INTRODUCTION TO GEOGRAPHICAL INFORMATION SYSTEMS

Kenna Matheson & Vicky Copley

INTRODUCTION TO GIS

1. What is a GIS?

A GIS consists of the following components:

- Hardware (i.e. screen display or personal computer)
- Software (e.g. IDRISI, MapInfo, ARCVIEW)
- Peripherals (e.g. digitising tablet, plotter)
- Data
- Information
- Spatial referencing
- Topology

What is the definition of a GIS?

"A computer system capable of holding and using data describing places on the earth's surface"

****

"An organised collection of computer hardware, software, geographic data., and personnel designed to efficiently capture. store, update, manipulate, analyse and display all forms of geographically referenced information"

As you can see, there is more than one definition of GIS and use of GIS tends to depend on experience, applications and user needs. As you gain more understanding of GIS, your use of GIS will alter. Applications vary and user needs are diverse, from creating cartographic products for presentation to building databases and spatial modelling.

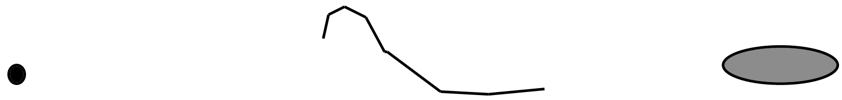
What does a GIS do?

A GIS allows you to carry out the following:

- Data management
- Data input
- Data storage
- Manipulation of data
- Analysis of data
- Data modelling
- Data output

What are geographic features?

Geographic features are all around you, both in the man-made and natural environment. Geographic features have real world locations**,** i.e. *x,y* co-ordinates. They are represented in abstract form on a map, using either points (dots). lines or areas.



point *line* *area*

Point

A point is used to represent, for example, a telegraph pole, a spot height, or a meteorological station. This will be represented by (*x,y*) co-ordinates and will also be assigned an attribute number. The scale of your map will determine whether you choose to represent certain features as points or area symbols. For example, on a town plan (large scale) buildings may be represented as individual point symbols; on a smaller scale map the urban extent may be represented as a polygon. On an even smaller scale map, a town would be represented as a point symbol.

Line

This is used to represent features such as boundaries, roads, railway lines or contours.

This will be represented by a series of (*x,y*) co-ordinates called vertices, which define the shape of the line. It will also be assigned an attribute number. A line is sometimes referred to as an *arc*. An *arc* will begin and end with a node. It will also be assigned an attribute number (or *identifier*) which is used to identify what it represents.

Area

This is used to represent, for example, an urban expanse, a soil type or a land use type. This will be represented by a series of (*x,y*) co-ordinates and will allow for the measurement of perimeter and area.

One or more arcs will outline an area.

2. Spatial Referencing

A spatial referencing system allows the data to be represented in space. The following section outlines the *three* most common spatial referencing systems.

Geographical co-ordinates

We know these as latitude and longitude, which are the only true geographical co-ordinates. They are represented in degrees, minutes and seconds, East-West for longitude and North-South for latitude.

The latitude of a point on the Earth's surface is the angle between that point and the Equator, measured North or South of the Equator.

The longitude of a point is the angle between that point and the Greenwich Meridian, measured East or West of Greenwich.

Rectangular Grid co-ordinates

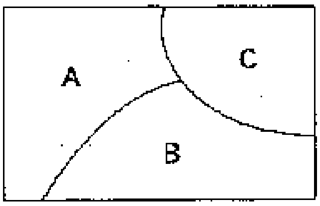
To represent the whole of the Earth's surface, or even parts of the globe, on a flat piece of paper or a computer screen, you need to use a *map projection.* There are a wide variety of map projections, and they all distort distance or angles in some way. Rectangular grid co-ordinates are derived from a transformation of the Earth's surface onto a flat surface, but it is sufficient to recognise that all "flat" maps are derived from a projection.

