7. GIS functionality

G.J. Meaden (FAO consultant, Canterbury, United Kingdom), J. Jenness (FAO consultant, Flagstaff, Arizona, United States of America), and S.Walker (NOAA Coastal Services Center, Charleston, United States of America).

This chapter discusses the functional tools and techniques offered by typical GIS software packages. The concern, therefore, is to outline all the major capabilities of GIS briefly in order to distinguish what types of spatial and nonspatial functions can be performed. The range of these functions varies enormously between one GIS and another, and many of the functions could be performed on non-GIS software. There is typically a relationship between GIS cost and the number of functions performed, and there may be numerous smaller software applications that can be acquired to perform specific applications. The terminology for functions will vary between GIS, and different functions may better be performed using either the raster or vector formats. No attempt is made in the technical paper to describe how to perform individual GIS-based functions (as this will vary between GIS software); the intention here is to describe typical functions that a GIS will carry out.

Initially, any digital data to be used for GIS purposes are likely to require some transformation or preprocessing. These are tasks or manipulations that aim to change the data as required to suit the area, scale, theme, etc., of the specific GIS task being undertaken. At the most basic level, these preprocessing functions, such as delete, crop, recode, dissolve, zoom, merge and rotate, are all functions that are important to mapping but that can be performed by many software packages. At a more sophisticated level, and with more direct relevance to GIS work, there is a further range of transformations, most of which are applied directly to digital mapping data. These include: (i) edge matching – allowing neighbouring mapped sheets to exactly match up; (ii) projection changes – so that all mapping conforms to a unified projection; (iii) coordinate transformations – allowing similar georeferencing systems to be used; (iv) rubber sheeting – allows map distortions to be eliminated; (v) geometric corrections – eliminates spatial distortions, usually in remotely sensed imagery; and (vi) structure conversions – allows GIS to toggle between working in raster or vector formats.

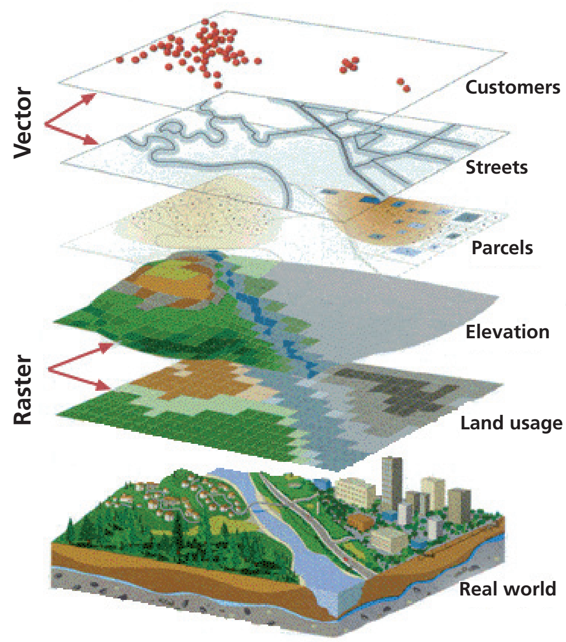
Through these preprocessing functions, the digital data held should be“fit for purpose”. However, there are still data transformation functions that may need performing, and two of these (map generalization and thematic classification) are required because of the huge scale variations that mapping may engender. It is easy to conceive that the scale of a map will greatly affect the amount of detail that can be shown, meaning that a GIS should be able to generalize real-world features so as to suit the scale requirements of individual projects. Generalization can be achieved through functions, such as the elimination of smaller features, the merging of groups of similar features, or the reduction in line complexity. Generalization should be performed on both the graphical features of a map and the named features, and generalization can only be effectively deployed when moving from large scale towards a smaller scale. Thematic classification is important to mapping and to GIS simply because the real world is too complex to show everything on a map. This means that, as with generalization, the map content must be simplified by being sorted into classes, for example, houses, factories, shops, etc., may all be classified as buildings or, when mapping fish species, they might simply be classed as pelagic fish or demersal fish. Classification is perhaps more frequently used in quantitative mapping. For instance, it is often desirable to show fish quantity variations in an area, and this will involve considerations of both the numerical range of the data and how many classes to divide this into. All GIS contain functions that scrutinize the data range and then make decisions on sensible numerical class boundaries. There will always be user options to override these decisions if preferred.

At a more complex level, there are two other basic data transformation functions that are important to most GIS work. The first of these is called "buffering". A buffer is a polygon created at some user-specified distance around a point, line or polygon feature. For instance, a buffer might be drawn around an aquaculture facility that has a notifiable disease so as to demarcate no-entry areas. Buffers can be of any user-defined width and there can be any number of buffers around an identified map feature. Buffers can be very useful in quantifying the extent of any subject of interest, e.g. the GIS can be asked the query "Show me (and quantify) the number of fishing boats registered at 10, 20, 30, 40 kilometres distance from cities x, y and z." The GIS will draw these distance buffers around the three cities, and then using a database of vessel registrations, it will calculate the answers to the query. The second basic transformation function is called "overlaying". Overlaying is the integration of two or more maps to form a new and enhanced map (Figure 7). For instance, a typical government topographic map can be derived from stored digital files containing separate mapped features, such as rivers, woodland, rail lines, roads and towns, and each of these files can be overlayed in desired combinations to create more detailed or preferred maps. To properly overlay maps, they must be in the same mapping projection and at the same scale, and overlaying can be performed using either the vector or raster data formats (though the two formats cannot be mixed for analytical purposes).

