Chapter 1:- A gentle introduction to GIS

1. Nature of GIS

The acronym GIS stands for geographic information system. As the name suggests, a GIS is a tool for working with geographic information.

Who use GIS?

- An urban planner might want to assess the extent of urban fringe growth in her/his city, and quantify the population growth that some suburbs are witnessing. S/he might also like to understand why these particular suburbs are growing and others are not
- A biologist might be interested in the impact of slash-and-burn practices on the populations of amphibian species in the forests of a mountain range to obtain a better understanding of long-term threats to those populations
- A natural hazard analyst might like to identify the high-risk areas of annual monsoon-related flooding by investigating rainfall patterns and terrain characteristics;
- A geological engineer might want to identify the best localities for constructing buildings in an earthquake-prone area by looking at rock formation characteristics;
- A mining engineer could be interested in determining which prospective copper mines should be selected for future exploration, taking into account parameters such as extent, depth and quality of the ore body, amongst others;
- A *geoinformatics engineer* hired by a telecommunications company may want to determine the best sites for the company's relay stations, taking into ac- count various cost factors such as land prices, undulation of the terrain;
- A *forest manager* might want to optimize timber production using data on soil and current tree stand distributions, in the presence of a number of operational constraints, such as the need to preserve species diversity in the area
- A *hydrological engineer* might want to study a number of water quality pa-rameters of different sites in a freshwater lake to improve understanding of the current distribution of *Typha* reed beds, and why it differs from that of a decade ago

All the professionals work with positional data; also called spatial data. Spatial data refers to where things are, or perhaps, where they were or will be. To be more precise, these professionals deal with questions Spatial data related to geographic space, which we define as having positional data relative to the Earth's surface.

 Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

The fundamental problem that we face in many uses of GIS is that of understanding phenomena that have a spatial or geographic dimension, as well as a temporal dimension. We are facing ‘spatio-temporal’ problems. This means that our object of study has different characteristics for different locations (the geographic dimension) and also that these characteristics change over time (the temporal dimension).

The El Nino event is a good example of such a phenomenon, because sea surface temperatures differ between locations, and sea surface temperatures change from one week to the next.

2. Study of El Nino as an example application of GIS

El Nino – “little boy” –

- is an aberrant pattern in weather and sea water temperature(every 4-9 years) in Pacific Ocean along equator.
- It is generally believed that El Nino has a considerable impact on global weather systems, and that it is the main cause for droughts in Wallacea and Australia, as well as for excessive rains in Peru and the southern U.S.A.
- It makes weather hotter, has impact on global weather systems and comes around Christmas

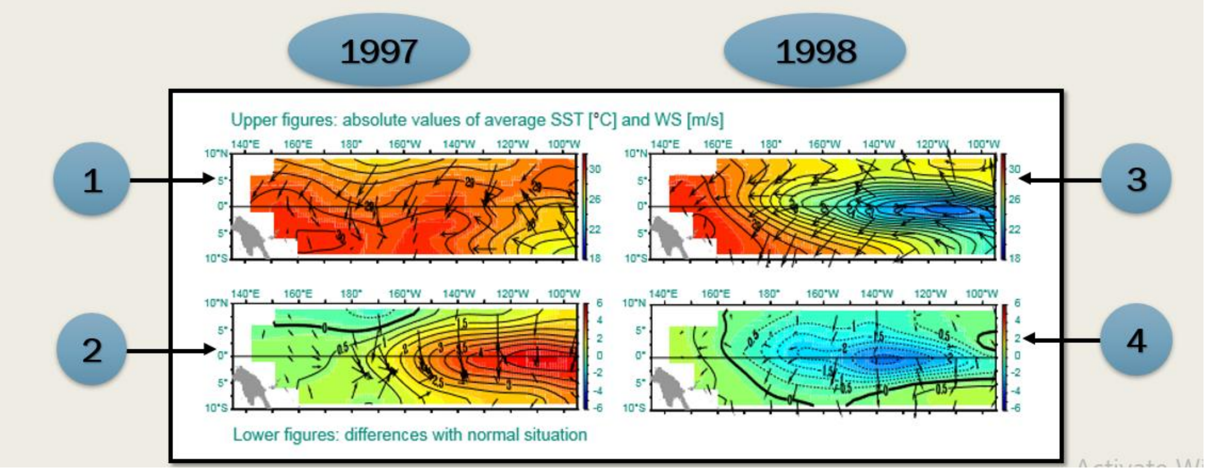
La Nina – “little girl” – pattern of colder temperature that occurs less frequently.