Non-co-ordinate systems

In areas of Canada and the Western USA, there is an alternative non-co-ordinate system which divides the land into "quarter sections". By knowing which section a point lies within, you can reference the point to the Earth's surface. However, you do not need to understand this system - only to be aware of its existence.

3. Raster and vector data storage

GIS may be raster-based or vector-based. IDRISI is a raster-based system, which stores maps as collection of grid cells or pixels, scanned onto the screen like a television picture or a satellite image. In contrast, MapInfo is a vector-based system; it stores maps as a database of (x,y) coordinates which are plotted on the screen and joined up to form an image.



Vector data storage

A	A	A	A	C	C	C	C
A	A	A	A	C	C	C	C
A	A	A	B	B	C	C	C
A	A	B	B	B	B	B	B
A	B	B	B	B	B	B	B

Raster data storage

Raster data are stored using rows and columns of grid cells of a uniform size. Vector uses simple points, lines and polygons (areas) with topological relationships. Both have their origin in the lower left. Raster data store real world co-ordinates of the origin and calculate other coordinates as required, whereas vector data store real-world *x,y* co-ordinates for all features. Storage requirements for raster data are generally larger than those for vector data as raster data are less compact. Raster data have an attribute (or identifier) value linked to their row and column location in the grid. Features in vector data have a unique identifier (attribute) linking them to descriptive attributes. Topological relationships are more easily represented in vector data. However, overlaying different data sets is easier, quicker and more efficient with raster data.

4. Geographical data

What are geographical data?

Geographical data have three *main* characteristics:

Position*:* the spatial referencing systems inherent in GIS give information about positions, i.e. they fix our data in space.

Topology*:*. is the procedure for defining spatial relationships and geometric properties. In relation to maps this defines the connections between features, i.e. adjacent polygons, sets of lines, proximity of features to each other, etc.

Attributes: features in the geographical landscape may be classified into individual classes, e.g. for roads. boundaries, etc.

Geographic data have two components: -

spatial data
attribute data

Which of the following are spatial and which are attribute data?

a dot on a map representing a spot height
a street, street name and city name
a polygon (area) and soil type

Remember one geographic feature (spatial data) can have more than one attribute. The attribute information is held in the *geographic database.*

5. Data Input and Data Sources

Data input is the process of converting data from their existing form into one that can be used by the GIS. The types of information which you may want to put into the GIS will include:

- Remotely sensed data (satellite imagery)
- Aerial photographs
- Topographic maps
- Thematic data
- Digital data which may require converting and importing
- Statistical information

Methods

There are several methods of converting *spatial* data to be used by a GIS:

- Scanning
- Digitising
- Keyboard to enter co-ordinates
- Purchase data on diskette, tape, etc.
- Converting and importing

Electronic scanning is a quick alternative input method to digitising, but you must be aware of several points. As an electronic scanner is highly sensitive it will scan all the information from a map sheet and convert it into a coverage - it is not selective in this process and can therefore scan paper crease marks, pen and pencil blots and any unwanted detail. Remember, as it scans everything, particular data sets cannot be selected to be scanned. It is worth remembering to carefully prepare a map sheet you want to be scanned as this will save on time and labour later. The best medium for this process is not paper, but a translucent polyester film. If you are using an original paper map, however, check it over carefully for marks and creases - it may be better to digitise instead. Lastly, some scanners will only produce what is known as *spaghetti* data i.e. data without any attributes, which will require a great deal of time-consuming editing and attribute updating.

There are other ways of converting information, including using flat files (text files) listing a series of co-ordinates, digital files, interpolation, on-screen drawing and many more. For example, in order to produce a land use map, you may use classified remotely sensed imagery, together with digitised topographic detail and point data from a flat file. You are not limited to one method or to one source of data.

Data considerations and scale

These processes may appear simple to you at first. But thorough investigation of your data sources, followed by careful planning of how you are going to input your data are imperative before you begin!