When overlaying vector-based maps, file sizes will be considerably increased and maps are likely to need careful editing. For instance, if a county boundary is following the course of a river and maps of these features are being overlayed, the two lines may not exactly match because the two files were created separately, perhaps from different data sources. The main purposes of overlaying are:

- to calculate any relationships that might exist between different mapped layers;
- to create any new or desirable combination of mapped features;
- to allocate weightings to features of variable importance to any mapping project; and
- to carry out map algebra or more complex geostatistical analyses.

FIGURE 7
Selected mapped layers for a specified area using points, lines, polygons, vectors and rasters



Source: Heard (2010).

The technical paper provides further details on each of these purposes for using overlaying functions. When used in combination, buffering and overlaying have the potential to provide a sophisticated range of GIS analyses.

This chapter now turns to the range of measurements that GIS can perform, including simple enumerations or counts, linear distances, areas, perimeters, volumes, directions and angles, plus a range of more sophisticated but less frequently used measurements. A caution is given as to the validity of GIS-based measurements made, this usually being a function of the fact that mapped data are a great simplification (or a generalization) of real-world situations. Measurements are performed very differently in raster formats than in vector formats. In the raster format, if the real-world size covered by one pixel is known, then measurements along the horizontal or vertical axes will be simple to calculate, including area calculations. However, most measurements using raster formats are complicated by the fact that measurements may need to be made along irregular lines in irregular directions, plus the fact that the real-world shape cannot correctly be projected onto a flat mapped surface. However, in general, the use of rasters can produce fairly accurate area and volumetric measurements using very basic algorithms. Measurements based on the vector format have the potential to be far more accurate because the vertices, segments and edges are precisely defined. However, this accuracy comes at the price of far greater computational demands. The technical paper provides further information on the Euclidian and Pythagorean geometry and the computing demands involved in vector-based measurement.

If GIS is to provide objective information on spatial relationships, then it requires examples of more sophisticated measurement techniques. Here, the three main techniques of measuring centrality, proximity and contiguity are briefly discussed. The concept of centrality is extremely important in many real-world situations, especially as it relates to optimizing economic locations. For instance, in the fisheries context, it is important that a fish processing plant can service as many customers as possible while minimizing accessibility costs. Most GIS software have the functionality to establish the mean centroid point of any polygon, and this centroid can also be a useful point from which to measure interpolygon distances. Looking at centrality from a different perspective, it may be of interest to know the theoretical area that might be served from each of many central points. An example might be that many aquaculture sites in a region are selling fish in their neighbouring area. Where ought the boundary lie between neighbouring producers? The GIS will produce so-called Thiessen polygons that cover the whole aquaculture region into idealized "fish market areas", with a fish farm being at the centre of each polygon.

Proximity analysis is concerned with the distances between different features, and this might be thought of in terms of: (i) numerical counts, or in terms of (ii) the spatial pattern revealed by the distribution of any features in an area. With respect to numerical counts, GIS will use a proximity analysis function to calculate the number of objects *a* that are within *b* kilometres of a specified line or point location, e.g. how many fish farms are located within 20 km of a fish processing plant? This form of proximity analysis mainly relies on the buffering function.

When analysing the spatial pattern revealed by a distribution, the researcher is seeking to identify the form of the distribution. It is easy to imagine that fishing boats at sea may be distributed in a point pattern that could be described as:

- random – where there is no discernible pattern in the distribution;
- uniform – where boats are very evenly spaced out across the marine area; or
- clustered – where vessels tend to be in one or more fairly tight clusters.

If one or other of these distribution patterns emerge, then it might be of value to investigate the cause. GIS uses what is called the "nearest neighbour" analysis to identify distribution patterns and this can produce an objective value describing each of the three distributional patterns. The final spatial relationship measurement technique described is that of contiguity analysis. This is similar to an analysis of point distributions, but here the concern is with measuring the dispersion of polygons in terms of their contiguity or spatial autocorrelation. What is therefore being identified here is "to what extent do similar classifications occupy adjacent cells?"In the case of a chessboard, it can easily be perceived that there is no relationship between neighbouring cells because cells alternate between black and white. However, in the real world, there may be a large amount of spatial autocorrelation because for all sorts of reasons similar features frequently occupy similar locations. Identifying marine or riverine areas having high adjacent spatial autocorrelation may be important in terms of the ease in which species may move from one area to another.

