance) surrounding a point or an area or along a line. The GIS may create buffers at any specified distance, and they can reflect elements like the maximum market radius surrounding a town, a legal no-go area, a noise-sensitive area, or a zone with certain economic rights. Although their usage may really be limited, mostly to isotropic surfaces, buffer zones are unquestionably very helpful in the construction of many geographic models.

I. Data Searching and Retrieval:

The two operations of obtaining and searching (querying) might be seen as one process for discussion's sake. Software must be able to seek and obtain necessary data using as many criteria as feasible in order for a GIS to complete any analytical process. These requirements will encompass attribute, numerical, or textual data in addition to mapped data (lines, points, and polygons). A specialized Standard Query Language (SQL) is used to search for data, and searches can be limited to a particular mapped region or subject. Additionally, there are other methods for retrieving data selectively.

Data can be categorised, for example, according to any subject, area, or class. Using Boolean logic to obtain data is a widespread practise. The basic operators "AND" "OR" "XOR," and "NOT" are used here to indicate the kinds of requirements that must be satisfied in order for the data to be obtained. Boolean logic commands for retrieval and search can be somewhat intricate. To "find all the marine areas having a mean water temperature of >20C, in combination with a depth of <50 metres, which are situated in the waters of both country "x" and "y" and in which quotas do not yet operate," for example, may be an instruction sent to the GIS. These kinds of complex queries might include any parameters for which the data is held.

2.2 Data Analysis:

The addition of analytical features is arguably what sets a real GIS apart from other types of mapping software. There has been criticism recently that many GIS software do not include a wide enough variety of analytical functions; however, this seems unjust given that most products these days contain at least a limited range of such functions. Additionally, connecting a GIS package to a specialised analytical software programme, where analysis is done before returning to the GIS software for mapping, is often a straightforward process these days. Furthermore, since the majority of them are only needed for study, it is typically not worth the time and money spent by a software company to incorporate a large number of analytical capabilities. The majority of GIS software's analytical features work with both attribute and geographic data, or a mix of the two. The majority of the studies that follow may be carried out on either vector or raster structured data, however the specific study being carried out will always determine which of these is most efficient. Although we have chosen to discuss the techniques under the headings of spatial, statistical, and measurement analyses, some authors have reviewed analytical capacities under headings that correspond to the types of data, i.e., point analysis, polygon analysis, and linear analysis, or vector analysis and raster analysis.

- Spatial Analysis
- Statistical Analysis
- Measurement

A. Spatial Analysis:

a. The usefulness of connectivity (or network) analysis lies in its ability to assess a site's level of connectivity across various communication channels. Consequently, it is possible to calculate a connectivity index that, for all towns (nodes) in a given region, indicates, for example, the proportion of roads, other communication channels, or pipeline connections (links) that connect each town to every other town in the area. In order to best allocate routes, connectivity can also be thought of in terms of distance, cost, or time.
b. Measuring and displaying distances between places or a location's degree of proximity to nearby sites are the basic purposes of proximity and contiguity investigations, respectively.
c. Using information from maps, intervisibility determines whether or not a straight line of sight may exist between any two spots on the map. Hence, a computation is

performed, considering the presence of elevated terrain indicated by contours, and determining whether or not hills or further elevated terrain will obstruct the view.

d. The method of creating a three-dimensional (3-D) representation of any chosen location using digitalized height data is known as digital terrain modelling. Since these models don't display actual volumetric data—only surface heights—they are often referred to as 2.5-D models. From a marine perspective, it would also be feasible to create visual models depicting the physical characteristics of certain sea floor regions using bathymetric data.
e. These days, location optimisation is a popular GIS-based technique that helps choose the best sites for any kind of activity. Larger commercial enterprises typically utilise this study to find locations for centralised distribution terminals or new retail stores, for example. In these situations, a variety of geographically variable economic and social indicators would need to be stored in a digital geo-referenced format, such as the population density and the social class structure of a region. The forestry and agricultural sectors employ comparable studies to optimise their operations, but in these cases, physical rather than economic factors may be more crucial.
f. A technique for determining if a generalised spatial surface—one that could be hidden by a wealth of detail in the actual world—exists is trend surface analysis. For example, there may be an overall "wealth" surface in any given nation, which may trend from east to west, but it may also be masked by many pockets of poverty or prosperity. Regarding the distribution of certain species, trend surfaces are probably present from a marine perspective. This means that the species would progressively move away from a biologically optimal region, but in an erratic and perhaps hidden manner. Thus, fitting a trend surface becomes an effective method of locating spatial anomalies—points or regions that are above or below the general trend.