Your first consideration is: "what do I actually require from the GIS?" You have probably noticed that most maps which represent a theme (e.g. rainfall or land use) display topographic information in addition to the theme itself. Your first problem is to decide what topographic detail is required in order to display your thematic information successfully. Topographic information includes features such as communication networks, e.g. roads, railways, major rivers, etc.; national and international boundaries; locations of towns and cities; and many more.

A selective amount of topographic information will enable the person reading your map to locate him/herself on the map easily. Reference positions allow for easier understanding of the map data being presented.

You must also decide at what scale you intend representing your spatial information; this will in Part dictate what topographic detail you should incorporate. *At this point you should remember several rules:*

(i) Always try to input your data at a larger scale than the scale at which it will be output. For example, if your topographical features are being derived from a map at a scale of *1: 50,000* (1 cm = 0. 5 km), then the output should be at a smaller scale such as 1: 100,000 (1 cm = 1 km).

(ii) If you are deriving thematic information from several map sources, check that the scales are similar. They do not necessarily have to be the same, but they should not be too different. You will generate inaccurate data if you attempt to input a land use map at 1:500,000 and overlay a topographic map series at 1:25,000!

(iii) Lastly, the most complex issue - projections. You cannot expect to input and overlay data accurately if your map sources are from different projections. You must attempt to have all your map data at the same projection.

Errors

Do remember that errors are easily produced and perpetuated in the world of GIS (and frequently go unnoticed!). By checking your data sources, processes and output carefully, these can be reduced to an acceptable level. Listed below are several sources of error which you may not be aware of. *Remember that the quality of your data output can be only as good as the quality of your input data.*

(i) if your original paper map data is old or has been stored badly, the paper may have stretched (usually due to damp conditions).

(ii) digitising always produces a certain amount of errors; these can be made worse by using stretched paper map, thus further reducing the level of accuracy.

(iii) vector to raster conversion (and *vice versa)* generates its own error, which is a complex issue.

(iv) typing mistakes are an example of the 'human' errors that are made frequently, e.g. when misreading a number from a list of (*x*,*y*) co-ordinates.

You will realise by now that a GIS is not just a computer system for making maps, nor is it simply a general purpose computer graphics system. Although a GIS can produce maps at different scales and projections, as well as selecting and retrieving data, it has also been designed to do much more.

6. Data Management

By the time you have carried out data inputing several times you will have amassed a great deal of information which should be useful, easily accessible and understandable. If the data are not managed correctly, then they will be none of these!! Data must be managed, archived, and stored efficiently and correctly. Management functions include:

- **Archive data** - write the data to store and read to retrieve
- **On-line storage** - organise data, preferably in a hierarchical fashion
- **Delete data** from the system - get rid of unwanted data

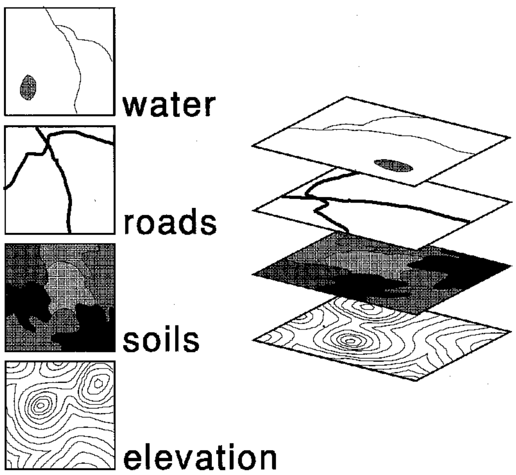
7. Processing

Before you can begin data analysis, interrogation and manipulation, your data now require editing, checking and updating. Map data may be stored in the database in the form of tables (e.g MapInfo) or they may be stored in *map layers.* Whichever way the data are stored you will need to use functions such as updating and editing of both your geographical data and your attribute data to obtain the desired results.