Progressing to more advanced GIS functions, the creation of statistical surfaces is next described. A statistical surface can be thought of as a surface area of land or the seabed that has numerical values attached to it indicating a value or coding on the vertical (z) axis. Values can relate to human or physical factors. A simple example would be that all places in a region or country have a current (or a mean) air temperature reading. These temperature readings could form the basis for constructing an isotherm map. Such a map can be constructed by GIS, it being based on temperature readings at sufficient representative sampled points. A temperature map is said to be spatially continuous because temperatures vary gradually over the surface, but other surfaces are spatially discrete because single values or codes may apply to wide areas before suddenly changing, e.g. geology or sea-bottom sediment types. Creating continuous surfaces by GIS relies on the software's ability to perform interpolation. This means that the GIS is required to estimate missing values for any data set based on the distribution of the known sampled data values. GIS will have various interpolation algorithms for achieving this based on the nature of the sampled distribution, the physical characteristics of an area, etc., and the technical paper discusses some of the interpolation methods available. If point sampled data show erratic distributions, then GIS might be used to create a general trend surface. For instance, fish counts of a specific species are likely to vary significantly from one sampling point to another. However, over a wide area it is likely that the sampling counts will gradually reduce (or increase) in a particular direction; therefore, the GIS can create a trend surface map to show the generalized distribution of the fish species across the study area.

The study of GIS functionality now turns from surfaces to linear distributions (or networks). It should be cautioned that not all GIS have the ability to undertake network analyses. Recall that networks are comprised of links and nodes, both of which can have values associated with them (see Chapter 5). Gravity modelling using GIS is important in establishing the relative importance of each link and node in an area based on size and/or distance differences. An example is given whereby towns of various sizes (each town is a node) are connected by a road network (the links). Given that larger towns are likely to attract more usage, but that people are going to use discretion in how far they wish to travel, then using so-called "gravity modelling" the GIS can calculate the number of people who are likely to support the facilities at each town, i.e. this will be the town's sphere of influence. Gravity modelling is useful in terms of where best to locate certain facilities, or in how far it may be worth travelling to catch certain fish species.

A second form of important network analysis is that concerned with optimizing travel or communication through networks (also called connectivity analysis). What is discussed here normally applies to human travel, but the logic could also apply to routes through communication systems, waterway networks, pipelines, etc. Again using the node and link model, it is easy for a GIS to calculate the shortest path through, for instance, a road network between any start and finish nodes. This calculation can be in terms of shortest travel distance, cost or time. Temporary impedance values can be introduced in any selected link for perhaps road works or a major traffic hold-up. This type of network analysis forms the basis on which in-vehicle navigation systems work. A variation on shortest path analysis is the so-called "travelling salesman problem", where the shortest/quickest route is the one that best connects many delivery points. From a fisheries perspective, it might be desirable for a GIS to establish the optimum daily route for the collection of landed fish catches. A second variation on shortest path analysis is the ability of GIS to determine, from any selected node, the number of other nodes that may be accessible in a given time or at a given cost. This might be useful for a fish farmer who wants to establish the number of potential customers that can be served from any nodal points.

An important final type of network analysis is that associated with natural water flow over terrestrial surfaces. Over much of the planet, a network of streams and rivers has naturally become established, with flows usually moving outwards from core highland areas towards low lying coastal areas and thence into the sea (though sometimes water flows are directed towards inland lakes). Based on any river network in a specific area, GIS can perform the following analyses, all of which may be useful to aquaculture site selection or to river fisheries:

- calculation of flow direction and water accumulation;
- derivation of river or stream catchment areas;
- detection of flood storage areas (or suitable dam site locations);
- inundation modelling to detect areas along waterways that are liable to varying degrees of flooding;
- rates of, and areas affected by, point source pollution or disease dispersal downstream through waterways;
- management of waterway extraction along streams or rivers;
- creating likely stream networks in areas that are poorly mapped; and
- peak flow prediction in rivers during storm events.

The basis on which any of these analyses works is described in some detail, especially that concerned with identifying the pour direction and routeways for natural water flows. The technical paper gives many clues to sources of information on all network analyses, including many Internet sites that provide online facilities allowing access to software packages for modelling hydrology.

The last type of GIS functionality discussed is topographic surface analyses. These were introduced in Chapter 5, but here the concern is providing information on some of the GIS analyses that are possible. Using either triangular irregular networks or digital elevation models mentioned in Chapter 5, together with their incorporated elevation data, GIS are able to calculate surface or submarine gradients for any given slope and they can produce complete gradient maps for any area. This information is useful in pond site identification for aquaculture. One associated factor that can be established is aspect, i.e. information on the direction in which slopes are facing, and another GIS function is that of visibility analysis, in which it is possible to determine all parts of an area that are visible from any chosen map location. This can be important if planning applications decree that for particular facilities the visual impact should be minimized (e.g. location of fish cage culture sites).

Any GIS will not be able to provide all desired functionality. To circumvent this, a good GIS will provide a means for users to create their own functions and tools. This is sometimes provided through a simple macro function, whereby a script can be written allowing for repetitive functions to be enacted. More advanced GIS software may allow users to write sophisticated tools using the most current software programming languages, such as C++, C#, Java, Python or VB.NET, and some of these GIS packages have enthusiastic user groups of people who write and share custom tools and who often provide them free online. With experience in the use of GIS, it seems that most functions are possible.