B. Statistical Analysis:

Similar to the geographical analyses, each given GIS may be capable of executing a vast array of statistical operations. We won't go into detail about many of these functions here because they are commonly performed by statistical packages, spreadsheet packages, or database packages, and aren't unique to GIS. However, we should note that many statistical packages can be linked with GIS software to enable the execution of statistical analyses. These consist of more intricate correlations and multi-variate analyses, as well as basic descriptive statistics displaying measures of centrality, frequency analyses, or measures of dispersion. These days, a lot of GISs can carry out intricate spatial statistical analyses like closest neighbor and spatial autocorrelation. One method for calculating the degree of contiguity between regions is spatial autocorrelation. We may need to look for reasons if a marine species were found to be dispersed in any of these ways. A relative estimate of the dispersion of points in a particular region is provided by nearest neighbor analysis; these points may tend towards clustering, randomness, or a uniform spread.

C. Measurement:

A GIS will be able to carry out a wide range of operations on one or more data layers under this category. Simple counts (enumeration), linear or curvilinear distance measures, area, perimeter, and volume calculations, as well as directional or angular measurements, are some examples of measurements. The calculated measurements can subsequently serve as the foundation for additional GIS software work, such as tabular or graphical presentations. It goes without saying that some data formats will lend themselves more readily to particular kinds of measurement. Therefore, if a raster format is used, computing area will be easy, but utilizing vector data might result in more precise distance measurements, if only because the central point of a pixel (in the raster structure), from which distances are measured, may not be the true starting point.

2.3 Data Display:

The primary purpose of any GIS is expected to be data presentation; that is, the display capability will correspond to the system's output as it was originally shown on the VDU. The ability of the GIS concept to display results at any point throughout the data processing process is one of its primary uses. Thus, the GIS offers the ability for maps to be created gradually, with the flexibility to make changes as needed along the way. Changes to the map's visual representation or to the data inputs used to create it can be considered modifications. For the purpose of producing a final product that makes sense, the GIS user can thus control, evaluate, or experiment at any point. A variety of visual display capabilities are included in any competent GIS applications to adjust things like label size, typefaces, color or shade ranges, line lengths, symbols, map feature placements, etc. Additionally, the display format is not limited to maps; it can be taking the shape of a table, graphic, or text display. The perceptual science of "visualization"—how humans look at maps, what information is being sent, how various individuals may interpret the same mapped landscape, how information is communicated, etc.—has received a lot of attention in the GIS area since the early 1990s. The actual data presentation may be either transient or permanent. What is recorded on the VDU is the temporary display. It is a functional user interface in that interactive exploration may be done for free, quickly, and in an almost limitless number of ways. The VDU displays the output of any instructions entered via the keyboard. Permanent output doesn't need to be obtained until the user is happy with any temporary screen display. Permanent output is often achieved by hardcopy display, which may be accessed using a range of devices. However, it can also be permanently stored on an internal hard drive, a transportable disc or tape, or delivered to a different place via networking capabilities. Hardcopy displays can be multicolored or black and white, and they are typically output to paper or film. The quality of the display will vary depending on the GIS's capacity, the data's detail, the mapping's scale, the paper's quality, whether vector or raster structuring is being used, and—most importantly—the printing resolution (in dpi) that is set.

Nowadays, computerized output of the highest quality can outperform manual approaches. It's possible that GIS output—and its entire functional range—will soon be able to be used and/or shown on appropriately outfitted fishing vessels. These days, a number of boats have highly advanced navigation systems that include radar, plotting capabilities, and electronic charts. These can show a range of static data, including the tracks of moving boats and bathymetry, land masses, restricted regions, and navigational landmarks. Extending this feature to enable other desired layers to be merged and shown in an interactive mode—that is, to enable the vessel to carry out a variety of necessary GIS activities while on board—will be a straightforward development.

2.4 Database Management:

A database is a sizable, ordered collection of connected data that is unrelated to any one application and has been arranged in an ordered manner. This data gathering can be kept digitally either internally in a database that is a component of the GIS programme or outside in a purpose-built computer software database package, such as Oracle or dBASE, for GIS applications.

While spatial data are often maintained within the GIS programme, attribute data are often stored elsewhere. Distributed databases are also increasingly being employed in GIS applications. In this instance, the data could be stored in many locations both inside and outside of an organization. It goes without saying that a GIS with access to these databases will have access to a wealth of additional data. It is simple to see how a GIS may be more useful in the context of fisheries if it had direct access to databases that were probably located elsewhere from the fisheries GIS, such as those related to oceanography, meteorology, and possibly the environment.

Anon (1993b) offers comprehensive details on the construction of the Regional Maritime Database (BDRM), a large database that will cover the maritime areas along the West African coast. Any or all of the ten participating nations will be able to access this database, and all data is properly geo-referenced to guarantee GIS functioning.