You can manipulate and analyse data to obtain the following:

- data that are easier to work with
- more useful data
- answers to a particular question
- solutions to a particular problem

Analysis may involve overlaying several map layers (see figure), selecting and retrieving data which meet particular conditions, and extracting certain measurements from the data. In IDRISI these functions are relatively easy to understand. However, in other GIS systems, various commands which perform the same functions may give you different results. Beware of these differences which may result from using different GIS system software! It may also be possible to ask the GIS system "what if?" questions, for example: when building a road, how much agricultural land or forest area will be lost by selecting alternative routes?



Overlaying map layers.

The questions a GIS can answer can be divided into 5 broad categories:

1. Locations	***What is at.....?***
2. Conditions	***Where is......?***
3. Trends	***What has changed.....?***
4. Patterns	***Which data are related.....?***
5. Models	***What if?***

1. Locations

By defining the location of an object / feature, you should be able to obtain its characteristics, i.e. its attributed information. There are several methods of specifying an object. You can point to it and select it with the mouse, type in locational (*x,y* co-ordinates) information, or type in a specific attribute type.

2. Conditions

This involves specifying a set of conditions, or defining the criteria by which you make a selection. If for example you want to find a prime site for say, a large new market, you might enter the following criteria:

- within 5 miles of a main town
- within 1 mile of a good road
- site for sale
- site must be 10 acres minimum

or, given the project, you might want to find a site to test varieties of a particular crop. In this case you might enter the following criteria:

- within 5 miles of the research institute
- within 1 mile of a good road
- the farmer grows the crop in question
- site must be 1 hectare minimum

Providing you have all the above criteria itemised and attributed in your data base, to each polygon or line, then the GIS will be able to calculate the available sites for you.

3. Trends

This third kind of question seeks to identify geographic occurrences that have changed or are in the process of changing, involving summarising data. This is more of a speculative question, based on assumptions, evidence and observations. Once a trend or theory has been established by the user, then the theory can be tested by the GIS. A basic example of this would perhaps be the change in land use from one type of crop to another. Providing you have the data sets covering this area available on an annual basis, you would establish what the trend is, whether it increasing or decreasing, and what the new trend is.

4. Patterns

In this case, the GIS uses existing data in new combinations to provide a better description of what is occurring. This usually requires a long-term set of observations and knowledge of potential data relationships.

5. Models

This involves using different data sets and enquiring on 'what if... ?' relationships. A model is generated using selection criteria and ranked responses build up a picture of new data relationships.

8. Output

There are several ways of displaying your output. You may be required to produce a "hard copy", for example, a paper map. This can be produced using an ink-jet printer, a simple pen plotting device or a more complex raster plotter. Both have advantages and disadvantages. The pen plotter will be quick and economic if you require only simple line work. However, if you require large areas to be filled with colour or shading a raster plotter will be quicker and more effective. It is however an expensive initial purchase.

Alternatively, you may require only a simple screen output displaying either your map(s) or a list of numerical results derived from analysis and displayed in a flat file.

Whichever way your map is displayed, it is important to emphasise the appropriate features, to select only the required features, and to attempt to use some "cartographic flair" to make your map look interesting! Don't overcrowd your map with unnecessary detail - you do not have to include every feature available!

If you are producing a paper map, remember to include the title, town or river names if necessary (be selective!), a scale (bar or ratio), acknowledgements, source information, etc. Remember that your data may be used by someone else in the future, so information such as projection parameters will be important for retaining accuracy!

APPENDIX

Digital Image Processing (adapted from IDRISI documentation)

Remotely sensed data are important to a broad range of disciplines. These data are very useful to environmental decision-makers and planners, particularly in the areas of landuse mapping and change detection.

The inherent raster structure of remotely sensed data makes it readily compatible with raster GIS, and many, including IDRISI, are designed to perform image processing tasks, and to facilitate the incorporation of the results into GIS analyses.

Digital image processing is largely concerned with four basic operations performed on raster data:

- image restoration
- image enhancement
- image classification
- image transformation

Image restoration is concerned with the correction and calibration of satellite images in order to achieve as faithful a representation of the earth's surface as possible.