Figure marked 1 & 2 – El Nino that occurred in the year 1997
Figure marked 3 & 4 – La Nina that occurred in the year 1998

In all figures, colour is used to indicate sea water temperature, while arrow lengths indicate wind speeds. The top figures provide information about absolute values, while the bottom figures are labelled with values relative to the average situation for the month of December. The bottom figures also give an indication of wind speed and direction.

SST – Sea Surface Temperature (oC) & WS – Average wind speed (m/s)

 Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By



3. Defining GIS

Functional definition of GIS

A GIS is a computer-based system that provides the following four sets of capabilities to handle georeferenced data:

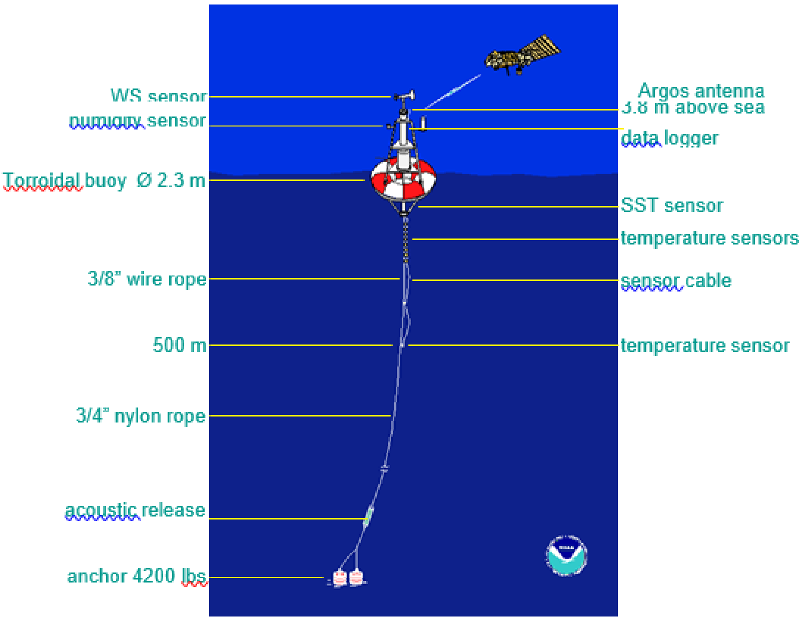
1. Data capture and preparation
2. Data management, including storage and maintenance
3. Data manipulation and analysis
4. Data presentation

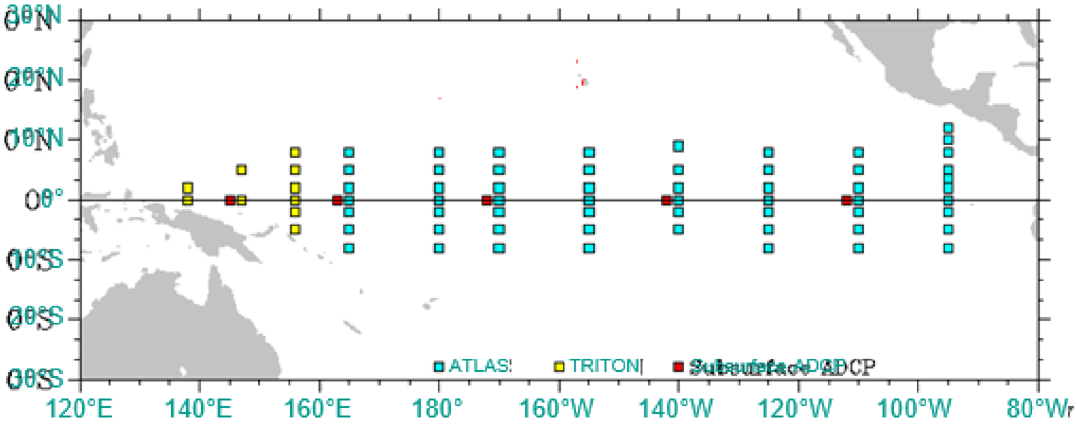
We look into these steps in more detail below, in the context of the El Nino example