A database management system (DBMS) is a computer application used to create, manage, and retrieve digital databases. Numerous commercial solutions are available to accomplish this. The GIS software, other data sources, graphics-enhancement software, and any actions the user would want to carry out are all connected by the DBMS. DBMSs contain various file formats, programming tools, and the ability to operate with a variety of data kinds, including characters, numbers, and dates. They also offer languages for defining, modifying, and querying the data.

There are a few very fundamental reasons to have a DBMS. First of all, it follows that maintaining the data in the sense of keeping it accurate, up to date, correctly organized, formatted, etc., is obviously necessary. It must also be managed to ensure that it is appropriately understood and accessible to the right people at the right time and place. To increase searching efficiency and decrease storing space, storage structures need to be continuously observed. Additionally, as most databases are always expanding, it may be necessary to add new fields or, conversely, to delete older entries.

Typically, a database manager is designated to oversee all of the aforementioned tasks as well as control database access, handle system issues when they arise, and connect databases to external databases as needed.

In addition, he or she could have to deal with the legal aspects of data management, such as ensuring that utilising his data won't likely lead to incorrect judgements being made and preventing copyright infringements. The data must be stored in an organised manner for a DBMS to function at its best. Generally speaking, there are four different kinds of relational, object-oriented, networked, and hierarchical DBMS structural/storage approaches.

A. Hierarchical:

Each record in the hierarchical model may have several ties to subordinate "levels," but only one link to a superior level. The "root" is the highest connection; "children" are the lower levels, and "parents" are the higher levels. When searching up and down is necessary, hierarchical data structures are easy to understand, update, and expand. However, they perform poorly when searching horizontally, i.e., when it may be necessary to locate all records at a single level because there are no connections at the same level.

B. Networked:

This is comparable to the hierarchical data model, but it allows for the possibility of many parents, which allows for the discovery of many-to-many connections. A communications network with several links connecting any number of centers can be compared to a networked database system. Rapid connections are feasible with this kind of data structure, which optimizes the data that is accessible but is challenging to establish and manage. These days networked and hierarchical systems are rarely utilized in GISs.

C. Relational:

In this case, the data is arranged into several tables, each containing a single kind of entry. Records are represented by the rows of the tables, and fields by the columns of the tables. Every table inside the database will be connected via a common field, often known as a key attribute or unique identifier. By specifying the relationship that is appropriate for the query being made, data is pulled from the database. If new tables need to be created, relational algorithms may be used in this process. Relational databases are simple to set up, operate, and maintain and offer a great deal of flexibility. Boolean logic and mathematical processes may be used to figure out almost any relationship, and adding more data to the database is a simple process. The primary disadvantage is that searches might take a long time, especially considering the numerous tables that may be required, the computations and other parameters involved. The relational DBMS is the type of database that GISs use the most commonly, and Standard Query Language (SQL) has been designed as the standard language for use with it.

D. Object Oriented:

By guaranteeing that object-oriented database management systems (DBMS) are highly efficient and that a global framework for their development exists, this group hopes to establish object-oriented DBMS as the standard for the future. This will enable the DBMS to function in all settings, on all hardware platforms, and under all operating systems. Because these databases adopt a complex perspective on spatial things, object-oriented database management systems (DBMS) are theoretically challenging to describe and comprehend. Essentially, the three core ideas of inheritance, class, and object are seen to be present in all things. Geographic items include things like "road," "port," "sea," and so on. Then, each object may have a class, such as "Fishing (port)," "Shallow (sea)," or "Secondary (road)," and they may also have sub-classes. any object may also be described as possessing certain qualities; for example, a fishing boat could have an owner, a value, a size, and so on. Additionally, any object can be made to execute specific activities, such as select, measure, find, alter, or draw. Because of inheritance, when a class is added to an object, all elements of the original class may be included in the new description. For example, all characteristics related to fishing may be associated with the class of "Fishing port." If one is looking for a more thorough explanation of object-oriented database management systems, Laurini and Thompson (1992) or Cooper (1993) provide comprehensible definitions.

2.5 References:

1. Joseph G. Fundamentals of Remote Sensing, Universities Press 2003
2. Otto Huisman and Rolf A., de Principles of Geographic Information Systems ITC, Enschede, The Netherlands, 2009
3. Ondieki C.M. and Murimi S.K. Environmental Monitoring, Vol. II, Kenyatta University, Kenya Copyright-Encyclopaedia of Life Support Systems (EOLSS)
4. https://docs.qgis.org/3.4/en/docs/index.html
5. https://www.geographyandyou.com/
6. http://www.geogra.uah.es/patxi/gisweb/GISModule
7. https://www.esri.com/library/bestpractices/what-is-gis.pdf
8. https://iasscore.in/current-affairs/prelims/geospatial-technology-for-monitoringnatural-resources