Image enhancement is predominantly concerned with the modification of images to optimise their appearance to the visual system.

Image classification refers to the computer-assisted interpretation of images -- an operation that is vital to GIS.

Image transformation refers to the derivation of new imagery as a result of some mathematical treatment of the raw image bands.

Many raster GIS systems now offer a mixture of GIS and image processing capabilities. In IDRISI for Windows, the modules found in the Analysis/Image Processing menu allow you to perform all the procedures described below.

Image Restoration

Broadly, image restoration can be broken down into the two sub-areas of *geometric restoration and radiometric restoration.*

Geometric Restoration

With commercial satellite imagery, such as Landsat and SPOT, most elements of geometric restoration associated with image capture are corrected by the distributors of the imagery. These include elements such as:

- *skew correction* to account for the fact that the earth is moving while an image is being captured;
- *scanner distortion correction* to account for the fact that the instantaneous field of view (IFOV) covers more territory at the ends of scan lines (where the angle of view is very oblique) than in the middle.

For mapping purposes, it is also essential that any form of remotely sensed imagery be accurately registered to the proposed map base. With satellite imagery, the very high altitude of the sensing platform results in minimal image displacements due to relief. As a result, registration can usually be achieved through the use of a systematic *rubber sheet* transformation process that gently warps an image (through the use of polynomial equations) based on the known positions of a set of widely dispersed control points. This capability is provided in IDRISI for Windows through the module RESAMPLE.

Radiometric Restoration

Radiometric restoration refers to the removal or diminishment of distortions in the degree of electromagnetic energy registered by each detector. A variety of agents can cause distortion in the values recorded for image cells. Some of the most common distortions for which correction procedures exist include:

- *uniformly elevated values,* due to atmospheric haze, which preferentially scatters short wavelength bands (particularly the blue wavelengths);
- *striping,* due to detectors going out of calibration;
- *random noise,* due to unpredictable and unsystematic performance of the sensor or transmission of the data;
- *scan line drop out,* due to signal loss from specific detectors.

Image Enhancement

Image enhancement is concerned with the modification of images to make them more suited to the capabilities of human vision. Regardless of the extent of digital intervention, visual analysis invariably plays a very strong role in all aspects of remote sensing. While the range of image enhancement techniques is broad, the following list indicates the major enhancement techniques:

Contrast stretch

Digital sensors have a wide range of output values to accommodate the strongly varying reflectance values that can be found in different environments. However, in any single environment it is often the case that only a narrow range of values will occur over most areas. Grey level distributions thus tend to be very skewed. Contrast manipulation procedures are thus essential to most visual analyses.

Composite generation

For visual analysis, colour composites make fullest use of the capabilities of the human eye. Depending upon the graphics system in use, composite generation ranges from simply selecting the bands to use, to more involved procedures of band combination and associated contrast stretch.

Digital filtering

Filters can be used to provide edge enhancement (also known as *crisping),* to remove image blur, and to isolate lineaments and directional trends.

Image Classification

The majority of image classification is based solely on the detection of the spectral signatures (i.e., spectral response patterns) of land cover classes. The success with which this can be done will depend on two things: 1) the presence of distinctive signatures for the land cover classes of interest in the band set being used; and 2) the ability to reliably distinguish these signatures from other spectral response patterns that may be present.

A vital step in the classification process is the assessment of the accuracy of the final images produced. This involves identifying a set of sample locations that are visited in the field. The land cover found in the field is then compared to that which was mapped in the image for the same location. Statistical assessments of accuracy may then be derived for the entire study area, as well as for individual classes.

There are two general approaches to image classification: *supervised and unsupervised.*

Supervised Classification

With supervised classification, we identify examples of the information classes (i.e., land cover types) of interest in the image. These are called *training sites.* The software system is then used to develop a statistical characterisation of the reflectances for each information class.