4. Steps in GIS

4.1 Data Capture and Preparation

- In the El Nino case, data capture refers to the collection of Sea Surface Temperature and wind speed measurements.
- This is achieved by placing buoys with measuring equipments at various places in the ocean.
- 70 buoys were placed to measure
- A buoy measures:- i) wind speed & direction ii) air temperature & humidity iii) sea water temperature at surface and depth upto 500 m.
- A buoy - an anchored float serving as a navigation mark. The buoys have been anchored, so they are stationary.





The array of positions of sea surface temperature and wind speed measuring buoys in the equatorial Pacific Ocean

 Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

4.2 Data management

For our example application, data management refers to
1) the storage and maintenance of the data transmitted by the buoys via satellite communication.
2) a decision to be made on how best to represent our data, both in terms of their spatial properties and the various attribute values which we need to store.
3) acquired data to be put in digital form, that is, it has been converted into computer-readable format, so that we can begin our analysis.

4.3 Data manipulation and analysis

Once the data has been collected and organized in a computer system, we can start analysing it.

Here, we look at how data generated at the buoys was processed before map production.
The initial (buoy) data have been generalized from 70 point measurements (onefor each buoy) to cover the complete study area.
1. For each buoy, average SST for each month calculated using daily SST measurements.
2. For each buoy, monthly average SST was taken with geographic location to obtain georeferenced list of averages.
3. From this georeferenced list, through a method of spatial interpolation, the estimated SST of other positions were computed. (Table below)

Buoy	Geographic position	Dec. 1997 avg. SST
B0789	(165° E, 5° N)	28.02 °C
B7504	(180° E, 0° N)	27.34 °C
B1882	(110° W, 7°30' S)	25.28 °C
...	...	...

Previous to the above steps, we obtained data about average SST for December for a series of years. This is spatially interpolated to obtain a normal situation for December data set.

GIS function: deriving an estimate value for a property for some location where we have not measured
Georeferenced: associated with some position on Earth's surface by using spatial reference system (latitude longitude co-ordinate system)
Interpolation: A statistical method of deriving a simple function from the given discrete data set such that the function passes through the provided data points.

4.4 Data presentation

1. Message :- What message to portray?

Here, we portra the El Nin˜o and La Nin˜a events, both in absolute figures, and also in relative figures, i.e. differences from a normal situation.

2. Audience :- Who is the audience?

The *audience* for this data presentation clearly were the readers of a book where this data is published or analysts.

3. Medium :- What form to present it? It can be either printed matter of A4 size or it can be published in a website.

4. Rules of aesthetics :- How will it look? The maps should be printed north-up; with clear georeferencing; with intuitive use of symbols et cetera.

5. Techniques:- This includes usage of a colour scheme , isolines etc.

5. Spatial Data and Geo information

A subtle difference exists between the terms data and information.

By data, we mean representations that can be operated upon by a computer.

More specifically, by spatial data we mean data that contains positional values, such as (x, y) co-ordinates. Sometimes the more precise phrase geospatial data is used as a further refinement, which refers to spatial data that is georeferenced.

By Geospatial data and information, we mean data that has been interpreted by a human being. Humans work with and act upon information, not data. Human perception and mental processing leads to information, and hopefully understanding and knowledge. Geoinformation is a specific type of information resulting from the interpretation of spatial data.

Key components of spatial data quality include

positional accuracy (both horizontal and vertical),

temporal accuracy (that the data is up to date),

attribute accuracy (e.g. in labelling of features or of classifications),

lineage (history of the data including sources),

completeness (if the data set represents all related features of reality), and

logical consistency (that the data is logically structured).

Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

6. Modelling

'Modelling' is a term used in many different ways and which has many different meanings. A representation of some part of the real world can be considered a model because the representation will have certain characteristics in common with the real world.
Specifically, those which we have identified in our model design. This then allows us to study and operate on the model itself instead of the real world in order to test what happens under various conditions, and help us answer 'what if' questions. We can change the data or alter the parameters of the model, and investigate the effects of the changes.

Models—as representations—come in many different flavours. In the GIS environment, the most familiar model is that of a map.

A map is a miniature representation of some part of the real world. Databases are another important class of models. A database can store a considerable amount of data, and also provides various functions to operate on the stored data.

A 'real world model' is a representation of a number of phenomena that we can observe in reality, usually to enable some type of study, administration, computation and/or simulation.

Most maps and databases can be considered static models. At any point in time, they represent a single state of affairs. Usually, developments or changes in the real world are not easily recognized in these models. Dynamic models or process models address precisely this issue. They emphasize changes that have taken place, are taking place or may take place sometime in the future. Dynamic models are inherently more complicated than static models, and usually require much more computation. Simulation models are an important class of dynamic models that allow the simulation of real world processes.

Observe that our El Nino system can be called a static model as it stores state-of- affairs data such as the average December 1997 temperatures. But at the same time, it can also be considered a simple dynamic model, because it allows us to compare different states of affairs.

7. Maps

Maps are perhaps the best known (conventional) models of the real world.

Maps have been used for thousands of years to represent information about the real world, and continue to be extremely useful for many applications in various domains. Their conception and design has developed into a science with a high degree of sophistication. A disadvantage of the traditional paper map is that it is generally restricted to two-dimensional static representations, and that it is always displayed in a fixed scale. The map scale determines the spatial resolution of the graphic feature representation. The smaller the scale, the less detail a map can show. The accuracy of the base data, on the other hand, puts limits to the scale in which a map can be sensibly drawn. Hence, the selection of a proper map scale is one of the first and most important steps in map design.

A map is always a graphic representation at a certain level of detail, which is determined by the scale. Map sheets have physical boundaries, and features spanning two map sheets have to be cut into pieces. Cartography, as the science and art of map making, functions as an interpreter, translating real world phenomena (primary data) into correct, clear and understandable representations for our use. Maps also become a data source for other applications, including the development of other maps.

With the advent of computer systems, analogue cartography developed into digital cartography, and computers play an integral part in modern cartography.

Alongside this trend, the role of the map has also changed accordingly, and the dominance of paper maps is eroding in today's increasingly 'digital' world.

The traditional role of paper maps as a data storage medium is being taken over by (spatial) databases, which offer a number of advantages over 'static' maps, as discussed in the sections that follow. Notwithstanding these developments, paper maps remain as important tools for the display of spatial information for many applications.

8. Databases

A database is a repository for storing large amounts of data. It comes with a number of useful functions:

1. A database can be used by multiple users at the same time—i.e. it allows concurrent use,

2. A database offers a number of techniques for storing data and allows the use of the most efficient one—i.e. it supports storage optimization,

3. A database allows the imposition of rules on the stored data; rules that will be automatically checked after each update to the data—i.e. it supports data integrity,

4. A database offers an easy to use data manipulation language, which allows the execution of all sorts of data extraction and data updates—i.e. it has a query facility,

5. A database will try to execute each query in the data manipulation language in the most efficient way—i.e. it offers query optimization.

 Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

For the El Nino project, one may assume that the buoys report their measurements on a daily basis and that these measurements are stored in a single, large table.

DAYMEASUREMENTS

Buoy	*Date*	*SST*	*WS*	*Humid*	*Temp10*	...
B0749	1997/12/03	28.2 °C	NNW 4.2	72%	22.2 °C	...
B9204	1997/12/03	26.5 °C	NW 4.6	63%	20.8 °C	...
B1686	1997/12/03	27.8 °C	NNW 3.8	78%	22.8 °C	...
B0988	1997/12/03	27.4 °C	N 1.6	82%	23.8 °C	...
B3821	1997/12/03	27.5 °C	W 3.2	51%	20.8 °C	...
B6202	1997/12/03	26.5 °C	SW 4.3	67%	20.5 °C	...
B1536	1997/12/03	27.7 °C	SSW 4.8	58%	21.4 °C	...
B0138	1997/12/03	26.2 °C	W 1.9	62%	21.8 °C	...
B6823	1997/12/03	23.2 °C	S 3.6	61%	22.2 °C	...
...	...	...	...	...	...	...

****

Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By