Once a statistical characterisation has been achieved for each information class, the image is then classified by examining the reflectances for each pixel and making a decision about which of the signatures it resembles most. There are several techniques for making these decisions, and these are often termed *classifiers.* Most image processing software will offer more than one classifier. IDRISI for Windows offers three commonly used supervised classifiers: Parallelepiped, Minimum Distance to Means, and Maximum Likelihood. Each of these employs a different logic for assigning pixels to classes.

Unsupervised Classification

In contrast to supervised classification, where we tell the system about the character (i.e., signature) of the information classes we are looking for, unsupervised classification requires no advance information about the classes of interest. Rather, it examines the data and breaks it into the most prevalent natural spectral groupings, or clusters, present in the data. The analyst then identifies these clusters as landcover classes through a combination of familiarity with the region and ground truth visits.

The logic by which unsupervised classification works is known as *cluster analysis,* and is provided in IDRISI for Windows by the CLUSTER module. The clusters that unsupervised classification produces are not information classes, but spectral classes (i.e., they group together features (pixels) with similar reflectance patterns). It is thus usually the case that the analyst needs to reclassify spectral classes into information classes. For example, the system might identify classes for asphalt and cement which the analyst might later group together, creating an information class called pavement.

Image Transformation

Digital image processing offers a limitless range of possible transformations on remotely sensed data. Two are mentioned here, specifically, because of their special significance in environmental monitoring applications.

Vegetation Indices

There are a variety of vegetation indices that have developed to help in the monitoring of vegetation. Most are based on the very different interactions between vegetation and electromagnetic energy in the red and near-infrared wavelengths. Reflectance in the red region (about 0.6 - 0.7μm) is low because of absorption by leaf pigments (principally chlorophyll). The infrared region (about 0.8 - 0.9 μm), however, characteristically shows high reflectance because of scattering by the cell structure of the leaves. A very simple vegetation index can thus be achieved by dividing the measure of infrared reflectance by that of the red reflectance. Areas of strong vegetation will thus result in a very high index value.

Although a number of variants of this basic logic have been tried, the one which has received the most attention is the *normalised difference vegetation index* (NDVI). It is calculated in the following manner:

NDVI = (NIR - R) / (NIR + R)
where NIR = Near Infrared and R = Red

This kind of calculation is quite simple for a raster GIS or Image Processing software system, and the result has been shown to correlate well with ground measurements of biomass. Many agencies have found the index to be useful as a relative measure for monitoring purposes. For example, the United Nations Food and Agricultural Organisation (FAO) Africa Real Time Information System (ARTEMIS), and the USAID Famine Early Warning System (FEWS) programs both use continental scale NDVI images derived from the NOAA-AVHRR system to produce vegetation index images for the entire continent of Africa every ten days.

Principal Components Analysis

Principal Components Analysis (PCA) is a linear transformation technique related to Factor Analysis. Given a set of image bands, PCA produces a new set of images, known as components, that are uncorrelated with one another and are ordered in terms of the amount of variance they explain from the original band set.

PCA has traditionally been used in remote sensing as a means of data compaction. For a typical multispectral image band set, it is common to find that the first two or three components are able to explain virtually all of the original variability in reflectance values. Later components thus tend to be dominated by noise effects. By rejecting these later components, the volume of data is reduced with no appreciable loss of information.

Recently, PCA has been shown to have special application to environmental monitoring. In cases where multispectral images are available for two dates, the bands from both images are submitted to a PCA as if they all came from the same image. In these cases, changes between the two dates tend to emerge in the later components. More dramatically, if a time series of NDVI (or a similar single-band index) images is submitted to the analysis, a very detailed analysis of environmental changes and trends can be achieved. In this case, the first component will show the typical NDVI over the entire series while each successive component illustrates change events in an ordered sequence of importance. By examining these images, along with graphs of their correlation with the individual bands in the original series, important insights can be gained into the nature of changes and trends over the time series.